

# ACTAS

## XI CONGRESO DE METODOLOGÍA DE LAS CIENCIAS SOCIALES Y DE LA SALUD

Málaga, 15-18 septiembre de 2009

El Congreso de Metodología de las Ciencias Sociales y de la Salud es de ámbito nacional y se organiza bianualmente con el auspicio de la Asociación Española de Metodología de las Ciencias del Comportamiento (AEMCCO) a la que pertenecen los docentes e investigadores adscritos al área de Metodología de las Ciencias del Comportamiento de la Facultad de Psicología. Esta área se caracteriza por abordar un contenido dinámico y transversal, con elevada capacidad de aplicación a cualquier ámbito dentro de las Ciencias Sociales y de la Salud. El Congreso, ya consolidado a través de las diez ediciones anteriores, persigue hacer patente los progresos producidos en los contenidos propios del área, a la vez que promover y fomentar el intercambio de conocimientos entre profesores, investigadores, profesionales y estudiantes interesados por cuestiones metodológicas.

ACTAS DEL XI CONGRESO DE METODOLOGÍA DE LAS CIENCIAS SOCIALES Y DE LA SALUD

Dpto. de Psicobiología y Metodología de las Ciencias del Comportamiento  
Facultad de Psicología  
Universidad de Málaga  
Asociación de Metodología de las Ciencias del Comportamiento





XI Congreso de Metodología de las Ciencias Sociales y de la Salud

Coordinadores: M. J. Blanca, R. Alarcón y D. López-Montiel

Maquetación: Andrade García, Carmen María  
Correcciones: Martín Tirado, Almudena

ISBN: 978-84-613-7589-9

Depósito Legal:

Edita UMA-Tecnolex (B93004000)

Año: 2010

Lugar: Málaga

**XI CONGRESO DE METODOLOGÍA  
DE LAS CIENCIAS SOCIALES Y DE LA SALUD**

**Comité de Honor**

Excma. y Magfca. Rectora de la Universidad de Málaga  
Adelaida de la Calle Martín

Ilmo. Sr. Decano del Colegio Oficial de Psicólogos de Andalucía Oriental  
Manuel Mariano Vera Martínez

Ilmo. Sr. Decano de la Facultad de Psicología  
Julián Almaraz Carretero

Ilmo. Sr. Director del Dpto. de Psicobiología y Metodología de las Ciencias del Comportamiento  
Jesús Miranda Páez

Sr. Presidente de *European Association of Methodology*  
José Muñoz Fernández

Sra. Presidenta de la Asociación Española de Metodología de las Ciencias del Comportamiento  
M. Teresa Anguera Argilaga

**Comité Científico**

Coordinadora: Juana Gómez Benito  
Manuel Ato García  
Inmaculada de la Fuente Solana  
Rafael San Martín Castellanos

**Comité Organizador**

Presidenta: María J. Blanca Mena  
Vicepresidenta: Dolores López Montiel  
Secretario: Manuel Pelegrina del Río  
Vicesecretario: Rafael Alarcón Postigo  
Tesorero: Jesús Miranda Páez  
Vocales: Marcos Ruiz Soler  
Roser Bono Cabré  
Catalina Pérez López  
Gema López Montiel  
Agustín Wallace Ruiz  
Rebecca Bendayan



# ÍNDICE

## SIMPOSIOS

	Nº Pág.
<b>Investigación longitudinal en Ciencias Sociales y de la Salud. Análisis, métodos y aplicaciones</b> <b>Coordinador: J. Arnau</b>	<b>17</b>
Aplicación de modelos lineales mixtos para el análisis de datos de seguimiento en el ámbito clínico J. Arnau, R. Bono y G. Vallejo	18
Selección de modelos anidados con datos de medidas repetidas usando criterios de información y pruebas de ajuste condicional G. Vallejo, J. Arnau, R. Bono y P. Fernández García	24
Comparación de cinco procedimientos para evaluar la efectividad de intervenciones en diseños de caso único R. Manolov, J. Arnau, A. Solanas y R. Bono	30
Análisis de datos longitudinales en un modelo de series temporales agrupadas multinivel J. F. Rosel, P. Torrente, E. F. Maldonado y J. J. Canales	34
<b>Investigación aplicada e investigación metodológica con diseños orientados a la evaluación del cambio</b> <b>Coordinadora: P. Fernández García</b>	<b>45</b>
Análisis de la robustez de cuatro estadísticos univariados en un diseño de medidas repetidas carente de esfericidad con y sin autocorrelación serial P. Fernández García, G. Vallejo, P. Livacic-Rojas y E. Tuero	46
Análisis de un diseño de medidas repetidas usando remuestreo <i>bootstrap</i> y permutación aleatoria G. Vallejo, P. Fernández García, E. Tuero y P. Livacic-Rojas	51
Estimación del efecto en series temporales cortas. Propuesta de un programa en SPSS como alternativa al procedimiento ARIMA J. Arnau, R. Bono, A. Solanas y R. Manolov	56
Análisis de la sintomatología menstrual: un modelo factorial autorregresivo de ecuaciones estructurales con series temporales agrupadas F. Herrero, P. Jara y J. F. Rosel	61

<b>Aplicaciones de la teoría de la medida y de los métodos de análisis de datos a la investigación en Psicología de las Organizaciones</b>	<b>68</b>
<b>Coordinadora: E. Torres</b>	
Ventajas de la aplicación de los modelos politómicos en la medición de la actitud emprendedora	70
E. Torres, M. Aranburu y D. Hermosilla	
¿Inferencia bayesiana o frecuentista en el análisis de datos?	76
E. I. de la Fuente, L. M. Lozano, G. R. Cañadas y C. Vargas	
Análisis clásico y bayesiano en el estudio diferencial de los ítems	83
L. M. Lozano, E. I. de la Fuente, M. Martín y G. R. Cañadas	
Análisis bayesiano de los niveles de <i>burnout</i> en diferentes colectivos profesionales	88
E. I. de la Fuente, L. M. Lozano, G. A. Cañadas y V. Sánchez Garzón	
Influencia de la interdependencia de objetivos individual y la autogestión y participación grupal sobre la satisfacción laboral y la conducta cooperativa: una aproximación multinivel	93
A. Aritzeta y N. Balluerka	
<b>Aproximaciones metodológicas para obtener evidencias de validez</b>	<b>100</b>
<b>Coordinadores: J. L. Padilla y N. Balluerka</b>	
Validación de la versión reducida en castellano del <i>Trait Meta-Mood Scale</i> (TMMS) para la población adolescente	101
N. Balluerka, J. Martín Salguero, P. Fernández-Berrocal y A. Aritzeta	
Evidencias de validez de contenido de la versión española rasgo del <i>State Trait Cheerfulness Inventory</i> (STCI)	107
E. Delgado y H. Carretero-Dios	
Obtención de evidencias de validez para un cuestionario sobre “apoyo familiar” mediante entrevistas cognitivas	112
M. Castillo, I. Benítez y J. L. Padilla	
Aplicación de las técnicas de DIF a los cuestionarios utilizados en las encuestas	117
J. Gómez-Benito, M. D. Hidalgo, J. L. Padilla e I. Benítez	
<b>Metodología para la traducción y adaptación de instrumentos de evaluación entre distintos grupos lingüísticos y culturales</b>	<b>122</b>
<b>Coordinadores: N. Balluerka y J. L. Padilla</b>	
Adaptación al euskera del EPQ-J ( <i>Cuestionario de Personalidad para Niños</i> )	123
A. Gorostiaga, N. Balluerka y X. Isasi	
Adaptación del <i>Athlete Burnout Questionnaire</i> (ABQ) del inglés al castellano	130
C. Arce, C. de Francisco, E. Andrade, J. Garrido, J. Torrado e I. Arce	

Análisis de la equivalencia del constructo “asertividad” en muestras mexicanas y españolas I. Benítez, M. Castillo, J. L. Padilla, J. Gómez-Benito, M. M. Flores-Galaz y R. Díaz-Loving	136
<b>Teoría de la Generalizabilidad: investigaciones aplicadas</b> <b>Coordinadores: A. Blanco y V. Morales</b>	<b>142</b>
Diseño de medida en la construcción de una herramienta para evaluar la calidad de los programas de voluntariado ambiental E. Chica, A. Hernández Mendo y V. Morales	143
Estimación de diseños para evaluar la calidad en los centros de atención infantil temprana (CAIT) R. P. Romero y V. Morales	149
Diseño de precisión de una herramienta para la evaluación de la calidad de una instalación municipal deportiva V. Morales y P. Gálvez	154
<b>Estimación de parámetros bajo el supuesto de incumplimiento de las condiciones de aplicación</b> <b>Coordinador: J. Guàrdia</b>	<b>160</b>
Aplicaciones de modelos de regresión bajo condiciones de muestra pequeña A. Gordovil y J. Guàrdia	162
El uso de la mediana como estadístico resistente. Coincidencias en la decisión a partir de diferentes estadísticos M. Però, J. Guàrdia, M. Freixa y J. Turbany	169
La fase de estimación de parámetros en los modelos de ecuaciones estructurales con variables categóricas S. Benítez, J. Guàrdia, M. Arch y A. Jarne	176
Efecto de los <i>missing data</i> en un diseño multivariante de grupos independientes con muestras pequeñas M. J. Carrera, G. Feixas y J. Guàrdia	183

## SESIONES PARALELAS

### Metodología experimental

Características y análisis de los diseños de medidas repetidas en la investigación en España en los últimos 10 años P. Fernández García, G. Vallejo, P. Livacic-Rojas y E. Tuero	193
Tendencias de la potencia de prueba en un diseño de medidas repetidas carente de esfericidad con y sin autocorrelación serial P. Fernández García, G. Vallejo, P. Livacic-Rojas y E. Tuero	199
Potencia del estadístico $f$ vs. $F$ cuando no se asume la normalidad J. F. García Pérez, E. Gracia, S. Murgui y M. del Castillo Fuentes	204
Evaluación del comportamiento de selectores de estructuras de covarianza en diseños de medidas repetidas P. Livacic-Rojas, G. Vallejo y P. Fernández García	208
Aplicación de procedimientos heteroscedásticos transformados al análisis de diseños factoriales desequilibrados G. Vallejo, M. Ato, P. Fernández García y E. Tuero	213
Análisis de diseños longitudinales con datos incompletos G. Vallejo, P. Fernández García, E. Tuero y P. Livacic-Rojas	219

### Metodología observacional, metodología cualitativa y complementariedad metodológica

Comparación de CAQDAS para la investigación psicosocial: ATLAS-TI <i>versus</i> NVivo M. T. Anguera y O. López Fernández	224
Análisis observacional de la fase ofensiva del fútbol de alto nivel C. A. Casal y T. Ardá	229
Aproximación integrada al análisis de marcos interpretativos en artículos periodísticos S. Fàbregues y M. H. Paré	239
Evaluación de un programa de habilidades de pensamiento en niños con Síndrome de Asperger M. Herrero, M. Frontera, P. Allueva y M. T. Anguera	245
Estudio comparativo de la percepción del riesgo entre el personal interno y externo de una obra de construcción E. López Fernández, L. Maria y O. López Fernández	253
Aplicación de herramientas observacionales para el análisis de la violencia en medios impresos venezolanos Z. Márquez y A. Hernández Mendo	257

Revisión y comparación de la aplicación de la metodología híbrida en tres revistas de referencia en ciencias del comportamiento J. F. Molina y O. López Fernández	262
Fuentes psicosociales de estrés en médicos residentes M. Ortiz, B. Valenzuela y M. Guillén	267
La calidad del registro en sesiones de larga duración: aplicación de la teoría de la generalizabilidad E. Rodríguez-Naveiras	273
Análisis de las experiencias subjetivas de personas con trastornos mentales graves en procesos de “recuperación” F. J. Saavedra	285
Prácticas de planificación, gestión, ejecución y control de la producción de alimentos en situaciones simuladas y reales M. Tenutto, C. Brutti y S. Algarañá	290

### **Psicometría**

El consenso entre expertos como método para evaluar la validez de contenido A. Andrés, J. Gómez-Benito y C. Saldaña	297
Calibración <i>online</i> en test adaptativos informatizados A. Antivilo, J. Olea y F. J. Abad	307
Elaboración de una escala de liderazgo en deportes de equipo C. Arce, J. Torrado, E. Andrade, J. Garrido, C. de Francisco e I. Arce	313
Elaboración de un cuestionario de accesibilidad de los inmigrantes al sistema sanitario I. Benítez, J. Gómez-Benito, J. L. Padilla, P. Tejero y M. Martín-Baranera	319
Detecting halo effects in performance-based examinations T. M. Bechger y G. Maris	330
Estructura factorial del Cuestionario de Violencia Escolar Cotidiana (CUVECO) M. J. Blanca, F. J. Fernández-Baena, M. Escobar y M. V. Trianes	334
Estructura factorial de la escala SSI sobre respuestas de estrés S. Cantos, F. J. Fernández-Baena, M. V. Trianes, M. Escobar y M. J. Blanca	340
¿Cómo evaluamos? Análisis de ítems de opción múltiple y su relación con errores en la construcción C. García García, V. Ponsoda y A. Sierra	344
Estudio experimental sobre la influencia de la redacción de los ítems en las respuestas de los sujetos M. R. Garrido, M. A. Fernández Borrero, E. Villalba, M. de Vega y P. J. Pérez Moreno	350

Sesiones supervisadas y no supervisadas en la administración de subtest de anclaje a través de Internet J. López-Cuadrado, T. A. Pérez Fernández, J. A. Vadillo y J. Gutiérrez	357
Análisis factorial confirmatorio de la <i>Statistical Anxiety Scale</i> L. Pérez-Gómez, J. Carmona y J. L. Sánchez Ramos	362
Discriminación de los ítems de la subescala de combinación de emociones del <i>Mayer-Salovey-Caruso-Emotional Intellingence Test</i> (MSCEIT) M. Sánchez-García, N. Extremera y P. Fernández-Berrocal	368
Evidencias de validez relacionadas con la estructura interna de los ítems de la subescala de combinación de emociones del <i>Mayer-Salovey-Caruso Emotional Intellingence Test</i> Version (MSCEIT) M. Sánchez García, N. Extremera y P. Fernández Berrocal	374
Validación de constructo factorial de un instrumento objetivo de tendencia al riesgo I. Sánchez-Iglesias y M. J. Sueiro	380
Validez y fiabilidad de una versión española del cuestionario SA-45: apoyo empírico preliminar A. Sánchez Prada y B. Palacios	387
<b>Innovación docente</b>	
LVE: laboratorio virtual de estadística J. Chacón, V. de la Fuente, T. Ruiz Gallego-Largo, F. Serradilla, M. J. Hernández-Lloreda y M. R. Martínez-Arias	394
Identificación de perfiles proambientales en el alumnado del Certificado de Aptitud Pedagógica en la Universidad de Málaga L. I. Estrada y A. Matas	401
Contenidos metodológicos en el Grado en Psicología en la Comunidad Autónoma de Andalucía. Descripción del caso de la Universidad de Almería J. García García y L. de la Fuente	407
La innovación educativa desde la metodología como disciplina transversal: un sistema para contrastar modelos teóricos en las clases de prácticas J. F. García Pérez, E. Gracia, M. S. Lila y M. del Castillo Fuentes	411
Dificultades y temores del alumnado universitario en su salida al mundo laboral P. García Soidán y H. Dios	417
Innovación docente en “Diseños experimentales y aplicados”: ECTS, campus virtual y evaluación continua O. López Fernández, M. Viader, A. Coscolluela, M. L. Honrubia y J. M. Malapeira	423

Estrategias de aprendizaje y su relación con el rendimiento académico previo de los estudiantes de nuevo ingreso en la titulación de Psicología D. López-Montiel, M. J. Blanca, G. López-Montiel, R. Alarcón y J. Miranda	429
Diferencias individuales en la autorregulación de los aprendizajes en estudiantes universitarios: estrategias motivacionales y aspectos volitivos de la personalidad D. López-Montiel, M. J. Blanca, G. López-Montiel, R. Alarcón y J. Miranda	436
La relación interpersonal docente-estudiante en la educación superior. Adaptación del QTI y resultados por ramas de conocimiento E. Mena y J. C. Tójar	441
Integración del inglés como segunda lengua en el aprendizaje de diseños experimentales J. C. Oliver y P. Salazar	447
Aplicación de la metodología de encuestas para determinar la actitud del profesor universitario ante el uso del ordenador. Caso: universidades públicas venezolanas H. Padrón y M. Ruiz Soler	454
¿Qué deberíamos cambiar en la docencia de análisis de datos? M. Ruiz Soler	460
Modelado estadístico con <i>R</i> : un entorno web para la docencia M. Ruiz Soler y A. Wallace	464
Intenciones del profesorado de educación física español hacia la salud y la calidad de vida F. Salinas, M. T. Miranda, A. Cocca, K. Mohamed y J. Viciano	469
Una experiencia de innovación docente en métodos de investigación en educación J. Sánchez Santamaría	475
Bases teóricas para una innovación docente: los esquemas científicos en asignaturas metodológicas J. Sánchez Santamaría y M. G. Ballester	483
Cómo medir la metacognición en estudiantes universitarios A. Sandoval y L. Franchi	491
El constructivismo, un nuevo significante para significados con historia M. Tenutto	498

### **Aplicaciones metodológicas**

Diferencias de género en la autopercepción de la voluntad de trabajo H. F. Attorresi, F. J. P. Abal, G. S. Lozzia, M. E. Aguerri y M. S. Galibert	503
Uso del análisis envolvente de datos para la evaluación de los programas de fomento de la I+D+I R. C. Barea, E. Cabrera, J. L. Pino y F. M. Solís	507

Aplicación del análisis de componentes principales al estudio de la conducta de ratones salvajes y carentes del receptor LPA <sub>1</sub> en el laberinto de hoyos E. Castilla, J. Sánchez López, C. Hoyo, E. Zambrana, E. Matas, J. Chun, F. Rodríguez de Fonseca, C. Pedraza, G. Estivill y L. J. Santín	516
Influencia de los factores del autoconcepto sobre cantidad y calidad de la actividad física en jóvenes españoles A. Cocca, J. Viciano, F. Salinas y K. Mohamed	522
Estudio de prevalencia de los trastornos internalizantes y externalizantes en una población infantil A. Cortés, J. Miranda y M. V. Trianes	530
Consideraciones metodológicas acerca de la detección de asociaciones mediante la inducción al consumo de sal A. Espinet, J. M. Caramés, R. Alonso, A. González Galán, A. Quero y T. Vaquero	536
El procesamiento de homógrafos interlingüales: datos de una simulación computacional A. J. Estudillo y J. García Orza	540
Predicción de los resultados en pista para piragüistas de alta competición J. L. García Soidán y P. García Soidán	544
Discrepancias entre los conocimientos auto-percibidos y los conocimientos reales en sexualidad E. García-Vega, E. Menéndez-Robledo, P. Fernández García y A. Menéndez	550
Evolución de la edad de inicio en el consumo de drogas y la demora en el tratamiento F. Herrero, P. Jara, M. Salvador y F. Verdú	556
Ansiedad y objetividad en las investigaciones del comportamiento S. Hurtado	562
Conducta alimentaria como factor de riesgo en los trastornos del comportamiento alimentario J. M. Llopis y M. V. Acosta	567
Estudio empírico del uso excesivo de las tecnologías de entretenimiento de los adolescentes de Barcelona O. López Fernández, M. L. Honrubia y M. Freixa	573
A definition of potential entrepreneur from a probabilistic point of view J. López Puga, J. García García, C. J. Cano, A. B. Gea y L. de la Fuente	577
Revisión de instrumentos de medida de cantidad-calidad de actividad física-salud en educación física K. Mohamed, B. Szabolcs, L. Lozano, A. Cocca, F. Salinas, J. C. Martínez Álvarez y J. Viciano	582
Diseños genéticamente informativos: el registro de gemelos de Murcia J. R. Ordoñana, I. Rebollo-Mesa, F. González-Javier, E. Carrillo, J. F. Sánchez-Romera, L. Colodro y J. M. Martínez-Selva	588

Modelado estadístico de variables asociadas a la accidentalidad en motocicleta P. Pérez-Fuster, M. F. Rodrigo, M. L. Ballestar y J. Sanmartín	594
Variables discriminantes del funcionamiento ejecutivo en niños y adolescentes con trastorno por déficit de atención con hiperactividad E. Verche, R. Martín, A. Machado, P. A. González-Pérez, I. Quintero, J. Bravo y S. Hernández Sánchez	602
Modelado estadístico de la latencia del P300 en memoria explícita e implícita A. Wallace, M. Pelegrina, A. Pelegrina y R. Marín	612
<b>ÍNDICE DE AUTORES</b>	<b>619</b>



# SIMPOSIOS



# INVESTIGACIÓN LONGITUDINAL EN CIENCIAS SOCIALES Y DE LA SALUD. ANÁLISIS, MÉTODOS Y APLICACIONES

**Coordinador: Jaume Arnau**  
Universidad de Barcelona

No es necesario insistir que la investigación longitudinal es la base de muchos estudios sociales a gran escala, así como de otros estudios de menor alcance propios del ámbito epidemiológico, psicológico o educativo. En ambos casos, lo característico es tomar información y obtener datos, a lo largo de un período de tiempo, de las mismas unidades de observación, por cuya razón la técnica utilizada es conocida por medidas repetidas. Los datos longitudinales suelen obtenerse a través de encuestas o mediante acceso directo a bases de datos administrativos de la población o datos censales. Los estudios longitudinales son instrumentos útiles para el estudio de los cambios individuales a través del tiempo. Esto contrasta con los estudios o datos transversales que proporcionan una instantánea de una población en un punto de corte en el tiempo. Sabemos que el tiempo es el factor principal del cambio, de modo que cuando es incorporado en el estudio nos ofrece una información más intensiva, rica y significativa que los estudios simplemente transversales o transeccionales. Si comparamos el enfoque longitudinal con el enfoque transversal, se constata que el primer enfoque es más eficiente, más robusto en la selección del modelo y con más potencia estadística (Edwards, 2000; Helms, 1992; Zeger y Liang, 1992). No nos debe extrañar que los investigadores de lo longitudinal hayan centrado su interés en estudiar las propiedades estadísticas de los estimadores del cambio a partir de modelos estadísticos modernos, utilizando para ello estudios basados en los procedimientos de simulación Monte Carlo. Espero que con este simposio conozcamos cual es la realidad y nivel que tiene la investigación sobre datos longitudinales dentro de nuestro contexto de trabajo.

# APLICACIÓN DE LOS MODELOS LINEALES MIXTOS PARA EL ANÁLISIS DE DATOS DE SEGUIMIENTO EN EL ÁMBITO CLÍNICO

Jaume Arnau<sup>1</sup>, Roser Bono<sup>1</sup> y Guillermo Vallejo<sup>2</sup>

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Universidad de Oviedo

Desde una perspectiva metodológica, los diseños longitudinales son instrumentos útiles para estudiar los procesos de cambio directamente asociados con el paso del tiempo y para analizar datos de carácter longitudinal. Uno de los procedimientos más recientemente utilizados son los modelos lineales mixtos (MLM).

Dentro del contexto de la prevención del consumo de sustancias, es frecuente el registro de medidas repetidas, lo que ha llevado a un avance en los métodos estadísticos utilizados. En este trabajo nos centraremos en los MLM aplicados, precisamente, a un estudio longitudinal sobre el conocimiento ante el consumo de alcohol y otras drogas en adolescentes españoles. El principal objetivo es comparar la eficacia de un programa de entrevista motivacional con otro estándar sobre el conocimiento de las drogas, examinando las posibilidades que tienen los MLM para este tipo de estudios. Con este fin, analizamos el indicador global de conocimiento mediante los siguientes modelos: a) el modelo incondicional lineal de curva de crecimiento, b) el modelo de crecimiento con variables asociadas a las personas y c) el modelo de curvas individuales con variables predictoras.

Los diseños longitudinales son instrumentos útiles para examinar los procesos de cambio directamente asociados con el paso del tiempo y para analizar datos de carácter longitudinal en la prevención del consumo de sustancias (Mackinnon y Lockwood, 2003). La principal ventaja de los MLM es la especificación de la correcta estructura de covarianza de los datos observados, lo que conlleva más potencia estadística al probar los efectos del estudio (Fernández, Livacic-Rojas y Vallejo, 2007; Gill, 2000; Kowalchuk, Keselman, Algina y Wolfinger, 2004; Wolfinger, 1996). Además, los MLM son robustos a la no proporcionalidad de los grupos y al desgaste de la muestra (Fernández, et al., 2007; Gill, 2000). Dadas las ventajas de los MLM, el objetivo de este estudio es conocer su eficacia con datos longitudinales del conocimiento de sustancias psicoactivas.

## Método

### *Participantes*

Se realizó un seguimiento en cuatro tiempos a lo largo de un año de una muestra de 113 adolescentes consumidores de sustancias psicoactivas, de edades comprendidas entre los 12 y 17 años (26 chicos y 87 chicas) que acudieron por primera vez a las consultas externas de un centro de salud mental, durante un período aproximado de 14 meses a partir de enero de 2004.

### *Instrumentos y variables predictoras*

Los sujetos fueron evaluados mediante un cuestionario sobre conocimientos de drogas elaborado *ad hoc* para esta investigación. De este cuestionario, se obtuvo la puntuación global sobre conocimiento en una escala de cero a 31 puntos, de poco a mucho conocimiento.

En cuanto a las variables que influyen en el conocimiento que tienen los adolescentes sobre el consumo de sustancias psicoactivas, la más relevante es el paso del tiempo. Otras posibles variables predictoras del conocimiento global sobre el consumo de drogas son la escolarización y el índice de gravedad de la adicción de los adolescentes en el área escolar (ASI-escuela) medida con una escala Likert de cinco puntos.

### *Procedimiento*

Los sujetos de la muestra fueron asignados, aleatoriamente, a dos condiciones experimentales: 60 jóvenes consumidores recibieron una Intervención Motivacional (IM) diseñada *ad hoc* para este estudio a fin de reducir o abandonar el consumo de drogas y 53 jóvenes consumidores recibieron una Intervención Estándar (IE). De ambos grupos se tomaron medidas pre y post-intervención del conocimiento de drogas. La evaluación pre-intervención se realizó tras la primera visita y las evaluaciones post-intervención después de un mes, a los seis meses y al año. En todos los casos, los adolescentes y sus padres firmaron un informe de consentimiento.

Los sujetos del grupo control recibieron la IE, mientras que el grupo experimental recibió, además de la intervención psiquiátrica habitual, una IM breve dirigida a disminuir el consumo de sustancias. Esta intervención consistió en a) una entrevista motivacional individual con el adolescente y b) una sesión informativo-educativa para los padres que podía realizarse en grupo o por separado, dependiendo de su disponibilidad. La IM fue llevada a cabo por un equipo especializado al que se remitieron los casos del grupo experimental.

## Resultados

### *Modelo incondicional lineal de curva de crecimiento*

Al primer nivel, se plantea la hipótesis que el CONOCIMIENTO sobre el consumo de drogas que tiene el adolescente  $j$  en la ocasión  $i$  es una función lineal del TIEMPO:

$$\text{CONOCIMIENTO}_{ij} = \pi_{0j} + \pi_{1j}\text{TIEMPO}_{ij} + e_{ij} \quad (1)$$

La parte fija del modelo de crecimiento individual de la ecuación (1) representa la trayectoria de cada sujeto en función del tiempo y está integrada por dos parámetros: el intercepto o nivel de conocimiento del individuo en el primer punto del tiempo ( $\pi_{0j}$ ) y la pendiente o tasa de crecimiento individual del conocimiento en función del tiempo ( $\pi_{1j}$ ). La parte aleatoria  $e_{ij}$  es el residual del modelo o término de error aleatorio intra-sujeto.

A un segundo nivel, lo que interesa es describir la variabilidad entre los sujetos en torno a los valores iniciales (interceptos) y a las tasas de crecimiento (pendientes). Esta información queda recogida en dos ecuaciones donde los parámetros  $\pi_{0j}$  y  $\pi_{1j}$  del nivel 1 se explican por un valor de población o media común ( $\beta_{00}$ ) y por una pendiente de población o valor de crecimiento común a todos los sujetos ( $\beta_{10}$ ):

$$\begin{aligned} \pi_{0j} &= \beta_{00} + u_{0j} \\ \pi_{1j} &= \beta_{10} + u_{1j} \end{aligned} \quad (2)$$

A su vez, las diferencias individuales en torno a los dos parámetros del nivel 1,  $\pi_{0j}$  y  $\pi_{1j}$ , constituyen los términos aleatorios residuales de las ecuaciones del segundo nivel ( $u_{0j}$  y  $u_{1j}$ ).

Sustituyendo los correspondientes valores  $\pi_{0j}$  y  $\pi_{1j}$  de las ecuaciones del segundo nivel (2) en la ecuación del primer nivel (1), se deriva el modelo combinado con la parte fija y la parte aleatoria:

$$\text{CONOCIMIENTO}_{ij} = [\beta_{00} + \beta_{10} \text{TIEMPO}_{ij}] + [u_{0j} + u_{1j} \text{TIEMPO}_{ij} + e_{ij}] \quad (3)$$

En este modelo de crecimiento, se tienen dos efectos fijos:  $\beta_{00} = 17,14$  que representa el valor de población cuando  $\text{TIEMPO} = 0$  y  $\beta_{10} = 1,08$  es la pendiente promedio (Tabla 1). Esto significa que cuando el conocimiento promedio inicial sobre el consumo de drogas se sitúa en 17,14 puntos, el individuo incrementa su conocimiento por sesión en 1,08. Ambos efectos son significativos ( $p < 0,001$ ), se rechaza que los parámetros sean cero en la población. En cuanto a las estimaciones de los componentes de covarianza, el valor estimado de la varianza residual intra  $\sigma^2$  es 8,43, y el resto de componentes toman los siguientes valores:  $\tau_{00} = 14,30$ ,  $\tau_{01} = -1,04$  y  $\tau_{11} = 0,72$ . De estos valores son significados la varianza residual intra ( $p < 0,001$ ), lo cual indica que la varianza entre las medidas repetidas es no cero en la población, la varianza de los errores entre interceptos individuales ( $p < 0,001$ ) y la varianza de los errores entre pendientes individuales ( $p = 0,049$ ). En cambio, la covarianza entre intercepto y pendiente no es significativa ( $p = 0,200$ ).

Tabla 1. Efectos fijos y componentes de covarianza estimados del primer modelo

Parámetro	Efectos fijos			
	$\beta$	$\beta$ DS	$t$	$p$
Intercepto	17,14	0,42	40,81	<0,001
Tiempo	1,08	0,15	7,2	<0,001
Parámetro	Componentes de covarianza			
	$\beta$	$\beta$ DS	Wald Z	$p$
Residual	8,43	0,84	10,03	<0,001
NE (0,0)	14,30	2,77	5,16	<0,001
NE (0,1)	-1,04	0,82	-1,27	0,200
NE (1,1)	0,72	0,37	1,95	0,049

Nota. NE: Matriz de varianza-covarianza no-estructurada.

Con estos resultados se concluye que, dada la significación de la variación de los interceptos y las pendientes, cabe la posibilidad de que dicha variación sea debida, en parte, a la presencia de alguna variable asociada a los sujetos.

#### Modelo de crecimiento con variables asociadas las personas

El modelo de análisis del nivel entre-sujetos se expresa, si se toma como variable asociada a las personas la ESCOLARIZACIÓN, de la siguiente forma:

$$\begin{aligned} \pi_{0j} &= \beta_{00} + \beta_{01} \text{ESCOLARIZACIÓN}_j + u_{0j} \\ \pi_{1j} &= \beta_{10} + \beta_{11} \text{ESCOLARIZACIÓN}_j + u_{1j} \end{aligned} \quad (4)$$

donde los parámetros  $\beta_{01}$  y  $\beta_{11}$  explican la diferencia entre los conocimientos iniciales y la diferencia entre las tasas de crecimiento atribuibles a la escolarización. Sustituyendo en la ecuación (1), las ecuaciones (4) se obtiene el siguiente modelo de efectos mixtos:

$$\text{CONOCIMIENTO}_{ij} = [\beta_{00} + \beta_{10} \text{TIEMPO}_{ij} + \beta_{01} \text{ESCOLARIZACIÓN}_j] + \beta_{11} (\text{TIEMPO}_{ij} \times \text{ESCOLARIZACIÓN}_j) + [u_{0j} + u_{1j} \text{TIEMPO}_{ij} + e_{ij}] \quad (5)$$

En la parte fija se ha incluido la variable ESCOLARIZACIÓN con el propósito de reducir la variación producida por las diferencias entre-sujetos. Por lo que respecta a la parte aleatoria,  $u_{0j}$  representa el efecto aleatorio asociado a cada sujeto,  $u_{1j}$  TIEMPO<sub>ij</sub> la variación aleatoria de la interacción entre sujetos y tiempo, y  $e_{ij}$  la variación aleatoria de los datos entre los sujetos.

En la Tabla 2 se observa que el intercepto tiene un valor de 7,77 estadísticamente significativo ( $p = 0,005$ ), lo cual indica el conocimiento promedio sobre el consumo de drogas de los sujetos, controlando su nivel escolar. El valor del parámetro asociado a la escolarización es significativo ( $p = 0,001$ ). Por tanto, explica en parte el conocimiento que tienen los sujetos sobre el consumo de drogas. Así mismo, la variable tiempo también influye en el conocimiento global ( $p = 0,004$ ). La interacción tiempo por escolarización no alcanza el nivel significación ( $p = 0,057$ ), de lo que puede inferirse que el efecto de la escolarización no varía con el tiempo. Nótese que el valor estimado de la varianza residual es equivalente al primero modelo. En cuanto a las estimaciones restantes de los componentes de covarianza, constatamos que la inclusión de la escolarización ayuda a reducir el tamaño de la varianza del intercepto. Así,  $(14,30-12,28) / 14,30 = 0,14$  significa que hay una reducción del 14%. Es decir, la covariable explica el 14% de la variación en los interceptos. De igual modo, en cuanto a la varianza de las pendientes, al comparar los dos modelos tenemos que  $(0,72-0,63) / 0,72 = 0,125$ , es decir, un 12,5% de reducción.

Tabla 2. Efectos fijos y componentes de covarianza estimados del segundo modelo

Parámetro	Efectos fijos			
	$\beta$	$\beta$ DS	$t$	$p$
Intercepto	7,77	2,68	2,90	0,005
Tiempo	3,02	1,02	2,96	0,004
Escolarización	0,99	0,28	3,53	0,001
Escolarización x Tiempo	-0,20	0,11	-1,82	0,057
Parámetro	Componentes de covarianza			
	$\beta$	$\beta$ DS	Wald Z	$p$
Residual	8,42	0,84	10,02	<0,001
NE (0,0)	12,28	2,51	4,89	<0,001
NE (0,1)	-0,63	0,76	-0,83	0,411
NE (1,1)	0,63	0,37	1,70	0,084

Nota. NE: Matriz de varianza-covarianza no-estructurada.

### Modelo de curvas individuales con variables predictoras

A razón de los cambios observados, tanto intra como entre-sujetos, nos preguntamos cuáles son las variables predictoras capaces de explicar las diferencias de los perfiles entre los individuos. A continuación, reformulamos el modelo de conocimiento sobre el consumo de drogas, incorporando la variable INTERVENCIÓN y también la variable ASI\_ESCUELA. Así, el modelo (5) queda reformulado en los términos siguientes:

$$\text{CONOCIMIENTO}_{ij} = [\beta_{00} + \beta_{10} \text{TIEMPO}_{ij} + \beta_{01} \text{INTERVENCIÓN}_j + \beta_{02} \text{ASI\_ESCUELA}_j + \beta_{03} \text{ESCOLARIZACIÓN}_j] + [u_{0j} + u_{1j} \text{TIEMPO}_{ij} + e_{ij}] \quad (6)$$

En este tercer modelo se ha partido de una estructura de covarianza de los residuales intra autorregresiva de primer orden (AR1), por ser la de mejor ajuste, según el criterio de información de Akaike (AIC).

La Tabla 3 muestra los valores estimados de los efectos fijos y los parámetros de covarianza del modelo más completo del estudio, habiendo fijado como matriz de covarianza del error la estructura AR(1) y, al mismo tiempo, manteniendo la especificación de los interceptos y las pendientes como variables aleatorias. En este caso, el conocimiento promedio sobre el consumo de drogas de los sujetos, controlando su escolarización, intervención y ASI-escuela es 8,26, que en poco difiere del modelo anterior donde no se ha introducido la estructura intra a la matriz global de varianza-covarianza. El valor del parámetro asociado al tiempo es significativo ( $p < 0,001$ ), así como los efectos debidos a la intervención ( $p = 0,014$ ) y escolarización ( $p = 0,005$ ). No es significativo, en cambio, el efecto atribuible al factor ASI-escuela ( $p = 0,307$ ).

Los parámetros de covarianza son la varianza residual ( $\sigma^2 = 11,91$ ) y la correlación entre cada ocasión de observación con la siguiente ( $\rho = 0,30$ ), ambos significativos ( $p < 0,001$  y  $p = 0,005$ ). A su vez, los parámetros de los efectos aleatorios,  $\tau_{00} = 8,14$ ,  $\tau_{01} = 0,04$  y  $\tau_{11} = 0$  expresan, respectivamente, el grado de variación del intercepto, la covarianza entre los valores iniciales de conocimiento global y tasas de crecimiento y, por último, la variación entre las pendientes. De estos parámetros de efectos aleatorios sólo es significativo el primero ( $p = 0,003$ ), que indica que hay una gran variación entre los conocimientos iniciales que tienen los individuos con respecto a las drogas.

Tabla 3. Efectos fijos y componentes de covarianza estimados del tercer modelo

		Efectos fijos			
Parámetro		$\beta$	$\beta$ DS	$t$	$p$
Intercepto		8,26	2,65	3,12	0,002
Tiempo		1,09	0,16	6,81	<0,001
Intervención		1,72	0,69	2,49	0,014
ASI-escuela		-0,27	0,27	-1,00	0,307
Escolarización		0,69	0,24	2,87	0,005
		Componentes de covarianza			
Parámetro		$\beta$	$\beta$ DS	Wald Z	$p$
Residual	AR1	11,91	1,70	7,00	<0,001
Correlación	AR1	0,30	0,11	2,73	0,005
NE (0,0)		8,14	2,76	2,95	0,003
NE (0,1)		0,04	0,59	0,07	0,940
NE (1,1)		0,00 <sup>a</sup>	0,00	.	.

Notas. NE: Matriz de varianza-covarianza no-estructurada; AR1: Matriz de varianza-covarianza autorregresiva de primer orden.

<sup>a</sup> Parámetro de covarianza redundante.

## Discusión

A la vista de los resultados obtenidos de cada uno de los modelos sucesivos aplicados, se concluye que cuando sólo se tiene en cuenta el modelo simple de crecimiento, los individuos muestran un conocimiento sobre el consumo de drogas que crece significativamente a lo largo del tiempo. En el segundo modelo, la escolarización explicó, en parte, la variación del conocimiento que los sujetos tienen en torno al valor promedio en la población, así como en torno a las pendientes de crecimiento a través del tiempo. Por último, se reformuló el modelo de conocimiento sobre el consumo de drogas al considerar las variables intervención y ASI-escuela, además de la escolarización. En este tercer modelo, se demostró que tanto el efecto debido a la intervención como el efecto de la escolarización eran significativos, no en cambio el relativo al índice ASI-escuela.

Agradecimientos. Este trabajo ha sido financiado mediante sendos proyectos de investigación concedidos por el MCI (Ref.: PSI2009-11136/PSIC y PSI2008-03624).

## Referencias

- Fernández, P., Livacic-Rojas, P. y Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. *International Journal of Clinical and Health Psychology*, 7, 153-175.
- Gill, P.S. (2000). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, 19, 975-987.
- Kowalchuk, R.K., Keselman, H.J., Algina, J. y Wolfinger, R.D. (2004). The analysis of repeated measurements with mixed-model adjusted F tests. *Educational and Psychological Measurement*, 64, 224-242.
- MacKinnon, D.P. y Lockwood, C.M. (2003). Advances in statistical methods for substance abuse prevention research. (*Prevention Science*, 4) 155-171.
- Wolfinger, R.D. (1996). Heterogeneous variance-covariance structures for repeated measurements. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 205-230.

# SELECCIÓN DE MODELOS ANIDADOS CON DATOS DE MEDIDAS REPETIDAS USANDO CRITERIOS DE INFORMACIÓN Y PRUEBAS DE AJUSTE CONDICIONAL

Guillermo Vallejo<sup>1</sup>, Jaime Arnau<sup>2</sup>, Roser Bono<sup>2</sup> y Paula Fernández García<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Barcelona

Este trabajo examinó el desempeño del test razón de verosimilitud (LRT) condicional y de múltiples criterios de información (IC) para seleccionar estructuras de medias y de covarianzas anidadas. Los resultados pusieron de relieve que los IC eficientes se comportaban mejor que sus homólogos consistentes cuando las matrices de dispersión usadas en la generación eran complejas y peor cuando eran simples. El desempeño del test LRT resultó superior al de los IC cuando se basaba en el estimador de máxima verosimilitud completa (FML), e inferior cuando se basaba en el estimador máxima verosimilitud restringida (REML). También se encontró que la estrategia sugerida en la literatura estadística de usar el estimador REML para seleccionar la estructura de covarianza y el estimador FML para seleccionar la de medias debería ser evitada.

Cada vez son más las disciplinas que utilizan enfoques basados en la teoría del modelo lineal mixto (MLM) para analizar datos jerarquizados. Su aceptación se ha visto favorecida por su incorporación dentro de los principales paquetes estadísticos, incluyendo la función *lme* en R/SPLUS, el módulo *Proc Mixed* en SAS, o los comandos *Mixed* y *xtmixed* en SPSS y STATA. Con datos *de corte longitudinal* estos modelos resultan muy útiles, ya que permiten realizar inferencias acerca de la estructura de medias (EM) y modelar la estructura de covarianza (EC) en términos de efectos aleatorios y puro error. Este enfoque, más que asumir EC parcas (p.e., la matriz de simetría compuesta típica del enfoque ANOVA) o generales (p.e., la matriz no estructurada típica del enfoque MANOVA), trata de buscar un equilibrio entre los criterios de flexibilidad y simplicidad científica (Fitzmaurice, Laird & Ware, 2004). Especificando un modelo excesivamente simple se corre el riesgo de efectuar inferencias erróneas, debido a la subestimación de los errores estándar (ES), mientras que especificando un modelo excesivamente complejo se corre el riesgo de efectuar inferencias ineficientes.

Cuando se modela de forma efectiva la EC y también la EM, dado que la forma de primera depende de la elección que se haga de la segunda (Fitzmaurice et al., 2004), las estimaciones de los parámetros son más exactas y precisas. Vallejo, Ato y Valdés (2008) confirman la importancia de identificar el verdadero proceso generador de los datos (VPGD). De hecho, en el trabajo citado, las tasas de error se mantuvieron próximas al valor nominal elegido cuando el VPGD se especificaba correctamente; sin embargo, los ES resultaban sesgados cuando el VPGD se especificaba erróneamente.

Aunque la selección del VPGD resulte central para interpretar correctamente los datos, dicho objetivo no es fácil de lograr, pues para una misma evidencia muestral existen múltiples modelos candidatos. Para facilitar el modelado de la EC, algunos programas incorporan un completo menú de estructuras. Por ejemplo, *SAS Proc Mixed* permite ajustar modelos de simetría compuesta (SC), de Huynh-Feldt (HF), autorregresivos (AR), de media móvil (MA), ARIMA, antedependientes (ANTE) y no estructurados (NE). Para una exhaustiva descripción de estas estructuras, véase Littell et al (2006) y Zimmerman y Núñez-Antón (2009).

Existen diversos criterios para determinar la bondad de ajuste del modelo elegido durante el proceso de selección. Para comparar modelos anidados, el criterio más usado es el test de razón de verosimilitudes (LRT) con la desviación obtenida a partir de la función de máxima verosimilitud completa (FML) o de máxima verosimilitud restringida/residual (REML), según se trate de elegir entre modelos con idéntica EC o entre modelos con idéntica EM. A su vez, para comparar modelos no anidados se suelen emplear herramientas estadísticas menos formales, tales como los Criterio de Información eficientes (por ejemplo, AIC, AICC) y los Criterio de Información consistentes (por ejemplo, BIC, CAIC, HQIC). La mayor parte de los programas de ordenador que ajustan MLM incorporan los criterios AIC y BIC. Otros criterios de selección, tales como el coeficiente de determinación, el coeficiente de correlación de concordancia (CCC) o la suma de cuadrados residual de predicción (PRESS), han recibido escasa atención. Wang y Schaalje (2009) informan que el desempeño de los criterios predictivos CCC y PRESS no supera al desempeño de los criterios de información AIC y BIC.

En el contexto de los modelos no anidados y escenarios de investigación relativamente simples, Gurka (2006) encontró que los IC basados en el método REML elegían la verdadera EM tan bien o mejor que los IC basados en el método FML. En la literatura estadística especializada se defiende ajustar la EM vía FML exclusivamente. Gurka también halló que el desempeño de los IC eficientes basados en REML mejoraba cuando se excluía del estimador el término constante (en adelante  $REML_2$ ) (para detalles consúltese también Gurka & Edwards, 2008); en cambio, el desempeño los criterios consistentes mejoraba cuando se mantenía dicho término ( $REML_1$ ). Los hallazgos de Gurka (2006) afectan de lleno al proceso de selección, dado que ponen de relieve diversas inconsistencias existentes en la literatura, en lo referido a la estimación de parámetros y a las fórmulas usadas para calcular los IC.

Sería clarificador investigar el comportamiento de los IC, contemplando las mejoras analíticas referidas, cuando se manipula la distribución del término de error, la complejidad de las estructuras y el número modelos incluidos en el proceso de selección. Consiguientemente, el objetivo fundamental del presente trabajo se centra en determinar cuán efectivos son los criterios AIC, AICC, BIC, CAIC y HQIC para descubrir el VPGD en una familia de modelos anidados. Estos criterios serán evaluados bajo estimación FML,  $REML_1$  y  $REML_2$ . Además, para proporcionar un punto de referencia para la comparación también utilizaremos el criterio de ajuste condicional LRT.

### Método de la simulación

Para dar respuesta de los objetivos planteados, llevamos acabo un estudio de simulación en que manipulamos simultáneamente diversas EM y EC. Utilizamos sendos diseños crossover con dos tratamientos, dos secuencias y seis periodos/doce periodos, en los que se violaba la normalidad y homogeneidad. Los participantes del primer grupo recibieron la secuencia de tratamiento 3(A)3(B)/6(A)6(B), mientras que los del segundo grupo recibieron la secuencia inversa para contrarrestar los posibles efectos residuales. La evaluación fue realizada bajo estimación FML y  $REML_1/REML_2$  cuando se manipulaban las variables siguientes:

(a) *Tipo de modelo usado para generar los datos.* La evaluación del desempeño implicaba seleccionar de un conjunto de nueve modelos candidatos el VPGD. En la Tabla 1 aparecen recogidos los modelos utilizados en la comparación, así como el valor de los parámetros de efectos fijos usados para generar los datos. Examinando detalladamente la Tabla 1 se aprecia que los modelos estaban anidados unos dentro de otros. Cuando el número de efectos fijos era idéntico, como sucedía con los modelos  $M_3 - M_4$  y  $M_6 - M_7$ , las estructuras de covarianza diferían unas de otras y se hallaban anidadas entre sí, en el sentido que AR(1) es un caso especial de ARH(1) la cual es a su vez un caso especial de ANTE(1)

Tabla 1. Conjunto de modelos de medias y de covarianza candidatos y valor de los parámetros de efectos fijos

$M_1$	$E(y_{ij}) = \beta_{00}$	$Var(y_{ij}) = V_i [AR(1)]$
$M_2$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j$	$Var(y_{ij}) = V_i [AR(1)]$
$M_3$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot$	$Var(y_{ij}) = V_i [AR(1)]$
$M_4$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot$	$Var(y_{ij}) = V_i [ARH(1)]$
$M_5$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot + \beta_{11} G_j \times T_{ij} \cdot$	$Var(y_{ij}) = V_i [ARH(1)]$
$M_6$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot + \beta_{11} G_j \times T_{ij} \cdot + \beta_{20} CT_{ij} \cdot$	$Var(y_{ij}) = V_i [ARH(1)]$
$M_7$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot + \beta_{11} G_j \times T_{ij} \cdot + \beta_{20} CT_{ij} \cdot$	$Var(y_{ij}) = V_i [ANTE(1)]$
$M_8^{\circ}$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot + \beta_{11} G_j \times T_{ij} \cdot + \beta_{20} CT_{ij} \cdot + \beta_{21} G_j \times CT_{ij} \cdot$	$Var(y_{ij}) = V_i [ANTE(1)]$
$M_9$	$E(y_{ij}) = \beta_{00} + \beta_{01} G_j + \beta_{10} T_{ij} \cdot + \beta_{11} G_j \times T_{ij} \cdot + \beta_{20} T_{ij}^2 + \beta_{30} CT_{ij} \cdot + \beta_{31} G_j \times CT_{ij} \cdot$	$Var(y_{ij}) = V_i [ANTE(1)]$
$\beta^2 = [\beta_{00}=1.00 \quad \beta_{01}=1.25 \quad \beta_{10}=-0.50 \quad \beta_{11}=0.50 \quad \beta_{20}=-0.50 \quad \beta_{21}=0.50]$		
$\beta^3 = [\beta_{00}=1.00 \quad \beta_{01}=1.25 \quad \beta_{10}=-1.00 \quad \beta_{11}=1.00 \quad \beta_{20}=-1.00 \quad \beta_{21}=1.00]$		

Nota:  $M_1 \subset M_2 \subset M_3 \subset M_4 \subset M_5 \subset M_6 \subset M_7 \subset M_8 \subset M_9$ ;  $M_8^{\circ}$  = modelo usado para generar los datos.

(b) *Tamaño de muestra.* Los tamaños usados fueron:  $N = 30$  (10-20) y  $N = 60$  (20-40).

(c) *Número de medida repetidas:* Los periodos manipulados fueron:  $t = 6$  y  $t = 12$ .

(d) *Igualdad de las matrices de dispersión:* El desempeño de las herramientas de selección fue evaluado cuando las matrices de covarianza grupales eran homogéneas y también cuando eran heterogéneas. En el primer caso, los elementos de las dos matrices de dispersión fueron iguales entre sí ( $\Sigma_2 = \Sigma_1$ ) mientras que en el segundo caso, los elementos de una de las matrices fueron cinco veces mayores que los de la otra ( $\Sigma_2 = 5\Sigma_1$ ).

(e) *Valor de los parámetros: grandes y pequeños.* En la Tabla 1 se muestran los valores usados.

(f) *Forma de la distribución de la población.* En orden a investigar los efectos que la forma de la distribución tiene en la robustez de los enfoques, generamos datos desde distribuciones normales y no normales. Además de la distribución normal ( $\gamma_1 = 0; \gamma_2 = 0$ ), también fueron investigadas una simétrica con un moderado grado de curtosis ( $\gamma_1 = 0; \gamma_2 = 3$ ) y dos asimétricas. Una moderadamente sesgada ( $\gamma_1 = 2; \gamma_2 = 6$ ) y otra fuertemente sesgada ( $\gamma_1 = 6; \gamma_2 = 114$ ).

Para realizar los cálculos se utilizó un MACRO escrito en lenguaje SAS/IML (2008) y las condiciones del diseño fueron replicadas 5000 veces.

## Resultados

En la Tabla 2 que se muestra más abajo aparece tabulado el porcentaje de veces que los criterios examinados elegían correctamente la estructura de medias y de covarianza, tanto bajo estimación FML como REML. Los datos denotan el porcentaje promedio de elecciones correctas a través de las variables tamaño de muestra, igualdad de las matrices de dispersión y forma de la distribución. Globalmente, los resultados indican lo siguiente:

Tabla 2. Porcentaje de veces que los criterios elegían correctamente el modelo de efectos fijos y aleatorios bajo estimación FML y REML

ME Global	Criterio	$t = 6$		$t = 12$		
		$\beta'_{(1)}$	$\beta'_{(2)}$	$\beta'_{(1)}$	$\beta'_{(2)}$	
FML	AIC <sup>(SAS, SPSS)</sup>	38.94 <sup>1</sup>	67.01 <sup>2</sup>	73.92	79.44	<b>64.8<sup>2</sup></b>
FML	AICC <sub>1</sub>	08.62	24.53	20.75	24.70	<b>19.7</b>
FML	AICC <sub>2</sub> <sup>(SPSS, SAS)</sup>	33.59 <sup>2</sup>	64.31	74.74 <sup>3</sup>	83.22 <sup>3</sup>	<b>64.0</b>
FML	HQIC <sub>1</sub>	20.74	55.14	69.62	89.10	<b>58.7</b>
FML	HQIC <sub>2</sub> <sup>(SAS)</sup>	30.64	64.23 <sup>3</sup>	77.03 <sup>2</sup>	84.56	<b>64.1<sup>3</sup></b>
FML	BIC <sub>1</sub> <sup>(SPSS)</sup>	05.54	25.84	33.32	70.72	<b>33.9</b>
FML	BIC <sub>2</sub> <sup>(SAS)</sup>	17.17	50.51	70.91	89.84 <sup>2</sup>	<b>57.1</b>
FML	CAIC <sub>1</sub> <sup>(SPSS)</sup>	02.74	16.98	22.82	57.16	<b>24.9</b>
FML	CAIC <sub>2</sub> <sup>(SAS)</sup>	09.52	34.71	55.95	85.54	<b>46.4</b>
FML	LRT	31.89 <sup>3</sup>	73.75 <sup>1</sup>	87.93 <sup>1</sup>	99.16 <sup>1</sup>	<b>73.2<sup>1</sup></b>
REML <sub>1</sub>	AIC <sub>1</sub>	81.51 <sup>1</sup>	82.87	88.59	89.36	<b>85.6<sup>1</sup></b>
REML <sub>2</sub>	AIC <sub>2</sub> <sup>(SAS, SPSS)</sup>	74.21 <sup>3</sup>	87.27 <sup>1</sup>	92.68 <sup>1</sup>	98.93 <sup>1</sup>	<b>88.3<sup>2</sup></b>
REML <sub>1</sub>	AICC <sub>1</sub>	77.34 <sup>2</sup>	81.38	88.25	89.39	<b>84.1</b>
REML <sub>1</sub>	AICC <sub>2</sub>	42.99	54.64	44.14	44.48	<b>46.6</b>
REML <sub>2</sub>	AICC <sub>1</sub> <sup>(SAS, SPSS)</sup>	63.58	85.08 <sup>2</sup>	90.91 <sup>2</sup>	98.91 <sup>2</sup>	<b>84.6<sup>3</sup></b>
REML <sub>2</sub>	AICC <sub>2</sub>	37.22	56.31	46.10	49.43	<b>47.3</b>
REML <sub>1</sub>	HQIC <sub>1</sub>	59.98	74.79	79.13	89.59	<b>75.9</b>
REML <sub>1</sub>	HQIC <sub>2</sub>	73.21	81.04	87.47	89.22	<b>82.7</b>
REML <sub>2</sub>	HQIC <sub>1</sub>	51.50	76.42	81.90	98.32	<b>77.1</b>
REML <sub>2</sub>	HQIC <sub>2</sub> <sup>(SAS)</sup>	60.76	84.18 <sup>3</sup>	91.03 <sup>3</sup>	98.87 <sup>3</sup>	<b>83.7</b>
REML <sub>1</sub>	BIC <sub>1</sub>	32.28	47.20	42.69	80.73	<b>50.7</b>
REML <sub>1</sub>	BIC <sub>2</sub>	54.90	73.73	80.09	87.12	<b>73.9</b>
REML <sub>2</sub>	BIC <sub>1</sub> <sup>(SPSS)</sup>	27.13	47.99	42.52	87.77	<b>51.4</b>
REML <sub>2</sub>	BIC <sub>2</sub> <sup>(SAS)</sup>	46.73	73.11	80.98	98.79	<b>74.9</b>
REML <sub>1</sub>	CAIC <sub>1</sub>	23.35	43.77	32.04	73.29	<b>43.1</b>
REML <sub>1</sub>	CAIC <sub>2</sub>	40.34	59.66	65.73	88.20	<b>63.5</b>
REML <sub>2</sub>	CAIC <sub>1</sub> <sup>(SPSS)</sup>	20.16	35.62	32.11	79.41	<b>41.8</b>
REML <sub>2</sub>	CAIC <sub>2</sub> <sup>(SAS)</sup>	34.51	58.58	63.99	97.62	<b>63.7</b>
REML <sub>2</sub>	LRT	28.33	71.11	79.59	96.89	<b>68.9</b>

Nota: Los datos denotan el porcentaje promedio de elecciones correctas a través del tamaño de muestra, forma de la distribución e igualdad de las matrices de dispersión.  $\beta'_{(1)} = [\beta_{00}=1.00 \beta_{01}=1.25 \beta_{10}=-0.50 \beta_{11}=0.50 \beta_{21}=-0.50 \beta_{21}=0.50]$ ;  $\beta'_{(2)} = [\beta_{00}=1.00 \beta_{01}=1.25 \beta_{10}=-1.00 \beta_{11}=1.00 \beta_{21}=-1.00 \beta_{21}=1.00]$ ,  $t$  = número de periodos de observación. La última columna representa el porcentaje promedio a través de los tres experimentos.

1. La ejecución de los criterios examinados dependía del método de estimación, valor de los parámetros y número de medidas repetidas. En la Tabla 2 aparece recogido, pero el tamaño de muestra también afectaba sustancialmente a la selección del VPGD. Sin embargo, la forma de la distribución e igualdad de las matrices de dispersión lo hacían moderadamente.
2. El desempeño de los IC fue mejor bajo estimación REML que bajo estimación FML. Promediando a través de las 1052 ( $64 \times 18$ ) condiciones manipuladas, el porcentaje de aciertos obtenidos vía REML fue del 69.7% (del cual el 65.2% corresponde a los IC consistentes y el 74.2% a los IC eficientes), mientras que el obtenido vía FML promediando a través de 576 ( $64 \times 9$ ) condiciones

manipuladas fue del 55.9% (del cual el 47.5% corresponde a los IC consistentes y el 64.3% a los IC eficientes).

3. Por su parte, el LRT eligió el VPGD en el 73.2% de las veces bajo FML y en el 68.9% bajo REML. Cuando el método de estimación usado era FML y  $t = 6$ , las diferencias existentes entre los IC y el LRT no excedían los 2 puntos porcentuales. Sin embargo, bajo estimación REML las diferencias favorecían a los IC y eran superiores a los 20 puntos. La situación se invertía cuando  $t = 12$ . En este caso las diferencias excedían los 10 puntos porcentuales.
4. El desempeño de los IC mejoraba si el estimador REML incluía el término constante, sobre todo, cuando el número de medidas repetidas era moderado.
5. El desempeño de los IC era superior cuando se usaba el número total de sujetos (nivel2), en vez del número total de observaciones (nivel1). Este hallazgo, apoya la estrategia seguida en el módulo *Proc Mixed* del SAS, como opuesta a la seguida en el comando *Mixed* del SPSS.

### Conclusión

Globalmente, los IC eficientes trabajaban mejor que los IC consistentes cuando la estructura de covarianza era compleja y, viceversa, cuando era sencilla. Los IC consistentes tendían a seleccionar modelos más parcos, generalmente de carácter estacionario, que los IC eficientes. En los estudios aplicados resulta habitual que la varianza de las observaciones sea heterogénea y que la correlación entre las mismas decrezca a lo largo del tiempo, de ahí que nos decantemos por el empleo de IC eficientes, en particular del AIC basado en el estimador  $REML_1$ , pues cumple mejor el objetivo de encontrar un equilibrio entre un modelo complejo y otro parco. Hecha esta recomendación, debemos advertir que los resultados son limitados a las condiciones examinadas, si bien conjeturamos que pueden ser generalizadas a un rango más amplio de condiciones; por ejemplo, a situaciones donde los modelos no se hallen anidados unos dentro de otros. Finalmente, en la investigación realizada el VPGD siempre pertenecía a la familia de modelos investigados. No obstante, cuando se trabaja con datos reales desconocemos si el VPGD pertenece a la clase de modelos considerados. Por lo tanto, sería deseable realizar una investigación donde el objetivo fuese comparar los IC en términos de seleccionar el modelo más próximo al VPGD, dado que éste no se haya incluido en el conjunto de modelos presentes en la comparación.

Agradecimientos. Este trabajo ha sido financiado mediante sendos proyectos de investigación concedidos por el MCI (Ref.: PSI2008-03624 y PSI2009-11136/PSIC).

### Referencias

- Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley.
- Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician*, 60, 19-26.
- Gurka, M. J. & Edwards, L. J. (2008). Mixed models. En C. R. Rao, J. P. Miller & D. C. Rao (Eds.): *Handbook of Statistics, Vol 27, Epidemiological and Medical Statistics* (pp. 253-280). New York: Elsevier Science.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. & Schabenberger, O. (2006). *SAS System for Mixed Models* (2nd edition). Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2008). *SAS/STAT® Software: Version 9.2*. SAS Institute Inc., Cary, NC.

Vallejo, G., Ato, M. & Valdés, T. (2008). Consequences of misspecifying the error covariance structure in linear mixed models for longitudinal data. *Methodology*, 4, 10-21.

Wang, J. & Schaalje, G. B. (2009). Model selection for linear mixed models using predictive criteria. *Communications in Statistics - Simulation and Computation*, 38, 788-801.

Zimmerman, D. L. & Núñez-Antón, V. (2009). *Antependence Models for Longitudinal Data*. London: Chapman & Hall/CRC.

# COMPARACIÓN DE CINCO PROCEDIMIENTOS PARA EVALUAR LA EFECTIVIDAD DE INTERVENCIONES EN DISEÑOS DE CASO ÚNICO

Rumen Manolov, Jaume Arnau, Antonio Solanas y Roser Bono  
Universidad de Barcelona

Para el análisis de datos provenientes de diseños de caso único se ha propuesto una gran variedad de técnicas (e.g., inspección visual, pruebas de aleatorización, ARIMA, índices de tamaño del efecto). En el presente estudio se compara el análisis de la regresión basado en mínimos cuadrados ordinarios con técnicas que incorporan métodos de control de variables de confundido como tendencia general y autocorrelación con un procedimiento de cálculo fácil y relacionado con el análisis visual. Se generan condiciones experimentales mediante simulación incorporando factores como la dependencia serial, tendencia general y diferentes tipos de efectos. Los resultados muestran que en presencia de autocorrelación o tendencia general en los datos hay correspondencia entre las tasas nominales y empíricas de error Tipo I solamente para algunos procedimientos. Las técnicas menos afectadas por estas características de los datos resultan ser las menos sensibles a efectos existentes del tratamiento. Es necesario modificar los procedimientos para mejorar su rendimiento y explorar las propiedades estadísticas de estas nuevas técnicas.

Para los datos obtenidos mediante diseños de caso único todavía no hay consenso sobre en qué deberían basarse las decisiones sobre la efectividad de los tratamientos. Históricamente, una primera propuesta se centra en el análisis visual (Johnston y Pennypacker, 2008; Parsonson y Baer, 1986) y este es el procedimiento aplicado con mayor frecuencia en las últimas décadas (Kratochwill y Brody, 1978; Parker, Hagan-Burke y Vannest, 2007). Sin embargo, se ha encontrado que mediante la inspección visual es probable que se omitan efectos existentes (Ottenbacher, 1990) y que se detecten efectos inexistentes (Fisch, 2001), además de la conocida falta de concordancia entre jueces (DeProspero & Cohen, 1979). El problema de las tasas de error Tipo I o Tipo II demasiado elevadas también se ha observado para procedimientos como el ANOVA (Scheffé, 1959; Toothaker, Banz, Noble, Camp y Davis, 1983) y las pruebas de aleatorización (Ferron y Ware, 1995; Sierra, Solanas y Quera, 2005). Los modelos ARIMA, que controlan la autocorrelación de forma explícita (Harrop y Velicer, 1985), requieren una gran cantidad de datos y además pueden llegar a ser excesivamente complejos para psicólogos no expertos en temas estadísticos. Además, algunas versiones simplificadas de ARIMA parecen no ser adecuadas (Huitema, 2004; Huitema, McKean y Laraway, 2007).

En este estudio se escogieron procedimientos relativamente simples y que pueden ser implementados mediante programas informáticos ampliamente disponibles. Cuatro de las técnicas estudiadas se basan en el análisis de la regresión. Tres de ellas implican corrección de los datos – eliminación de la autocorrelación (en mínimos cuadrados generalizados; Simonton, 1977) o eliminación de la tendencia (a través de una regresión previa o a través de diferenciación; Gorsuch, 1983). El quinto procedimiento incluido tiene una base diferente – la cantidad de solapamiento que hay entre los datos (Scruggs, Mastropieri y Casto, 1987).

## Método

### *Generación de datos*

Los procedimientos se compararon en el contexto de condiciones experimentales definidas por simulación Monte Carlo, estudiando tres longitudes de serie, con énfasis series cortas, de 10, 15 ó 20 observaciones. El modelo de generación de datos permite establecer si los datos presentan tendencia, cambio de nivel y de pendiente tal y como se propuso en Huitema y McKean (2007):  $y_t = \beta_0 + \beta_1 \cdot T_t + \beta_2 \cdot D_t + \beta_3 \cdot SC_t + e_t$ . En la literatura científica todavía existen directrices claras sobre los valores concretos que deberían tener los parámetros beta. En este estudio, se han escogido tamaños de efecto que puedan ser detectados por los procedimientos, teniendo en cuenta que un cambio de pendiente progresivo de uno reflejaría un tratamiento más potente que un cambio de nivel constante de uno.

En el término de error se especificaron varios niveles de autocorrelación de primer orden mediante la expresión  $\varepsilon_t = \phi_1 \cdot \varepsilon_{t-1} + u_t$ , concretamente los más comunes en datos reales:  $\phi_1 = -.3(.3).9$ . La variable aleatoria  $u_t$  se generó según tres distribuciones: exponencial, normal y uniforme (debido a que representan diferentes grados de asimetría y kurtosis), controlando que en cada caso la media y la desviación estándar sean las mismas, a través de los momentos de primer y segundo orden.

### *Análisis de datos*

Para los procedimientos basados en la regresión se utilizaron dos modelos – uno que especificaba cambio de nivel y otro cambio de pendiente. Estos modelos se aplicaron con o sin corrección de los datos originales, según la técnica. En el caso de datos sin efecto de tratamiento se usaron los dos modelos, mientras que en presencia de efecto sólo se aplicó el modelo correcto: el que incluía el efecto simulado. Un error Tipo I se definió como un coeficiente de regresión estadísticamente significativo en ausencia de efecto. La detección de efectos existentes también se basó en la significación del coeficiente de regresión. En ambos casos la hipótesis nula se basaba en un alfa de 5%.

En el caso de PND no fue posible estimar tasas de error Tipo I y Tipo II propiamente dichos, debido a que no se dispone de p valores. En cambio, un resultado significativo fue definido como un porcentaje por encima del 90%, como los creadores del procedimiento sugirieron (Scruggs y Mastropieri, 1998). Tanto en el caso de los procedimientos de regresión, como del PND, las estimaciones se hicieron en base a diez mil iteraciones.

## Resultados

En series independientes y sin tendencia, las tasas de falsas alarmas concuerdan con el alfa nominal para casi todos los procedimientos de regresión, excepto el *Trend Analysis*. De hecho, la corrección de los datos que realiza el *Trend Analysis* hace que la hipótesis nula no se rechace casi nunca. El PND producía porcentajes mayores que el 90 en muy pocas ocasiones. Estos resultados son comunes para las diferentes longitudes de serie y distribuciones de la variable aleatoria. Cuando se introduce autocorrelación en los datos, los procedimientos que estiman a través de mínimos cuadrados ordinarios o generalizados muestran tasas de error Tipo I inaceptables. El control de la dependencia serial mejora en series más largas, pero sigue siendo insuficiente. El *Differencing Analysis* mantiene la tasa de falsas alarmas muy cercana al 5% nominal y el PND tampoco parece verse afectada por la dependencia serial.

En el caso de la tendencia general, los resultados son parecidos a los anteriormente comentados. Los procedimientos más afectados son los mínimos cuadrados ordinarios y generalizados. En el caso

de esta última técnica, cuando se dispone de más de diez observaciones, el efecto de la tendencia se elimina para el modelo que especifica cambio de nivel. Los porcentajes del PND también incrementan en presencia de tendencia general. El *Differencing Analysis* es la técnica que mantiene tasas aceptables de falsas alarmas.

La única comparación con sentido que se puede realizar en términos de potencia estadística es para el caso en que las tasas de error Tipo I son controladas (i.e., en ausencia de autocorrelación y dependencia serial). La estimación por mínimos cuadrados es la que conlleva mayor sensibilidad a efectos de la intervención, sobre todo cuando se trata de un cambio de pendiente, de acuerdo con los parámetros de la simulación. El procedimiento que mostraba un mejor comportamiento en cuanto a las falsas alarmas, resultó tener potencia bastante baja – detectaba efectos alrededor del 10% de las ocasiones.

### Discusión

La combinación de los resultados sobre las tasas de error Tipo I y II sugiere que los procedimientos de regresión estudiados podrían ser poco apropiados en series cortas de  $N = 1$ . En el caso del control de la tendencia, parece que éste implica una sobrecorrección de los datos y se elimina no sólo la tendencia, sino también los efectos del tratamiento. En cuanto al control de la autocorrelación, tal vez la contrastación y eliminación de la dependencia serial deberían ser iterativas en vez de aplicarse una sola vez. El PND tampoco parece ser la solución a los problemas de los procedimientos basados en la regresión, sobre todo debido a la distorsión introducida por la tendencia.

Por otra parte, los resultados variaban mínimamente en función de la distribución de la variable aleatoria. En cualquier caso, consideramos que los datos no deberían generarse únicamente utilizando la distribución normal como se suele hacer, para que los resultados puedan generalizarse a una cantidad mayor de situaciones. Sin embargo, las condiciones experimentales estudiadas en este trabajo podrían haber sido incluso más amplias, generando datos también a partir de procesos de medias móviles y no sólo autorregresivos. Además, se deberían estudiar diferentes tipos de efecto de tratamiento (e.g., demorados y transitorios) y otras estructura de diseño. Sobre todo, hay que remarcar que no se han estudiado todas las técnicas basadas en regresión, sino sólo las que se consideraron más fáciles de aplicar e interpretar por parte de los profesionales.

Estudios futuros deberían mejorar el funcionamiento de los procedimientos basados en regresión. Una opción es utilizar las estimaciones de los coeficientes de regresión en vez de su significación estadística como criterio para contrastar la efectividad de un tratamiento. Por lo tanto, habría que explorar la calidad de estas estimaciones, por ejemplo, a través de error cuadrático medio. También se debería contrastar la utilidad de técnicas con otro tipo de fundamento, siempre teniendo en cuenta que los procedimientos no deberían ser excesivamente complicados. Otro aspecto que podría potenciar el uso de los procedimientos es su disponibilidad en programas informáticos.

### Referencias

- DeProspero, A. y Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, 12, 573-579.
- Ferron, J. y Ware, W. (1995). Analyzing single-case data: The power of randomization tests. *The Journal of Experimental Education*, 63, 167-178.

- Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes*, 54, 137-154.
- Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, 5, 141-154.
- Harrop, J. W. y Velicer, W. F. (1985). A comparison of alternative approaches to the analysis of interrupted time-series. *Multivariate Behavioral Research*, 20, 27-44.
- Huitema, B. E. (2004). Analysis of interrupted time-series experiments using ITSE: A critique. *Understanding statistics*, 3, 27-46.
- Huitema, B. E. y McKean, J. W. (2007). An improved portmanteau test for autocorrelated errors in interrupted time-series regression models. *Behavior Research Methods*, 39, 343-349.
- Huitema, B. E., McKean, J. W. y Laraway, S. (2007). Time series intervention analysis using ITSACORR: Fatal flaws. *Journal of Modern Applied Statistical Methods*, 6, 367-379.
- Johnston, J. M. y Pennypacker, H. S. (2008). *Strategies and tactics of behavioral research* (3rd ed.). New York, NJ: Routledge.
- Kratochwill, T. R. y Brody, G. H. (1978). Single subject designs: A perspective on the controversy over employing statistical inference and implications for research and training in behavior modification. *Behavior Modification*, 2, 291-307.
- Ottenbacher, K. J. (1990). When is a picture worth a thousand p values? A comparison of visual and quantitative methods to analyze single subject data. *Journal of Special Education*, 23, 436-449.
- Parker, R. I., Hagan-Burke, S. y Vannest, K. (2007). Percentage of all non-overlapping data: An alternative to PND. *Journal of Special Education*, 40, 194-204.
- Parsonson, B. S. y Baer, D. M. (1986). The graphic analysis of data. En A. Poling y R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 157-186). New York: Plenum Press.
- Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.
- Scruggs, T. E. y Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification*, 22, 221-242.
- Scruggs, T. E., Mastropieri, M. A. y Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, 8, 24-33.
- Sierra, V., Solanas, A. y Quera, V. (2005). Randomization tests for systematic single-case designs are not always appropriate. *The Journal of Experimental Education*, 73, 140-160.
- Simonton, D. K. (1977a). Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychological Bulletin*, 84, 489-502.
- Toothaker, L. E., Banz, M., Noble, C., Camp, J. y Davis, D. (1983). N = 1 designs: The failure of ANOVA-based tests. *Journal of Educational Statistics*, 4, 289-309.

# ANÁLISIS DE DATOS LONGITUDINALES EN UN MODELO DE SERIES TEMPORALES AGRUPADAS MULTINIVEL

Jesús F. Rosel<sup>1</sup>, Pedro Torrente<sup>1</sup>, Enrique F. Maldonado<sup>2</sup> y Juan J. Canales<sup>3</sup>

<sup>1</sup>Universidad 'Jaume I' de Castellón

<sup>2</sup>Universidad de Málaga

<sup>3</sup>Instituto de Investigación Cavanilles, Universidad de Valencia

El objeto del presente trabajo es comprobar qué modelo Box-Jenkins de serie temporal sigue un conjunto de datos temporales agrupados. A 19 participantes normales (9 hombres y 10 mujeres) se les analiza el logaritmo de una variable salivar (NS). Para el análisis exploratorio, se hizo una función de autocorrelación simple y parcial (ACF y PACF) de los datos; el modelo sale AR(1), con constante aleatoria significativa, pero hay un error de estimación al no 'separar' los datos de cada participante. Se expone cómo calcular la ACF y la PACF de manera correcta. En el caso de hacer la estimación dejando los huecos entre los participantes (correctamente), da un modelo AR(3), con constante aleatoria no significativa.

Por tanto, la NS sigue un proceso mediante el cual, para predecir cuál será el nivel de la misma en un momento determinado, se ha de conocer cuál fue el nivel de NS en la hora anterior, dos horas y tres horas antes. Se insiste en la importancia de: a) organizar los datos adecuadamente, b) calcular correctamente las ACF y PACF de los datos obtenidos, y b) en el análisis de los residuales.

Cada vez será más frecuente el análisis de datos 'híbridos' en psicología, en el sentido de que se utilizará metodología de análisis de diferentes tipos dentro de un mismo sistema, en nuestro caso utilizaremos 'análisis de series temporales agrupadas, con regresión multinivel', y aunque la teoría está bien asentada (Hedeker & Gibbons, 2006; Walls & Schafer, 2006), el analista ha de prestar atención a aspectos relacionados con: (i) la organización de los datos, (ii) los supuestos de los modelos utilizados y (iii) los análisis de los residuales de los diferentes modelos implicados (en nuestro caso: series temporales y regresión multinivel). Analizaremos unos datos reales para indicar los errores comunes que pueden cometerse y su sistema de solución.

El objeto del presente estudio es analizar qué modelo estadístico de serie temporal sigue un conjunto de datos temporales agrupados. La muestra está constituida por 19 participantes normales (9 hombres y 10 mujeres), que asistieron cada uno por separado al laboratorio en días diferentes, a los que se midió la variable 'el logaritmo neperiano de la actividad enzimática de la alfa-amilasa salivar (NS)' cada hora (entre las 9 horas y las 21 horas del mismo día, es decir, con 13 mediciones en total para cada participante).

Se parte del supuesto de que todos los participantes siguen el mismo proceso temporal de NS, teniendo cada uno de ellos sus peculiaridades (en la media y en los coeficientes), por lo que se plantea la hipótesis de que los datos siguen un modelo autorregresivo y multinivel, siendo nivel 1: NS, nivel 2: participante (Goldstein, 2003; Hoepfner, Goodwin, Velicer & Heltshe, 2007; Moore, Osgood, Larzelere & Chamberlain, 1994; Muthén & Asparouhov, 2009; Plewis, 2005; Raudenbush & Bryk, 2002; Sun & Li, 2002; Soliday, Moore & Lande, 2002; Velicer & Colby, 1997; Velicer & Fava, 2003).

## Método

*Participantes*

Se ha medido a 19 estudiantes voluntarios en un laboratorio de psicobiología de la Universidad de Málaga (9 hombres y 10 mujeres), cada uno de ellos gozaba de buena salud y no fumaba.

*Procedimiento*

La técnica de recogida de datos utilizada se conoce como protocolo de rutina constante. Cada hora (de las 9 a las 21 hr) se les tomaba una muestra de saliva que era posteriormente analizada para calcular su NS.

Los datos analizados se presentan en la Tabla 1, en la que puede apreciarse cómo la actividad media de la NS sigue una tendencia ascendente y con varianza estable (Nater, Rohleder, Schlotz, Ehlert & Kirschbaum, 2007; Rohleder, Wolf, Maldonado & Kirschbaum, 2006); habiendo un total de 247 registros (13 para cada uno de los 19 participantes).

Tabla 1

*Datos de NS para cada participante, según la hora. Se añaden las medias y desviaciones típicas por participante ( $\overline{NS}_{j,\bullet}$  y  $s_{j,\bullet}$ ) y por hora ( $\overline{NS}_{\bullet,j}$  y  $s_{\bullet,j}$ )*

Participante	9 horas	10 horas	11 horas	12 horas	13 horas	14 horas	15 horas	16 horas	17 horas	18 horas	19 horas	20 horas	21 horas	$\overline{NS}_{j,\bullet}$	$s_{j,\bullet}$
1	3.18	3.66	4.38	4.69	4.91	4.53	3.91	4.80	4.83	4.57	4.83	4.86	5.31	4.50	0.59
2	2.48	3.56	4.08	3.69	3.09	3.89	4.11	4.49	4.39	3.09	4.39	3.89	3.99	3.78	0.59
3	0.69	0.00 <sup>a</sup>	0.00 <sup>a</sup>	1.39	1.39	0.69	1.39	1.10	1.39	1.95	1.39	1.79	2.08	1.17	0.66
4	3.04	3.50	3.69	3.85	4.42	4.67	4.65	5.38	5.16	4.30	5.16	5.20	4.69	4.44	0.73
5	3.64	3.69	4.66	4.92	4.74	4.11	4.34	4.39	4.32	4.67	4.32	4.84	3.47	4.32	0.47
6	1.95	1.95	3.14	2.94	2.64	2.89	3.26	2.64	2.77	3.40	2.77	3.83	4.33	2.96	0.66
7	2.83	3.00	3.00	2.77	2.83	3.00	3.30	3.58	3.22	3.43	3.22	3.18	3.00	3.10	0.24
8	3.66	4.49	3.89	3.76	2.94	4.37	4.84	5.27	3.95	4.30	3.95	4.38	4.06	4.15	0.58
9	1.39	1.61	2.94	2.56	3.33	3.22	3.33	2.48	2.20	4.04	2.20	3.99	4.44	2.90	0.94
10	3.69	3.18	3.91	3.26	4.04	3.53	4.29	4.41	3.83	3.91	3.83	4.47	4.55	3.91	0.44
11	2.20	2.89	2.77	3.40	3.26	2.30	3.53	4.22	3.99	4.64	3.99	4.16	4.16	3.50	0.78
12	3.40	4.55	5.25	5.63	5.14	5.57	5.81	5.78	5.75	5.07	5.75	4.97	5.21	5.22	0.67
13	2.30	3.09	2.48	3.37	2.89	4.28	3.95	3.47	4.23	2.83	4.23	4.74	5.32	3.63	0.91
14	3.64	3.66	4.30	3.37	4.19	3.58	3.93	4.81	4.54	5.00	4.54	5.42	3.18	4.17	0.68
15	3.47	2.20	2.48	3.50	2.20	3.50	3.33	3.04	2.08	3.00	2.08	2.30	4.84	2.92	0.81
16	2.40	2.20	2.08	2.20	2.83	2.56	2.48	2.83	2.89	3.09	2.89	2.64	3.09	2.63	0.34
17	2.48	3.53	4.17	4.16	4.28	4.03	4.37	4.55	4.50	4.72	4.50	4.93	4.67	4.22	0.63
18	3.64	3.37	4.84	4.39	4.85	3.87	5.42	4.84	5.91	4.16	5.91	5.00	5.09	4.72	0.80
19	4.29	4.74	4.74	4.36	4.23	4.34	4.65	5.12	5.24	4.96	5.24	5.15	5.00	4.78	0.37
$\overline{NS}_{\bullet,j}$	2,86	3,10	3,52	3,59	3,59	3,63	3,94	4,06	3,96	3,96	3,96	4,20	4,24		
$s_{\bullet,j}$	0,91	1,13	1,24	1,00	1,04	1,06	1,01	1,20	1,25	0,88	1,25	1,04	0,91		

<sup>a</sup> Los valores 0 significan que el valor original bruto de NS es de 1 (Ln 1 = 0).

## Resultados

Los datos serán analizados mediante dos procedimientos, que llamaremos Análisis de datos 1 y 2, respectivamente.

Análisis de datos 1

Los datos de la Tabla 1 se han organizado mediante un sistema de registro consistente en ‘participante por fila’, pero para un análisis autorregresivo y multinivel, hemos cambiado su configuración a un sistema en el cual cada registro es llevado a filas separadas, mediante un sistema de ‘dato por registro’ (Singer & Willett, 2003), tal como se muestra en la Figura 1.

1	Hora	Participant	NS	NS_1	NS_2	NS_3	NS_4	NS_5	NS_6	NS_7	NS_8	NS_9	NS_10	NS_11	NS_12
2	9	1	3,18												
3	10	1	3,66	3,18											
4	11	1	4,38	3,66	3,18										
5	12	1	4,69	4,38	3,66	3,18									
6	13	1	4,91	4,69	4,38	3,66	3,18								
7	14	1	4,53	4,91	4,69	4,38	3,66	3,18							
8	15	1	3,91	4,53	4,91	4,69	4,38	3,66	3,18						
9	16	1	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18					
10	17	1	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18				
11	18	1	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18			
12	19	1	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18		
13	20	1	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18	
14	21	1	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18
15	9	2	2,48	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66
16	10	2	3,56	2,48	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38
17	11	2	4,08	3,56	2,48	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69
18	12	2	3,69	4,08	3,56	2,48	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91
19	13	2	3,09	3,69	4,08	3,56	2,48	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53
20	14	2	3,89	3,09	3,69	4,08	3,56	2,48	5,31	4,86	4,83	4,57	4,83	4,80	3,91
21	15	2	4,11	3,89	3,09	3,69	4,08	3,56	2,48	5,31	4,86	4,83	4,57	4,83	4,80
22	16	2	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48	5,31	4,86	4,83	4,57	4,83
23	17	2	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48	5,31	4,86	4,83	4,57
24	18	2	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48	5,31	4,86	4,83
25	19	2	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48	5,31	4,86
26	20	2	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48	5,31
27	21	2	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48
28	9	3	0,69	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56
29	10	3	0,00	0,69	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08
30	11	3	0,00	0,00	0,69	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69
31	12	3	1,39	0,00	0,00	0,69	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09
32	13	3	1,39	1,39	0,00	0,00	0,69	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89
33	14	3	0,69	1,39	1,39	0,00	0,00	0,69	3,99	3,89	4,39	3,09	4,39	4,49	4,11
34	15	3	1,39	0,69	1,39	1,39	0,00	0,00	0,69	3,99	3,89	4,39	3,09	4,39	4,49
35	16	3	1,10	1,39	0,69	1,39	1,39	0,00	0,00	0,69	3,99	3,89	4,39	3,09	4,39
36	17	3	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,00	0,69	3,99	3,89	4,39	3,09
37	18	3	1,95	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,00	0,69	3,99	3,89	4,39
38	19	3	1,39	1,95	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,00	0,69	3,99	3,89
39	20	3	1,79	1,39	1,95	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,00	0,69	3,99
40	21	3	2,08	1,79	1,39	1,95	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,00	0,69
41	9	4	3,04	2,08	1,79	1,39	1,95	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,00
42	10	4	3,50	3,04	2,08	1,79	1,39	1,95	1,39	1,10	1,39	0,69	1,39	1,39	0,00
43	11	4	3,69	3,50	3,04	2,08	1,79	1,39	1,95	1,39	1,10	1,39	0,69	1,39	1,39
44	12	4	3,85	3,69	3,50	3,04	2,08	1,79	1,39	1,95	1,39	1,10	1,39	0,69	1,39
45	13	4	4,42	3,85	3,69	3,50	3,04	2,08	1,79	1,39	1,95	1,39	1,10	1,39	0,69
46	14	4	4,67	4,42	3,85	3,69	3,50	3,04	2,08	1,79	1,39	1,95	1,39	1,10	1,39
47	15	4	4,65	4,67	4,42	3,85	3,69	3,50	3,04	2,08	1,79	1,39	1,95	1,39	1,10
48	16	4	5,38	4,65	4,67	4,42	3,85	3,69	3,50	3,04	2,08	1,79	1,39	1,95	1,39
49	17	4	5,16	5,38	4,65	4,67	4,42	3,85	3,69	3,50	3,04	2,08	1,79	1,39	1,95
50	18	4	4,30	5,16	5,38	4,65	4,67	4,42	3,85	3,69	3,50	3,04	2,08	1,79	1,95
51	19	4	5,16	4,30	5,16	5,38	4,65	4,67	4,42	3,85	3,69	3,50	3,04	2,08	1,79
52	20	4	5,20	5,16	4,30	5,16	5,38	4,65	4,67	4,42	3,85	3,69	3,50	3,04	2,08
53	21	4	4,69	5,20	5,16	4,30	5,16	5,38	4,65	4,67	4,42	3,85	3,69	3,50	3,04
54	9	5	3,64	4,69	5,20	5,16	4,30	5,16	5,38	4,65	4,67	4,42	3,85	3,69	3,50
55	10	5	3,69	3,64	4,69	5,20	5,16	4,30	5,16	5,38	4,65	4,67	4,42	3,85	3,69
56	11	5	4,66	3,69	3,64	4,69	5,20	5,16	4,30	5,16	5,38	4,65	4,67	4,42	3,85
57	12	5	4,92	4,66	3,69	3,64	4,69	5,20	5,16	4,30	5,16	5,38	4,65	4,67	4,42
58	13	5	4,74	4,92	4,66	3,69	3,64	4,69	5,20	5,16	4,30	5,16	5,38	4,65	4,67
59	14	5	4,11	4,74	4,92	4,66	3,69	3,64	4,69	5,20	5,16	4,30	5,16	5,38	4,65
60	15	5	4,34	4,11	4,74	4,92	4,66	3,69	3,64	4,69	5,20	5,16	4,30	5,16	5,38

Figura 1. Datos organizados por registro y retardados, para cada participante

La correlación intraclase (ICC), es .662, lo cual indica que aproximadamente el 66% de la variabilidad de los resultados en NS es debido a la variabilidad entre los participantes.

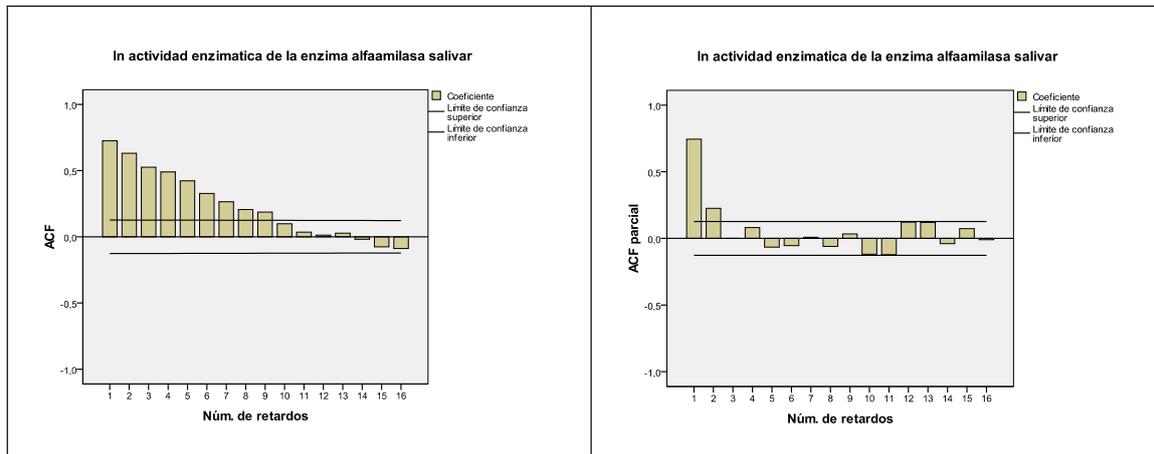


Figura 2. ACF y PACF de los datos de la Figura 1

Al hacer la ACF y PACF (Figura 2), se ha comprobado que los datos pueden seguir un modelo autorregresivo de segundo orden (AR(2)). Se ha llevado a cabo una regresión multinivel en la que la DV:  $NS_{j,t}$ , y como variable independiente (IV): NS la hora anterior ( $NS_{j,t-1}$ ), y de dos horas antes ( $NS_{j,t-2}$ ).

Al efectuar la regresión multinivel, se comprueba que el modelo que mejor ajusta es el de la ecuación correspondiente a la Tabla 2, al que a partir de ahora llamaremos M1. El modelo de la Tabla 2 indica que los datos siguen un proceso en el cual la constante tiene un valor fijo para todos los individuos de la muestra (2.731) y también un efecto aleatorio propio para cada individuo (su desviación respecto de la media), siendo el valor autorregresivo de  $NS_{j,t-1}$  fijo para todos los participantes (.271).

Al hacer la ACF y la PACF de estos residuales y el ANOVA de los residuos respecto de la hora de medición, se comprueba que no son ‘ruido blanco’. Por lo tanto, no se puede aceptar M1.

*Tabla 2*  
*Resultados de la regresión multinivel de MI. Variable dependiente: NS<sub>t</sub>*

*Criterios de información*

-2 log de la verosimilitud	529.062
Criterio de información de Akaike (AIC)	537.062
Criterio de Hurvich y Tsai (AICC)	537.228
Criterio de Bozdogan (CAIC)	555.083
Criterio bayesiano de Schwarz (BIC)	551.083

*Estimaciones de efectos fijos*

Parámetro	Estimación	Error típico	g.l.	t	Sig.
Intersección	2.731108	.266290	77.568	10.256	.000
NS <sub>t-1</sub>	.271052	.057050	243.209	4.751	.000

*Estimaciones de parámetros de covarianza*

Parámetro	Estimación	Error típico	Wald Z	Sig.
Residuos	.407223	.038632	10.541	.000
Intersección	.452877	.176361	2.568	.010

De cualquier forma, este análisis no es correcto, pues tal como se observa en la Figura 1, el registro 14 está mal, porque se ha cruzado el último dato del primer participante (5.31) con el primer dato del segundo participante (2,48). Este mismo error se comete en los datos 27, 40, ..., 235, por lo que el análisis queda invalidado (en el cálculo de las ACF y PACF de los valores directos, de la regresión multinivel, y de las ACF y PACF de los residuales), puesto que se cruzan los valores de un participante con los del siguiente. Sólo el cálculo de la ICC y de la comparación de medias interparticipantes es correcto.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Hora	Participant	NS	NS_1	NS_2	NS_3	NS_4	NS_5	NS_6	NS_7	NS_8	NS_9	NS_10	NS_11	NS_12
2	9	1	3,18												
3	10	1	3,66	3,18											
4	11	1	4,38	3,66	3,18										
5	12	1	4,69	4,38	3,66	3,18									
6	13	1	4,91	4,69	4,38	3,66	3,18								
7	14	1	4,53	4,91	4,69	4,38	3,66	3,18							
8	15	1	3,91	4,53	4,91	4,69	4,38	3,66	3,18						
9	16	1	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18					
10	17	1	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18				
11	18	1	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18			
12	19	1	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18		
13	20	1	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18	
14	21	1	5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66	3,18
15				5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38	3,66
16					5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69	4,38
17						5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91	4,69
18							5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53	4,91
19								5,31	4,86	4,83	4,57	4,83	4,80	3,91	4,53
20									5,31	4,86	4,83	4,57	4,83	4,80	3,91
21										5,31	4,86	4,83	4,57	4,83	4,80
22											5,31	4,86	4,83	4,57	4,83
23												5,31	4,86	4,83	4,57
24													5,31	4,86	4,83
25														5,31	4,86
26															5,31
27	9	2	2,48												
28	10	2	3,56	2,48											
29	11	2	4,08	3,56	2,48										
30	12	2	3,69	4,08	3,56	2,48									
31	13	2	3,09	3,69	4,08	3,56	2,48								
32	14	2	3,89	3,09	3,69	4,08	3,56	2,48							
33	15	2	4,11	3,89	3,09	3,69	4,08	3,56	2,48						
34	16	2	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48					
35	17	2	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48				
36	18	2	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48			
37	19	2	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48		
38	20	2	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48	
39	21	2	3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56	2,48
40				3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08	3,56
41					3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69	4,08
42						3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09	3,69
43							3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89	3,09
44								3,99	3,89	4,39	3,09	4,39	4,49	4,11	3,89
45									3,99	3,89	4,39	3,09	4,39	4,49	4,11
46										3,99	3,89	4,39	3,09	4,39	4,49
47											3,99	3,89	4,39	3,09	4,39
48												3,99	3,89	4,39	3,09
49													3,99	3,89	4,39
50														3,99	3,89
51															3,99
52	9	3	0,69												
53	10	3	0,00	0,69											
54	11	3	0,00	0,00	0,69										
55	12	3	1,39	0,00	0,00	0,69									
56	13	3	1,39	1,39	0,00	0,00	0,69								
57	14	3	0,69	1,39	1,39	0,00	0,00	0,69							
58	15	3	1,39	0,69	1,39	1,39	0,00	0,00	0,69						
59	16	3	1,10	1,39	0,69	1,39	1,39	0,00	0,69						
60	17	3	1,39	1,10	1,39	0,69	1,39	1,39	0,00	0,69					

Figura 3. Datos organizados de manera correcta para su análisis por registro y retardados con 12 huecos, para cada participante

Análisis de datos 2

Con el fin de preparar correctamente los datos para los análisis exploratorios y la autorregresión multinivel, se deben organizar los datos de modo que no se cruce ningún dato de un participante con los de otro participante, por lo que se ha de dejar un numero de ‘huecos’ entre dos participantes diferentes. En la Figura 3 se ha hecho: (a) dejar 12 espacios entre cada participante y (b) se ha retrasado la serie (NS\_1, NS\_2,... , NS\_12). Observar que en la Figura 1 hay 259 registros ([13 datos × 19 participantes] + 12 retardos), mientras en la Figura 3 hay 475 registros ([13 datos + 12 huecos] × 19 participantes).

Tabla 3

*Correlaciones simples y parciales de  $NS_t$  con su correspondiente retardo, la probabilidad exacta ( $p$ ) y número de datos válidos ( $N$ ), habiendo dejado 12 ‘huecos’ en las filas entre cada participante*

Correlaciones simples

	$NS_{t-1}$	$NS_{t-2}$	$NS_{t-3}$	$NS_{t-4}$	$NS_{t-5}$	$NS_{t-6}$	$NS_{t-7}$	$NS_{t-8}$	$NS_{t-9}$	$NS_{t-10}$	$NS_{t-11}$	$NS_{t-12}$
$NS_t$	.813	.805	.769	.771	.766	.777	.772	.696	.757	.645	.524	.403
$p$	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.087
$N$	228	209	190	171	152	133	114	95	76	57	38	19

Correlaciones parciales

	$NS_{t-1}$	$NS_{t-2}$	$NS_{t-3}$	$NS_{t-4}$	$NS_{t-5}$	$NS_{t-6}$	$NS_{t-7}$	$NS_{t-8}$	$NS_{t-9}$	$NS_{t-10}$	$NS_{t-11}$	$NS_{t-12}$
$NS_t$	.813	.417	.144	.138	.145	.181	.120	-.151	.244	-.204	-.295	<sup>a</sup>
$p$	.000	.000	.048	.075	.079	.041	.217	.161	.045	.164	.127	---
$N$	228	209	190	171	152	133	114	95	76	57	38	19

<sup>a</sup> No hay suficiente número de grados de libertad para estimar la correlación.

Para llevar a cabo correctamente una ACF y una PACF de un conjunto de series agrupadas, no se puede hacer con un programa ‘ómnibus’ (SPSS, SAS, BMDP, ...) por un motivo fundamental: estiman los valores perdidos mediante máxima verosimilitud. Así, si se pide una ACF y PACF de la variable  $NS_{j,t}$  de los datos de la Figura 3, los resultados son muy parecidos a los de la ACF y PACF de la Figura 1, por lo que para estimar correctamente la ACF y la PACF de la variable  $NS_{j,t}$ , se han de correlacionar mediante sintaxis los valores de esa variable con cada una de las variables retardadas ( $NS_{j,t-1}, NS_{j,t-2}, NS_{j,t-3}, \dots, NS_{j,t-12}$ ), tal como están organizados en la Figura 3; haciendo lo mismo, para los valores de las PACF (los ‘viejos’ analistas de datos es lo que hacían cuando tenían un programa de regresión que no incluía un apartado de modelos Box-Jenkins; Box, Jenkins & Reinsel, 1994). Los resultados se muestran en la Tabla 3.

La importancia de los resultados de las correlaciones simples y parciales expuestas en la Tabla 3 es que son correctas. Al observar los resultados de la Tabla 3 se comprueba que los datos posiblemente se ajustan a un modelo AR(3).

Tabla 4

Tabla 4

Resultados de M2. Variable dependiente:  $NS_{j,t}$ *Criterios de información*

-2 log de la verosimilitud	322.873
Criterio de información de Akaike (AIC)	332.873
Criterio de Hurvich y Tsai (AICC)	333.199
Criterio de Bozdogan (CAIC)	354.108
Criterio bayesiano de Schwarz (BIC)	349.108

*Estimaciones de efectos fijos*

Parámetro	Estimación	Error típico	gl	t	Sig.
Intersección	.676	.153	190	4.408	.000
$NS_{j,t-1}$	.314	.074	190	4.245	.000
$NS_{j,t-2}$	.401	.074	190	5.599	.000
$NS_{j,t-3}$	.147	.073	190	2.007	.046

*Estimaciones de parámetros de covarianza*

Parámetro	Estimación	Error típico	Wald Z	Sig.
Residuos	.320285	.032861	9.747	.000

Al hacer la regresión multinivel (estimada mediante ML) de la DV  $NS_{j,t}$  en función de las IVs  $NS_{j,t-1}$ ,  $NS_{j,t-2}$  y  $NS_{j,t-3}$ , la ecuación que mejor explica los datos es la de la Tabla 4, a la que llamaremos M2, siendo todos sus parámetros significativos (de ajuste global y de distintos coeficientes); es un modelo AR(3) sin constante aleatoria.

Tabla 5

*Correlaciones simples y parciales de los residuales de M2 con su correspondiente retardo, la probabilidad exacta (p) y número de datos válidos (N)*

#### Correlaciones simples

	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
Residuales de M2	-.060	-.103	-.102	-.148	-.053	.059	.197	-.199	.332	--
p	.435	.207	.245	.115	.612	.612	.143	.232	.164	--
N	171	152	133	114	95	76	57	38	19	0

#### Correlaciones parciales

	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
Residuales de M2	-.060	-.109	-.113	-.162	-.076	.044	.187	.315	-.164	--
p	.435	.184	.198	.089	.473	.713	.189	.345	.379	--
N	171	152	133	114	95	76	57	38	19	0

De cualquier forma, puesto que el modelo de regresión multinivel de M1 daba todos los parámetros significativos, haremos un análisis de residuales del modelo M2, estando los resultados en la Tabla 5. Nótese cómo ningún valor de la ACF ni de la PACF es significativo. También se ha hecho un ANOVA de los residuos en función de la hora de medición, y da un resultado no significativo, luego los residuales de M2 son ‘ruido blanco’ y, por tanto, se acepta M2 como modelo válido de explicación de la variabilidad de  $NS_{j,t}$ .

#### Discusión

En este trabajo se ha insistido en que al analizar series temporales agrupadas (datos temporales de diferentes participantes) se han de organizar los datos de manera adecuada: (i) pasando de ‘datos por participante’ a ‘datos por registro’, (ii) dejando, por lo menos, tantos espacios de separación entre participante y participante como retardos se deseen analizar; (iii) no fiándose de los sistemas ‘automáticos’ de análisis (ACF, PACF, ARIMA,...) de los programas ‘ómnibus’ (SPSS, SAS, ...), porque por defecto, estiman los datos perdidos como si fuesen provenientes del mismo participante y mediante procedimientos que tienen tendencia hacia la media, pero en series temporales agrupadas se tienen múltiples participantes, por lo que se obtienen relaciones espurias en los resultados; la solución correcta es hacer los análisis manualmente o mediante sintaxis; (iv) otro problema de los programas automáticos de estimación del ACF y del PACF, es que llevan a cabo el cálculo de ambos estadísticos mediante el procedimiento asintótico, en lugar del muestral, y (v) realizando los análisis de residuales también manualmente o mediante sintaxis, efectuando los correspondientes contrastes y correlaciones con otras variables temporales (la hora o el día de registro, en su caso).

Con nuestros datos, organizándolos de manera adecuada, se ha conseguido un modelo de ajuste

AR(3) sin constante aleatoria; mientras que al organizar los datos de modo inadecuado, se conseguía un (incorrecto) modelo AR(1), con constante aleatoria. Estos resultados AR(3) indican que para pronosticar el valor de NS, en un momento cualquiera y para un participante determinado ( $NS_{j,t}$ ), sólo se necesita saber cuál es el valor de la variable NS para ese participante 'j' en la hora inmediatamente anterior ( $NS_{j,t-1}$ ), dos horas antes ( $NS_{j,t-2}$ ), y tres horas antes ( $NS_{j,t-3}$ ), es decir, en términos funcionales:  $NS_{j,t} = f(NS_{j,t-1}, NS_{j,t-2}, NS_{j,t-3})$ .

Cabe esperar que el incremento de investigación intensiva en diferentes personas mediante telemetría u otras técnicas, y los métodos de análisis más adecuados para este tipo de datos supongan un avance para la comprensión científica, la prevención de trastornos y la mejora de la calidad de vida humana.

### Referencias

- Box, G.E.P., Jenkins G.M. & Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Goldstein, H. (2003). *Multilevel Statistical Models*. London: Edward Arnold.
- Hedeker, D. & Gibbons, R.D. (2006). *Longitudinal data analysis*. Hoboken: John Wiley and Sons.
- Hoepfner, B., Goodwin, M., Velicer, W. & Heltshe, J. (2007). An applied example of pooled time series analysis: Cardiovascular reactivity to stressors in children with autism. *Multivariate Behavioral Research*, 42, 707-727.
- Moore, K., Osgood, D., Larzelere, R. & Chamberlain, P. (1994). Use of pooled time series in the study of naturally occurring clinical events and problem behavior in a foster care setting. *Journal of Consulting and Clinical Psychology*, 62, 718-728.
- Muthén, B. & Asparouhov, T. (2009). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J.K. Roberts (Eds), *The Handbook of Advanced Multilevel Analysis*. London: Taylor and Francis.
- Nater, U.M., Rohleder, N., Schlotz, W., Ehlert, U. & Kirschbaum, C. (2007). Determinants of the diurnal course of salivary alpha-amylase. *Psychoneuroendocrinology*, 32, 392-401.
- Plewis, I. (2005). Modelling behavior with multivariate multilevel growth curves. *Methodology*, 1, 71-80.
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Rohleder, N., Wolf, J.M., Maldonado, E.F., Kirschbaum, C. (2006). The psychosocial stress-induced increase in salivary alphaamylase is independent of saliva flow rate. *Psychophysiology* 43, 645-652.
- Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. New York: Oxford University Press.
- Soliday, E., Moore, K. & Lande, M. (2002). Daily reports and pooled time series analysis: Pediatric

psychology applications. *Journal of Pediatric Psychology*, 27, 67-76.

Sun, Y. & Li, Y. (2002). Children's well-being during parents' marital disruption process: A pooled time-series analysis. *Journal of Marriage and Family*, 64, 472-488.

Velicer, W. F. & Colby, S. M. (1997). Time series analysis for prevention and treatment research. In K. J. Bryant, M. Windle, & S. G. West (Eds.) *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research* (pp. 211-249). Washington, DC: American Psychological Association.

Velicer, W.F. & Fava, J.L. (2003). Time series analysis. In I.B. Weiner (Series Ed.), J. Schinka & W.F. Velicer, (Vol. Eds.). *Handbook of Psychology : Vol. 2. Research Methods in Psychology* (pp. 581-606). New York: John Wiley & Sons.

Walls, T.A. & Schafer, J.L. (2006). *Models for intensive longitudinal data*. New York: Oxford University Press.

# INVESTIGACIÓN APLICADA E INVESTIGACIÓN METODOLÓGICA CON DISEÑOS ORIENTADOS A LA EVALUACIÓN DEL CAMBIO

**Coordinadora: Paula Fernández García**  
Universidad de Oviedo

Los términos investigación metodológica e investigación aplicada suponen un binomio indiscutiblemente imprescindible para el avance de la ciencia y la solución de problemas. Cuando estos dos términos los conjugamos con otros como diseño longitudinal y diseños de medidas repetidas, estamos tomando posición del lado de la investigación que se lleva a cabo utilizando enfoques flexibles de recogida de datos y técnicas de análisis de los mismos que se pueden aplicar a una amplia gama de temas que impliquen cambios en el tiempo. Y es que la popularidad e importancia de los diseños de investigación en los cuales se efectúan mediciones repetidas de cada uno de los sujetos se justifica por ser los únicos que permiten describir, pronosticar y explicar los procesos (biológicos, psicológicos, sociales, etcétera) que se producen como consecuencia del paso del tiempo, estudiar el cambio que se produce tras la aplicación clínica de un tratamiento y su evolución prospectiva, evaluar el impacto de la aplicación de algún programa social sobre la conducta de los individuos.... Así pues, los diseños que aquí nos ocupan, en función de la naturaleza de la/s variable/s intra-sujeto (variables independientes -diferentes tratamientos-, variables independientes asignadas -tiempo, edad-), de la muestra (sean uno o pocos sujetos, sea uno o varios grupos de sujetos definidos por los niveles de la/s variable/s entre-sujetos -variables independientes o de clasificación-), de la regla de asignación (si los niveles de la/s variable/s intra y entre-sujetos han sido elegidos y asignados de modo aleatorio o no), del número de medidas realizadas, de la amplitud de los intervalos de tiempo entre los registros y de la cantidad de variables dependientes, se despliegan en un amplio abanico de diseños particulares orientados a la solución de problemas concretos tanto de carácter clínico como social. En los últimos 20 años, el desarrollo de técnicas de análisis apropiadas para acometer todos los objetivos concretos que se pueden llevar a cabo con estos diseños, y para hacer frente a problemas adversos que pueden surgir en una investigación de estas características, ha sido espectacular, a la par que ha sido espectacular el desarrollo de *software* para llevarlas a cabo y por ende, su aplicación al análisis de datos reales también es esperable que sea espectacular. El conjunto de trabajos que aquí se presentan va orientado tanto a investigaciones sobre el método como a las aplicaciones del mismo en investigaciones experimentales y cuasi-experimentales en las que se busca describir, explicar o predecir.

# ANÁLISIS DE LA ROBUSTEZ DE CUATRO ESTADÍSTICOS UNIVARIADOS EN UN DISEÑO DE MEDIDAS REPETIDAS CARENTE DE ESFERICIDAD CON Y SIN AUTOCORRELACIÓN SERIAL

Paula Fernández García<sup>1</sup>, Guillermo Vallejo<sup>1</sup>, Pablo Livacic-Rojas<sup>2</sup> y Ellián Tuero<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Santiago de Chile

En este trabajo examinamos el comportamiento de cuatro estadísticos univariados para analizar los datos de un diseño de medidas parcialmente repetidas. Los cuatro asumen que la matriz de desviación subyacente es no esférica. Sin embargo, existe una clara distinción entre dos alternativas, dos de ellos presuponen que la correlación entre los datos es arbitraria y los otros dos asumen que existe autocorrelación serial de primer orden. Mediante una investigación Montecarlo son comparados con respecto a su robustez en ausencia de esfericidad bajo autocorrelación serial y bajo correlación arbitraria. Se han manipulado cuatro variables: el tamaño de muestra total, la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión, la estructura de la matriz de covarianza de la población y la forma de distribución de los datos.

Cuando un investigador se enfrenta a la tarea de analizar los datos recogidos mediante un diseño de medidas parcialmente repetidas opta por elegir una de las muchas técnicas (univariadas y multivariadas) disponibles (véase, v.g. Fernández, Livacic-Rojas & Vallejo, 2007). Asumiendo que decide utilizar una técnica univariada, quizás, el mayor problema con el que puede toparse sea la ausencia de esfericidad, y ésta puede suceder tanto cuando la correlación entre las observaciones es arbitraria como cuando sigue un patrón determinado.

Son muchas las investigaciones que se han llevado a cabo sobre la robustez de los procedimientos univariados que corrigen la ausencia de esfericidad sin reparar en la existencia de una estructura de correlación definida, y también aquellas que han puesto a prueba el MLM (revisiones excelentes pueden consultarse en Keselman, Algina, & Kowalchuck, 2001; Blanca Mena, 2004 y Fernández et al., 2007). Sin embargo, son escasas las investigaciones sobre los procedimientos univariados orientados a corregir la ausencia de esfericidad atendiendo a los efectos de la dependencia serial. Recientemente dos de ellos, los procedimientos de Hearne, Clark & Hatch (1983) y Jones (1985), han sido estudiados junto con otros procedimientos univariados (Fernández, Vallejo & Livacic-Rojas, 2008a, 2008b; Fernández, Vallejo & Livacic-Rojas, enviado) asumiendo un diseño balanceado y homogeneidad entre las matrices de dispersión. Por esta razón, el objetivo de esta investigación ha sido evaluar la robustez de los procedimientos de Greenhouse & Geisser (1959), corrección de Lecoutre (1991) realizada sobre el procedimiento de Huynh & Feldt (1976), Hearne et al., (1983) y Jones (1985), cuando en los datos subyace ausencia de homogeneidad entre las matrices de dispersión de los grupos que conforman la variable entre sujetos, tanto cuando el diseño es balanceado como cuando no lo es, y también cuando en los datos subyace distribución normal y no normal. La robustez de los procedimientos bajo las condiciones anteriores se examinó en ausencia de esfericidad, tanto bajo correlación serial como bajo correlación arbitraria.

## Método

Se realizó un estudio de simulación Montecarlo para un diseño de medidas repetidas (3x4) (modelo aditivo) donde las matrices de covarianza grupales fueron heterogéneas en la relación  $\Sigma_1=1/3\Sigma_2$  y  $\Sigma_3=5/3\Sigma_2$ .

### *Procedimientos que se someten a investigación*

Greenhouse & Geisser (1959), corrección de Lecoutre (1991) del estadístico de Huynh & Feldt (1976), Hearne et al., (1983) y Jones (1985), a los cuales haremos referencia como GG, LEC, HCH y JN respectivamente. La formulación de las dos primeras se puede consultar en (v.g., Vallejo, 1991; Fernández et al. 2007) y las de las dos últimas en Hearne et al. (1983) y Jones (1985) respectivamente.

### *VARIABLES MANIPULADAS:*

1. *Tamaño de muestra total.* Se utilizaron dos tamaños (N): N=15 y 30. El coeficiente de variación muestral fue 0 cuando el diseño estaba equilibrado ( $n_j=5$  y 10 respectivamente en los tamaños N anteriores). Cuando fue 0.33 el diseño estaba no equilibrado del siguiente modo (N=15:  $n_j=3, 5$  y 7; N=30:  $n_j=6, 10$  y 14).

2. *Patrones de covarianza empleados para generar los datos.* El interés prioritario en esta investigación ha sido observar qué influye más en la estimación del error de Tipo I empírico ( $\hat{\alpha}$ ) de los procedimientos, si la magnitud de la autocorrelación serial de primer orden ( $\rho$ ) o la cuantía de la desviación de la esfericidad ( $\varepsilon$ ). Para ello generamos los datos utilizando dos estructuras de covarianza no estacionarias desestructuradas, ARAH[1] y NE, ambas con dos desviaciones de la esfericidad ( $\varepsilon=.56$  y  $\varepsilon=.75$ ). En las matrices ARAH[1] existe autocorrelación serial de primer orden y por lo tanto la correlación entre la  $k$ th y la  $k'$ th observación es  $\rho^{|k-k'|}$ ,  $\rho = [0.2:0.8: (0.2)]$ . Las Matrices NE presentan una estructura de correlación arbitraria, y, como consecuencia, la autocorrelación serial es cero ( $\rho=0$ ).

3. *Emparejamiento de las matrices de covarianza y el tamaño de los grupos.* Razón C/H. H0 (diseño equilibrado y heterogeneidad entre grupos), H+ y H- (diseño no equilibrado y relación entre los tamaños de los grupos y el de la matriz de dispersión positiva y negativa respectivamente).

4. *Forma de distribución de la variable de medida.* Los procedimientos se han observado bajo dos formas de distribución: distribución normal ( $\gamma_1=0$  y  $\gamma_2=0$ ) y distribución exponencial ( $\gamma_1=2$  y  $\gamma_2=6$ ). Los datos han sido generados mediante las distribuciones  $g$  y  $h$  introducidas por Tukey (1977).

## Resultados

Las tasas de  $\hat{\alpha}$  se han obtenido dividiendo el número de veces que cada estadístico excedió el valor crítico ( $\alpha=.005$ ) entre el número de replicaciones efectuadas ( $10^4$ ). Hemos utilizado el criterio de robustez de Bradley, por lo tanto, un procedimiento será robusto si su  $\hat{\alpha}$  no excede el intervalo (0.025-0.075). La interpretación de los resultados podría ser distinta utilizando otros criterios de robustez. En la Tabla 1 se muestran los resultados.

*Matriz subyacente ARAH[1]:* Cuando la distribución es normal los procedimientos GG y LEC ajustan el  $\hat{\alpha}$  al nivel nominal para toda razón C/H con independencia del resto de variables manipuladas. Cuando la distribución es exponencial la tasa de error se incrementa de modo pronunciado abandonando la robustez cuando  $\varepsilon=.50$ ,  $n_j=5$  y las razones C/H son H0 y H-.

Bajo normalidad, HCH y JN son liberales (más JN) cuando  $\varepsilon=.50$ , en mayor medida cuando la razón C/H es H- y cuanto menor es  $\rho$  para todo  $n_j$ . Cuando la distribución es exponencial  $\hat{\alpha}$  se incrementa en ambos procedimientos, alcanzando en éste caso el procedimiento JN una posición liberal también para  $\varepsilon=.75$ ,  $n_j=5$  y 10 y  $\leq 40$ .

Tabla 1.- Tasas de error empíricas asociadas al Tratamiento Intra-sujeto para los cuatro procedimientos en un diseño Split-Plot (3x4)  
Matrices de desviación ARAH [1] y NE

Distribución normal ( $\gamma_1=0$ y $\gamma_2=0$ )												Distribución Exponencial ( $\gamma_1=2$ y $\gamma_2=6$ ).													
GG	LEC	GG			LEC			$\rho$	$n_j$	$\rho$	GG			LEC			GG	LEC	GG	LEC					
		.50	.75	.50	.75	.50	.75				.50	.75	.50	.75	.50	.75					.50	.75			
.055	.048	.062	.057	.047	.042	.052	.053	.060	.048	.067	.057	.20	.20	.087	.048	.096	.060	.075	.048	.082	.060	.093	.053	.100	.064
.052	.045	.060	.056	.048	.040	.054	.050	.059	.048	.067	.058	.40	.40	.079	.050	.087	.059	.073	.043	.081	.054	.088	.053	.096	.063
.054	.045	.061	.055	.047	.040	.054	.050	.059	.054	.065	.064	.60	.60	.074	.044	.081	.053	.067	.038	.072	.048	.080	.047	.086	.055
.054	.047	.061	.057	.050	.041	.057	.052	.060	.053	.067	.065	.80	.80	.073	.036	.079	.045	.062	.033	.068	.041	.073	.044	.078	.054
.056	.054	.064	.061	.046	.046	.052	.054	.061	.056	.068	.065	.00	.00	.054	.047	.060	.054	.044	.045	.048	.053	.057	.051	.063	.060
.053	.047	.055	.051	.053	.047	.055	.051	.057	.047	.060	.053	.20	.20	.069	.058	.072	.063	.066	.051	.069	.057	.072	.055	.075	.060
.054	.048	.057	.052	.048	.042	.050	.047	.052	.050	.056	.056	.40	.40	.074	.049	.075	.052	.066	.050	.068	.054	.075	.054	.073	.059
.053	.048	.055	.053	.053	.047	.056	.051	.057	.048	.060	.053	.60	.60	.065	.049	.067	.052	.060	.043	.061	.047	.069	.050	.072	.055
.052	.048	.055	.054	.045	.044	.048	.049	.054	.049	.056	.053	.80	.80	.060	.043	.062	.048	.055	.044	.057	.048	.062	.047	.064	.051
.052	.049	.055	.053	.039	.032	.048	.050	.052	.050	.053	.054	.00	.00	.052	.049	.055	.051	.051	.049	.053	.053	.056	.053	.058	.057
H0												H+													
H0						H+						H-						H+							
.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	$\epsilon$	$\epsilon$	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75
.091	.068	.098	.070	.080	.058	.086	.063	.098	.071	.108	.088	.20	.20	.120	.072	.117	.073	.102	.067	.109	.067	.124	.081	.128	.090
.086	.060	.089	.066	.075	.054	.078	.055	.097	.067	.102	.075	.40	.40	.106	.068	.112	.066	.102	.058	.105	.061	.120	.077	.129	.082
.081	.054	.089	.060	.073	.049	.078	.050	.087	.069	.092	.068	.60	.60	.098	.058	.104	.059	.091	.051	.104	.046	.106	.067	.116	.065
.074	.052	.074	.057	.071	.045	.070	.048	.087	.063	.088	.070	.80	.80	.095	.045	.097	.059	.084	.042	.087	.050	.100	.060	.100	.066
.096	.075	.098	.083	.079	.070	.081	.072	.101	.086	.103	.092	.00	.00	.093	.075	.097	.082	.080	.075	.085	.075	.099	.089	.105	.094
.087	.063	.091	.066	.083	.061	.086	.065	.091	.066	.095	.070	.20	.20	.098	.075	.100	.068	.094	.067	.098	.065	.099	.074	.104	.080
.086	.059	.086	.061	.081	.051	.082	.051	.088	.066	.089	.066	.40	.40	.100	.060	.097	.064	.091	.061	.096	.061	.105	.068	.105	.066
.077	.055	.081	.053	.078	.052	.081	.052	.085	.055	.089	.056	.60	.60	.087	.059	.092	.054	.081	.051	.083	.049	.093	.062	.098	.060
.072	.050	.071	.053	.066	.045	.062	.047	.074	.053	.077	.059	.80	.80	.081	.050	.076	.056	.075	.048	.083	.050	.083	.055	.085	.061
.089	.075	.089	.075	.083	.071	.082	.071	.090	.078	.089	.079	.00	.00	.086	.073	.087	.074	.082	.075	.082	.075	.095	.081	.097	.085
HCH						JN						HCH						JN							

Leyenda:  $\rho = .20, .40, .60, .80 =$  autocorrelación serial positiva de primer orden en ARAH [1];  $\rho = 0 =$  ausencia de autocorrelación serial (correlación arbitraria), matriz NE. El resto de etiquetas no conducen a error dado que están igualmente definidas en el texto.

*Matriz subyacente* NE: En la misma Tabla podemos observar que los procedimientos GG y LEC son siempre robustos en ambas distribuciones. Un incremento de N mejora sustancialmente el resultado.

Los procedimientos HCH y JN son siempre liberales cuando  $\varepsilon=.50$  para toda razón C/H, y para  $\varepsilon=.75$  cuando la razón es H- (más JN). Un incremento de N apenas consigue variar el resultado anterior.

### Discusión y conclusión

Respecto de una distribución normal, bajo una distribución exponencial los cuatro procedimientos se comportan de diferente modo en función de la estructura de la matriz de desviación subyacente. Los cuatro procedimientos incrementan su error empírico si la matriz es ARAH[1], y permanecen casi inalterables cuando en los datos subyace una matriz NE.

GG y LEC siempre son robustos cuando  $\varepsilon=.75$ . HCH y JN son liberales en algunas ocasiones sobre todo cuando la relación C/H es H- y la distribución es exponencial.

Bajo distribución normal GG y LEC son insensibles al valor de  $\rho$ , sin embargo, HCH y JN son muy sensibles a su tamaño, beneficiándose cuanto mayor es su valor. Cuando la distribución es no normal todos ellos tienden a ajustar mejor la tasa de error cuanto mayor es  $\rho$ .

El incremento de N favorece la estimación empírica del error en todos ellos, sin embargo, HCH y JN son menos sensibles a esta variable.

Todos los procedimientos manifiestan mayor  $\hat{\alpha}$  cuando la razón C/H es H-, y menor error cuando la razón C/H es H+.

En respuesta a la pregunta, ¿Qué influye más sobre la tasa de error empírico en estos procedimientos, el tamaño de  $\varepsilon$ , o la magnitud de  $\rho$ ? A tenor de estos resultados, cuando existe autocorrelación serial de primer orden, a mayor  $\rho$ , mejor es la estimación, sin embargo es más determinante el valor de  $\varepsilon$ , porque los cuatro procedimientos son liberales cuando  $\varepsilon=.50$  y, aunque no del mismo modo, los cuatro mantienen el error dentro de los límites de robustez cuando  $\varepsilon=.75$ . Estos resultados están en sintonía con los obtenidos en las investigaciones aquí referenciadas.

Cuando existe autocorrelación serial, la razón C/H tiene su efecto, sin embargo, no hace variar significativamente los resultados; cuando no hay correlación también sucede lo mismo, con una excepción, y es que tanto HCH como JN abandonan sistemáticamente la robustez para  $\varepsilon=.75$  y H-.

Como denominador común, estos resultados nos llevan a afirmar que, de los procedimientos puestos a prueba en esta investigación, la corrección de Lecoutre (1991) es el procedimiento que mejor y más uniforme ajusta el error de Tipo I al valor nominal en las condiciones sometidas a estudio.

Agradecimientos. Este trabajo ha sido realizado con la ayuda concedida por el MCI (Ref.: PS-I2008-03624).

## Referencias

- Blanca Mena, M<sup>a</sup>. (2004). Alternativas de análisis estadístico en los diseños de medidas repetidas. *Psicothema*, 16(3), 509-518.
- Fernández, P., Livacic-Rojas, P. & Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. *International Journal of Clinical and Health Psychology*, 7(1), 153-175.
- Fernández, P., Vallejo, G. & Livacic-Rojas, P. (2008 a). Comparison of the robustness of the SPSS MIXED procedure with regard to another three univariate statistics in repeated measures designs with specified departures from sphericity. *Paper presented at the III European Congress of Methodology*, Oviedo (Spain). July 8-12, 2008.
- Fernández, P., Vallejo, J., Livacic-Rojas, P., Herrero, J. & Cuesta, M. (2008 b). Comparative robustness of six tests in repeated measures designs with specified departures from sphericity. *Quality & Quantity*, DOI: 10.1007/S11135-008-9198-3.
- Fernández, P., Vallejo, J. & Livacic-Rojas, P. (Enviado). Robustez de cinco estadísticos univariados para analizar diseños de medidas repetidas en condiciones adversas. *Revista Latinoamericana de Psicología*.
- Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Hearne, E.M., Clark, G.M. & Hatch, J.P. (1983). A test for serial correlation in univariate repeated-measures analysis. *Biometrics*, 39, 237-243. Huynh, H. & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69-82.
- Jones, R. H. (1985). Repeated measures, interventions, and time series analysis. *Psychoneuroendocrinology*, 10(1), 5-14.
- Keselman, H. J., Algina, J. & Kowalchuck, R. K. (2001). The analysis of the repeated measures design: A review. *British Journal of Mathematical and Statistical Psychology*, 54, 1-20.
- Lecoutre, B. (1991). A correction for the  $\epsilon$  approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics*, 16, 371-372.
- Tukey, J.W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southern Massachusetts University (North Dartmouth, MA).
- Vallejo, G. (1991). *Análisis univariado y multivariado de los diseños de medidas repetidas de una sola muestra y de muestras divididas*. Barcelona, PPU.

# ANÁLISIS DE UN DISEÑO DE MEDIDAS REPETIDAS USANDO REMUESTREO *BOOTSTRAP* Y PERMUTACIÓN ALEATORIA

Guillermo Vallejo<sup>1</sup>, Paula Fernández García<sup>1</sup>, Ellián Tuero<sup>1</sup> y Pablo Livacic-Rojas<sup>2</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Santiago de Chile

En la presente comunicación se evaluó la robustez de cuatro procedimientos diseñados para analizar datos de medidas repetidas cuando se violaban los supuestos de normalidad y esfericidad del tradicional modelo de análisis de varianza (ANOVA) de Scheffé. En concreto, se examinó vía Montecarlo el desempeño de dos métodos de remuestreo, pruebas de permutación y de *bootstrap*, con el desempeño de los enfoques ANOVA y modelo lineal mixto (MLM) implementado en el programa SAS. Cuando se incumplía el supuesto de normalidad y de esfericidad, los resultados numéricos pusieron de relieve que la prueba de permutación se comportaba mejor que los enfoques *bootstrap* y MLM. El enfoque ANOVA, por su parte, fue severamente afectado por la falta de esfericidad y escasamente por la ausencia de normalidad.

Diversas soluciones se hallan disponibles hoy en día, para vencer el impacto negativo que la ausencia de normalidad ejerce sobre la robustez de los enfoques ANOVA y MANOVA cuando se analizan diseños de medidas repetidas. Entre las alternativas propuestas destacan las cuatro que siguen: (a) las encaminadas a lograr la normalidad de los datos utilizando alguna transformación de la familia Box-Cox; (b) las basadas en utilizar procedimientos no paramétricos (p.e., Beasley, 2002); (c) las orientadas a sustituir los usuales estimadores de tendencia central y variabilidad por estimadores robustos, tales como medias recortadas y varianzas winsorizadas (Lix, Algina & Keselman, 2003) y; (d) las caracterizadas por emplear la metodología del modelo mixto generalizado, la cual permite especificar explícitamente una estructura de error no normal. La viabilidad de estas soluciones no es la misma, ya que unas adolecen de mayores defectos que otras.

Afortunadamente, diversos trabajos, incluyendo los de Berkovits, Hancock y Nevitt (2000) y Vallejo, Cuesta, Fernández y Herrero (2006), sugieren que el enfoque basado en el remuestreo *bootstrap* puede constituir una alternativa viable para abordar los problemas referidos. En el contexto específico de los diseños de medidas repetidas sencillos, Berkovits et al. (2000) muestran que el desempeño del método *bootstrap-F* era generalmente satisfactorio cuando los datos incumplían los supuesto de normalidad y de esfericidad. Resultados similares fueron obtenidos por Vallejo et al. (2006) con diseños de medidas repetidas en ausencia de esfericidad multimuestral. También cabe la opción de generar la distribución de probabilidad empírica de un estadístico de contraste usando las pruebas de permutación introducidas por Fisher (1935). En el contexto de los diseños experimentales con factores no repetidos, se ha encontrado que esta técnica estadística limita el número de errores al valor nominal (Anderson & ter Braak, 2003; Jung, Jhun & Song, 2006).

El objetivo fundamental del presente trabajo se centra en examinar la robustez de la prueba *F* tradicional cuando los valores críticos se obtienen mediante de permutación estocástica aproximada y remuestreo *bootstrap*, en vez de calcularlos analíticamente desde la teoría normal. Además, con fines comparativos también examinaremos el comportamiento del modelo lineal general (MLG) y del MLM usando un diseño de medidas parcialmente repetidas.

## Definición del procedimientos estadísticos que motivaron la investigación

### *Prueba de aleatorización*

Para determinar el grado de significación estadística de una hipótesis mediante esta técnica de análisis es preciso, definir con claridad las hipótesis a contrastar, seleccionar la prueba estadística que refleje adecuadamente la diferencia entre los datos observados y la situación nula, elegir el esquema de permutación que genere la distribución de probabilidad del test de acuerdo a lo sucedido en el estudio y concretar el tipo de datos usados para realizar las permutaciones. A continuación, se resumen los pasos a seguir:

1. Se especifica el patrón de aleatorización adoptado en el experimento para asignar los tratamientos a las unidades de muestreo. En nuestro caso,  $n_j$  unidades diferentes fueron asignadas al azar a cada uno de los  $J$  niveles del factor  $A$ . Después, dentro de cada nivel del factor  $A$ , los  $K$  niveles del factor repetido  $B$  se asignaron al azar a cada una de las unidades. Las acciones referidas son críticas, pues el esquema de permutación utilizado para generar distribución probabilidad empírica se debe corresponder con el patrón seguido en la planificación del estudio. El esquema de permutación usado será uno u otro, dependiendo de si la naturaleza del factor  $B$  es activa o asignada. Las respuestas se asumen intercambiables si el factor  $B$  es experimental y dependientes cuando no lo es. Adviértase que bajo los esquemas de aleatorización completa y restringida resultan  $[N!/(n_1!n_2!\dots n_J!)](K!)^N$  y  $[N!/(n_1!n_2!\dots n_J!)](1!)^N$  permutaciones, respectivamente.
2. Se lleva a cabo un muestro aleatorio sin reposición desde la población de permutaciones obtenida tras aplicar la fórmula correspondiente al esquema de aleatorización completa y se eligen  $M$  matrices independientes de dimensión  $N \times K$ ,  $Y^*(1), \dots, Y^*(M)$ . En vista de la ingente cantidad de cálculo que implicaría trabajar con cada uno de los posibles ordenamientos de los datos originales, las pruebas de aleatorización están basadas en  $M$  replicaciones. Para evitar que se puedan obtener valores  $p$  diferentes usando los mismos datos, algunos autores (p.e., Edgington & Onghena, 2007; Manly, 2007) recomiendan seleccionar un número de permutaciones que no sea inferior a 1000.
3. Se especifica la prueba estadística que se va a utilizar y se obtiene su valor para los  $M$  conjuntos de datos permutados. En concreto, para comprobar que los posibles efectos diferenciales de  $A$ ,  $B$  y  $AB$  no están determinados por los tratamientos y, por ende, sería razonable que ocurriesen bajo cualquier reordenamiento de los datos originales, se halla  $M$  veces  $F_c^{*m}$  ( $c = A, B, \text{ o } AB$ ), donde  $F_c$  es el valor del estadístico definido en (2).
4. Se determina la excepcionalidad del estadístico  $F_c$  utilizando los valores  $F_c^{*m}$  de las  $M$  permutaciones. De acuerdo con Efron y Tibshiriani (1993), el nivel de significación alcanzado por el test  $F_c$  se calcula mediante la expresión:  $p_{alea} = M^{-1} \sum_{m=1}^M I[F_c^* > F_c]$ , donde  $I[F_c^* > F_c]$ , la usual función indicador, vale 1 si  $F_c^* > F_c$  y 0 si es menor. La proporción de valores que superan al valor representa el valor  $p$ .

### *Enfoque bootstrap-F*

Los pasos seguidos para determinar el grado de significación mediante el enfoque *bootstrap-F* son convenientemente descritos en el trabajo de Vallejo et al (2006).

## Método de la simulación

En orden a evaluar la robustez de los enfoques definidos en el apartado anterior cuando los valores críticos se obtienen mediante valores teóricos y mediante técnicas de computación intensiva, llevamos a cabo un estudio de simulación usando un diseño de medidas parcialmente repetidas no equilibrado carente de normalidad con  $J = 3$  y  $K = 4$ . Para ello fueron manipuladas las cinco variables siguientes:

1. *Tamaño de muestra total.* El desempeño fue investigado usando dos tamaños de muestra distintos:  $n = 30$  y  $n = 45$ . Estos tamaños grupales son representativos de los encontrados frecuentemente en las investigaciones psicológicas. Para  $n = 30$ , los tamaños grupales fueron: 6, 10 y 14; mientras que para  $n = 45$ , los tamaños grupales fueron: 9, 15 y 21.
2. *Patrones de covarianza empleados para generar los datos.* Los patrones utilizados para generar los datos fueron dos, a saber: una matriz de simetría compuesta (CS) y otra no estructurada (UN) con una desviación del patrón de esfericidad de moderado tirando a severo ( $\epsilon = .50$ ).
3. *Emparejamiento de las matrices de covarianza y el tamaño de los grupos.* La forma de relacionar el tamaño de los grupos y el tamaño de las matrices de dispersión pueden tener diferentes efectos en las pruebas estadísticas. Cuando el diseño está equilibrado, la relación entre el tamaño de las matrices de dispersión y el tamaño de los grupos es nula. Cuando el diseño está desequilibrado, la relación puede ser positiva o negativa.
4. *Forma de la distribución de la variable de medida.* Para investigar el efecto que ejerce la forma de la distribución en el desempeño de las técnicas analíticas, generamos datos desde distribuciones normales y no normales mediante las distribuciones  $g$  y  $h$  introducidas por Tukey (1977). Además de la distribución normal también investigamos otras tres: (a) una distribución que tiene el mismo grado de sesgo y de curtosis que la exponencial doble o de Laplace; (b) una distribución que tiene el mismo grado de sesgo y de curtosis que la exponencial; y (c) una distribución que tiene el mismo grado de sesgo y de curtosis que la distribución lognormal.

## Resultados

El procedimiento más directo para decidir si un determinado enfoque es o no robusto consiste en identificar todas aquellas tasas que excedan significativamente el valor nominal de alfa ( $\alpha$ ) en más/ menos dos errores estándar. De acuerdo con este criterio, para el nivel de significación nominal usado en esta investigación ( $\alpha = 5\%$ ) el intervalo utilizado para definir la robustez de las pruebas fue  $4.4 \leq \hat{\alpha} \leq 5.6$ . Se excusa decir que de haber utilizado otros criterios, diferentes interpretaciones de los resultados son posibles.

La Tabla 1 contiene las tasas de error empíricas correspondientes a la interacción entre los grupos y tratamientos. El patrón de resultados hallado para los efectos principales entre y dentro de los grupos era cualitativa y cuantitativamente similar al descrito. Por consiguiente, en orden a evitar redundancias tan sólo nos centraremos en la fuente que usualmente acapara el interés de los investigadores en este tipo de estudios. Globalmente, los resultados de la Tabla 1 indican lo siguiente:

1. Cuando los datos fueron generados usando un modelo excesivamente parco (i.e., matriz SC), sólo la prueba de aleatorización limitaba el número de errores al valor nominal establecido. Los enfoques basados en derivar los valores críticos analíticamente a partir de la teoría normal (MLG y MLM) controlaban aceptablemente las tasas de error. El enfoque *bootstrap-F* tendía a producir tasas de error inferiores al nivel nominal, particularmente bajo las distribuciones asimétricas (i.e., exponencial y lognormal).

2. Cuando los datos fueron generados usando un modelo completamente general (i.e., matriz UN), el patrón de resultados hallado con los enfoques MLM, de aleatorización y *bootstrap-F* fue muy similar al obtenido bajo la condición de esfericidad. Sin embargo, el enfoque MLG evidenció un comportamiento excesivamente liberal. Observando detenidamente la Tabla 1, se aprecia que el patrón de resultados obtenido con el enfoque MLG no se ve afectado por la forma distribución. Como cabía esperar, la variable crítica resultó ser la ausencia de esfericidad. Por consiguiente, cuando las matrices de dispersión son homogéneas, resulta factible corregir la liberalidad del enfoque MLG ajustando los grados de libertad mediante alguno de los múltiples correctores tipo Box existentes.

Tabla 1. Porcentaje de veces que los procedimientos rechazaban la hipótesis nula referida a la interacción AB cuando las matrices de dispersión eran homogéneas (NS = 5%)

N	$n_j/\Sigma_j$	GH	Matriz CS ( $\epsilon = 1.00$ )				Matriz UN ( $\epsilon = 0.50$ )			
			GLM	MLM	Aleat.	Bootst.	GLM	MLM	Aleat.	Bootst.
Distribución Normal ( $\gamma_1 = 0$ & $\gamma_2 = 0$ )										
30	=	1:1:1	5.28	<b>5.96</b>	5.42	<b>3.22</b>	<b>10.12</b>	4.96	5.02	<b>4.34</b>
30	≠	1:1:1	5.36	<b>5.74</b>	5.32	<b>3.88</b>	<b>9.80</b>	<b>5.86</b>	5.00	5.32
45	=	1:1:1	4.92	<b>5.94</b>	4.88	<b>4.06</b>	<b>10.00</b>	<b>5.90</b>	5.00	4.44
45	≠	1:1:1	4.94	5.16	5.08	<b>4.26</b>	<b>9.30</b>	4.88	4.98	5.10
Distribución tipo Laplace ( $\gamma_1 = 0$ & $\gamma_2 = 3$ )										
30	=	1:1:1	4.59	5.60	4.78	<b>2.66</b>	<b>8.48</b>	<b>5.86</b>	4.62	<b>4.12</b>
30	≠	1:1:1	5.00	<b>6.62</b>	4.96	<b>3.34</b>	<b>8.94</b>	5.14	4.82	<b>4.36</b>
45	=	1:1:1	5.16	5.14	5.32	<b>3.64</b>	<b>9.54</b>	4.66	5.00	4.86
45	≠	1:1:1	5.10	5.06	5.00	<b>3.96</b>	<b>9.56</b>	5.30	4.84	4.58
Distribución tipo Exponencial ( $\gamma_1 = 2$ & $\gamma_2 = 6$ )										
30	=	1:1:1	<b>3.82</b>	4.54	4.40	<b>1.54</b>	<b>9.34</b>	<b>3.82</b>	5.00	<b>3.00</b>
30	≠	1:1:1	4.76	4.58	5.04	<b>1.76</b>	<b>8.80</b>	<b>4.32</b>	4.78	<b>3.42</b>
45	=	1:1:1	4.52	4.90	4.86	<b>2.18</b>	<b>9.74</b>	<b>3.96</b>	5.04	<b>3.58</b>
45	≠	1:1:1	4.98	5.60	4.86	<b>2.12</b>	<b>9.58</b>	4.64	4.90	<b>3.40</b>
Distribución tipo Lognormal ( $\gamma_1 = 6.18$ & $\gamma_2 = 114.94$ )										
30	=	1:1:1	<b>4.06</b>	<b>3.66</b>	4.74	<b>1.16</b>	<b>8.88</b>	<b>3.70</b>	4.98	<b>2.20</b>
30	≠	1:1:1	4.94	<b>4.34</b>	4.66	<b>1.16</b>	<b>9.16</b>	<b>4.16</b>	5.02	<b>2.50</b>
45	=	1:1:1	<b>3.98</b>	<b>4.32</b>	4.42	<b>1.52</b>	<b>9.54</b>	<b>3.96</b>	5.02	<b>2.74</b>
45	≠	1:1:1	5.10	4.54	4.86	<b>1.78</b>	<b>9.18</b>	4.58	5.00	<b>2.94</b>

Leyenda: NS = nivel de significación;  $n_j/\Sigma_j$  = relación entre tamaño de los grupos y el tamaño de las matrices de covarianza; GH = grado de heterogeneidad;  $\gamma_1$  = índice de asimetría;  $\gamma_2$  = índice de curtosis.

### Discusión y conclusión

Nuestros resultados ponen de relieve que, excepto la prueba de aleatorización, ninguno de los enfoques evaluados se ha mostrado efectivo bajo todas las condiciones manipuladas. Globalmente, el enfoque MLG con valores críticos obtenidos mediante remuestreo *bootstrap* tendía a comportarse de manera conservadora a medida que se incrementaba el sesgo de los datos, mientras que el clásico enfoque MLG tendía a comportarse de manera excesivamente liberal. El enfoque MLM, por su parte, controlaba aceptablemente las tasas de error, en especial cuando el tamaño de muestra aumentaba. El patrón de resultados descrito no hace más que confirmar y generalizar los encontrados en otros estudios similares (Vallejo & Livacic-Rojas, 2005; Vallejo et al., 2006; Fernández, Livacic-Rojas & Vallejo, 2007).

Para concluir queremos resaltar que resultados son limitados a las condiciones examinadas en el estudio. Por consiguiente, antes de proceder a su generalización sería muy conveniente examinar el comportamiento de los procedimientos de remuestreo, en particular de la prueba de aleatorización, cuando las varianzas sean heterogéneas y la forma de la distribución asimétrica.

Reconocimientos: Este trabajo ha sido financiado mediante el proyecto de investigación concedido por el MCI (Ref.: PSI-2008-03624).

## Referencias

- Anderson, M. J. & ter Braak, C. J. F. (2003). Permutation tests for multifactorial analysis of variance. *Journal of Statistical Computation and Simulation*, 73, 85-113.
- Beasley, T. M. (2002). Multivariate aligned rank test for interactions in multiple group repeated measures designs. *Multivariate Behavioral Research*, 37(2), 197-226.
- Berkovits, I., Hancock, G. R. & Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement*, 60, 877-892.
- Edgington, E. S. & Onghena, P. (2007). *Randomization Tests* (4th Edition). London: Chapman & Hall/CRC.
- Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Fisher, R. A. (1935). *The Design of Experiments*, Oliver & Boyd, Edinburgh.
- Fernández, P., Livacic-Rojas, P. & Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. *International Journal of Clinical and Health Psychology* 7, 153-175.
- Jung, B. C., Jhun, M. & Song, S. H. (2006). A new random permutation test in ANOVA models. *Statistical Papers*, 48, 47-62.
- Lix, L. M., Algina, J. & Keselman, H. J. (2003). Analysing multivariate repeated measures designs: A comparison of two approximate degrees of freedom procedures. *Multivariate Behavioral Research*, 38, 403-431.
- Manly, B. F. J. (2007). *Randomization, Bootstrap and Monte Carlo Methods in Biology* (3rd Edition). London: Chapman & Hall/CRC.
- Tukey, J. W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southern Massachusetts University (North Dartmouth, MA).
- Vallejo, G., Cuesta, M., Fernández, P. & Herrero, J. (2006). A comparison of the Bootstrap-F, Improved General Approximation, and Brown-Forsythe multivariate approaches in a mixed repeated measures design. *Educational and Psychological Measurement* 66, 35-62.
- Vallejo, G. & Livacic-Rojas, P. E. (2005). A comparison of two procedures for analyzing small sets of repeated measures data. *Multivariate Behavioral Research*, 40, 179-205.

# ESTIMACIÓN DEL EFECTO EN SERIES TEMPORALES CORTAS. PROPUESTA DE UN PROGRAMA EN SPSS COMO ALTERNATIVA AL PROCEDIMIENTO ARIMA

Jaume Arnau, Roser Bono, Antonio Solanas y Rumen Manolov  
Universidad de Barcelona

Una posible solución alternativa a los análisis basados en modelos de series temporales es el uso de modelos más simples y de menos difícil identificación. Estos modelos han de tener en cuenta el problema de la autocorrelación de los datos, pero sin que para ello sea necesaria la identificación de un modelo autorregresivo integrado de medias móviles (ARIMA). Siendo esto así, ha habido intentos de aplicar distintos procedimientos de Mínimos Cuadrados Generalizados (MCG), en situaciones donde se generan datos autocorrelacionados. Estos procedimientos transforman los datos para eliminar la autocorrelación entre los residuales. Cuando los residuales, obtenidos a partir de esta transformación, no presentan correlación alguna entonces se puede operar con los procedimientos estadísticos tradicionales.

En virtud de estas consideraciones, presentamos un programa en SPSS que desarrolla el modelo de la regresión de dos etapas para estudios de sujeto único. En la primera etapa, una vez estimados los parámetros se obtienen los residuales a los que se aplica una prueba de independencia. Según el resultado de esta prueba, se corrigen los datos originales por la autocorrelación y se vuelven a estimar los parámetros de la regresión. Gracias a este modelo, es posible obtener una estimación más precisa del efecto del tratamiento.

A partir del año 1975, se constata un creciente interés por el grado de correlación que muestran los datos de caso único y por la propuesta de modelos estadísticos adecuados a esta clase de datos (Glass, Willson y Gottman, 1975; Hartmann et al., 1980; Jones, Vaught y Reid, 1975; Jones, Vaught y Weinrott, 1977). Así, los datos de caso único plantean graves dificultades cuando se aplican las técnicas de análisis de series temporales o modelos ARIMA (Sharpley, 1987). En cuanto al ajuste del modelo, es aconsejable que las series tengan entre 50 y 100 observaciones para poder identificar su estructura latente. Siendo esto así, difícilmente se logrará identificar el modelo con datos de diseños conductuales. Ahora bien, como ha señalado Sharpley (1987), el problema de la longitud de la fase o período no es el principal escollo en la identificación del modelo. Por lo que respecta a las pruebas estadísticas de inferencia del efecto, ha de tenerse en cuenta la probabilidad de cometer errores de Tipo I o II. El problema está en saber para qué cantidad de observaciones la probabilidad de cometer un error de Tipo I es mayor que la comúnmente aceptada del 5% (valor nominal).

El tema de la longitud de la fase sigue siendo un grave problema para la identificación de los modelos ARIMA. Aunque Sharpley (1987) concluya, en función del análisis de los estudios publicados por Gottman (1981), Velicer y Harrop (1983), y Harrop y Velicer (1985), que la correcta identificación del modelo no parece ser una cuestión primordial en la prueba de fiabilidad de los efectos, el tema permanece ahí y hay que darle una solución.

La alternativa a los análisis basados en modelos de series temporales es el uso de modelos más simples y de menos difícil identificación. Estos modelos han de tener en cuenta el problema de la autocorrelación de los datos, pero sin que para ello sea necesario la identificación de un modelo ARIMA. En el presente estudio, utilizaremos técnicas de análisis basadas en el modelo lineal y que han sido adaptadas a la situación secuencial de los datos.

## Modelos de Mínimos Cuadrados Generalizados (MCG)

Con datos de diseños de series temporales es posible aplicar modelos de análisis estadísticos basados en el modelo lineal, como el análisis de la varianza y el análisis de la regresión. Estos modelos asumen que cada observación es independiente. Dicho supuesto es difícil de sostener con datos conductuales, ya que suelen estar afectados por la dependencia serial. Este fenómeno, propio de los datos de series temporales, genera una estructura de dependencia serial que plantea un grave problema a la utilización de las técnicas de análisis convencionales, el problema de la autocorrelación.

Analicemos, en primer lugar, las consecuencias que se derivan de la presencia de correlación entre los errores cuando se aplican las técnicas estadísticas clásicas o convencionales. La consecuencia más directa es que el uso de estos modelos estadísticos (como el análisis de la varianza y el análisis de la regresión) para verificar el efecto del tratamiento o intervención, tienden a inflar el estadístico de la prueba y hay un alto riesgo o probabilidad de cometer errores de Tipo I y de Tipo II (Gottman y Glass, 1978; Scheffé, 1959). Cuando los términos de error están positivamente autocorrelacionados, las estimaciones de los errores estándar de los coeficientes tienden a ser más pequeños, con lo que obtienen valores de  $F$  positivamente sesgados. Por este motivo, ha habido intentos de aplicar distintos procedimientos de MCG a datos autocorrelacionados. Cada uno de ellos transforma los datos para corregir la autocorrelación y requiere que las autocorrelaciones entre los residuales, después de la transformación, sean aproximadamente cero. Cuando los residuales, obtenidos a partir de esta transformación están exentos de correlación, entonces se puede operar con los procedimientos estadísticos tradicionales.

### Modelación de la autorregresión

El procedimiento de análisis propuesto, inspirado en los MCG, se aplica a observaciones con residuales autocorrelacionados y consiste en corregir directamente los datos. De este modo, se asume que la autocorrelación de los residuales en la población es cero, después de la corrección. En esta línea, se sitúan los enfoques propuestos por Simonton (1977) y Gorsuch (1983).

Con datos de series temporales, el problema básico a plantear se refiere a la autocorrelación. Según Gottman (1981), con autocorrelación en los datos es más difícil determinar si un cambio es resultado de la intervención, o simplemente es resultado del comportamiento normal de la serie. Por esa razón, Gottman (1981) sugiere, como alternativa a los modelos ARIMA, que se aplique un modelo lineal autorregresivo. Con ello se consigue, en primer lugar, transformar la serie temporal en un residual que sea la realización de un proceso de ruido blanco (la autocorrelación de los residuales es cero) y estimar los parámetros por el criterio de mínimos cuadrados. Así, mediante este procedimiento, es posible determinar la significación del efecto de tratamiento. En otras palabras, los datos se modelan mediante una estructura autorregresiva y se añade a los datos de la fase de tratamiento un parámetro de intervención, en términos de un proceso asintótico. Ello permite, según Gottman (1981), transformar los datos en un proceso de ruido blanco y, al mismo tiempo, obtener la estimación mínima cuadrada de un coeficiente de la regresión lineal cuya significación es verificable por el estadístico  $t$ .

Simonton (1977), siguiendo a Kmenta (1971), recomienda la aplicación de la regresión generalizada o procedimiento de estimación basado en MCG (Hibbs, 1974). Se modela la serie temporal en términos de los parámetros de la regresión que representan los cambios teóricamente esperados, tanto de nivel como de pendiente. Para estimar los valores del modelo, se procede como sigue:

1. Se regresa la variable dependiente (puntos de observación seriales,  $Y$ ) sobre la independiente (variable dummy, o ausencia y presencia de tratamiento,  $X$ ). A continuación se recuperan los residuales de los datos.

2. Se estima la autocorrelación entre los residuales y se aplica el estadístico  $d$  (Durbin y Watson, 1951). El parámetro autorregresivo se estima a partir de la formula  $\rho_1 \approx 1 - d/2$ .

3. Se utiliza el coeficiente de la autorregresión para transformar la variable dependiente con la siguiente regla:

$$\begin{aligned} Y'_{it} &= Y_{it} \sqrt{1 - r^2}, & \text{si } t = 1 \\ Y'_{it} &= Y_{it} - \rho_1 Y_{it-1}, & \text{si } t > 1 \end{aligned} \quad (1)$$

donde  $i = 1, 2, 3, \dots, N$ .

4. Se aplica a los valores transformados de la variable dependiente ( $Y'_{it}$ ) un segundo análisis de la regresión por mínimos cuadrados ordinarios. Los errores estándar resultantes son no sesgados y, en consecuencia, pueden usarse las pruebas  $F$  o  $t$  para interpretar los coeficientes de la regresión.

Según Gorsuch (1983), para esta transformación debería usarse el parámetro autorregresivo de la población y no su estimación ( $r_1$ ). Cuando su valor no es disponible, la mejor aproximación es la autocorrelación estimada a partir del estadístico  $d$  de Durbin-Watson. Esta transformación de los datos originales de la serie temporal es menos necesaria cuando el valor de correlación es muy pequeño, ya que su efecto, en cuanto a la estimación de error estándar, es imperceptible.

Gorsuch (1983) presenta un procedimiento similar al de Simonton (1977) que consiste en estimar la autocorrelación para transformar las variables  $Y$  y  $X$ . Una vez transformadas, las variables son introducidas en el modelo. De esta forma, la ecuación original es reemplazada por un modelo de variables transformadas, mediante el procedimiento autorregresivo, como sigue:

$$Y'_t = b'_0 + b'_1 X'_t + e'_t \quad (2)$$

De lo anterior, se deriva que, con frecuencia, los residuales de una serie secuencial de datos reflejan un proceso autorregresivo. Nótese, sin embargo, que no es lo mismo autocorrelación que autorregresión. Los residuales autorregresivos muestran autocorrelación, pero no todos los residuales autocorrelacionados son residuales autorregresivos.

### Presentación del programa

A modo de ejemplo de diseño de caso único de dos réplicas, o diseño A-B-A-B (diseño de retirada de cuatro fases, DR4F), utilizamos el trabajo publicado por Murphy, Hutchinson y Bailey (1983). En este estudio se pretende probar si la aplicación de juegos organizados a niños escolarizados reduce el número de actos agresivos, en los períodos de descanso. Un total de 300 niños de jardín de infancia, de primer y segundo grado, sirvieron de sujetos. Se registró la cantidad de conductas inapropiadas en períodos de juego de 20 minutos. En base al gráfico del trabajo de Murphy et al., (1983), extrajimos los datos de cada fase del diseño. Estos datos sirvieron de base para la aplicación del programa realizado para evaluar el impacto del tratamiento, teniendo en cuenta la tendencia general y las tendencias parciales por fase.

El programa tiene dos partes, en cada una de las partes se aplica el modelo de la regresión donde se tiene en cuenta el tiempo total, el tiempo parcial (tendencia por fase) y el tratamiento. El paso del primer modelo de la regresión al segundo dependen del valor del estadístico  $d$ . Este estadístico varía entre 0 y 4 y toma el valor 2 cuando los residuales son independientes. Valores menores que 2 indican dependencia entre los residuales. Podemos asumir independencia entre los residuales cuando  $d$  toma valores entre 1,5 y 2,5. De acuerdo con este criterio, utilizamos como valores de dependencia valores igual o menores que 1,5.

Cuando el resultado del primer modelo tiene un valor de  $d$  igual o menor que 1,5, el programa automáticamente calcula la autocorrelación de primer orden,  $\rho_1 = 1 - d/2$ , se corrigen los datos por la autorrelación, según la regla de la expresión (1) y se pasa el segundo modelo de la regresión con los datos ya corregidos. Por último, se interpreta el efecto del tratamiento.

La sintaxis del programa en SPSS para un diseño A-B-A-B es la siguiente:

```
GET FILE='k:\Murphy1.sav'.
PRESERVE.
SET TVAR NAMES TNUMBERS VALUES.
OMS
  /SELECT TABLES
  /IF COMMANDS = ['Regression'] SUBTYPES=['Model Summary' 'Resumen del modelo']
  /DESTINATION FORMAT=SAV OUTFILE='k:\temp1.sav'.
REGRESSION_1
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI R ANOVA CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT dep
  /METHOD=ENTER temptot /METHOD=ENTER temparc /METHOD=ENTER tratam
  /RESIDUALS DURBIN .
OMSEND.
GRAPH
  /LINE(SIMPLE)=VALUE( dep ) BY tratam .
RESTORE.
MATCH FILES FILE=*
  /FILE='k:\temp1.sav'
  /RENAME (durbinwatson=durbinwatson)
  /DROP command_ TO Sig.FChange.
RMV
  /d=MEAN(durbinwatson ALL).
DO IF d<1.5.
COMPUTE lagdep = LAG(dep) .
IF (temptot = 1) depcor = dep*SQRT(1-(1-d/2)**2).
IF (temptot > 1) depcor = dep-(1-d/2)*lagdep.
END IF.
EXECUTE.
REGRESSION_2
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI R ANOVA CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT depcor
  /METHOD=ENTER temptot /METHOD=ENTER temparc /METHOD=ENTER tratam
  /RESIDUALS DURBIN .
EXECUTE .
SET ERRORS=OFF.
GRAPH
  /LINE(SIMPLE)=VALUE( depcor ) BY tratam .
```

## Referencias

- Durbin, J. y Watson, G. S. (1951). Testing for serial correlation in least squares regression II. *Biometrika*, 38, 159-178.
- Glass, G. V., Willson, V. L. y Gottman, J. M. (1975). *Design and analysis of time series experiments*. Boulder, CO: Colorado Associated University Press.
- Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, 5, 141-154.
- Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge: Cambridge University Press.
- Gottman, J. M. y Glass, G. V. (1978). Analysis of interrupted time-series experiments. En T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change*. New York: Academic Press.
- Harrop, J. W. y Velicer, W. F. (1985). A comparison of alternative approaches to the analysis of interrupted time-series. *Multivariate Behavioral Research*, 20, 27-44.
- Hartmann, D. P., Gottman, J. H., Jones, R. R., Gardner, W., Kazdin, A. E. y Vaught, R. S. (1980). Interrupted time-series analysis and its application to behavioral data. *Journal of Applied Behavior Analysis*, 13, 543-559.
- Hibbs, D. A. (1974). Problems of statistical estimation and causal inference in time-series regression models. En H. L. Costner (Ed.), *Sociological methodology, 1973-1974*. San Francisco: Jossey-Bass.
- Jones, R. R., Vaught, R. S. y Reid, J. B. (1975). Time series analysis as a substitute for single subject analysis of variance designs. En G. R. Patterson, I. M. Mark, J. D. Matarazzo, R. A. Myuers, G. E. Schwartz y H. H. Strapp (Eds.), *Behavior change*. Chicago: Aldine.
- Jones, R. R., Vaught, R. S. y Weinrott, M. (1977). Time-series analysis in operant research. *Journal of Applied Behavior Analysis*, 10, 151-166.
- Kmenta, J. (1971). *Elements of econometrics*. New York: Macmillan.
- Murphy, H. A., Hutchinson, J. M. y Bailey, J. S. (1983). Behavioral school psychology goes outdoors: The effect of organized games on playground aggression. *Journal of applied Behavior Analysis*, 16, 29-35.
- Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.
- Sharpley, C. F. (1987). Time-series analysis of behavioural data: An update. *Behaviour Change*, 4, 40-4.
- Simonton, D. K. (1977). Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychological Bulletin*, 85, 489-502.
- Velicer, W. F. y Harrop, J. W. (1983). The reliability and accuracy of the time-series model identification. *Evaluation Review*, 7, 551-560.

# ANÁLISIS DE LA SINTOMATOLOGÍA MENSTRUAL: UN MODELO FACTORIAL AUTORREGRESIVO DE ECUACIONES ESTRUCTURALES CON SERIES TEMPORALES AGRUPADAS

Francisco Herrero, Pilar Jara y Jesús F. Rosel  
Universidad 'Jaume I' de Castellón

El estudio de la sintomatología relacionada con el ciclo menstrual ha permitido conocer que ésta puede agruparse dentro de tres grandes bloques de síntomas: fisiológicos, psicológicos y conductuales. 51 mujeres sanas, rellenaron diariamente un cuestionario sobre síntomas de ciclo menstrual, con un total de 8040 registros. El trabajo que presentamos tiene un doble objetivo: (a) Desde los modelos ARIMA, trata de explicar individualmente el comportamiento sintomatológico desde una perspectiva global. (b) Desde los modelos factoriales dinámicos (y para toda la muestra), trata de comprobar, descriptivamente, si la factorización (mediante variables latentes) temporal de las variables observadas día a día (bloques de síntomas) sigue un modelo autorregresivo multinivel. Se comprueba que cada participante ajusta a un modelo Box-Jenkins (simple o estacional) y, al agrupar los datos en una '*pooled time series*' análisis, se obtiene: (a) un factor diario de síntomas (fisiológicos, psicológicos y conductuales), (b) un modelo dinámico con un componente autorregresivo simple, lo cual indica que la conducta psicofisiológica femenina presenta dependencia temporal respecto a un día atrás, y (c) un modelo dinámico con un componente autorregresivo estacional de 26, 27 y 28 días.

Las cuatro fases que conforman el ciclo reproductivo femenino son cronológicamente: menstrual, post-menstrual (de menstruación a ovulación), ovulatoria, y premenstrual (de ovulación a menstruación). La duración del ciclo puede variar, Ganong (1982) plantea un término medio de 28 días. La recurrencia de fases es el resultado de la alternancia del sistema hipotálamo-hipófisis-ovarios, con los cambios funcionales y estructurales de los órganos reproductivos. La interacción de esta triada modula las cantidades hormonales que implicará la aparición (o la desaparición) de sintomatología asociada al ciclo menstrual (Lenz, 2007).

Ya sea por los niveles hormonales, los neurotransmisores, factores nutricionales, o factores psicológicos (Bäckström y Mattson, 1975; Abplanalp, 1983; Abraham, 1984; Dalton, 1984; Barnard et al., 2000; Yonkers et al., 2008), el ciclo menstrual comporta síntomas que por su incidencia, y las fases del ciclo menstrual, se consideran mayoritariamente padecidos por las mujeres durante su vida reproductiva. Entre ellos: dolor corporal (cabeza, abdomen,...) hinchazón (pies, manos y mamas), flatulencia, aumento de peso, acné, problemas digestivos, irritación, insomnio, depresión, agresividad, miedo, cambios de humor, o imposibilidad para la actividad cotidiana (Webster, 1978).

Estudios de análisis factorial han establecido los diferentes grupos de síntomas que aparecen reiterativamente en los casos estudiados (Stephenson et al., 1986; Van der Ploeg, 1990). Concretamente, son tres bloques, síntomas: fisiológicos, psicológicos y conductuales (Jara et al., 1998).

Desde esta perspectiva, los objetivos de este trabajo son:

- Conocer la génesis de la sintomatología del ciclo menstrual en mujeres jóvenes. Se considera que la sintomatología dependerá de la padecida el día anterior y de la estacionalidad acorde con el ciclo menstrual particular.

- Establecer la funcionalidad de la sintomatología menstrual desde el enfoque de síntomas fisiológicos, psicológicos y conductuales. Se considera que los tres bloques de síntomas (día a día) serán la manifestación de una variable latente (factor de sintomatología general); y se estudiará el modelo autorregresivo factorial (simple y estacional) del factor de sintomatología general.

Las series temporales son una estrategia adecuada para estudiar el comportamiento de la sintomatología menstrual (Jara, 1995), por las características que, inherentemente, posee la estructura de las series temporales. Los modelos ARIMA permiten descomponer el contenido de la información en tres componentes antes de transformar la serie en un ruido blanco, objetivo del análisis; es decir, desentrañamos la información de un proceso (sintomatología menstrual), hasta conseguir que los valores residuales no contengan ningún tipo de información inter-correlacionada.

Abundando en el estudio temporal, el Análisis Factorial Dinámico comprueba, descriptivamente, si la factorización (mediante variables latentes) temporal de las variables observadas diariamente (síntomas) sigue un modelo autorregresivo. El análisis factorial dinámico pretende el análisis del tipo “Unidad\*Variable (cuantitativa)\*Ocasión”, donde la ocasión se identifica como *tiempo* (día), realizando así el orden temporal de los datos (Curran y Bollen, 2001; Frees, 2004; Moore et al., 1994; Plewis et al., 2006; Soliday et al., 2002; Sun y Li, 2002; Walls y Schafer, 2006).

## Método

### *Participantes*

Consideramos 51 mujeres sanas, voluntarias, con 8040 registros de sintomatología padecida, diariamente.

Tenían edades entre 18 y 31 años (media=21,96,  $S_x=2,64$ ). Y menstruación regular (78,4%), siendo 28 días los más referidos por lo que respecta a su ciclicidad (23,5%), oscilando entre 21-35 días. El 68% manifestó tener síntomas relacionados con la menstruación.

### *Procedimiento*

El cuestionario de registro de síntomas diario que, basado en el “Menstrual Distress Questionnaire Manual” de Moos (1968), puede consultarse en: <https://spreadsheets.google.com/viewform?formkey=cDcwa2wxcGtqZHVSzAwQVkyRVBaUXc6MA>

Permitió recoger los tres bloques de sintomatología. Cada mujer rellenó, vía internet, al final de cada día, los síntomas padecidos, indicando el inicio de su menstruación, para detectar las fases de su ciclo menstrual.

Tras los registros, para abordar el primer objetivo, se generó una variable sumando los síntomas diarios, conformando así la serie temporal individual para el análisis ARIMA del comportamiento de la sintomatología global.

Para el segundo objetivo, consideramos un fichero único, con los datos de todas las mujeres, en el que se consideró el modelo básico de series temporales obtenido las mujeres individualmente, para esto, se dejaron 32 huecos entre los registros de cada mujer, y retardar las tres series de datos (bloques de síntomas) de cada mujer: desde 1 hasta 32 retardos.

## Análisis

Para el primer objetivo consideramos los modelos ARIMA, propuestos por Box y Jenkins (1970) y Box et al. (2008), dado que los datos recogidos individualmente superan los 90 días.

Para el segundo objetivo, consideramos los modelos de análisis factorial dinámico para verificar si se reproduce el modelo subyacente a la génesis del ciclo menstrual en el que tanto la sintomatología general, como los tres bloques, se va modificando paulatinamente de acuerdo con la diferente evolución (hipotálamo-hipófisis-ovario).

Estadísticamente, se ha mantenido (en puntuaciones directas) la invarianza de las cargas factoriales de cada respectivo factor (día) con relación a las tres variables observables (constricciones 1, a y b), también invarianza de las respectivas varianzas de los errores de cada variable (constricciones c, d y e), y la invarianza de las covarianzas entre los respectivos errores de pronóstico (constricciones f, g y h).

## Resultados

Respecto al primer objetivo, se obtuvieron los modelos ARIMA de la Tabla 1.

Tabla 1. Cuadro-resumen con los modelos ARIMA(p,d,q)(P,D,Q)S obtenidos para cada participante

Modelos	SIMPLES		ESTACIONALES		
	AR	ARMA	AR	MA	ARMA
n (%)	n=9 (17,6%)	n=8 (15,7%)	n=20 (39,2%)	n=10 (19,6%)	n=4 (7,8%)
Rango Coeficientes	(0,32- 0,76)	AR (0,31-0,82) MA(-0,12)-(-0,53)	(0,16-0,76)	(-0,28) -0,78)	AR(0,14-0,37) MA(-0,36)-(-0,73)

Por ejemplo, considerando el modelo estacional (mujer 22), vemos que,

$$\text{Sínt}_{(t)} = 0,29 + 0,24 \cdot \text{Sínt}_{(t-1)} + 0,25 \cdot \text{Sínt}_{(t-29)} - 0,24 \cdot 0,25 \cdot \text{Sínt}_{(t-30)} + a_t$$

su sintomatología (en un día cualquiera) es función de la padecida el día anterior ( $\text{Sínt}_{(t-1)}$ ), y, además, depende de la padecida 29 días antes ( $\text{Sínt}_{(t-29)}$ ), y 30 días antes ( $\text{Sínt}_{(t-30)}$ ).

Respecto al segundo objetivo, la Figura 1 muestra el resultado del análisis factorial dinámico obtenido, los valores están en puntuaciones estandarizadas. Los valores de los estadísticos de ajuste de conjunto son: chi-cuadrado = 3966, g.l. = 377, p = .000; Bentler-Bonett Normed Fit Index = .931; Bentler-Bonett Non-Normed Fit Index = .933; Comparative Fit Index = .937; RMSEA = .044 (90% IC: .042 - .045), indicando muy buen ajuste de los datos al modelo. Los parámetros individuales del modelo (efectos, varianzas y covarianzas) son todos significativos al nivel 'alfa'=.05.

Se pusieron a prueba los valores de los coeficientes multinivel de: las constantes de las variables observables, los coeficientes autorregresivos, los valores de la fase premenstrual sobre las variables observables, y los efectos de los factores sobre cada variable observable, pero no resultaron significativos.

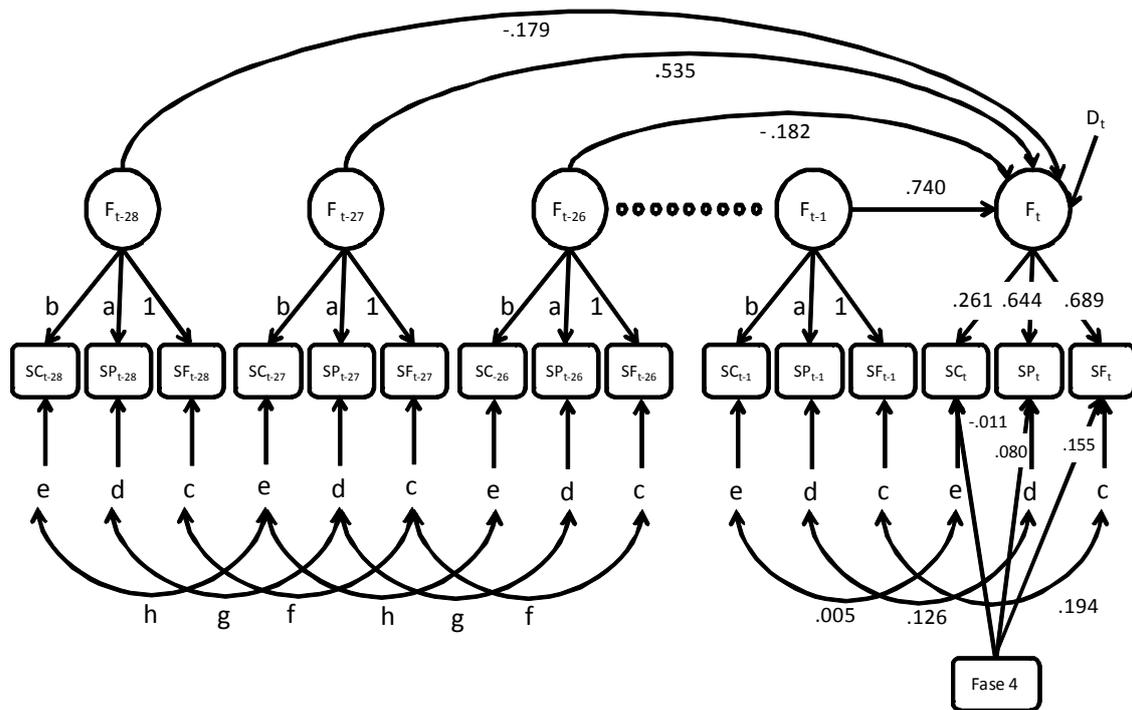


Figura 1. Modelo factorial dinámico de la sintomatología menstrual, (puntuaciones estandarizadas)

### Discusión

Las series temporales individuales permiten aprehender la génesis de la sintomatología menstrual. Que todos los modelos contengan un componente autorregresivo de orden uno, indicaría que la sintomatología se va modulando con una cadencia diaria. Los coeficientes obtenidos en modelos con componentes autorregresivos de orden 1 son, mayoritariamente, los de mayor intensidad, indicando la importancia de la memoria de la sintomatología entre días consecutivos.

Los modelos que poseen componentes estacionales están, en su mayoría, en el rango de 27-30 días, dentro de los valores en que, mayoritariamente, se encuentra la ciclicidad menstrual. Estos modelos representan el mayor porcentaje (66,66%) en la muestra, de lo que puede desprenderse la adecuación de los modelos ARIMA como instrumento sensible a la recurrencia de la sintomatología menstrual.

El modelo factorial dinámico (incluyendo a todas las participantes, analizando los datos mediante ‘pooled time series’ multinivel) indica la tendencia temporal general de la muestra, corroborando la existencia de los tres bloques de sintomatología (fisiológicos, psicológicos y conductuales), reforzando la idea de autocovariación entre cada uno de ellos, habiendo un sistema general (factor latente) que genera los síntomas, siendo el mejor indicador los síntomas fisiológicos (carga factorial estandarizada: 0,689), seguidos de los psicológicos (0,644) y de los conductuales (0,261). Los valores positivos de las cargas indican que cuando un tipo de síntoma aumenta, los otros también tienden a aumentar (y viceversa).

Se observa una autorregresión de retardo uno (la sintomatología presentada el día anterior determina la que se padecerá al día siguiente), con un valor del efecto entre las variables latentes de 0,740 ( $F_t = 0,740 \cdot F_{t-1}$ ), asumiendo los demás valores fijos.

En la muestra, hay una tendencia general autorregresiva estacional con retardos significativos en los días 26, 27 y 28, siendo el efecto sobre un día cualquiera (manteniendo los demás valores

constantes):  $F_t = -0,182 \cdot F_{t-26} + 0,525 \cdot F_{t-27} - 0,179 \cdot F_{t-28}$ , siendo los demás retardos no-significativos. Esto indica que la estacionalidad más influyente en nuestra muestra es de 27 días.

En los retardos 26 y 28 los coeficientes son negativos, pero ambos quedan mitigados por el peso de los coeficientes de los retardos 1 y 27, siendo sus valores absolutos bastante mayores. Si la regresión total de la puntuación factorial es:

$$F_t = 0,740 \cdot F_{t-1} - 0,182 \cdot F_{t-26} + 0,525 \cdot F_{t-27} - 0,179 \cdot F_{t-28},$$

en el caso de que las puntuaciones factoriales fuesen iguales (mediante el procedimiento estadístico de ‘punto de equilibrio’ (Huckfeldt, Kohfeld y Likens, 1982):  $F_{t-1} = F_{t-26} = F_{t-27} = F_{t-28} = F^*$ , entonces:  $F_t = 0,740 \cdot F^* - 0,182 \cdot F^* + 0,525 \cdot F^* - 0,179 \cdot F^* = 0,904 \cdot F^*$ , es decir, el conjunto de la sintomatología previa (en las variables latentes) tiene un peso esperado de 0,904 desviaciones típicas respecto a la media de síntomas presentados previamente. Nótese que la serie tiene una elevada ‘memoria’, pues casi se repiten los valores previos ( $F_t = 0,904 \cdot F^*$ ).

También se incluyeron las fases del ciclo, como variables ‘dummy’, influyendo en cada valor  $Y_t$  de los tres bloques de síntomas, dando sólo significativo el efecto de la fase premenstrual sobre las tres variables observables (Figura 1: Fase 4). La fase premenstrual aumenta los síntomas fisiológicos y psicológicos, pero disminuye los conductuales. Durante las demás fases, la variación en las tres variables está influida (estadísticamente, ‘memorizada’) sólo por los retardos 1, 26, 27 y 28, mientras durante cualquier día, en fase premenstrual, los síntomas presentados son función de los mismos retardos (1, 26, 27 y 28), más por el efecto de los coeficientes de la fase premenstrual.

Los resultados no confirmados sobre los valores aleatorios de los coeficientes multinivel indican (pese a la gran variabilidad entre mujeres) que el modelo autorregresivo dinámico encontrado posee “alta memoria”, y, por sí solo, puede pronosticar adecuadamente los valores de cada mujer.

Desde la perspectiva estadística, los modelos ARIMA tienen el inconveniente de permitir sólo el modelado de una única variable en función de valores de la misma variable retardada, pero los modelos autorregresivos (de regresión múltiple o de ecuaciones estructurales) realizados ‘artesanalmente’ poseen la ventaja de que permiten incluir tanto valores anteriores de la misma variable, como otras variables temporales o transversales. El desarrollo de la investigación longitudinal hará que los analistas de datos presten cada vez más atención a este tipo de modelos factoriales dinámicos.

## Referencias

- Abplanalp, J.M. (1983). Psychological components of the premenstrual syndrome. Evaluating the research and choosing the treatment. *Journal of Reproductive Medicine*, 28, 517-524.
- Abraham G.E. (1984). Nutrition and the premenstrual tension syndromes. *Journal of Applied Nutrition*, 36, 103-24.
- Bäckström, T. y Mattson, B. (1975). Correlation of symptoms in premenstrual tension o estrogen and progesterone concentrations in blood plasma. *Neuropsychobiology*, 1, 80-86.
- Barnard, N.D., Scialli, A.R., Hurlock, D. y Bertron, P. (2000). Diet and sex-hormone binding globulin, dysmenorrhea, and premenstrual symptoms. *Obstetrics & Gynecology*, 95, 245-250.

- Box, G.E.P. y Jenkins, G.M. (1970) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day
- Box, G.E.P., Jenkins G.M. y Reinsel, G.C. (2008). *Time Series Analysis: Forecasting and Control* (4th ed.). New York: John Wiley.
- Curran, P. J. y Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. En L. Collins and A. Sayer (Eds.), *New Methods for the Analysis of Change* (pp. 105-136). Washington, DC: American Psychological Association.
- Dalton, K. (1984). *The Premenstrual Syndrome and Progesterone Therapy*. (2ª.ed). Chicago: Williams Heineman Medical Book.
- Ganong, W.F. (1982). *Fisiología Médica* (8ªEd). México: El Manual Moderno.
- Jara, P. (1995). *Las series temporales: Aplicación de los modelos ARIMA para el análisis de la sintomatología del ciclo menstrual*. Tesis Doctoral. Universitat Jaume I. Castellón.
- Jara, P., Rosel, J. y Palmero, F. (1998) Análisis de la evolución de la sintomatología del ciclo menstrual mediante modelos ARIMA. *Psicothema*, 10, 193-205.
- Frees, E.W. (2004). *Longitudinal and panel data: analysis and applications in the social sciences*. Cambridge: Cambridge University Press.
- Huckfeldt, R.R., Kohfeld, C.W. y Likens, T.W. (1982). *Dynamic modelling*. Newbury Park, CA: Sage.
- Lentz, G.M. (2007). Primary and secondary dysmenorrhea, premenstrual syndrome, and premenstrual dysphoric disorder: etiology, diagnosis, management. En: Katz VL, Lentz GM, Lobo RA, Gershenson DM. *Comprehensive Gynecology*. (5th ed). Philadelphia, PA: Mosby Elsevier.
- Moore, K., Osgood, D., Larzelere, R. y Chamberlain, P. (1994). Use of pooled time series in the study of naturally occurring clinical events and problem behavior in a foster care setting. *Journal of Consulting and Clinical Psychology*, 62, 718-728.
- Moos, R.H. (1968). The development of a menstrual distress questionnaire. *Psychosomatic Medicine*, 30, 853-867.
- Plewis, I., Vitaro, F. y Tremblay, R. (2006). Modelling repeated ordinal reports from multiple informants. *Statistical Modelling*, 6, 251-263.
- Soliday, E., Moore, K. y Lande, M. (2002). Daily reports and pooled time series analysis: Pediatric psychology applications. *Journal of Pediatric Psychology*, 27, 67-76.
- Stephenson, L.A., Denney, D.R. y Aberger, E.W. (1986). Factor structure of the menstrual symptoms questionnaire: Relation to oral contraceptives, neuroticism and life stress. *Behavioral Research Therapy*, 21, 129-135.
- Sun, Y. y Li, Y. (2002). Children's well-being during parents' marital disruption process: A pooled time-series analysis. *Journal of Marriage and Family*, 64, 472-488.

- Van der Ploeg, H.M. (1990). The factor structure of the Menstrual Distress Questionnaire Dutch. *Psychological Reports*, 66, 707-714.
- Walls, T.A. y Schafer, J.L. (2006). *Models for intensive longitudinal data*. New York: Oxford University Press.
- Webster,SK. (1978). Problem for diagnosis of spasmodic and congestive dysmenorrhea. En A.Dan, E.Graham y C.Beecher (Ed.): *The Menstrual Cycle: A Synthesis of Interdisciplinary research*. New York: Springer.
- Yonkers, K., O'Brien, P. y Erikson, E. (2008). Premenstrual syndrome. *Lancet*, 371, 1200-1210.

# APLICACIONES DE LA TEORÍA DE LA MEDIDA Y DE LOS MÉTODOS DE ANÁLISIS DE DATOS A LA INVESTIGACIÓN EN PSICOLOGÍA DE LAS ORGANIZACIONES

**Coordinadora: Esther Torres**  
Universidad del País Vasco

En las últimas décadas se han desarrollado importantes avances en la teoría de la medida y en los métodos para el análisis de datos, que contribuyen a incrementar la validez de las investigaciones y responder a nuevas necesidades que se plantean en las investigaciones aplicadas que se llevan a cabo dentro las ciencias sociales. En Psicología se han aplicado en las distintas áreas substantivas como en psicología clínica, educativa, del deporte, psicología social o psicología de las organizaciones.

Uno de los procedimientos más extendido para la recogida de datos en las ciencias sociales y de la conducta es el desarrollo y aplicación de cuestionarios, tests o inventarios, es decir por medio de preguntas estructuradas. Los análisis de las respuestas dadas por los participantes y las conclusiones extraídas, históricamente han sido tratados desde la perspectiva clásica, en la que los resultados la medición dependen de las características del cuestionario empleado y de los sujetos de los responden. Esta dependencia mutua, ha sido ampliamente criticada por no cumplir los criterios de la teoría de la medida (Mitchell, 1997). Por lo tanto en el área de la medición de las variables del comportamiento tanto individual como grupal, el principal problema que se intenta resolver es obtener medidas psicológicas (variables inobservables directamente) independientes del instrumento empleado e instrumentos de medida independientes de los sujetos que lo responden. Esto es posible debido a los avances provenientes de la teoría de la medición, de la aplicación de modelos matemáticos y de los métodos de análisis de datos. Testimonio de estos avances se encuentran publicados en numerosos libros y distintas revistas de reconocido prestigio como *Psychometrika*, *Applied Psychological Measurement*, *Journal of Educational Measurement*, *Psychological Methods*, *Multivariate Behaviour Research*, *Educational and Psychological Measurement*, *Journal of Behavioral Statistics* o *Journal of Applied Psychology* entre otras.

Desde la psicometría se intenta dar respuesta a esta demanda con los modelos desarrollados especialmente dentro de la Teoría de respuesta al ítem (TRI). La importancia y las posibilidades que ofrecen estos modelos, junto con el desarrollo del software cada vez más específico y manejable, ha hecho que su aplicación y utilidad se extienda más allá de las variables cognitivas e ítems de respuesta dicotómica (verdadero – falso) en las que las respuestas son consideradas como una medida indirecta de una habilidad latente, y sean aplicados también para la medición de constructos de contenido actitudinal o de personalidad que requieren más de una dimensión para ser explicados; y con distintos formatos de respuesta, preferentemente ítems categóricos. Los modelos politómicos de respuesta al ítem están siendo aplicados en la medición de variables provenientes de las ciencias de la salud, ciencias sociales o de la economía.

El propósito de este simposium es poner de relieve la importancia de la aplicación de modelos de medida y métodos de análisis de datos más sofisticados en la investigación de la Psicología de las Organizaciones, con la intención de contribuir a extraer más información de los datos que los métodos más simples, cuando los problemas o cuestiones a los que se quiere dar respuesta son también más complejos. Evidentemente esta posición no va en detrimento de la aplicación de métodos más simples, que en ocasiones son los necesarios y los suficientes, pues no hay que olvidar el principio de la

parsimonia, y que pueden ser el primer paso en el proceso de la investigación. El investigador aplicado en Psicología del Trabajo y de las Organizaciones se enfrenta a necesidades cada vez más complejas y requiere evaluar modelos teóricos, describir, predecir y entender la conducta individual y grupal para una mejor toma de decisiones.

Los procedimientos para medir y analizar las respuestas de los participantes en distintos contextos están mejorando debido a los avances alcanzados por los psicómetras y estadísticos aplicados, entre otros. En la medición de las variables, tanto discretas como continuas, se ha avanzado en los procedimientos orientados a asegurar la calidad de los ítems, de los tests y escalas, para evaluar el sesgo de la medida, el funcionamiento diferencial del ítem y del test, la fiabilidad y el error de medida y la dimensionalidad. Mejoras orientadas para un mejor cumplimiento de los requerimientos de la teoría de la medición.

En lo referente a la calidad de los ítems se han publicado numerosos trabajos sobre la calidad de su redacción (Moreno, Martínez y Muñiz, 2006) o el número de alternativas más apropiado (López Pina, 2005). En psicometría se han llevado a cabo investigaciones que aplican procedimientos de la TRI en temas con implicaciones substantivas relacionadas con la medición de la personalidad (Gray-Little, Williams y Hancock, 1997) y con medición de las actitudes (Roberts y Laughlin, 1996, 2004). Entre los aplicados se encuentran los modelos de respuesta a ítems politómicos: como los de respuesta nominal, respuesta graduada, de clasificación o crédito parcial (Andersen, 1977, Andrich 1978, Bork, 1972, Rasch, 1961, Samejima, 1969, 1972, 1997). Estos modelos, a parte, de calibrar conjuntamente la calidad del ítem y el nivel del rasgo para determinar la calidad y utilidad del ítem, tienen implicaciones sobre la construcción de los tests y la forma de entender las propiedades psicométricas como el funcionamiento diferencial del ítem o el ajuste del error (Fox y Glas, 2003). Igualmente tienen implicaciones en la interpretación de las puntuaciones de los sujetos.

Del mismo modo, en las investigaciones aplicadas, resultan de máxima utilidad los modelos estructurales de variables latentes para analizar la relación entre múltiples variables y poner a prueba los modelos teóricos analizando simultáneamente varias ecuaciones de regresión o para medir el cambio en el individuo. Los procedimientos de ecuaciones de modelos estructurales generalmente han sido desarrollados para el análisis de medidas continuas, pero los datos reales son muchas veces de naturaleza discreta. Esta situación ha cambiado incluyendo en los paquetes de software opciones para analizar variables discretas.

En otras situaciones prácticas interesa analizar la conducta del individuo que está interrelacionada, acoplada en un equipo o grupo. En este caso analizar la conducta del individuo en conjunción con el grupo requiere un nivel de análisis que diferencie entre los niveles en la organización, por lo que son útiles los modelos jerárquicos.

Otro avance metodológico importante es la simulación cuando el investigador tiene dificultad para obtener los datos empíricos. Esta técnica permite llegar a nuevos conocimientos a partir de potenciales relaciones entre las variables.

En resumen, en este simposio se pretende por un lado, poner de relieve, con algunos casos prácticos, la estrecha relación que existe entre los avances en las técnicas de investigación y el progreso a nivel de explicación y teorización de las áreas aplicadas en Psicología. Y por otro, cómo esta vinculación sirve de estímulo a los investigadores en metodología para seguir avanzando en las técnicas para dar respuesta a las demandas que provienen de los investigadores aplicados.

# VENTAJAS DE LA APLICACIÓN DE LOS MODELOS POLITÓMICOS EN LA MEDICIÓN DE LA ACTITUD EMPRENDEDORA

Esther Torres, Mikel Aranburu y Daniel Hermosilla  
Universidad del País Vasco

Los avances alcanzados en la medición de los constructos psicológicos son propiciados, la mayor parte de ellos, por los modelos psicométricos derivados desde Teoría de la Respuesta al Ítem (TRI). Estos modelos, aunque en un principio eran más apropiados para la medición de constructos cognitivos y unidimensionales, más recientemente se han desarrollando y adecuado para la medición de variables sociales como las actitudes y constructos de personalidad de naturaleza multidimensional. Estos modelos son más acordes para analizar ítems con formato de respuesta politómica, y que en general es el formato más utilizado en ese tipo de variables. Sin embargo, a pesar de reconocer las ventajas de estos modelos en la medición de las variables psicológicas todavía es escasa su aplicación a datos empíricos, predominando el uso de la Teoría Clásica de los Test (TCT) en las investigaciones aplicadas y reservándose los modelos derivados de la TRI para la simulación y en el ámbito académico.

El objetivo de este trabajo es poner de manifiesto las ventajas de la aplicación del Modelo de Respuesta Graduada (MRG) (Samejima, 1969) a un cuestionario de nueva creación que tiene como objetivo evaluar la Actitud Emprendedora en Jóvenes.

La persona emprendedora sigue siendo objeto de estudio desde las perspectivas económica, sociológica y psicológica. La sociología trata de identificar los factores culturales, familiares y situacionales (Herron y Robinson 1993; Van de Ven 1993) que pueden influir en la iniciación de la actividad emprendedora. Pero es evidente que bajo circunstancias parecidas no todas las personas se convierten en emprendedoras. Desde la psicología se incorporación modelos psicológicos que destacan los rasgos de la personalidad como factor importante en el proceso de emprendizaje (Durham University Business School, 1988; Koh, 1996; Hill y McGowan, 1999; Cromie y Johns, 1983). En los análisis de las características de la persona destacan los estudios que ponen de relieve la importancia de algunos atributos por manifestarse de forma predominante en las personas consideradas emprendedoras, como son la propensión al riesgo, la necesidad de logro o el locus de control (Bromiley y Curley, 1992; McClland, 1961; Torres, 2006). Estos atributos se relacionan con la tendencia a explorar nuevas oportunidades, con la creación de ideas y desarrollo de las mismas. Mientras que las personas que los manifiestan en menor grado tienen más difícil en alcanzar dichas oportunidades, Corner y Pavlovich (2007), Michael, Harris, y Gibson. (2008).

Desde un enfoque conductual, el emprendizaje se considera como una actitud y sería el resultado de un proceso interactivo entre la conducta y el ambiente.

Generalmente los cuestionarios elaborados para evaluar la actitud emprendedora presentan ítems con un formato de respuesta tipo Likert y en los análisis los datos han sido analizados desde el paradigma de la TCT, por lo las puntuaciones son tratadas como una variable continua y se asume una escala de intervalos iguales, aunque de hecho sean una variable discreta.

Los ítems politómicos requieren funciones específicas para estimar la probabilidad de responder a una categoría. Esta función de respuesta al ítem (FRI) dicotomiza la respuesta en el umbral de la categoría. En los modelos politómicos se obtiene la función de información de respuesta a la categoría

(FIRC). Para los modelos de respuestas politómicas el poder discriminativo de cada función de respuesta categórica del ítem (FRC) depende de la combinación de la pendiente del ítem y de los parámetros de los umbrales.

El objetivo de este trabajo es analizar las propiedades psicométricas de un cuestionario de nueva creación orientado a la evaluación de la Actitud Emprendedora en Jóvenes universitarios (CAEJ) aplicando el Modelo de Respuesta Graduada (MRG) de Samejima (1969).

## Método

### *Procedimiento*

Esta investigación se llevó a cabo en el Campus Universitario de Gipuzkoa de la Universidad del País Vasco durante el curso 2007-2008. A los participantes se les presentó un cuadernillo bilingüe euskera-castellano y respondían en el idioma acorde a su dominio.

En la primera parte, se presentan las cuestiones referidas a las variables sociodemográficas y las consideradas más relevantes para el estudio del emprendizaje en jóvenes. En la segunda parte se presentaba el cuestionario de Actitud Emprendedora en Jóvenes universitarios (CAEJ) formado por 44 ítems.

### *Participantes*

Se seleccionó una muestra estratificada aleatoria por titulación, campo científico, grado y curso del Campus de Gipuzkoa. No se tuvieron en cuenta los estudiantes de doctorado ni postgrados. Respondieron al cuestionario 1070 sujetos, y el error muestral es del tres por ciento.

En las licenciaturas se recogieron datos de los estudiantes matriculados en 2º curso y en los últimos cursos 4º o 5º, según la licenciatura fuera de 4 o 5 años; en las diplomaturas respondieron al cuestionario los de 1º y 3º curso. La media de la edad fue 22,12 años y la desviación típica 4,07. En género estaba repartido entre el 63,08 de varones y el 36,82 de mujeres.

### *Instrumento*

El cuestionario CAEJ es una adaptación del Cuestionario del Perfil Emprendedor (CPE) (Torres, Aranburu, Ayestarán, 2004) a la población estudiantil. El CPE es un instrumento destinado a la evaluación de la actitud emprendedora.

Para la adaptación se eliminaron aquellas cuestiones que por su contenido no correspondían al ámbito de los estudiantes. El cuestionario resultante lo formaban 44 cuestiones a las que había que responder utilizando una escala de cinco categorías de respuesta ordenadas (1 = Totalmente en desacuerdo a 5 = Totalmente de acuerdo).

## Resultados

El análisis psicométrico del cuestionario se realiza en dos fases. En la primera, se aplica el modelo de TCT para realizar una primera depuración del mismo. En la segunda fase se aplicó el MRG. Se empleó el programa MULTILOG 7 (Thissen, Chen y Bock, 2003) empleando el método de máxima verosimilitud.

*Aplicación de la TCT:* El cuestionario presenta un coeficiente de fiabilidad ( $\alpha$ ) de 0,89. Se retiraron cuatro ítems por no obtener un índice de homogeneidad superior a 0,25.

El Análisis Factorial Exploratorio se realizó a partir de la matriz de correlaciones policóricas obtenidas con el programa FACTOR 7.0. La adecuación de la matriz está comprobada por el determinante = 0,000135 y por la prueba KMO = 0.89423. El método de extracción de factores fue el de mínimo cuadrado no ponderado, con rotación Varimax. Se obtuvieron ocho factores que explicaban el 57,55% de la varianza total. Los estadísticos de ajuste obtenidos fueron el: Simplicity Index de Bentler =: 0,59313 y el Loading Simplicity Index =: 0,35830; en el análisis de residuos se obtuvo RMSR = 0.0223 que era menor que el valor esperado para ser el modelo aceptable = 0,0335, según el criterio de Kelly. De la solución factorial obtenida se conservaron los ítems con una carga factorial  $a_{jF}$  > 0.35. El cuestionario resultante quedó reducido a 30 ítems. Los 10 ítems eliminados se repartieron entre los distintos factores definidos a priori.

Se comprobó la dimensionalidad del cuestionario con los criterios de Lord (1980), Carmines y Zaller (1979) y Rackase (1979) y según los resultados obtenidos podemos considerar que el cuestionario es unidimensional. El primer factor explica el 25,15%, el segundo el 6,19% y el tercero el 5,92% de la varianza.

*Aplicación de modelos de TRI politómicos:* estos modelos permiten estimar la distancia entre las categorías de respuesta, verificar el rango del rasgo medido, el nivel de discriminación y la función de información de los ítems y del test. En este modelo el poder discriminativo viene dado por la función de respuesta a cada categórica del ítem (FRC). La respuesta discreta del sujeto depende de la posición relativa de su valor  $s$  con respecto a una serie de  $K - 1$  umbrales de paso.

En este modelo la probabilidad de escoger una determinada categoría de un ítem depende de la combinación de la pendiente del ítem y de los parámetros de los umbrales entre las categorías. Puede calcularse a partir de la siguiente función logística:

$$P_{jk}(\theta) = \frac{1}{1 + e^{[Da_j(\theta - b_{jk})]}} - \frac{1}{1 + e^{[Da_j(\theta - b_{jk+1})]}}$$

donde  $a_j$  es el parámetro de la pendiente para cada ítem,  $b_{jk}$  es el parámetro de cada punto de corte entre categorías para cada ítem y  $D$  es la constante 1,7.

*Estimación de los parámetros de los ítems:* se agruparon las categorías 1 y 2, debido a que la categoría 1 fue elegida por menos del 5% en casi todos los ítems. Quedando los ítems con cuatro categorías.

El análisis se realizó con los 887 patrones de respuesta. En la tabla 1 se presentan los parámetros  $\alpha$  y los tres umbrales de los 30 ítems.

Tabla 1. parámetro  $a$  y umbrales  $b$ 

	$a$ (SE)	$b_1$ (SE)	$b_2$ (SE)	$b_3$ (SE)
<b>I01</b>	1.42 (.11)	-1.81 (.15)	0.00 (.07)	1.81 (.14)
<b>I02</b>	1.46 (.11)	-1.52 (.12)	-0.24 (.07)	1.58 (.13)
<b>I03</b>	0.72 (.09)	-2.48 (.33)	-0.27 (.14)	3.06 (.37)
<b>I04</b>	0.92 (.09)	-2.05 (.23)	-0.33 (.11)	1.84 (.20)
<b>I05</b>	0.94 (.09)	-1.04 (.14)	0.65 (.11)	2.87 (.29)
<b>I06</b>	1.00 (.09)	-1.97(.20)	-0.04(.10)	2.11 (.21)
<b>I07</b>	1.23 (.10)	-1.63 (.15)	0.18 (.08)	2.43 (.20)
<b>I08</b>	1.13 (.10)	-1.70 (.16)	0.72 (.10)	2.71 (.24)
<b>I09</b>	0.81 (.09)	-3.05 (.35)	-0.35 (.12)	2.87 (.32)
<b>I10</b>	0.88 (.09)	-1.76 (.21)	0.49 (.12)	3.62 (.37)
<b>I11</b>	1.01(.09)	-2.71 (.28)	-0.68 (.11)	1.90 (.18)
<b>I12</b>	0.75 (.08)	-1.45 (.21)	0.29 (.13)	3.17 (.35)
<b>I13</b>	1.48 (.11)	-2.48 (.20)	-0.47 (.08)	2.17 (.16)
<b>I14</b>	<b>1.54 (.11)</b>	-1.60 (.13)	-0.09 (.07)	2.14 (.15)
<b>I15</b>	1.23 (.10)	-2.54 (.22)	-0.87 (.10)	2.03 (.17)
<b>I16</b>	0.88 (.09)	-2.44 (.27)	-0.58 (.12)	1.67 (.19)
<b>I17</b>	1.28 (.11)	-2.60 (.22)	-0.97 (.10)	1.68 (.15)
<b>I18</b>	1.23 (.11)	-2.69 (.24)	-1.23 (.12)	0.93 (.11)
<b>I19</b>	1.07 (.10)	-2.02 (.20)	-0.38 (.10)	2.09 (.20)
<b>I20</b>	1.18 (.10)	-1.16 (.12)	0.15 (.08)	2.00 (.17)
<b>I21</b>	0.71 (.09)	-1.50 (.23)	0.92 (.16)	3.77 (.44)
<b>I22</b>	0.96 (.09)	-1.40 (.16)	0.08 (.10)	2.00 (.20)
<b>I23</b>	0.75 (.09)	-3.23 (.41)	-0.65 (.15)	1.82 (.23)
<b>I24</b>	0.84 (.09)	-3.31 (.37)	-1.12 (.16)	1.90 (.23)
<b>I25</b>	0.82 (.09)	-2.72 (.32)	-0.46 (.13)	2.75 (.30)
<b>I26</b>	0.70 (.08)	-1.04 (.18)	1.10 (.18)	4.23 (.47)
<b>I27</b>	0.83 (.09)	-3.30 (.40)	-1.24 (.17)	1.21 (.18)
<b>I28</b>	0.79 (.09)	-1.91 (.24)	0.95 (.15)	3.03 (.33)
<b>I29</b>	<b>0.54 (.08)</b>	-2.34 (.40)	0.93 (.21)	3.95 (.58)
<b>I30</b>	1.08 (.10)	-3.07 (.29)	-1.10 (.12)	2.00 (.19)

Podemos observar que la capacidad de discriminación de los ítems no es demasiado alta. El parámetro  $a$  varía entre 0.54 del ítem 29 y 1.54 del ítem 14 y el 53,33% de los ítems tienen valores inferiores a 1. Sin embargo los errores de estimación son relativamente pequeños. En cuanto a los valores de los parámetros  $b_k$  se encuentran distanciados en el rasgo, pero  $b_1$  y  $b_2$ , que representan la transición entre las categorías 1 – 2 y 2 - 3 indican niveles bajos en  $\theta$ . Sin embargo el tránsito entre la categoría 3 y 4 ( $b_3$ ) está muy espaciado respecto al  $b_2$  y el 66.66% de los ítems presentan un valor en  $b_3$  mayor o igual a 2.

*Función de Información:* este cuestionario ofrece mayor información para niveles inferiores a  $\theta = 0$ . En la tabla 2 se observa que es más preciso para el intervalo de  $\theta$  -1.4 y 0.0, es decir entre la media y una desviación y media por debajo.

Tabla 2. Función de información para los intervalos de  $\theta$ 

$\theta$	FUNCIÓN DE INFORMACIÓN							
-3.0 - -1.6	7.114	7.678	8.209	8.690	9.108	9.453	9.722	9.915
<b>-1.4 - 0.0</b>	<b>10.041</b>	<b>10.114</b>	<b>10.146</b>	<b>10.150</b>	<b>10.127</b>	<b>10.075</b>	<b>9.991</b>	<b>9.877</b>
0.2 - 1.6	9.745	9.616	9.513	9.455	9.445	9.474	9.515	9.532
1.8 - 3.0	9.485	9.341	9.082	8.706	8.229	7.681	7.092	

### Discusión

El objetivo de este trabajo ha sido comprobar las características psicométricas del cuestionario de nueva creación Actitud Emprendedora en Jóvenes aplicando el MRG, para obtener información sobre la distancia entre las cuatro categorías de respuesta y la información que ofrece.

Teniendo en cuenta los resultados obtenidos con la TCT y del análisis factorial se retiraron 14 ítems el cuestionario. El cuestionario resultante de 30 ítems se consideró unidimensional al cumplir de forma satisfactoria los criterios de Lord (1980), Carmines y Zaller (1979) y de Rackase, (1979).

Según los resultados obtenidos de aplicar el MRG, el test ofrece información para niveles bajos del rasgo ( $\theta = -1.4 - 0.0$ ). Los ítems poseen, en general, parámetros de discriminación bajos lo que indica que se diferencian poco las categorías de respuesta y en consecuencia una persona de nivel bajo podría seleccionar una categoría alta.

Los parámetros  $b_k$  están en secuencia cómo exige el modelo. Estos parámetros los señalan niveles del rasgo donde las FRCs alcanzan la máxima probabilidad de elección de una categoría. En nuestro caso, como se observa en la tabla 1, los parámetros  $b_1$  y  $b_2$ , que representan la transición entre las categorías 1 y 2 y entre la 2 y 3 tienen valores muy bajos y bajos en todos los ítems, mientras que pasar de la categoría 3 a la 4 requiere mayor nivel en el rasgo y resulta más difícil puntuar alto en los ítems.

A la luz de estos resultados es conveniente revisar el número de categorías de respuesta y el nivel de discriminación para conseguir que el cuestionario ofrezca información para intervalos más amplios en la actitud.

Agradecimientos. Los datos son parte del Proyecto “Evaluación de la actitud, cultura y de las habilidades emprendedoras en los estudiantes universitarios del Campus de Gipuzkoa” subvencionado por la Diputación Foral de Gipuzkoa en la convocatoria de Noviembre de 2007.

### Referencias

- Bromiley, P. y Curley, S. (1992). Individual differences in risk taking. En J.Yates (Eds.), *Risk taking behavior* (pp. 87-132). New York: Wiley.
- Carmines y Zaller (1979). *Reliability and Validity assessment*. Londres. Sage.
- Corner P. y Pavlovich K. (2007) Entrepreneurship Research: Follow the Yellow-Brick Road? *Journal of Management & Organization* (2007) 13, 288-294.

- Cromie, S. y Johns, S. (1983). Irish entrepreneurs: Some personal characteristics. *Journal of Organizational Behaviour*, 4, 317-324.
- Durham University Business School. (1988). *The General Enterprising Tendency test*. Durham, UK: Author.
- Herron, L. y Robinson, R.B. (1993). A structural model of the effects of entrepreneurial characteristics on venture performance. *Journal of Business Venturing*, 8, 282-294.
- Hill, J. y McGowan, P. (1999). Small business and enterprise development: questions about research methodology. *International Journal of Entrepreneurial Behaviour & Research*. Vol. 5 No 1, pp 5-18.
- Koh, H.C. (1996). Testing hypotheses of entrepreneurial characteristics. *Journal of Management Psychology*, 11, 12-25.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA.
- Lorenzo-Seva, U. y Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers*, 38(1), 88-91.
- McClelland D (1961). *The Achieving Society*. Princeton, NJ: Van Norstrand.
- Michael L. Harris y Shanan G. Gibson. (2008). Examining the entrepreneurial attitudes of US business students. *Education and Training*. Vol. 50 (7), 568 – 581.
- Rackase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics*, 4, 207-230.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, n° 17.
- Thissen, D., Chen, W-H, y Bock, R. D. (2003). *MULTILOG 7 para windows. Multiple-category item analysis and test scoring using item response theory*. Lincolnwood, IL: Scientific Software International, Inc.
- Torres, E. (2006). Perfil de la Persona Emprendedora. Cursos de Verano de San Sebastian, Julio (paper).
- Torres, E., Aranburu, M. y Ayestaran, S. (2004). Perfil de la Persona Emprendedora. SPRI. Gobierno Vasco (inédito).
- Van de Ven, A. (1993). The development of an infrastructure for entrepreneurship. *Journal of Business Venturing*, 8, 211-230.

# ¿INFERENCIA BAYESIANA O FRECUENTISTA EN EL ANÁLISIS DE DATOS?

Emilia I. de la Fuente, Luis M. Lozano, Gustavo R. Cañadas y Cristina Vargas  
Universidad de Granada

La Estadística Bayesiana propone una serie de procedimientos que permiten realizar el análisis de los datos de una investigación y cuyo rango de opciones posibles es tan amplio como el de las técnicas estadísticas clásicas. Es cierto que existen diferencias filosóficas importantes entre ambas escuelas estadísticas, no obstante, también lo es que son dos filosofías perfectamente compatibles. Esta última afirmación tiene como base que, en muchas ocasiones, la información inicial que se tiene es poco informativa y/o el tamaño muestral utilizado es muy grande. En esas situaciones, las conclusiones a las que se llega por ambos procedimientos son muy similares. Si las circunstancias son que existe conocimiento previo sobre el fenómeno que se estudia o las muestras son pequeñas, hay razones que apoyan el uso de los procedimientos bayesianos. En este trabajo se presentan las inferencias obtenidas utilizando ambas filosofías en situaciones de datos obtenidos tras la simulación de las poblaciones correspondientes y en datos reales procedentes de investigaciones en curso.

El Análisis de Datos de investigación que se utiliza en Ciencias Sociales, combina la perspectiva de Neyman y Pearson con el planteamiento Fisheriano del p-valor, y permite, a partir de dicha combinación, hablar de una estructura frecuentista, en la que las probabilidades se obtienen como aproximación de las correspondientes frecuencias, tras una serie infinita de pruebas o réplicas con idéntico diseño de investigación.

Algunas características que asume el investigador que utiliza la estructura frecuentista, en el análisis de los datos de su investigación, son las siguientes:

- Los parámetros de la población que interesa estudiar o estimar son siempre fijos y no pueden cambiar, por ejemplo, a lo largo del tiempo.
- Nunca se tiene en investigación información previa ya existente, con la que poder contar en el análisis de la situación correspondiente.
- Sólo tienen valor real en sus conclusiones, después de un análisis estadístico, los experimentos, pruebas clínicas o ensayos clínicos controlados, y tan solo en aquellas situaciones en las que se respetan todos los requisitos metodológicos –muestreo, representatividad de la muestra, control de variables, etc–.
- Interesa más la “significación” que el “tamaño del efecto”.
- Los grandes computadores están, poco o nada, al alcance de los investigadores.

Si por el contrario, el investigador considera de interés otro tipo de cuestiones, puede considerar si le interesan las características que subyacen al uso de una estructura analítica bayesiana que son, entre otras, las siguientes (Bernardo y Smith, 2004; Gill, 2008):

- Existen supuestos básicos flexibles en los modelos de análisis.

- Se obtienen conclusiones probabilísticas acerca de las cantidades teóricas –parámetros- que resultan de interés en los modelos de análisis.
- Son procedimientos que permiten adaptar estas afirmaciones probabilísticas cuando se tiene nueva información –aprender de la experiencia-. Permiten incorporar el conocimiento previo sobre el tema, cuando exista.
- Se trabaja desde una perspectiva en la que se admite que las cantidades que nos interesa estudiar –parámetros poblacionales-, pueden ser cambiantes a lo largo del tiempo, en la población de interés.

En realidad, cada investigador debe plantearse cuál sería el marco en el que tiene más sentido desarrollar su trabajo y cuál de los dos bloques de características anteriores encaja mejor con la planificación, desarrollo y conclusión de su investigación. No parece acorde con la realidad investigadora defender la estaticidad de los parámetros poblacionales; afirmar que la información previa procedente de nuestro propio trabajo o de otras investigaciones nunca existe parece poco acorde con la realidad; que el experimento no es la única forma de investigación válida, parece claro; la defensa de la significación estadística como lo primordial en investigación, quizás en el momento actual, es, cuando menos, desfasado. Tampoco tienen ya demasiado sentido, afirmaciones obsoletas como las que se planteaban hace décadas y que iban en la línea de la inviabilidad de un análisis por las necesidades computacionales del mismo.

Como contrapartida, cada vez hay una apertura mayor hacia procedimientos con supuestos básicos más flexibles; “se agradece” cualquier información que se le ofrezca al investigador, añadida a la mera afirmación de si hay o no diferencias estadísticamente significativas entre sus grupos; frente a posturas estáticas de todo tipo, se encuentran alternativas en las que los modelos son dinámicos y, generalmente, jerárquicos, y permiten la incorporación de lo investigado previamente. Las conclusiones de tipo probabilístico, que eran interpretaciones erróneas en el esquema de análisis clásico, en el esquema bayesiano son realmente las que corresponde hacer. Es cierto que las necesidades computacionales son importantes, pero hoy se dispone, cada vez más, de la infraestructura adecuada en ese sentido.

Llegado este punto, se hace necesaria la pregunta: ¿por qué el uso de la perspectiva clásica estadística ha sido y aún en este momento es mayor que el de la filosofía bayesiana? Se quiere resaltar que ambas posturas metodológico-analíticas, se utilizan cada vez más, como corresponde, de forma complementaria. Tradicionalmente se ha argumentado en contra de la postura bayesiana en el sentido de: (1) muchos investigadores tienen prejuicios porque consideran que la subjetividad que lleva inherente el uso de la filosofía bayesiana parece dificultar su manejo, (2) parece que el formalismo matemático que conllevan los procedimientos bayesianos es mayor y, (3) la obtención de soluciones bayesianas en la estimación, ajuste y validación de modelos, es difícil si no se dispone de grandes posibilidades computacionales. En la actualidad parece claro que ninguna de estas razones se puede mantener.

El objetivo de este trabajo es presentar y comparar los resultados obtenidos en la resolución de algunos problemas de inferencia básica, mediante procedimientos estadísticos frecuentistas y bayesianos. Se presentan ejemplos de estimación del parámetro “proporción”, así como la resolución de un contraste de hipótesis, tanto con poblaciones de datos simuladas –estudio 1–, como con muestras de datos reales –estudios 2 y 3–.

## ESTUDIO 1

### Método

Se presenta un estudio de simulación en el que se realizan las estimaciones del parámetro prevalencia de un trastorno mediante el método de Máxima Verosimilitud (MV), el método Bayes con distribución inicial no informativa (No Inf) y el método Bayes con distribución inicial informativa (Inf).

### *Procedimiento*

Población Binomial (presencia o ausencia del trastorno). Se simulan cuatro poblaciones con prevalencias respectivas 0,1; 0,3; 0,5 y 0,7.

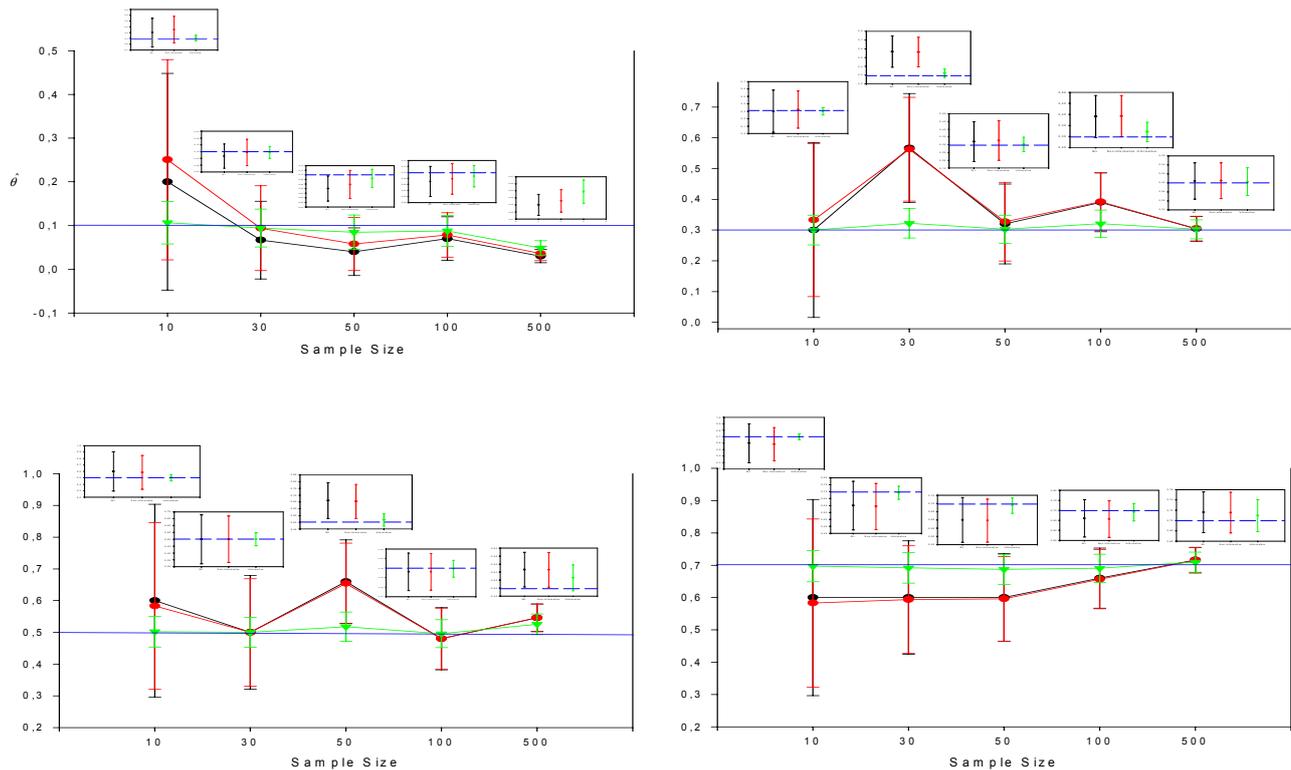
Con el programa S -PLUS 8.0 se simularon muestras de tamaños 10, 30, 50, 100 y 500 de cada una de las poblaciones anteriores. Se trata de un diseño con 20 niveles de simulación.

### Resultados

Se realiza la estimación clásica por el procedimiento de máxima verosimilitud e intervalo de estimación en todas las condiciones, con el programa STATGRAPHIC PLUS. Se realiza la estimación bayesiana de la prevalencia, puntual y mediante intervalo, por dos procedimientos diferentes. De una parte se calcula la estimación bayesiana con una distribución inicial no informativa Beta(1, 1). De otra, se calcula la estimación bayesiana de dicho parámetro, con información inicial conocida. Se programó como información inicial un intervalo de valores de cuatro desviaciones típicas alrededor del verdadero valor de prevalencia programado en cada una de las poblaciones, siguiendo las sugerencias de Congdon (2003). Las estimaciones bayesianas se obtuvieron con el programa WinBugs (Lunn, Thomas, Best y Spiegelhalter, 2000).

En éste, como en los estudios posteriores, la distribución final se obtiene transformada mediante el teorema de Bayes, a partir de la distribución inicial. Para la obtención de dicha distribución se emplean 99000 remuestreos Gibbs con eliminación de los 1000 primeros para evitar el problema de autocorrelación. Se utiliza la función de pérdida cuadrática, por lo que el estimador Bayes es la media de la distribución final.

Las estimaciones puntuales, así como los intervalos de estimación en las diferentes condiciones de simulación se representan en las Figuras 1 a 4. En dichas figuras se señala con una línea horizontal el valor real de la prevalencia programada en la simulación para la población.



Figuras 1 a 4: Estimación puntual y por intervalos de  $\theta$ . Línea negra: Estimación Frecuentista, Línea roja: Estimación bayesiana no Informativa, Línea verde: Estimación Bayesiana Informativa

## Conclusión

Como se esperaba, los resultados más similares, tanto en la estimación puntual como en la amplitud de intervalo de estimación, se producen cuando la muestra es grande ( $n=500$ ). En tamaños muestrales inferiores, las amplitudes de los intervalos de estimación difieren de unos métodos a otros de forma que en algunas de las poblaciones simuladas, los intervalos ni tan siquiera incluyen en su rango el verdadero valor del parámetro poblacional.

## ESTUDIO 2

Se obtienen las estimaciones e intervalos de estimación de la prevalencia de ansiedad en niños, mediante los tres procedimientos utilizados en el estudio 1, -método de Máxima Verosimilitud (MV), método Bayes con distribución inicial no informativa (No Inf) y método Bayes con distribución inicial informativa (Inf)-. En este caso se utilizan datos reales.

### Participantes

La muestra está compuesta por 500 niños que cursan sus estudios en centros de primaria del Principado de Asturias. Los centros fueron seleccionados al azar respetando la proporción de centros públicos y concertados. El rango de edad va desde los 8 hasta los 12 años con una media de 10.33 años y una Desviación Típica de 1.45. Del total de la muestra se seleccionaron aleatoriamente muestras de 10, 30, 50, 100 y 500 niños.

## Material

A todos los niños se les aplicó el CECAD destinado a evaluar ansiedad, depresión, inutilidad, irritabilidad y problemas de pensamiento. Para este trabajo sólo se tienen en cuenta las puntuaciones obtenidas en la subescala de ansiedad.

## Resultados

Se utiliza el programa SPSS 15.0, para realizar la estimación máximo verosímil, así como su intervalo de confianza asociado, del parámetro prevalencia de ansiedad.

Se utiliza el programa WinBugs para obtener la estimación bayesiana de la prevalencia de ansiedad, puntual y mediante intervalo, tanto en el caso no informativo, como en el caso de información inicial conocida. Se sigue idéntico procedimiento al expuesto en el estudio 1.

Se utiliza la distribución inicial no informativa Beta(1,1) y se genera la distribución inicial informativa con el procedimiento sugerido por Congdon (2003), a partir de la prevalencia del trastorno ansioso, informada por Bragado, Carrasco y Sánchez-Bernardos (1995) para la población de niños de estas edades.

Las estimaciones, tanto puntuales como mediante intervalos se incluyen en la tabla siguiente.

*Tabla de Prevalencia de sintomatología ansiosa con datos reales, en niños de 8 a 12 años, por tres procedimientos de estimación*

		<b>n=10</b>	<b>n=30</b>	<b>n=50</b>	<b>n=100</b>	<b>n=500</b>
Máxima Verosimilitud	Estimación	0,100	0,333	0,160	0,250	0,200
	Límite Inferior	-0,086	0,165	0,058	0,165	0,165
	Límite Superior	0,286	0,502	0,262	0,335	0,235
Bayesiano No-Informativo	Estimación	0,167	0,344	0,173	0,255	0,201
	Límite Inferior	0,023	0,192	0,084	0,175	0,167
	Límite Superior	0,413	0,514	0,286	0,343	0,237
Bayesiano Informativo	Estimación	0,225	0,239	0,219	0,235	0,211
	Límite Inferior	0,179	0,193	0,176	0,194	0,183
	Límite Superior	0,274	0,288	0,265	0,278	0,240

## Conclusión

Las estimaciones puntuales para la prevalencia poblacional, obtenidas a partir de las diferentes muestras consideradas, oscilan en un rango mayor cuando se trata del método máximo verosímil, presentan unos valores más próximos entre sí al ser estimados por el método Bayes sin información previa y son muy semejantes cuando se utiliza la información inicial y el método bayesiano informativo.

## ESTUDIO 3

Se comparan las puntuaciones medias en ansiedad en tres grupos de niños de diferentes niveles educativos, según el procedimiento frecuentista y el bayesiano.

### *Participantes*

La muestra está compuesta por 1574 niños de tres cursos diferentes (centros de primaria del Principado de Asturias). Los centros fueron seleccionados al azar respetando la proporción de centros públicos y concertados. El rango de edad va desde los 8 hasta los 12 años con una media de 9.45 años y una Desviación Típica de 0.986.

### *Material*

Al igual que en el estudio 2, se utilizan los datos recogidos con el CECAD; más concretamente, las puntuaciones obtenidas en la subescala de ansiedad.

## Resultados

Se utiliza el programa SPSS 15.0, para realizar el ANOVA comparando los tres grupos considerados.

Se utiliza el programa WinBugs para realizar el ANOVA bayesiano con distribución inicial no informativa en el que se comparan los tres grupos considerados y se siguen las pautas ya expuestas en los dos estudios anteriores.

Los resultados obtenidos indican que existen diferencias estadísticamente significativas entre los tres grupos considerados ( $F_{2,1571} = 17,639$ ;  $p < 0,0001$ ). El estudio clásico de diferencias entre los grupos, mediante la prueba de Scheffé indica que existen diferencias estadísticamente significativas entre las puntuaciones promedio en ansiedad en los cursos 3º y 5º ( $F_{2,1571} = 4,122$ ;  $p < 0,001$ ) y 4º y 5º ( $F_{2,1571} = 2,927$ ;  $p < 0,001$ ). No existen diferencias estadísticamente significativas en ansiedad entre los niños de los cursos 3º y 4º.

Los resultados del análisis bayesiano llevan a establecer diferencias con probabilidad 1 entre las medias de ansiedad de los cursos 3º y 5º y los cursos 4º y 5º. Así mismo, dichos análisis establecen que la probabilidad de que la ansiedad media de los niños de 3º supere a la media de los de 4º, es de 0,9524.

## Discusión y conclusión

Se ha utilizado una muestra muy grande y un análisis clásico frente al análisis bayesiano no informativo, lo que significa que los resultados, como era de esperar según indica la teoría, son muy parecidos. El análisis clásico indica que existen diferencias estadísticamente significativas entre los niveles 3 y 5, y entre los niveles 4 y 5. No hay diferencia estadísticamente significativa en ansiedad entre los niños de los niveles 3 y 4.

El análisis bayesiano indica que con probabilidad 1, se dan diferencias entre los niveles 3 y 5, y entre los niveles 4 y 5. El procedimiento bayesiano da información añadida; establece que la probabilidad de que la media en ansiedad de los niños del nivel 3 supere a la ansiedad media de los del nivel 4, vale 0,9524.

Los resultados obtenidos son acordes con la teoría. Cuando se utiliza una muestra grande se obtienen estimaciones puntuales muy semejantes y rangos de estimación de tamaño muy similar. Sin

embargo, el uso de muestras pequeñas conlleva cambios en las inferencias finales tanto en el caso de estimación como en el de contraste de hipótesis.

En cuanto a la estimación con datos reales, hay que resaltar la estabilidad de las estimaciones puntuales obtenidas mediante el método bayesiano, sobre todo, en el caso de utilizar información previa. El valor del estimador puntual prácticamente no cambia, independientemente del valor del tamaño muestral. Además, el rango de los intervalos de estimación es sistemáticamente más pequeño en el procedimiento bayesiano, sobre todo, en muestras pequeñas.

Por último, parece destacable el interés que tiene la información que se obtiene de tipo complementario, cuando se utiliza el contraste bayesiano. En este caso, más que la simple decisión dicotómica, hay o no diferencias significativas, se nos informa de la probabilidad de las hipótesis en litigio y es tarea del investigador decidir si la probabilidad asociada a cada hipótesis es o no relevante.

Agradecimientos. El trabajo ha sido financiado por el proyecto SEJ2006-13009 del Ministerio de Educación y Ciencia y por el proyecto de excelencia P07HUM-02529 de la Junta de Andalucía.

### Referencias

- Bernardo, J.M. & Smith, A.F.M. (2004). *Bayesian Theory*. New York: Wiley.
- Bolstad, W.M. (2007). *Introduction to Bayesian Statistics*. New York: Wiley.
- Bragado, C. Carrasco, M.L. y Sánchez-Bernardos, M.L. (1995). Prevalencia de los trastornos psicológicos en niños y adolescentes. Resultados preliminares. *Clínica y Salud*, 6, 67-82.
- Congdon, P. (2003). *Bayesian Statistical Modelling*. New York: Wiley.
- De la Fuente, E.I., Cañadas, G.R., Guàrdia, J. y Lozano, L.M. (2009). Hypothesis Probability or Statistical Significance? *Methodology*, 5(1), 35-39.
- Gill, J. (2008). *Bayesian Methods*. Boca Raton: Chapman & Hall.
- Guàrdia, J. De la Fuente, E.I. y Lozano, L.M. (2008). Bayesian inference for binomial populations. bayesian estimation for child depression prevalence. *Advances and Applications in Statistics*, 9(1), 13-35.
- Lunn, D.J. Thomas, A. Best, N. y Spiegelhalter, D. (2000). WinBUGS. A Bayesian Modelling Framework: concepts, structure and extensibility. *Statistics and Computing*, 10, 325-337.

# ANÁLISIS CLÁSICO Y BAYESIANO EN EL ESTUDIO DIFERENCIAL DE LOS ÍTEMS

Luis M. Lozano<sup>1</sup>, Emilia I. de la Fuente<sup>1</sup>, María Martín<sup>2</sup> y Gustavo R. Cañadas<sup>1</sup>

<sup>1</sup>Universidad de Granada

<sup>2</sup>Universidad de Jaén

El estudio del Funcionamiento Diferencial de los Ítems es un paso básico que se debe realizar para aseverar que las inferencias que se extraen de un cuestionario son válidas. Una de las técnicas más flexibles a la hora de estudiar dicho funcionamiento es la regresión logística que permite, la detección del Funcionamiento Diferencial Uniforme así como del No Uniforme. Así mismo ofrece información sobre el tamaño del efecto de de dicho funcionamiento diferencial. Desde la perspectiva Bayesiana se puede realizar este análisis incorporando la información previa disponible (en caso de tenerla) y obteniendo estimaciones más precisas cuando el tamaño muestral es pequeño. La diferencia fundamental entre la perspectiva bayesiana y frecuentista en este análisis reside en el procedimiento de estimación de los parámetros así como en la toma de decisiones final a partir de los resultados.

Dentro de los múltiples tests con los que el psicólogo se enfrenta a su quehacer diario debe escoger el más adecuado para poder realizar inferencias, clasificaciones... del modo más correcto posible. El hecho de que un cuestionario no esté sesgado es de vital importancia para poder seleccionar y aplicar dicho test ya que si no cumple esta condición la validez de las inferencias del mismo se verá afectada y algún grupo puede ser evaluado de forma errónea (injusta), ya que se estarán cometiendo errores sistemáticos contra él. Por tanto el estudio del sesgo es de vital importancia para poder seleccionar los cuestionarios que poseen una mayor calidad.

Para poder detectar el funcionamiento diferencial de los ítems (DIF) existen múltiples técnicas. Desde la teoría Clásica de los Tests se pueden citar Mantel-Haenszel (Mantel, 1963; Mantel y Haenszel, 1959), la estandarización, el uso del SIBTEST (Shealy y Stout, 1993a,b), la regresión logística y desde la perspectiva de la Teoría de la Respuesta a los Ítems las medidas del área entre dos Curvas características del Ítem (Raju, 1988), ji-cuadrado de Lord, la comparación de los modelos...(véase Muñiz, 1996). Muchas de estas técnicas se utilizan en función del tipo de respuesta que tenga el cuestionario (dicotómico o politómico).

Este trabajo se centra en la técnica de la regresión logística para la detección del DIF (Swaminathan y Rogers, 1990). Este procedimiento tiene ciertas ventajas respecto al resto de técnicas anteriormente expuestas; no es necesario categorizar las variables cuando estas son continuas, puede detectar tanto el DIF uniforme como el no uniforme y se puede generalizar el procedimiento a ítems politómicos mediante una regresión logística ordinal (Zumbo, 1999).

Desde la perspectiva frecuentista el procedimiento para analizar el DIF empleando la regresión logística trata exclusivamente de ir incorporando de forma jerárquica variables predictoras a la ecuación. En primer lugar se debe realizar la regresión introduciendo el ítem analizado como variable dependiente y la puntuación total en el cuestionario como variable predictora, se observa el valor de  $\chi^2_1$  y de  $R^2_1$  (el subíndice 1 se refiere a que se analiza el primer modelo). Posteriormente se debe introducir en la regresión la variable en la que se sospeche que el ítem está actuando de forma diferencial, atendiendo a los valores de  $\chi^2_2$  y de  $R^2_2$ . Por último, se introduce la interacción entre la puntuación total en el cuestionario y la variable de agrupación obteniendo los valores de  $\chi^2_3$  y de  $R^2_3$ .

$$\begin{aligned}
 1) \ln \left[ \frac{p}{1-p} \right] &= b_0 + b_1 \text{Total} \\
 2) \ln \left[ \frac{p}{1-p} \right] &= b_0 + b_1 \text{Total} + b_2 \text{Grupo} \\
 3) \ln \left[ \frac{p}{1-p} \right] &= b_0 + b_1 \text{Total} + b_2 \text{Grupo} + b_3 (\text{Total} \times \text{Grupo})
 \end{aligned}$$

Para comprobar la existencia de DIF se restan las cantidades  $\chi^2_3$  y  $\chi^2_1$  siendo el resultado otra  $\chi^2$  con dos grados de libertad. Si el resultado obtenido es estadísticamente significativo se puede defender la existencia de DIF ya sea Uniforme o No Uniforme. Para comprobar qué tipo de DIF ocurre y en que cuantía sólo hay que comparar los valores de  $R^2$  de los tres modelos.

$$R_3^2 - R_1^2 \rightarrow \text{Tamaño del Efecto del DIF}$$

$$R_2^2 - R_1^2 \rightarrow \text{Tamaño del Efecto del DIF Uniforme}$$

$$R_3^2 - R_2^2 \rightarrow \text{Tamaño del Efecto del DIF No uniforme}$$

El análisis del Funcionamiento Diferencial de los Ítems también se puede realizar desde la perspectiva Bayesiana. Ésta permite la incorporación de información previa que se puede poseer y realiza estimaciones más precisas que la frecuentista cuando el tamaño muestral es pequeño (Guàrdia, de la Fuente y Lozano, 2008). Estas ventajas pueden ser de gran utilidad a la hora de construir cuestionarios. El hecho de no necesitar muestras excesivamente grandes puede permitir a los investigadores no “quemar” excesiva muestra en los estudios iniciales de la calidad de los ítems. Por otro lado, la posibilidad de incluir información inicial puede permitir realizar un trabajo más refinado, ya que la información final de diferentes estudios puede ser la inicial de otro, lo que permitirá ir aproximándose a estimaciones más precisas de los parámetros a estudio.

El modelo de regresión logística puede representarse como:

$$Y_i \sim \text{binomial}(\pi_i, N_i); \quad \log \left( \frac{\pi_i}{1-\pi_i} \right) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$$

La diferencia fundamental entre ambos enfoques se encuentra en el procedimiento de estimación de los parámetros de la regresión que se utiliza así como en la toma de decisiones final a partir de los resultados que se obtienen.

En el presente trabajo se ha empleado la distribución binomial, si bien otras distribuciones son empleadas con relativa frecuencia como la probit o los modelos log-log o clog-log (Ntzoufras, 2009).

El objetivo del presente trabajo es la comparación de ambos procedimientos en la detección del Funcionamiento Diferencial de los Ítems. Además se tratará de comprobar si el tamaño muestral del que se disponga para realizar el análisis puede afectar a los resultados obtenidos.

## Método

### *Procedimiento y Análisis*

Se utiliza el programa WINGEN (Han, 2007, Han y Hambleton, 2007) para simular la respuesta de 400 casos a 20 ítems. Uno de los 20 ítems se simuló siguiendo un Funcionamiento Diferencial Uniforme.

Del total de la muestra, 200 casos constituyen el grupo de referencia y los 200 restantes el grupo focal.

Para comprobar si el tamaño muestral afecta a los resultados obtenidos se seleccionan de forma aleatoria con la rutina programada para ello en el SPSS 15.0 200 casos (100 del grupo de referencia y 100 del grupo focal), 100 (50 del grupo de referencia y 50 del grupo focal) y 50 casos (25 del grupo de referencia y 25 del grupo focal).

Los análisis frecuentistas fueron realizados con el SPSS 15.0 y los bayesianos mediante el WinBugs (Spiegelhalter, Thomas, Best y Gilks, 1994a, 1994b). En el análisis bayesiano se utilizan 28000 remuestros Gibbs eliminando los primeros mil como "burn-in", para evitar la autocorrelación entre los diferentes remuestros.

## Resultados

En las Tablas 1 y 2 se muestran los resultados obtenidos cuando se analiza exclusivamente el ítem que fue simulado con DIF uniforme desde la perspectiva clásica y bayesiana.

*Tabla 1: Resultados obtenidos desde la perspectiva frecuentista*

<b>N</b>	<b>R<sup>2</sup><sub>1</sub></b>	<b>R<sup>2</sup><sub>2</sub></b>	<b>R<sup>2</sup><sub>3</sub></b>	<b><math>\chi^2</math></b>	<b>p</b>	<b>DIF</b>	<b>DIF U</b>	<b>DIF N U</b>
<b>400</b>	0.16	0.35	0.35	69.32	>0.001	0.19	0.19	0.00
<b>200</b>	0.09	0.28	0.28	32.99	>0.001	0.19	0.19	0.00
<b>100</b>	0.13	0.13	0.32	16.96	>0.001	0.19	0.00	0.19
<b>50</b>	0.07	0.22	0.28	9.31	0.01	0.21	0.15	0.06

*Tabla 2: Resultados obtenidos desde la perspectiva bayesiana*

<b>N</b>	<b>R<sup>2</sup><sub>1</sub></b>	<b>R<sup>2</sup><sub>2</sub></b>	<b>R<sup>2</sup><sub>3</sub></b>	<b>DIF</b>	<b>DIF U.</b>	<b>DIF N U</b>
<b>400</b>	0.12	0.28	0.28	0.16	0.16	0.00
<b>200</b>	0.06	0.22	0.22	0.16	0.16	0.00
<b>100</b>	0.09	0.25	0.25	0.16	0.16	0.00
<b>50</b>	0.03	0.13	0.16	0.14	0.10	0.04

## Conclusión

Los valores  $R^2_1$ ,  $R^2_2$  y  $R^2_3$  son los coeficientes de determinación de los modelos 1, 2 y 3;  $\chi^2$  es la diferencia entre el  $\chi^2$  del modelo 3 menos el del modelo 1; p es la probabilidad asociada al estadístico; DIF es el tamaño del efecto total del Funcionamiento Diferencial del Ítem; DIF U es el tamaño del efecto del DIF que se explica mediante un DIF Uniforme y DIF N U es el tamaño del efecto del DIF que se explica mediante un DIF No Uniforme.

En la perspectiva frecuentista se realiza primero la comparación entre el  $\chi^2$  del tercer modelo menos el del primero obteniendo un  $\chi^2$  con dos grados de libertad. Se puede observar que en todos los tamaños muestrales este  $\chi^2$  es estadísticamente significativo, señalando que el ítem funciona diferencialmente. Posteriormente, y siguiendo el método de comparación del  $R^2$  de cada modelo expuesto anteriormente, se obtiene el tamaño del efecto del DIF (tanto uniforme como no uniforme).

En la perspectiva Bayesiana los resultados que se obtienen son los tamaños del efecto. No hay una prueba inicial en la que se señale si existe DIF o no. Esta decisión queda en manos del investigador

en el que debe analizar si dicho tamaño del efecto es lo suficientemente grande o no como para poder determinar la existencia de funcionamiento diferencial.

Como se puede apreciar en el tamaño muestral 100 (50 casos de grupo de referencia y 50 de grupo focal) el análisis clásico detecta DIF pero al estudiar los coeficientes de determinación muestral detecta, de forma errónea, un funcionamiento diferencial no uniforme. De hecho el tamaño del efecto del DIF Uniforme es de 0. Esto no ocurre con la perspectiva bayesiana en la que detecta correctamente el DIF Uniforme en todos los tamaños muestrales.

### Discusión

Como se puede apreciar en el estudio presentado, ambas alternativas detectan el DIF de igual forma cuando los tamaños muestrales son grandes. Esto era previsible ya que con tamaños muestrales grandes los resultados frecuentistas y bayesianos tienden a converger. Por el contrario, se puede apreciar, cómo con tamaños muestrales pequeños (50 y 100 casos) la regresión logística bayesiana proporciona mejores resultados que la frecuentista. Se puede observar que con 50 casos el valor de la probabilidad asociada al estadístico, desde el marco frecuentista, es de 0,01. Si se trabajase a un Nivel de Confianza del 99% este resultado sería conflictivo. Con un tamaño muestral de 100 casos la regresión logística frecuentista mostraría, de forma errónea, que el DIF es No Uniforme. En ninguna de las anteriores condiciones la regresión bayesiana falla. En todas las condiciones experimentales probadas la regresión logística bayesiana ofrece un rendimiento tan bueno o mejor que la frecuentista.

Es reseñable el hecho de que desde la perspectiva Bayesiana es el propio investigador el que debe decidir si hay DIF o no. Parece sensato pensar que salir de la rigidez de la decisión dicotómica de si es estadísticamente significativo o no puede permitir al investigador enfrentarse a situaciones en las que su propio criterio debe indicarle si son susceptibles de estudio o no lo son.

Es destacable el hecho de que en este caso la regresión logística bayesiana fue realizada sin información inicial, es decir, la condición más desfavorable posible para realizar las estimaciones. Es de esperar que si se hubiese utilizado información inicial las estimaciones serían mejores, si bien esto deberá ser probado en siguientes investigaciones.

Agradecimientos. El trabajo ha sido financiado por el proyecto SEJ2006-13009 del Ministerio de Educación y Ciencia y por el proyecto de excelencia P07HUM-02529 de la Junta de Andalucía

### Referencias

- Guàrdia, J., de la Fuente, E.I. y Lozano, L.M. (2008). Bayesian inference for binomial populations. Bayesian estimation for child depression prevalence. *Advances and applications in statistics*, 9(1), 13-35.
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement*, 31(5), 457-459.
- Han, K. T. y Hambleton, R. K. (2007). User's Manual: WinGen (*Center for Educational Assessment Report No. 642*). Amherst, MA: University of Massachusetts, School of Education.
- Mantel, N. (1963) Chi-square tests with one degree of freedom: extensions of the Mantel Haenszel procedure. *Journal of American Statistics Association*, 58, 690-700.

- Mantel, N. y Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institution*, 22, 719-748.
- Muñiz, J. (1996). *Psicometría*. Editorial Universitas: Madrid.
- Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. New Jersey: Wiley & Sons.
- Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 492-502.
- Shealey, R. y Stout, W. (1993a). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58, 159-194.
- Shealey, R. y Stout, W. (1993b). An item response theory model for test bias and differential test functioning. En W.P. Holland y H. Wainer (Eds.). *Differential Item Functioning* (pp. 197-240). Hillsdale, NJ: LEA.
- Spiegelhalter, D., Thomas, A., Best, N. y Gilks, W. (1994a). *BUGS: Bayesian inference using Gibbs sampling*. Cambridge: Available from MRC Biostatistics Unit.
- Spiegelhalter, D., Thomas, A., Best, N. y Gilks, W. (1994b). *BUGS: Examples*. Cambridge: Available from MRC Biostatistics Unit
- Swaminathan, H. y Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

# ANÁLISIS BAYESIANO DE LOS NIVELES DE *BURNOUT* EN DIFERENTES COLECTIVOS PROFESIONALES

Emilia I. de la Fuente<sup>1</sup>, Luis M. Lozano<sup>1</sup>, Guillermo A. Cañadas<sup>1</sup> y Vicente Sánchez Garzón<sup>2</sup>

<sup>1</sup> Universidad de Granada

<sup>2</sup> Comité Nacional de Riesgos Laborales. Sindicato Unificado de Policía

El síndrome de *Burnout* es un problema social, que afecta tanto a los profesionales que lo desarrollan como a la organización en la que realizan su trabajo. Cuando se padece se manifiestan niveles altos de “Cansancio Emocional, y de “Despersonalización” y bajos niveles de “Realización Personal”. El objetivo de este trabajo ha sido comparar los niveles de *Burnout* de varias muestras de profesionales, entre los que se encuentran, policías, docentes, sanitarios de dos tipos, y un grupo de población general o mixto, con diferentes profesiones. Se presenta así mismo el procedimiento de comparación de los grupos citados anteriormente, en cada una de las tres dimensiones del Síndrome de *Burnout* desde la perspectiva analítica clásica y bajo el marco teórico que ofrece la filosofía bayesiana.

El *Burnout* es un problema que afecta a muchas personas, por ello se hace necesario saber más sobre él para reducir la prevalencia y la incidencia del mismo (Burke y Richardsen, 2001) y cada vez afecta a más profesiones –hasta 25 campos profesionales aparecen citados ya por Silverstein (1986). Los datos epidemiológicos sobre el síndrome hablan de un problema de tal magnitud, que conlleva consecuencias personales y laborales negativas y que justifican por sí mismas que el desarrollo investigador haya crecido de manera considerable en estos últimos (Hakanen, Bakker y Schaufeli, 2006; Halbesleben y Buckley, 2004; Van den Broeck, Vansteenkiste, De Witte, H. y Lens, 2008).

Maslach y Jackson (1981) entienden que el *Burnout* es un síndrome tridimensional caracterizado por “agotamiento emocional” (hace referencia a las sensaciones de sobreesfuerzo físico y hastío emocional que se produce como consecuencia de las continuas interacciones que los trabajadores deben mantener entre ellos y para con los clientes), “despersonalización” (supone el desarrollo de actitudes y respuestas cínicas hacia las personas a quienes los trabajadores prestan sus servicios) y “baja realización personal” (conlleva la pérdida de confianza en la realización personal y la presencia de un negativo autoconcepto como resultado de las situaciones ingratas). Definición ésta que, en contra de lo habitual, no es teórica sino la consecuencia empírica de las investigaciones que desarrollan estas autoras (Maslach y Jackson, 1984) y que da lugar al “*Maslach Burnout Inventory*” (MBI), cuestionario de valoración del grado de padecimiento del síndrome. En esta misma época, Pines, Aronson y Kafry (1981) dan su propia definición, también basada en soporte empírico y que da lugar a su propio inventario de evaluación. Tras la mayor aceptación de los planteamientos de Maslach y Jackson, surgen diferentes definiciones y trabajos que ubican el *Burnout* en la línea de los planteamientos de estas autoras con ligeras variantes (García Izquierdo, Castellón, Alvadalejo y García, 1993; García Izquierdo y Velandrino, 1992; Moreno, Oliver y Aragoneses, 1991).

En definitiva, tras la definición de Freudenberger (1974), con su “existencia gastada”, y tras las aportaciones de Maslach y Jackson (1981) con la tridimensionalidad del síndrome: agotamiento emocional, despersonalización y reducida realización personal, o las de Pines y Kafry, (1981) con el triple agotamiento: emocional, mental y físico, han sido pocas las aportaciones originales a los planteamientos y definición del *Burnout*, girando todas alrededor de estas tres, marco teórico en el que se desarrolla este trabajo.

El presente trabajo tiene un doble objetivo. De una parte, comparar los niveles de Burnout (Cansancio Emocional, Despersonalización y Realización Personal) en los cinco grupos de profesionales participantes. De otra, presentar los resultados de dicha comparación, cuando el análisis se realiza tanto desde la perspectiva clásica como la bayesiana.

## Método

### Participantes

Colaboran 417 participantes; población general (17%), docentes (17,7%), policías nacionales (24,5%) y dos tipos de profesionales sanitarios, una categoría general (22,8%) y sanitarios que atienden a personas mayores (18%).

## Resultados

Se utiliza el programa SPSS 15.0, para realizar el ANOVA comparando los cinco grupos considerados. Se utiliza el programa WinBugs (Spiegelhalter, Thomas, Best y Gilks, 1994a, 1994b) para realizar el ANOVA bayesiano con distribución inicial no informativa en el que se comparan los cinco grupos considerados. Se emplean 99000 remuestros Gibbs con eliminación de los 1000 primeros para evitar el problema de autocorrelación y distribución inicial no informativa.

Se comprueban los supuestos previos de la técnica de análisis utilizada. El incumplimiento de alguno de los supuestos queda subsanado por el tamaño muestral utilizado así como por la petición del estadístico de Brown-Forsyth al realizar el contraste principal del análisis de varianza. Existen diferencias estadísticamente significativas en Cansancio Emocional entre los diferentes grupos profesionales ( $F_{4,405}=7,885$ ;  $p<0,000$ ). De igual forma se encuentran diferencias estadísticamente significativas en la dimensión Despersonalización, entre los grupos considerados en el análisis ( $F_{4,405}=11,226$ ;  $p<0,000$ ). También se encuentran diferencias estadísticamente significativas en la dimensión Realización Personal, entre los diferentes tipos de profesionales que participan en la investigación ( $F_{4,403}=140,105$ ;  $p<0,000$ ). La corrección de Brown-Forsyth lleva a idénticos resultados en cuanto a significación estadística.

Se realiza el ANOVA bayesiano mediante la programación en WinBUGS, según la codificación que se incluye en la Tabla 1 siguiente.

Tabla 1

```

model...
{
  for (j in 1:5)
  {
    a[j] ~ dnorm(0.0, 1.0E-6)
  }

  tau ~ dgamma(0.001, 0.001)

  for (i in 1:401)
  {
    mean[i] <- a[tra[i]]
    Y[i] ~ dnorm(mean[i], tau)
  }

  m1_m2 <- step(mean[1]-mean[2])
  m1_m3 <- step(mean[1]-mean[3])
  m1_m4 <- step(mean[1]-mean[4])
  m1_m5 <- step(mean[1]-mean[5])
  m2_m3 <- step(mean[2]-mean[3])
  m2_m4 <- step(mean[2]-mean[4])
  m2_m5 <- step(mean[2]-mean[5])
  m3_m4 <- step(mean[3]-mean[4])
  m3_m5 <- step(mean[3]-mean[5])
  m4_m5 <- step(mean[4]-mean[5])
}

```

Se realizan las comparaciones dos a dos de los grupos, tanto del análisis clásico como del análisis bayesiano. Los resultados obtenidos mediante ambos procedimientos, para la dimensión Cansancio Emocional, se incluyen en la Tabla 2.

Tabla 2

		SPSS	WinBugs
		Sig.	Probabilidad
Sanitario 3ª Edad	Policía Nacional	0.939	0.229
	Personal Sanitario	0.010	0
	Población general	0.561	0.844
	Profesor	0.985	0.1087
Policía Nacional	Personal Sanitario	0.059	0
	Población general	0.120	0.962
	Profesor	1.000	0.271
Personal Sanitario	Población general	0.000	1
	Profesor	0.065	0.999
Población general	Profesor	0.263	0.015

De igual modo, se comparan los resultados por ambos procedimientos para la dimensión Despersonalización (ver Tabla 3).

Tabla 3

		SPSS	WinBugs
		Sig.	Probabilidad
Sanitario 3ª Edad	Policía Nacional	0.000	0
	Personal Sanitario	0.359	0.021
	Población general	0.996	0.665
	Profesor	0.843	0.893
Policía Nacional	Personal Sanitario	0.110	0.998
	Población general	0.000	1
	Profesor	0.000	1
Personal Sanitario	Población general	0.179	0.992
	Profesor	0.029	0.999
Población general	Profesor	0.969	0.785

Los resultados en la dimensión Realización Personal para las comparaciones de grupos dos a dos se incluyen en la Tabla 4 siguiente.

Tabla 4

		SPSS	WinBugs
		Sig.	Probabilidad
Sanitario 3ª Edad	Policía Nacional	0.000	1
	Personal Sanitario	0.000	1
	Población general	0.000	1
	Profesor	0.000	1
Policía Nacional	Personal Sanitario	0.000	1
	Población general	0.101	0.998
	Profesor	0.001	1
Personal Sanitario	Población general	0.028	4.4E-4
	Profesor	0.576	0.025
Población general	Profesor	0.674	0.890

### Conclusión

Nuestros datos nos permiten concluir en dos sentidos, cuestiones de tipo metodológico así como conclusiones acerca de los niveles de *burnout* en las tres dimensiones del síndrome, en los colectivos profesionales considerados. En relación con los aspectos metodológicos se puede resaltar lo siguiente:

- El ANOVA clásico sólo identifica que hay diferencias significativas y en los grupos en que se produce.
- El ANOVA bayesiano identifica con probabilidades 1 las diferencias significativas entre los grupos, como el clásico.
- El ANOVA bayesiano incorpora la probabilidad de todas las hipótesis, incluso en las comparaciones entre grupos que no son significativas.

En relación con la valoración acerca de los niveles de *Burnout* observados en los diferentes grupos participantes en esta investigación, cabe resaltar que:

- El perfil que resulta más favorecido en nuestro análisis es el de sanitarios que atienden a personas mayores.
- Las puntuaciones más altas en despersonalización las presentan los participantes que pertenecen al colectivo de policías nacionales.
- Los profesionales sanitarios son los que obtienen puntuaciones más altas en Cansancio Emocional.
- A excepción de los sanitarios que atienden a personas mayores, los policías nacionales obtienen la puntuación más alta en R.P.

Agradecimientos. El trabajo ha sido financiado por el proyecto SEJ2006-13009 del Ministerio de Educación y Ciencia y por el proyecto de excelencia P07HUM-02529 de la Junta de Andalucía.

### Referencias

- Burke, R.J. y Richardsen, A.M. (2001). *Psychological burnout in organizations. Research and Interventions*. En R.T. Golembiewski (Ed.) 2ª ed. New York: Marcel Dekker.
- Freudenberger, H.J. (1974). Staff burnout. *The Journal of Social Issues*, 30(1), 159-166.
- García Izquierdo, M., Castellón, M., Albadalejo, B. y García, A.L. (1993). Relaciones entre burnout, ambigüedad de rol y satisfacción laboral en personal de banca. *Psicología del Trabajo y Organizaciones*, 11(24), 17-26.
- García Izquierdo, M. y Velandrino, A.P. (1992). EPB: Una escala para la evaluación del burnout profesional de las organizaciones. *Anales de Psicología*, 8(1-2), 131-138.
- Hakanen, J.J., Bakker, A.B. y Schaufeli, W.B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology*, 43(6), 495-513.
- Halbesleben, J.R.B. y Buckley, M.R. (2004). Burnout in Organizational Life. *Journal of Management*, 30(6), 859-879.
- Maslach, C. y Jackson, S.E. (1981). *MBI: Maslach Burnout Inventory*. Manual. Palo Alto: University of California, Consulting Psychologists Press.
- Maslach, C. y Jackson, S.E. (1984). Burnout in organizational setting. *Applied Social Psychology Annual*, 5, 133-154.
- Moreno, B., Oliver, C. y Aragoneses, A. (1991). El burnout, una forma específica de estrés laboral. En G. Buela-Casal y V.E. Caballo (Comps.), *Manual de Psicología Clínica Aplicada* (pp. 271-284). Madrid: Siglo XXI.
- Silverstein, R. (1986). *How experienced psychotherapists cope with burnout at a state mental hospital*. Folleto informativo. Carbondale: South Illinois University. Citado en García Izquierdo (1991).
- Spiegelhalter, D., Thomas, A., Best, N., y Gilks, W. (1994a). BUGS: Bayesian inference using Gibbs sampling. Cambridge: Available from MRC Biostatistics Unit.
- Spiegelhalter, D., Thomas, A., Best, N. y Gilks, W. (1994b). *BUGS: Examples*. Cambridge: Available from MRC Biostatistics Unit
- Pines, A., Aronson, E. y Kafry, D. (1981). Burnout: From tedium to personal growth. En C. Cherniss (Ed.), *Staff burnout: Job stress in the human services*. Nueva York: The Free Press.
- Pines, A. y Kafry, D. (1981). The experience of life tedium in three generations of professional women. *Sex Roles*, 7, 117-134.
- Van den Broeck, A., Vansteenkiste, M., De Witte, H. & Lens, W. (2008). Explaining the relationships between job characteristics, burnout, and engagement: The role of basic psychological need satisfaction. *Work & Stress*, 22, 277-294.

# INFLUENCIA DE LA INTERDEPENDENCIA DE OBJETIVOS INDIVIDUAL Y LA AUTOGESTIÓN Y PARTICIPACIÓN GRUPAL SOBRE LA SATISFACCIÓN LABORAL Y LA CONDUCTA COOPERATIVA: UNA APROXIMACIÓN MULTINIVEL

Aitor Aritzeta y Nekane Balluerka  
Universidad del País Vasco

Siguiendo los supuestos de la Teoría Organizacional Multinivel (Klein y Kozlowsky, 2000), el objetivo del presente estudio consistió en examinar la influencia de los equipos de trabajo autónomos sobre el comportamiento de las personas que los integran en una organización industrial. Para ello, se examinó la relación existente entre distintos indicadores individuales y grupales en una muestra de 232 trabajadores y trabajadoras que formaban 25 equipos de trabajo utilizando modelos de análisis multinivel. Los resultados pusieron de manifiesto que el incremento de la interdependencia de objetivos individual aumenta la satisfacción laboral y las conductas cooperativas. La autogestión y la participación grupal mejoran la satisfacción pero no la cooperación interpersonal. Se constata que si bien el análisis multinivel constituye una aproximación muy útil para predecir el comportamiento en contextos organizacionales, la adecuada estimación de los efectos y la identificación de interacciones transnivel requiere una gran cantidad de unidades en los niveles superiores, lo que puede limitar su utilidad.

Los equipos de trabajo constituyen unidades compuestas por varias personas que poseen conocimientos y habilidades complementarios para el logro de objetivos laborales. Hoy en día, estas unidades son consideradas imprescindibles para aumentar los niveles de innovación y competitividad de las organizaciones (Gil, Alcover y Peiro, 2005).

Se ha demostrado que los equipos de trabajo efectivos aumentan simultáneamente la satisfacción laboral (dado que generan mayor identificación y comunicación grupal) y los comportamientos cooperativos (dado que promueven la integración de conocimientos), lo que aumenta la productividad (Yeatts y Hyten, 1997).

La autoridad para tomar decisiones (autogestión) y la conducta proactiva en la búsqueda de los objetivos del equipo (participación) han sido consideradas dos de las variables grupales más importantes en la predicción de la satisfacción laboral (Langan-Fox, Code, Gray y Langfield-Smith, 2002) y, a nivel individual, una de las dimensiones más importantes es la denominada interdependencia de objetivos, es decir, la conciencia de que dependemos de los demás para alcanzar nuestros propios objetivos (Aritzeta y Balluerka, 2006). Por ello, en este trabajo, esperábamos que el incremento de la autogestión y la participación grupal, así como de la interdependencia de objetivos individual, aumentarían la satisfacción laboral.

Por otro lado, se ha demostrado que los equipos incrementan su eficiencia debido a la influencia que ejercen sobre las conductas cooperativas (Locke y Latham, 1984). Desde un punto de vista longitudinal, los procesos de aprendizaje que se producen dentro de los equipos incrementan las conductas cooperativas (Russ-Eft, Preskill y Sleezer, 1996). La interdependencia de objetivos constituye uno de los factores que aumentan los comportamientos cooperativos (Tjosvold y De Dreu, 1997) y cabe esperar que a mayor autogestión y participación grupal, los niveles de cooperación en el equipo sean mayores (Ely, 2004).

## Método

### *Participantes y Procedimiento*

En el estudio participaron 232 trabajadores/as pertenecientes a 25 equipos de trabajo. El 90,5% eran hombres y el 9,5% mujeres. En cuanto a la edad, el 59,8% tenían entre 20 y 30 años, el 30,8% entre 31 y 40 años y el 9,4% más de 41 años. Los participantes respondieron a los cuestionarios en dos ocasiones separadas por un intervalo de un año. En la primera ocasión (tiempo 1) los equipos de trabajo eran de reciente creación y en la segunda (tiempo 2) llevaban trabajando como equipos durante 11 meses.

### *Instrumentos*

- Escala de Satisfacción Laboral (Warr, Cook y Wall, 1979). Está compuesta de 15 ítems que evalúan la satisfacción laboral. La escala de respuesta es de tipo Likert de 6 puntos, donde 1 corresponde a “totalmente insatisfecho” y 6 a “totalmente satisfecho”. El alpha de Cronbach para la escala global es de .91.
- Características de los Equipos de Trabajo (Campion, Medsker y Higgs, 1993). Evalúa la autogestión grupal, la participación grupal y la interdependencia de objetivos, a través de 9 ítems. La escala de respuesta es de tipo Likert de 6 puntos, donde 1 corresponde a “totalmente en desacuerdo” y 6 a “totalmente de acuerdo”. Los índices de fiabilidad de las dimensiones son de .76, .88 y .78, respectivamente.
- Conductas ante los conflictos (MODE, Thomas y Kilmann, 1974). Este instrumento mide 5 estilos de manejo de conflictos (sumisión, cooperación, compromiso, competición y evitación) mediante 5 ítems por estilo. La escala de respuesta es de tipo Likert de 6 puntos, donde 1 corresponde a “totalmente en desacuerdo” y 6 a “totalmente de acuerdo”. Los índices de fiabilidad de las dimensiones son de .76, .89, .78, .83 y .74, respectivamente.

### *Análisis de datos*

Los datos se analizaron mediante el software MlwiN 2.10. Se tomaron como variables criterio la satisfacción laboral y la cooperación y como variables explicativas de nivel individual (nivel trabajadores) la interdependencia de objetivos y de nivel grupal (nivel equipos) la autogestión y la participación grupal. Las variables explicativas de nivel grupal fueron agregadas siguiendo el modelo de cambio de referente de Chan (1998).

Los modelos que se sometieron a análisis fueron modelos de interceptos aleatorios y pendientes fijas. En primer lugar (modelo 0) se elaboró un modelo en el que no se introdujeron variables explicativas. Se trata de un modelo similar al ANOVA de efectos aleatorios. Este modelo se tomó como referencia con el objetivo de examinar el cambio que se producía en el ajuste a medida que se introducían nuevas variables predictoras. En segundo lugar (modelo 1) se introdujo la variable de nivel individual, a saber, la interdependencia de objetivos. Por último, se añadieron al modelo (modelo 2) las variables de nivel grupal, es decir, la autogestión y la participación grupal.

Las ecuaciones 1 y 2, donde las  $x$  hacen referencia a variables individuales y las  $z$  a variables grupales, representan los tipos de modelos que utilizamos para someter a prueba las hipótesis del estudio:

$$\begin{aligned} \text{(Ecuación 1) Nivel 1: } & Y_{ij} = B_{0j} + B_{1j} x_{1ij} + r_{ij} \\ \text{(Ecuación 2) Nivel 2: } & B_{0j} = \gamma_{00} + \gamma_{0j} z_j + u_{0j} \end{aligned}$$

Donde:

$Y_{ij}$  = Puntuación del trabajador  $i$  perteneciente al equipo  $j$  en la variable criterio

$B_{0j}$  = Intercepto del equipo  $j$  (nivel promedio que presenta en la variable criterio el equipo  $j$ )

$B_{1j}$  = Coeficiente de regresión que expresa la relación entre la variable explicativa  $x$  y la variable criterio

$r_{ij}$  = Término residual de nivel individual

$\gamma_{00}$  = Promedio de las medias de los equipos en la variable criterio

$u_{ij}$  = Término residual de nivel grupal

## Resultados

En la tabla 1 se presentan las medias, las desviaciones típicas, las correlaciones y las diferencias entre las medias de las variables sometidas a análisis en el tiempo 1 y en el tiempo 2.

*Tabla 1. Análisis descriptivos: Medias, DT, Correlaciones, t-tests y d de Cohen de las variables sometidas a análisis en el tiempo 1 (diagonal inferior) y en el tiempo 2 (diagonal superior)*

	M(DT)	t	d Cohen	1	2	3	4	5
M				3.55	4.44	3.01	3.32	4.36
(DT)				(.90)	(.91)	(1.13)	(1.17)	(.61)
Alpha de Cronbach				.91	.49	.70	.77	.79
1. Satisfacción Laboral	3.41 (.78)	-1.72	.16		.31**	.34**	.28**	.24**
2. Interdependencia Obj.	4.38 (.92)	-.21	.02	.20*		.21**	.16*	.31**
3. Autogestión	2.91 (1.13)	-.87	.008	.29**	.13*		.73**	.10
4. Participación	3.19 (1.07)	-1.25	.11	.34**	.11	.55*		.15*
5. Cooperación	4.45 (.66)	-.51	.05	.21**	.26*	.05	.07	

En la tabla 2 pueden observarse los resultados para la satisfacción laboral en el tiempo 1 y en el tiempo 2.

*Tabla 2. Resultados del análisis multinivel tomando la interdependencia de objetivos individual y la autogestión y participación grupal como variables predictoras y la satisfacción laboral como variable criterio*

	Tiempo 1 (MODELOS)			Tiempo 2 (MODELOS)		
	0	1	2	0	1	2
<b>Media Global (<math>\beta_{0ij}</math>)</b>	<b>3.439 (.078)</b>	<b>3.379 (.159)</b>	<b>3.385 (.144)</b>	<b>3.534 (.085)</b>	<b>3.513 (.079)</b>	<b>3.507 (.063)</b>
<b>Nivel Individual</b>						
<b>Interdependencia Ob</b>		.132* (.056)	.135* (.053)		.243** (.004)	.217** (.058)
<b>Nivel Grupal</b>						
<b>Autogestión</b>			.177 (.114)			.352* (.148)
<b>Participación</b>			.455* (.124)			.137 (.159)
<b>Componentes de la varianza</b>						
<b>Nivel Individual</b>	512 (.050)	.499 (.051)	.500 (.051)	.665 (.065)	.623 (.061)	.626 (.061)
<b>Nivel Grupal</b>	.089 (.043)	.069 (.037)	.000 (.016)	.148 (.060)	.105 (.048)	.032 (.030)
<b>Ajuste del modelo</b>						
<b>Desviación (D)</b>	526.119	479.480	461.163	622.018	601.065	586.119
<b><math>\Delta</math> model 0 (<math>\Delta</math> D)</b>		46.639**			20.953**	
<b><math>\Delta</math> model 1 (<math>\Delta</math> D)</b>			18.317**			14.946**
<b><math>\Delta</math> df</b>		3	2		5	2
<b>Nivel Individual R<sup>2</sup></b>		2.5%	-.02%		6.7%	-.04%
<b>Nivel Grupal R<sup>2</sup></b>		22.47%	77.53%		29.5%	69.5%
<b>Correlación</b>						
<b>Intraclase(p)</b>	.1480	.1214	.000	.1820	.1442	.0486

Como cabe observar en la tabla 2, la interdependencia de objetivos se asoció a niveles superiores de satisfacción laboral tanto en el tiempo 1 como en el tiempo 2, siendo los efectos más intensos en el tiempo 2 ( $z=60.75$ ). La varianza grupal no explicada se redujo en un 22,5 % en el tiempo 1 y en un 29,5% en el tiempo 2. A su vez, la participación grupal, en el tiempo 1, y la autogestión grupal, en el tiempo 2, también se asociaron a un incremento en la satisfacción laboral reduciendo en un 77,53% y en un 69,5%, respectivamente, la varianza grupal no explicada.

En la tabla 3 se presentan los resultados para la cooperación en el tiempo 1 y en el tiempo 2.

*Tabla 3. Resultados del análisis multinivel tomando la interdependencia de objetivos individual y la autogestión y participación grupal como variables predictoras y la cooperación como variable criterio*

	Tiempo 1 (MODELOS)			Tiempo 2 (MODELOS)		
	0	1	2	0	1	2
<b>Media Global (<math>\beta_{oij}</math>)</b>	<b>4.441 (.060)</b>	<b>4.421 (.137)</b>	<b>4.421 (.137)</b>	<b>4.364 (.048)</b>	<b>4.373 (.044)</b>	<b>4.373 (.043)</b>
<b>Nivel Individual</b>						
<b>Interdependencia Ob</b>		.167* (.049)	.165* (.049)		.197* (.041)	.188* (.042)
<b>Nivel Grupal</b>						
<b>Autogestión</b>			.049 (.131)			-.050 (.101)
<b>Participación</b>			.037 (.143)			.169 (.109)
<b>Componentes de la varianza</b>						
<b>Nivel Individual</b>	.402 (.039)	.394 (.040)	.393 (.040)	.348 (.034)	.325 (.032)	.322 (.031)
<b>Nivel Grupal</b>	.043 (.025)	.032 (.023)	.032 (.023)	.028 (.019)	.012 (.014)	.011 (.014)
<b>Ajuste del modelo</b>						
<b>Desviación (D)</b>	463.520	422.541	422.091	447.859	424.698	421.515
<b><math>\Delta</math> model 0 (<math>\Delta D</math>)</b>		40.979**			23.161**	
<b><math>\Delta</math> model 1 (<math>\Delta D</math>)</b>			0.450			3.183
<b><math>\Delta</math> df</b>		3	2		3	2
<b>Nivel Individual R<sup>2</sup></b>		1.99%	0.02%		6.61%	0.09%
<b>Nivel Grupal R<sup>2</sup></b>		25.58%	0%		57.14%	8.33%
<b>Correlación</b>						
<b>Intraclase (<math>\rho</math>)</b>	.0966	.0751	.0752	.0744	.0356	.0330

La interdependencia de objetivos fue la única variable asociada a comportamientos cooperativos tanto en el tiempo 1 como en el 2, reduciendo la varianza grupal no explicada en un 25,85% en el tiempo 1 y en un 57,14% en el tiempo 2.

## Discusión

La primera de las predicciones se confirma parcialmente. Así, la interdependencia de objetivos incrementó los niveles de satisfacción tanto en el tiempo 1 como en el tiempo 2. Sin embargo, la autogestión grupal sólo mostró un efecto estadísticamente significativo en el tiempo 2.

Cuando los trabajadores pasan de un contexto laboral individual a uno grupal, disponen inmediatamente de nuevas posibilidades de participación en la toma de decisiones dado que los procesos de comunicación interpersonal aumentan. Sin embargo, la autogestión requiere de un proceso de aprendizaje más largo (Cordey, Mueller y Smith, 1991). El efecto de la autogestión se observa únicamente en el tiempo 2, una vez los equipos han aprendido a tomar decisiones de forma autónoma.

La participación no aumentó la satisfacción en el tiempo 2. Este resultado puede ser explicado partiendo de que los equipos de trabajo incrementan rápidamente los niveles de exigencia respecto a su participación en la toma de decisiones y, en ocasiones, superan los niveles de participación que los gestores están dispuestos a permitir (Jackson y Mullarkey, 2000).

La interdependencia de objetivos fue la única variable que mejoró la conducta cooperativa. Ni la autogestión ni la participación se asociaron con la cooperación, probablemente porque ésta requiere de un aprendizaje específico o de un contexto objetivo como el que proporciona la interdependencia de objetivos en los equipos (Johnson y Johnson, 1989).

Para finalizar cabe señalar que el análisis multinivel nos ha permitido tomar en consideración tanto la especificidad de los individuos como los contextos en los que se ubican (equipos) y considerar tales contextos como muestras aleatorias extraídas de poblaciones más amplias aumentando la capacidad de generalización de los resultados. En consecuencia, entendemos que este tipo de análisis resulta muy útil para examinar conjuntamente las relaciones existentes entre los distintos niveles que se hallan presentes en el sistema organizacional. Sin embargo, para poder observar efectos en las pendientes es importante disponer de un número amplio de unidades de nivel superior. El hecho de contar sólo con 25 equipos de trabajo, constituye una importante limitación para poder observar variabilidad en las pendientes. Por ello, adoptamos la alternativa de construir modelos donde únicamente se permite variar aleatoriamente a los interceptos de los distintos grupos, mientras que las pendientes son fijas para todos ellos, limitando las ventajas que presenta la regresión multinivel frente a modelos de análisis más clásicos.

#### Referencias

- Aritzeta, A. y Balluerka, N. (2006). Cooperation, competition and goal interdependence in work teams: A multilevel approach. *Psicothema*, 18(4), 757-765.
- Campion, M. A., Medsker, G. J. y Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, 46, 823-850.
- Cordey, J. L., Mueller, W. S. y Smith, L. M. (1991). Attitudinal and behavioral effects of autonomous group working: A longitudinal field study. *Academy of Management Journal*, 34(2), 464-476.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83, 234-246.
- Ely, R. J. (2004). A field study of group diversity, participation in diversity education programs, and performance. *Journal of Organizational Behavior*, 25(6), 755-780.
- Gil, F., Alcover, C.M. y Peiro, J.M. (2005). Work team effectiveness in organizational contexts: Recent research and applications in Spain and Portugal. *Journal of Managerial Psychology*, 20(3-4), 193-218.
- Jackson, P.R. y Mullarkey, S. (2000). Lean production teams and health in garment manufacture. *Journal of Occupational Health Psychology*, 5(2), 231-245.
- Johnson, D. W. y Johnson, F. P. (1989). *Cooperation and competition: Theory and research* (5th. ed.), Edina, MN: Inter-action Book Company.

- Klein, K. J. y Kozlowski, S. W. J. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions and new directions*, San Francisco: Jossey-Bass.
- Langan-Fox, J., Code, S., Gray, R. y Langfield-Smith, K. (2002). Supporting employee participation: Attitudes and perceptions in trainees, employees and teams. *Group Processes & Intergroup Relations*, 5(1), 53-82.
- Locke, E. A. y Latham, G. P. (1984). *Goal setting: A motivational technique that works*, Englewood Cliffs, NJ: Prentice-Hall.
- Russ-Eft, D., Preskill, H. y Sleezer, C. (1996). *Human Resource Development Review*, London: Sage.
- Thomas, K. M. y Kilmann, R. H. (1974). *The Thomas-Kilmann conflict mode instrument*, Tuxedo, NY: Xicom.
- Tjosvold, D. y De Dreu, C. (1997). Managing conflict in Dutch organizations: A test of the relevance of Deutsch's cooperation theory. *Journal of Applied Social Psychology*, 27(24), 2213-2227.
- Warr, W., Cook, L. y Wall, P. (1979). Scale for the measurement of some work attitudes and aspects of psychological well-being. *Journal of Occupational Psychology*, 52, 129-148.
- Yeatts, D. E. y Hyten, C. (1998). *High-performing self-managed work teams: A comparison of theory and practice*, Thousand Oaks, CA: Sage.

# APROXIMACIONES METODOLÓGICAS PARA OBTENER EVIDENCIAS DE VALIDEZ

**Coordinadores: José Luis Padilla<sup>1</sup> y Nekane Balluerka<sup>2</sup>**

<sup>1</sup> Universidad de Granada

<sup>2</sup> Universidad del País Vasco

La validez se ha convertido en la consideración más importante para la evaluación de la calidad de las mediciones aportadas por tests y cuestionarios. La última edición de los Standards (AERA, APA, NCME, 1999) y la bibliografía especializada ratifican dicha valoración. La sustitución de las tradicionales “categorías de validez”, e incluso, de las “estrategias de validación” por el concepto de “fuentes de evidencias”, plantea el desafío de aportar nuevas estrategias metodológicas o de enfocar las estrategias tradicionales a la búsqueda de evidencias que apoyen la interpretación deseada de las mediciones. La elaboración de “argumentos de validez” sustentados en supuestos examinados mediante evidencias de validez orientará los estudios de validación en un futuro inmediato. Las comunicaciones agrupadas en este simposio intentan situarse ante estos nuevos desafíos. También se discutirá en el simposio las razones para la persistencia de estudios de validación, cuyo lenguaje y enfoque no coinciden con las versiones más recientes y aceptadas de la Teoría de la Validez. La primera comunicación presenta un trabajo elaborado por Nekane Balluerka, José Martín Salguero, Pablo Fernández-Berrocal y Aitor Aritzeta para obtener evidencias de validez sobre la dimensionalidad así como evidencias de validez discriminante y de relación con otras variables, de las mediciones aportadas por la versión en castellano de la *Trait Meta-Mood Scale* en la población adolescente española. El segundo estudio, llevado a cabo por Elena Delgado y Hugo Carretero, presenta los resultados obtenidos en la búsqueda de evidencias relacionadas con el contenido del *State Trait Cheerfulness Inventory*. En el tercer trabajo, Miguel Castillo, Isabel Benítez y José Luis Padilla, presentarán un trabajo en el que se utilizó el procedimiento de entrevista cognitiva para conseguir evidencias de validez sobre los procesos de respuesta a un cuestionario sobre funcionamiento familiar. El cuarto trabajo, llevado a cabo por Juana Gómez Benito, María Dolores Hidalgo, José Luis Padilla e Isabel Benítez, aborda las aplicaciones de las técnicas del funcionamiento diferencias de los ítems en cuestionarios utilizados en las encuestas.

# VALIDACIÓN DE LA VERSIÓN REDUCIDA EN CASTELLANO DEL *TRAIT META-MOOD SCALE* (TMMS) PARA LA POBLACIÓN ADOLESCENTE

Nekane Balluerka<sup>1</sup>, José Martín Salguero<sup>2</sup>, Pablo Fernández-Berrocal<sup>2</sup> y Aitor Aritzeta<sup>1</sup>

<sup>1</sup>Universidad del País Vasco

<sup>2</sup>Universidad de Málaga

El *Trait Meta-Mood Scale* (TMMS) es una medida de rasgo referido al meta-conocimiento sobre los estados emocionales; un aspecto de la inteligencia emocional percibida. Aunque el TMMS ha sido utilizado ampliamente en población adulta, hasta la fecha no se han aportado datos acerca de su validez y utilidad en población adolescente. El presente trabajo analizó las propiedades psicométricas de la versión reducida del TMMS en una muestra de 1497 adolescentes de 12 a 17 años. Los resultados del análisis factorial confirmatorio corroboraron la estructura de tres factores (atención, claridad y reparación emocional) de la escala original. Además, estas dimensiones mostraron adecuada fiabilidad y correlacionaron entre sí de la forma esperada. Por otra parte, se mostraron evidencias de validez discriminante con los Cinco Grandes rasgos de personalidad y se analizaron las diferencias en las dimensiones del TMMS en función de la edad y el sexo de los participantes.

El *Trait Meta-Mood Scale* –TMMS- (Salovey, Mayer, Goldman, Turvey y Palfai, 1995) es una medida de la inteligencia emocional (IE) percibida que incluye las creencias y actitudes, estables en el tiempo, que las personas poseen acerca de la importancia de atender a sus emociones y sentimientos, de su capacidad para comprender las propias emociones y de su habilidad para regular los estados emocionales negativos y prolongar los positivos.

El TMMS ha mostrado adecuadas propiedades psicométricas y evidencias de una estructura de tres factores tanto en su versión original (Salovey y cols., 1995) como en diferentes adaptaciones a otras poblaciones. En España, una versión abreviada del TMMS, el TMMS-24 (Fernández-Berrocal, Extremera y Ramos, 2004), mostró una alta fiabilidad para cada componente así como evidencias de estabilidad temporal para un intervalo de 4 semanas. Asimismo, diferentes estudios han encontrado evidencias de validez discriminante (Extremera y Fernández-Berrocal, 2005; Otto, Döring-Seipel, Grebe y Lantermann, 2001; Palomera y Brackett, 2006) y predictiva (para una revisión ver Fernández-Berrocal y Extremera, 2008).

Pese a ello, hasta la fecha ningún trabajo ha analizado las propiedades psicométricas del TMMS en la población adolescente. Contar con una medida válida y fiable que evalúe la inteligencia emocional percibida en esta población puede ser útil por varios motivos. En primer lugar, aumentaría el banco de instrumentos disponible para medir aspectos asociados a la IE en estas edades. En segundo lugar, nos permitiría analizar la forma en que evolucionan los procesos de atención, percepción y regulación emocional a lo largo del desarrollo y, por último, podría ayudarnos a explicar las diferencias en la adaptación psicosocial de los adolescentes y a determinar la importancia que posee, en la adolescencia, la competencia emocional percibida.

Teniendo en cuenta lo anterior, el principal objetivo del presente trabajo fue analizar las características psicométricas de la versión reducida del TMMS en una muestra de adolescentes españoles.

## Método

### *Participantes*

En el estudio participaron 1497 adolescentes (840 mujeres y 657 varones), con edades comprendidas entre los 12 y los 17 años ( $M = 14.42$ ;  $DT = 1.56$ ). Todos ellos pertenecían a diferentes centros de enseñanza secundaria de la provincia de Málaga (España).

### *Instrumentos*

- Versión española reducida del *Trait Meta-Mood Scale* (Fernández-Berrocal y cols., 2004). Incluye 24 ítems (8 de cada subescala) seleccionados a partir de los 48 ítems de la versión adaptada al castellano del TMMS. Presenta la misma estructura factorial así como buenos índices de consistencia interna y fiabilidad temporal.
- Versión española del *Big Five Inventory-44* (BFI-44; Benet-Martínez y John, 1998). Incluye 44 ítems que evalúan los Cinco Grandes Factores de la personalidad. Ha mostrado buena consistencia interna y fiabilidad temporal así como evidencias de validez convergente y discriminante y una clara estructura factorial.

## Resultados

### *Análisis factorial confirmatorio*

Con el objetivo de comprobar si en la muestra compuesta por adolescentes se corroboraba el modelo de tres factores encontrado en la versión española reducida del TMMS en población adulta, se llevó a cabo un análisis factorial confirmatorio utilizando el LISREL 8.80 (Jöreskog y Sörbom, 1993) y empleando el método de estimación de máxima verosimilitud. Los pesos factoriales fueron superiores a .35 para todos los ítems excepto para el ítem 23, poniendo de manifiesto que todos ellos son relevantes para definir sus correspondientes dimensiones. El ajuste del modelo fue evaluado mediante los siguientes índices: el  $\chi^2$ , el GFI y el AGFI (Jöreskog y Sörbom, 1989), el NNFI (Tucker y Lewis, 1973), el CFI (Bentler, 1990) y el RMSEA (Steiger, 1990). El valor observado en el  $\chi^2$ ,  $\chi^2(249, N = 1497) = 1667.09$ ,  $p = 0.01$ , indicó una ausencia de ajuste del modelo. Sin embargo, la dependencia de este índice con respecto al tamaño muestral nos llevó a valorar el ajuste del modelo mediante los índices arriba señalados, los cuales mostraron un buen ajuste, con valores GFI (.90), AGFI (.89), NNFI (.94) y CFI (.95) próximos a 1.00 y RMSEA (.065) ubicado entre .05 (buen ajuste) y .08 (ajuste aceptable).

### *Fiabilidad*

La consistencia interna de cada una de las dimensiones de la escala fue estimada mediante el coeficiente  $\alpha$  de Cronbach. Dicho índice adoptó valores de 0.84, 0.82 y 0.81 para las subescalas de Atención, Claridad y Reparación, respectivamente. Estos valores permiten concluir que la consistencia interna es elevada en todas las dimensiones.

En la tabla 1 se muestran, junto a los estadísticos descriptivos, las correlaciones entre las tres dimensiones del TMMS-24 en población adolescente. Como cabe observar en dicha tabla, existen correlaciones de magnitud considerable entre la Atención y la Claridad y entre la Claridad y la Reparación, siendo más pequeña la correlación observada entre la Atención y la Reparación.

Tabla 1. Media, desviación típica y correlaciones entre las dimensiones del TMMS-24 y del BFI-44

	1	2	3	4	5	6	7	8
1 Atención	--							
2 Claridad	.38**	--						
3 Reparación	.15**	.38**	--					
4 Extraversión	.14**	.15**	.12**	--				
5 Agradabilidad	.15**	.17**	.20**	.25**	--			
6 Consciencia	.07*	.14**	.12**	.04	.30**	--		
7 Neuroticismo	.24**	-.09**	-.20**	-.05	-.17**	-.14**	--	
8 Apertura	.18**	.17**	.18**	.21**	.21**	.19**	-.04	--
M	24.35	24.44	21.83	3.37	3.67	3.14	2.93	3.35
DT	6.46	6.36	6.16	.67	.55	.60	.60	.60

\*  $p < 0.05$ \*\*  $p < 0.01$ 

### Asociación entre las dimensiones del TMMS-24 y del BFI-44

A fin de obtener evidencia sobre la validez discriminante del instrumento, se calcularon los coeficientes de correlación de Pearson entre las puntuaciones de las dimensiones del TMMS-24 y del BFI-44 (véase Tabla 1). Como cabía esperar, las puntuaciones de los sujetos en las dimensiones de Atención, Claridad y Reparación del TMMS-24 no mostraron correlaciones de magnitud importante con ninguno de los Cinco Grandes rasgos de personalidad (entre .07 y .24), indicando que no existe solapamiento entre los constructos.

### Diferencias en las dimensiones del TMMS-24 en función del sexo y la edad

Con el objetivo de obtener evidencia adicional sobre la validez del TMMS-24 en la población adolescente, se examinaron las posibles diferencias existentes entre sus dimensiones en función de las variables género y edad de los participantes.

El análisis de la varianza llevado a cabo tomando el sexo y la edad como variables predictoras y la Atención como variable criterio mostró la existencia de un efecto estadísticamente significativo tanto en el caso del sexo,  $F(1, 1491) = 81.42$ ;  $p = 0.0001$ , como de la edad,  $F(2, 1491) = 5.49$ ;  $p = 0.004$ . No obstante, en ambos casos el tamaño del efecto fue de escasa magnitud ( $\eta^2$  sexo = 0.05;  $\eta^2$  edad = 0.007). Lo mismo cabe decir del tamaño del efecto asociado a la interacción ( $\eta^2$  interacción = 0.002), en cuyo caso no se alcanzó la significación estadística. Con el objetivo de examinar si en alguno de los rangos de edad establecidos existían diferencias en Atención emocional en función del género, se calcularon los efectos simples. Los resultados pusieron de manifiesto que en los rangos de edad de 12-13 años,  $F(1, 1491) = 15.67$ ;  $p = 0.0001$ , de 14-15 años,  $F(1, 1491) = 32.57$ ;  $p = 0.0001$  y de 16-17 años,  $F(1, 1491) = 36.04$ ;  $p = 0.0001$ , existían diferencias estadísticamente significativas entre hombres y mujeres. De acuerdo con tales resultados, el tamaño del efecto correspondiente a la diferencia de medias entre hombres y mujeres en el rango de edad de 16-17 años, adoptó un valor de magnitud considerable ( $g$  de Hedges<sub>16-17 años</sub> = 0.64). En el rango de edad de 14-15 años, se aproximó a una magnitud media ( $g$  de Hedges<sub>14-15 años</sub> = 0.45), mientras que en el rango de edad de 12-13 años, adoptó un valor entre pequeño y moderado ( $g$  de Hedges<sub>12-13 años</sub> = 0.37).

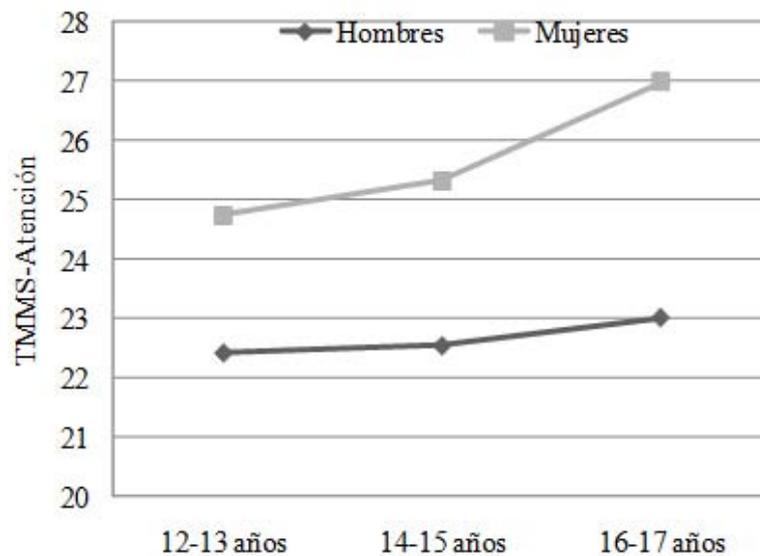


Figura 1. Diferencias en Atención emocional en función de la edad y del sexo

En lo que respecta a los análisis de la varianza llevados a cabo tomando el sexo y la edad como variables predictoras y la Claridad y la Reparación emocional como variables criterio, ninguno de los efectos observados fue de magnitud considerable.

### Discusión

La literatura científica en torno al TMMS ha mostrado evidencias de que es una prueba válida y fiable para evaluar las diferencias individuales en inteligencia emocional percibida, de las personas adultas. En el presente trabajo hemos tratado de ampliar tales evidencias analizando las características psicométricas de la versión abreviada del TMMS en una muestra de población adolescente.

Con respecto a la validez de constructo, los resultados encontrados corroboran la existencia de una estructura de tres factores correspondientes a las dimensiones Atención, Claridad y Reparación emocional, de acuerdo con la estructura original de la escala (Salovey y cols., 1995) y con su adaptación española (Fernández-Berrocal y cols, 2004). De los 24 ítems del TMMS-24, tan sólo el ítem 23 no obtuvo una saturación clara con su dimensión correspondiente, la subescala de Reparación, por lo que fue eliminado.

El análisis de la consistencia interna de la versión de 23 ítems del TMMS-24 reveló valores satisfactorios en las tres dimensiones, de forma similar a lo encontrado en otras adaptaciones (Fernández-Berrocal y cols., 2004; Otto y cols., 2001; Queirós, Fernández-Berrocal, Extremera, Cancela y Queirós, 2005). Además, las tres subescalas correlacionaron de acuerdo a lo esperado, correlaciones de mayor magnitud entre Atención y Claridad y entre Claridad y Reparación, lo cual sugiere, tal y como se ha mostrado en adultos (Palmer y cols., 2003), la existencia de una secuencia funcional en el proceso de regulación emocional, siendo necesario cierto nivel de atención para poder comprender los estados emocionales y cierto nivel de claridad para poder moderarlos o regularlos.

Con respecto a la relación entre el TMMS-24 y el Big-Five, los resultados muestran evidencias de validez discriminante del TMMS-24 con correlaciones de moderadas a bajas entre sus tres dimensiones y los Cinco Grandes factores de personalidad. En cuanto a las diferencias en las

dimensiones del TMMS-24 en función de la edad y el sexo, los análisis realizados nos han permitido obtener un resultado interesante. Algunos estudios han mostrado que, en población adulta, las mujeres presentan una mayor tendencia a atender a sus emociones y una menor claridad y reparación en comparación con los hombres (Thayer, Rossy, Ruiz-Padial y Johnsen, 2003). Nuestros resultados indican que algunas de estas diferencias comienzan a aparecer en la adolescencia. Así, las mujeres obtienen mayores puntuaciones en Atención emocional que los hombres. Además, tal diferencia se hacen más pronunciada a medida que avanza la edad de los adolescentes.

Algunos autores han sugerido que las diferencias en el proceso de regulación emocional entre hombres y mujeres están en la base de la mayor prevalencia de problemas emocionales o del uso de estrategias de afrontamiento ineficaces, como la rumiación, por parte de éstas (Nolen-Hoeksema, 2003; Thayer y cols., 2003). Los resultados encontrados muestran que tales diferencias comienzan ya a establecerse a lo largo de la adolescencia, principalmente, en su período final. De confirmarse en estudios longitudinales, estos resultados pueden ser de gran utilidad para adecuar los programas de intervención dirigidos al incremento de competencias socioemocionales en población adolescente, dado que muestran que estas intervenciones han de considerar la variable género para alcanzar sus objetivos. En este sentido, si se pretende mejorar la capacidad de regulación emocional en mujeres, tales programas deberían potenciar estrategias que reducen la rumiación derivada de una atención excesiva a las emociones.

Aunque nuestro estudio ha proporcionado evidencias de la validez del TMMS-24 como una medida de IE autopercebida en población adolescente, futuras investigaciones deben añadir nuevos datos referidos principalmente a la validez convergente y predictiva de la prueba.

### Referencias

- Benet-Martínez, V. y John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729-750.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Extremera, N. y Fernández-Berrocal, P. (2005). Inteligencia emocional percibida y diferencias en el meta-conocimiento de los estados emocionales: una revisión de los estudios con el TMMS. *Ansiedad y Estrés*, 11(2-3), 101-122.
- Fernández-Berrocal, P. y Extremera, N. (2008). A review of trait meta-mood research. *International Journal of Psychology Research*, 2(1), 39-67.
- Fernández-Berrocal, P., Extremera, N. y Ramos, N. (2004). Validity and reliability of Spanish modified version of the Trait Meta-Mood Scale. *Psychological Report*, 94, 47-59.
- Jöreskog, K. G. y Sörbom, D. (1989). *LISREL 7 User's reference guide*. Mooresville, IN.: Scientific Software, International.
- Jöreskog, K. G. y Sörbom, D. (1993). *LISREL 8 User's reference guide*. Chicago, IL.: Scientific Software, International.

- Nolen-Hoeksema, S. (2003). The response styles theory. En C. Papageorgiou & A. Wells (eds.), *Depressive rumination: Nature, theory and treatment of negative thinking in depression* (pp. 107-123). New York: Wiley.
- Otto, J. H., Döring-Seipel, E., Greb, M. y Lantermann, E. D. (2001). Entwicklung eines Fragebogens zur Erfassung der wahrgenommenen emotionalen Intelligenz Aufmerksamkeit auf, Klarheit und Beeinflussbarkeit von Emotionen. *Diagnostica, 4*, 178-187.
- Palmer, B., Gignac, G., Bates, T. y Stough, C. (2003). Examining the structure of the Trait Meta-Mood Scale. *Australian Journal of Psychology, 55*, 154-159.
- Palomera, R. y Brackett, M. (2006). Frecuencia del afecto positivo como posible mediador entre la inteligencia emocional percibida y la satisfacción vital. *Ansiedad y Estrés, 12(2-3)*, 231-239.
- Queirós, M. M., Fernández-Berrocal, P., Extremera, N., Cancela, J. M. C. y Queirós, P. S. (2005). Validação e fiabilidade da versão portuguesa modificada da Trait Meta-Mood Scale. *Revista de Psicologia, Educação e Cultura, 9*, 199-216.
- Salovey, P., Mayer, J.D., Goldman, S.L., Turvey, C. y Palfai, T.P. (1995). Emotional attention, clarity, and repair: exploring emotional intelligence using the Trait Meta-Mood Scale. En J.W. Pennebaker (ed.), *Emotion, Disclosure y Health* (pp.125-151). Washington: American Psychological Association.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.
- Thayer, J. F., Rossy, L. A., Ruiz-Padial, E. y Johnsen, B. H. (2003). Gender Differences in the Relationship between Emotional Regulation and Depressive Symptoms. *Cognitive Therapy and Research, 27*, 349-364.
- Tucker, L. R. y Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

# EVIDENCIAS DE VALIDEZ DE CONTENIDO DE LA VERSIÓN ESPAÑOLA RASGO DEL *STATE TRAIT CHEERFULNESS* *INVENTORY* (STCI)

Elena Delgado y Hugo Carretero-Dios  
Universidad de Granada

La validez de contenido se refiere al grado en el que los ítems de un test recogen las particularidades conceptuales del constructo objeto de medición (Rubio, Berg-Weger, Tebb, Lee y Rauch, 2003). Más específicamente, se trata de obtener evidencias acerca de si los aspectos de contenido que sirven para definir a un constructo, están adecuadamente representados por los ítems de un test (Haynes, Richard y Kubany, 1995). Tal y como fácilmente se desprende de la definición presentada, resulta más que relevante que dentro de todo proceso de construcción/adaptación de tests se trabaje en la dirección de aportar evidencias de validez de contenido. Por ello, el objetivo de este trabajo será mostrar los aspectos más importantes a tener en cuenta en el proceso de obtención de evidencias de validez de contenido. Para ello, se presentarán los datos obtenidos en tal sentido para la adaptación española de la versión rasgo del *State Trait Cheerfulness Inventory*, STCI (Ruch, Khöler y van Thriel, 1996), instrumento que tiene como objetivo la evaluación de las bases afectivas y cognitivas del sentido del humor (alegría, seriedad y mal humor).

Una de las áreas de estudio a tener en cuenta en el proceso de obtención de evidencias validez de las puntuaciones de un test es la referida a la validez de contenido. Se han propuesto diferentes definiciones (AERA, APA y NCME, 1999; Sirecci, 1998), sin embargo, la mayoría de las ellas enfatiza que la validez de contenido es el grado en el que los elementos de un instrumento de evaluación son representativos y relevantes del constructo que se pretende evaluar en una propuesta de evaluación particular (Haynes, Richard y Kubany, 1995). La validez de contenido es un componente fundamental de la validez de constructo, pues además de ayudar a refinar el constructo (Smith y McCarthy, 1995), es la base para una adecuada explicación de la varianza en las puntuaciones obtenidas (Haynes et al., 1995). Sin embargo, y pesar de lo que se acaba de señalar, quizá lo que confiere valor a la validez de contenido es que aporta evidencias acerca del grado en el que los elementos de un instrumento de evaluación son representativos y relevantes del constructo que se pretende evaluar (Haynes, et al., 1995). Por ello, resulta más que relevante que dentro de todo proceso de construcción/adaptación de instrumentos de evaluación se trabaje en la dirección de aportar evidencias de validez de contenido. Así, el objetivo de este trabajo será mostrar los aspectos más importantes a tener en cuenta en el proceso de obtención de evidencias de validez de contenido. Para ello, se presentarán los datos obtenidos en tal sentido para la adaptación española de la versión rasgo del *State Trait Cheerfulness Inventory*, STCI-T (Ruch, Khöler y van Thriel, 1996), instrumento de 106 ítems que tiene como objetivo la evaluación de las bases afectivas y cognitivas del sentido del humor (alegría, seriedad y mal humor) en sus manifestaciones rasgo.

*Obtención de evidencias de validez de contenido: una perspectiva aplicada*

## 1. Delimitación conceptual del constructo a evaluar

Un paso indiscutible a la hora de emprender el proceso de validación del contenido de un instrumento de evaluación es realizar la delimitación conceptual del constructo a evaluar. La conceptualización del constructo comprende la concreción inicial de las facetas o componentes operativos de éste (definición semántica de la variable), junto con la valoración a través de expertos de dicha definición (Carretero-Dios y Pérez, 2005; Carretero-Dios, Pérez, y Buela-Casal, 2006).

Existen numerosos trabajos que pueden servir de ayuda para realizar la concreción conceptual (Haynes et. al., 1995; Murphy y Davidshofer, 1994; Osterlind, 1989; Walsh, 1995). En el ejemplo que se presenta en este trabajo, situado dentro de un proceso más general centrado en la adaptación del STCI-T, se ha asumido, tras la adecuada revisión, la propuesta teórica realizada por los autores de la escala original. En la tabla 1 se presentan las definiciones de los componentes evaluados por el STCI-T (*Alegría, Mal humor y Seriedad*) y de cada una de las facetas delimitadas para éstos (Ruch, Köhler y Van Thriel, 1996).

*Tabla 1. Delimitación conceptual de los constructos alegría, mal humor y seriedad*

---

#### *ALEGRÍA (AL)*

Disposición o tendencia del estado de ánimo caracterizada, entre otras cosas, por la presencia habitual de un sentimiento de entusiasmo, regocijo, jovialidad, etc. Concretamente, la alegría como disposición del estado de ánimo quedaría definida a través de sus distintos componentes, que son:

- AL<sub>1</sub> Predominio de una afectividad alegre, de un estado de ánimo “chisposo”, vivaz, etc.
  - AL<sub>2</sub> Bajo umbral para la sonrisa y la risa, gran facilidad para reírse y expresar diversión.
  - AL<sub>3</sub> Visión positiva de las adversidades de la vida, tendencia a ver lo bueno de los acontecimientos negativos, a abordarlos de manera optimista.
  - AL<sub>4</sub> Visión de un extenso rango de estímulos cotidianos como divertidos, con la propiedad de hacer reír, o facilidad para ver lo divertido o gracioso de las situaciones rutinarias, mostrándose tendencia a la búsqueda y goce de todo lo que conlleve diversión, alegría, etc.
  - AL<sub>5</sub> Estilo de interacción generalmente alegre, gusto por hacer reír o reírse con los demás, por compartir celebraciones, reuniones, etc., donde la diversión, risas, etc., estén presentes.
- 

#### *MAL HUMOR (MH)*

Disposición o tendencia del estado de ánimo caracterizada, entre otras cosas, por la presencia habitual de un malestar afectivo general, un sentimiento de enojo, enfado, etc. Concretamente, el mal humor como disposición del estado de ánimo quedaría definida a través de sus distintos componentes, que son:

- MH<sub>1</sub> Predominio de un estado difuso de mal humor, que se caracteriza por la presencia de sensaciones generales de malestar afectivo, incomodidad, desagrado, etc.
  - MH<sub>2</sub> Predominio de la tristeza (abatimiento, decaimiento, desgana, etc.).
  - MH<sub>3</sub> Dificultad para disfrutar o para mostrar manifestaciones de alegría debido a que predomina un estado de abatimiento o tristeza, ocurriendo esto incluso ante situaciones alegres o humorísticas.
  - MH<sub>4</sub> Sentirse frecuentemente malhumorado, enfadado (gruñón, brusco, cascarrabias, irritable, etc.)
  - MH<sub>5</sub> Comportamientos y actitudes malhumoradas ante circunstancias alegres y evocadoras de humor, y hacia las situaciones, personas, objetos y roles envueltos.
- 

#### *SERIEDAD (SE)*

Actitud caracterizada, entre otras cosas, por la formalidad, solemnidad, sobriedad, etc., con la que se consideran y afrontan la mayoría de los eventos y situaciones de la vida. Concretamente, la seriedad como actitud o visión del mundo quedaría definida a través de sus distintos componentes, que son:

- SE<sub>1</sub> Predominio de un estado de seriedad, reflexión, solemnidad, formalidad, responsabilidad, etc.
- SE<sub>2</sub> Percepción de los sucesos de la vida diaria como importantes y tendencia a considerarlos minuciosa y cuidadosamente.
- SE<sub>3</sub> Tendencia a planear anticipadamente un amplio rango de objetivos, y alcanzar un estado lo más cercano a la “armonía” y al bienestar personal con las decisiones y acciones que envuelven a la consecución de esos objetivos.

- SE<sub>4</sub> Tendencia a preferir actividades para las que puedan darse razones racionales y concretas, y a considerar las actividades que no tienen un objetivo o razón clara, como un desperdicio, una pérdida de tiempo o como algo sin sentido.
- SE<sub>5</sub> Preferencia por un estilo de comunicación sobrio, orientado al objeto, diciendo lo que exactamente uno quiere decir, sin exageraciones o matices sarcásticos o irónicos.
- SE<sub>6</sub> Actitud de rechazo hacia la diversión, hacia las personas, comportamientos, situaciones, acciones, roles, etc., que se relacionan con la diversión, el bullicio, jolgorio, etc.

## 2. Construcción de ítems

Una vez que se cuenta con un constructo claramente definido y delimitado en cuanto a sus facetas y/o componentes, y tras la pertinente revisión de la definición propuesta, debe iniciarse la tarea de construcción de ítems. Al iniciar el proceso de construcción/adaptación de ítems se recomienda partir de una *tabla de especificaciones de los ítems*. Dicha tabla facilitará que se obtengan ítems más relacionados con los intereses de partida, además de posibilitar la creación de ítems por profesionales distintos a los autores de la construcción/adaptación, aumentando la cantidad, variedad y calidad de éstos (Carretero-Dios, Pérez y Buena-Casal, 2009; Osterlind, 1989).

El proceso de adaptación de una prueba no debe entenderse como la mera traducción de los ítems originales. De esta forma, y aunque los ítems originales podrían servir de anclaje para iniciar el proceso de adaptación, se recomienda contar con nuevos ítems a partir de la definición semántica del constructo original, lo que permitirá la adecuada representación de éste para el nuevo contexto cultural de evaluación.

Para este trabajo, los ítems originales (38 ítems *Alegría*; 31 ítems *Mal humor*; 37 ítems *Seriedad*) fueron sometidos a un proceso de traducción inversa. Posteriormente, los autores de este trabajo, por separado y teniendo en cuenta la definición semántica del constructo, elaboraron un nuevo pool de ítems para cada una de las facetas del constructo. La tabla 2 muestra algunos ejemplos de ítems traducidos y nuevos para algunas de las facetas de *Seriedad*. Finalmente, los autores tras una puesta en común y discusión de los ítems nuevos, seleccionaron a aquellos que representaban en mejor medida las facetas del constructo y que cumplían los aspectos formales a tener en cuenta en la redacción de ítems (Moreno, Martínez y Muñiz, 2006). Esto dio lugar a una versión del STCI-T constituida por 188 ítems (66 ítems *Alegría*; 53 ítems *Mal humor*; 69 ítems *Seriedad*).

Tabla 2. Ejemplo de ítems traducidos y nuevos para algunas facetas de *Seriedad*

Facetas	Ítems originales y traducidos	Ítems nuevos
<b>SE1:</b> Predominio de un “estado de seriedad”	<i>I am a serious person</i> Soy una persona seria	Suelo estar serio
<b>SE3:</b> Tendencia a planear anticipadamente un amplio rango de objetivos, y alcanzar un estado lo más cercano a la armonía, con las decisiones y acciones que envuelven a esos objetivos.	<i>I tend to plan far in advance and to set long-term goals for myself</i> Normalmente planifico las cosas con mucho tiempo de antelación, poniéndome metas a largo plazo.	Para sentirme bien conmigo mismo necesito hacer todo lo que tenía planificado para ese día.

## 3. Juicio de expertos de la batería inicial de ítems

Una vez constituida la batería preliminar de ítems, la siguiente fase consistió en contactar con un grupo de expertos para que juzgasen la calidad formal de éstos a la vez que su representatividad y pertenencia en función de la faceta objetivo para la que fueron creados.

## Método

### *Participantes*

Se seleccionaron un total de 6 evaluadores para cada una de las dimensiones recogidas en el STCI-T (alegría, mal humor y seriedad). Los expertos seleccionados eran estudiantes de doctorado o impartían clase en asignaturas del área de metodología de las ciencias del comportamiento. Respecto a los estudiantes de doctorado, todos ellos se encontraban realizando su tesis doctoral en programas del Departamento de Psicología Social y Metodología de las Ciencias del Comportamiento de la Universidad de Granada, y tenían experiencia en el proceso de adaptación/construcción de tests. A todos ellos se les invitó a participar voluntariamente en el estudio.

### *Material*

Se elaboró un cuadernillo donde aparecían las instrucciones de la tarea, la delimitación conceptual del constructo y facetas, ítems de esa dimensión ordenados al azar, y aspectos a juzgar de cada ítem (representatividad, pertenencia, comprensión, interpretación y claridad) con su escala de respuesta. Además, debajo de cada ítem había un apartado por si querían formular una redacción alternativa al ítem o se quería formular alguna aclaración.

### *Procedimiento*

Tras solicitar la participación personalmente, se les entregaba el cuadernillo que debían cumplimentar en un período máximo de una semana. Una vez finalizada la tarea, los participantes entregaban el cuadernillo al investigador principal del estudio.

## Resultados

Se calculó el Índice de Validez de Contenido (IVC) y el Coeficiente de Kappa para cada uno de los ítems y para el instrumento en general. Según Fleiss (1971), el mínimo de IVC y Kappa para obtener evidencias de validez de contenido debe ser de 0,70 y 0,40, respectivamente. Todo el instrumento debería ser extensamente revisado si los IVC general y Kappa no fueran satisfactorios. Así, todos los ítems que no alcanzaron los valores apuntados fueron eliminados o modificados tras la puesta en común entre los autores de este trabajo. Los comentarios cualitativos realizados para cada uno de los ítems guiaron las acciones que finalmente se llevaron a cabo, teniendo siempre en cuenta como criterio definitivo de decisión la definición semántica del constructo presentada en el apartado 1 de este trabajo.

Tras los análisis, se eliminaron un total de 60 ítems: 16 para *Alegría*, 24 para *Seriedad* y 20 para *Mal humor*. Por tanto, los análisis de validez de contenido dieron lugar a una versión del STCI-T con 128 ítems y que será la usada en los siguientes pasos del proceso de adaptación del STCI-T.

## Comentarios finales

Los datos obtenidos ponen de manifiesto como el proceso de validación lleva a una depuración de los ítems creados, tanto en lo que concierne a los aspectos formales de redacción como la relevancia-pertenencia teórica de dichos ítems. Así, la obtención de evidencias de validez de contenido permite

desde el inicio aportar datos empíricos que respalden el proceso de construcción de la prueba, lo que facilitará los siguientes análisis empíricos destinados a seguir los estudios de la escala o prueba en construcción/adaptación. El presente trabajo se ha abordado desde una perspectiva “tradicional” de cara a la obtención de validez de contenido. Por ello, resultaría interesante que futuras investigaciones comparasen los resultados facilitados por las diferentes estrategias.

### Referencias

- AERA, APA y NCME (1999). *Standards for educational and psychological tests*. Washington DC: American Psychological Association, American Educational Research Association, National Council on Measurement in Education.
- Carretero-Dios, H. y Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5, 521-551.
- Carretero-Dios, H., Pérez, C. y Buela-Casal, G. (2006). Dimensiones de la apreciación del humor. *Psicothema*, 18, 465-470.
- Carretero-Dios, H., Pérez, C. y Buela-Casal, G. (2009). Content validity and metric properties of a pool of items developed to assess humor appreciation. *Spanish Journal of Psychology*, 12, 773-787.
- Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378-382.
- Haynes, S.N., Richard, D.C.S. y Kubany, E.S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7, 238-247.
- Moreno, R., Martínez, R.J. and Muñiz, J. (2006). New guidelines for developing multiple-choice items. *Methodology*, 2, 65-72.
- Murphy, K.R. y Davidshofer, C.O. (1994). *Psychological testing: Principles and applications* (3ª ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Osterlind, S.J. (1989). *Constructing Test Items*. Londres: Kluwer Academic Publishers.
- Rubio, D.M., Berg-Weger, M., Tebb, S.S., Lee, E.S. y Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27, 94-104.
- Ruch, W., Köhler, G. y Van Thriel, H. (1996). Assessing the “humorous temperament”: construction of the facet and standard trait forms of the State-Trait-Cheerfulness-Inventory—STCI. *Humor: International Journal of Humor Research* 9, 303-339.
- Sireci, S.G. (1998). Gathering and analyzing content validity data. *Educational Measurement*, 5, 299-321.
- Smith, G.T. y McCarthy, D.N. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, 7, 300-308.
- Walsh, W.B. (1995). *Tests and assessment*. Nueva York: Prentice-Hall.

# OBTENCIÓN DE EVIDENCIAS DE VALIDEZ PARA UN CUESTIONARIO SOBRE “APOYO FAMILIAR” MEDIANTE ENTREVISTAS COGNITIVAS

Miguel Castillo, Isabel Benítez y José Luis Padilla  
Universidad de Granada

Dada la demanda actual sobre la realización de análisis empíricos y teóricos de los procesos de respuesta, con el fin de obtener evidencias de validez sobre el ajuste entre el constructo que se supone mide el test o el cuestionario y las respuestas dadas por los participantes; y el creciente uso de las escalas psicológicas en los cuestionarios de las encuestas. Los métodos de pretest, tales como la entrevista cognitiva, poseen una potencial utilidad para dar respuesta a esta situación. El objetivo del estudio fue mostrar cómo la entrevista cognitiva puede ser usada para la obtención de evidencias de validez de las respuestas de una escala psicológica y, por tanto, contribuir a la optimización de los datos recogidos a través de las encuestas que usan este tipo de escalas. Para alcanzar este objetivo se llevaron a cabo 21 entrevistas cognitivas en el marco del pretest de la escala psicológica APGAR. Los resultados mostraron que los procesos de respuesta de los participantes se ajustaban a las interpretaciones y las conductas esperadas por los autores de la encuesta; a excepción, del concepto clave “familia” cuya interpretación no coincidía con la esperada.

Las escalas psicológicas están siendo frecuentemente utilizadas en las encuestas, con el fin de evaluar constructos psicológicos o de salud. Ejemplos de ello son, el Cuestionario General de Salud en la Encuesta Nacional de Salud, las escalas SF-12 en la Encuesta Europea de Salud, o el uso de grupos de ítems politómicos en diversas rondas de la Encuesta Social Europea. Este creciente uso de las escalas psicológicas en las encuestas plantea interrogantes sobre como analizar las características métricas de las escalas psicológicas en el contexto de la investigación por encuesta.

A esta situación hay que añadir en el contexto de la evaluación psicológica, la demanda reciente para realizar análisis empíricos y teóricos de los procesos de respuesta, con el fin de obtener evidencias de validez sobre el ajuste entre el constructo que se supone mide el test o el cuestionario y las respuestas dadas por los participantes. Esta demanda está claramente reflejada en los *Standards for Educational and Psychological Testing* (APA, AERA, NCME, 1999). De acuerdo con los *Standards for Educational and Psychological Testing* (APA et al., 1999): “la evaluación de las estrategias de ejecución, o estrategias de respuesta de los encuestados a las preguntas, puede ofrecer evidencia que enriquezca la definición del constructo, y estas evidencias de validez basadas en los procesos de respuesta generalmente provienen del análisis de las respuestas individuales” (p. 12).

Hay números estudios que muestran la utilidad de los métodos de pretest, como la entrevista cognitiva, en la optimización de la información obtenida en las encuestas (Willis, 2004; Conrad & Blair, 2004). Por tanto existe una potencial utilidad a la hora de obtener información de las estrategias de respuesta de los encuestados y, por tanto, ofrecer información sobre el ajuste entre el constructo y las respuestas de los participantes.

El presente estudio tiene como objetivo mostrar la utilidad de la entrevista cognitiva para la obtención de evidencias de validez de los procesos de respuesta de estas escalas psicológicas.

## Método

### *Participantes*

En este estudio participaron 21 personas, 10 hombres y 11 mujeres con edades comprendidas entre 20 y 67 años. Se seleccionaron potenciales participantes que podrían ser entrevistados para la encuesta de salud. Las variables de distribución que se tuvieron en cuenta en la selección de participantes fueron las siguientes: a) equivalente distribución de género; b) rango de edad entre 18-70 años; c) equivalente distribución del estado civil; d) equivalente distribución del nivel de estudios y e) tipo de vivienda, a saber, participantes en vivienda uni-personal (personas que viven solas) y vivienda multi-personal. La Tabla 1 ofrece la distribución de valores de las variables nivel de estudios y tipo de vivienda.

*Tabla 1. Características de los Participantes*

Tipo de Vivienda	Nivel de Estudios			Total
	Básico	Medio	Superior	
Unipersonal	3	3	3	9
Multipersonal	6	3	3	12
Total	9	6	6	21

Es necesario resaltar que la razón de la utilización de la variable “tipo de vivienda” se basó en la temática de la escala APGAR, en la unidad de muestreo de la encuesta de salud y en la validez de la información proporcionada por participantes provenientes de ambos tipos de viviendas. La captación de los participantes fue realizada mediante el procedimiento de “bola de nieve”, a través de los contactos de miembros del equipo de investigación de la Universidad de Granada.

### *Material*

Los dos instrumentos utilizados en el estudio fueron, por una parte, la escala psicológica a ser evaluada y, por otra parte, la guía de la entrevista; es decir, la escala APGAR y el Protocolo de Entrevista.

### *La escala APGAR*

La escala APGAR fue diseñada por Smilkestein en 1978 para evaluar la percepción de los miembros de la familia sobre el apoyo familiar, examinando su satisfacción con las relaciones familiares. Esta escala psicológica ha sido incluida en las prácticas clínicas por médicos de familia, como una herramienta útil para obtener información sobre la situación familiar y su posible papel en el origen y solución de conflictos (Bellón, Delgado, Lunay y Lardelli, 1996). La escala APGAR es una técnica rápida y sencilla que sirve para explorar la función familiar evaluando 5 componentes: adaptabilidad (*adaptability*), cooperación (*partnertship*), desarrollo (*growth*), afectividad (*affection*) y capacidad resolutive (*resolve*) (Smilkstein, 1978). Se compone de cinco ítems con tres alternativas de respuesta (p. ej. “¿Está satisfecho con la ayuda que recibe de su familia cuando tiene un problema? Casi Nunca, A Veces o Casi Siempre”). Su traducción al español se llevó a cabo mediante el método de traducción inversa. El valor de Alpha de Cronbach de las respuestas de la escala APGAR en los estudios que se ha utilizado se sitúa alrededor de .84, así como, los análisis factoriales aplicados muestran una solución de un solo factor (Bellón et al., 1996).

## Protocolo de Entrevista

El protocolo de entrevista es la guía para la entrevista cognitiva. Este material incluye información y descripción del objetivo de la entrevista cognitiva y las pruebas a aplicar, las instrucciones que se le deben dar al participante al comenzar la entrevista cognitiva y las pruebas de indagación.

### *Procedimiento*

En este estudio se decidió aplicar un diseño retrospectivo. Es decir, en primer lugar se administró por completo la escala APGAR y tras la última respuesta del participante al último ítem de la escala comenzaron a aplicarse las diversas pruebas de indagación. Las entrevistas se grabaron en audio y vídeo tras el consentimiento de los participantes, y tuvieron lugar en un laboratorio cognitivo equipado de un sistema de grabación y digitalización de audio y vídeo en la Facultad de Psicología de la Universidad de Granada.

### *Esquema de Análisis*

Los autores de la encuesta de salud, en la que fue insertada la escala APGAR, establecieron las interpretaciones previstas para las mediciones aportadas por dicha escala. Estas son: a) los encuestados deberían mantener en mente un concepto constante de “familia” mientras responden a todos los ítems de la escala, y b) la interpretación prevista para el concepto “familia” es aquella que incluye a las personas con las que se convive, es decir, las respuestas de los entrevistados a los ítems del APGAR sólo debían tener en cuenta e incluir a aquellas personas con las que conviven los entrevistados. De este modo el esquema de análisis de las entrevistas cognitivas estaba totalmente enfocado a la obtención de evidencias de validez de estas interpretaciones previstas y, por consiguiente, a la identificación de los potenciales errores de respuesta de la escala APGAR.

La Tabla 2 muestra las fuentes potenciales de error, las interpretaciones previstas de la escala APGAR y las pruebas designadas para indagar sobre la presencia de dichos errores. Las potenciales fuentes de error identificadas por los expertos son: conceptos clave como “familia”, “recibir ayuda de la familia” o “tomar decisiones juntos”; la capacidad de los entrevistados para graduar su respuesta y, por tanto, ajustarla al sistema politómico de respuesta de la escala; la comprensión del objetivo de medición de la escala y la aceptación del rol como entrevistado.

*Tabla 2. Esquema de Análisis de la Entrevista Cognitiva*

Fuentes Potenciales de Error	Interpretación Prevista	Pruebas de Indagación
Concepto de Familia	Se refiere a las personas con las que convive	¿Cuántas personas forman tu familia más cercana?
Recibir ayuda de la familia	Menciona problemas, situaciones o eventos específicos	¿Qué has entendido por “recibir ayuda de la familia cuando tiene un problema”?
	Se refiere a las personas con las que convive	¿De qué miembros de su familia recibe ayuda cuando tiene un problema?
Tomar decisiones importantes juntos	Menciona problemas, situaciones o eventos específicos	Dime, con tus propias palabras, ejemplos de decisiones que se tomen en conjunto en la casa
	Se refiere a las personas con las que convive aunque no sea a todos	¿Qué miembros de su familia toman en conjunto esas decisiones?
Función del objetivo de la pregunta	Comprende el objetivo de medición de las preguntas	¿Cuál cree que es el objetivo de estas preguntas sobre su familia?
Perspectiva y rol del entrevistado	No expresa malestar	¿Cómo se ha sentido respondiendo a estas preguntas sobre su familia?

El análisis de los datos de las entrevistas cognitivas se llevó a cabo utilizando el programa AQUAD, versión 6.8 (Huber, 2008).

## Resultados

### *Interpretaciones Generales*

El análisis de las respuestas de los entrevistados a las pruebas generales de indagación mostró que dichos entrevistados habían pensado, durante la administración de la escala APGAR, en diversas situaciones con su familia, en las relaciones dentro de la familia, en los miembros de la misma, en problemas (accidentes, problemas laborales, etc.) en los que recibieron ayuda de su familia, etc. En el Ejemplo 1 se muestran algunos extractos de transcripciones.

#### *Ejemplo 1. Interpretaciones Generales*

- Entrevistador: “*Por favor, dígame en que ha pensado mientras respondía a estas preguntas sobre apoyo afectivo*”
- Entrevistado: “*Mayormente he pensado después de mi accidente. Antes del accidente teníamos una buena relación familiar pero después, fue mucho mejor. Para mí es como si me hubiera tocado la lotería con mi familia, ellos me dan una ayuda incondicional...*”
- Entrevistador: “*Por favor, dígame en que ha pensado mientras respondía a estas preguntas sobre apoyo afectivo*”
- Entrevistado: “*Bueno... pues no sé, estaba pensando en mi situación familiar... mi hijo mayor está casado y nos visita todos los sábados, el menor se va a casar este viernes... somos una familia unida... intentamos mantenernos en contacto...*”

Los informes verbales ilustran que los participantes, en términos generales, han comprendido la temática del texto de los ítems e incluso comienzan a ofrecer información sobre la interpretación de algunos conceptos clave en el texto de los ítems, como es el caso del tercer entrevistado.

### *Informes verbales sobre conceptos claves, objetivo y rol del entrevistado*

En primer lugar, con respecto a la interpretación prevista para el concepto de “familia”, los análisis de los informes verbales han mostrado que los participantes mantienen el concepto de familia constante a lo largo de las respuestas de la escala. Sin embargo, llevan a cabo una interpretación del concepto distinta a la que previamente se había establecido. Es decir, los participantes incluyen en su concepto de familia no sólo a las personas con las que conviven sino que también incluyen a otros parientes e incluso amigos. Este efecto es mucho mayor en los participantes provenientes de hogares uni-personales.

En segundo lugar, con respecto a los conceptos “recibir ayuda de la familia” o “tomar decisiones juntos”, el discurso de los participantes demostró que sus procesos de respuesta coincidían con las interpretaciones previstas para cada uno de estos conceptos. Los participantes mencionaron problemas en los que habían recibido apoyo familiar o situaciones específicas en las que se tomaron decisiones en conjunto en la familia.

En tercer lugar, en relación a la comprensión del objetivo de medición de la escala y la aceptación del rol como entrevistado, los participantes comprendieron el objetivo de medición de la escala y mostraron haberse sentido muy cómodos contestando a los ítems de la escala.

## Discusión

El objetivo del presente estudio fue mostrar cómo la utilidad de la entrevista cognitiva para la obtención de evidencias de validez de los procesos de respuesta de estas escalas psicológicas.

A través del método de la entrevista cognitiva se han obtenido evidencias de que la interpretación prevista para el concepto de “familia” no coincide con la que llevan a cabo los entrevistados. Los entrevistados no sólo incluyen en su concepto de familia a las personas con las que conviven sino que además incluyen a otros parientes cercanos e incluso a amigos. Con respecto al resto de las potenciales fuentes de error, los datos han mostrado que los participantes llevan a cabo las interpretaciones previstas, tales como, mencionar problemas en los que habían recibido apoyo familiar o situaciones específicas en las que se tomaron decisiones en conjunto en la familia, comprenden el objetivo de medición de la escala y muestran haberse sentido muy cómodos contestando a las preguntas de la escala.

Desde un punto de vista metodológico este estudio ha expuesto la utilidad de la entrevista cognitiva para obtener evidencias de validez de los procesos de respuesta de los participantes, consiguiendo con ello un mayor ajuste entre el constructo que se quiere medir y las estrategias de respuesta que llevan a cabo los participantes durante la medición de dicho constructo.

## Referencias

- American Psychological Association, American Educational Research Association & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association.
- Bellón, J. A., Delgado, J., Luna, P. & Lardelli, P. (1996). Validez y fiabilidad del cuestionario de función familiar Apgar-familiar. *Atención Primaria*, 18, 289-296.
- Conrad, F.G. & Blair, J. (2004). Data Quality in Cognitive Interviews: The Case of Verbal Reports. En S. Presser, J. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J.Martin & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questions* (pp. 67-87). Nueva Jersey: John Wiley & Sons.
- Huber, G.L. (2008). *AQUAD El Programa para Analizar Datos Cualitativos (versión 6.8.1.1)*. Universidad de Tübingen, Alemania: Ingeborg Huber Verlag.
- Smilkstein, G. (1978). The Family APGAR: A proposal for family function test and its use by physicians. *Journal of Family Practice*, 6, 1231-1239.
- Willis, G.B. (2004). Cognitive Interviewing Revisited: A Useful Technique, in Theory? En S. Presser, J. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J.Martin & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questions* (pp. 23-43). Nueva Jersey: John Wiley & Sons.

# APLICACIÓN DE LAS TÉCNICAS DE DIF A LOS CUESTIONARIOS UTILIZADOS EN LAS ENCUESTAS

Juana Gómez-Benito<sup>1</sup>, M. Dolores Hidalgo<sup>2</sup>, José Luis Padilla<sup>3</sup> e Isabel Benítez<sup>3</sup>

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Universidad de Murcia

<sup>3</sup>Universidad de Granada

Los conceptos y los métodos psicométricos pueden ayudar a incrementar la validez de los cuestionarios utilizados en las encuestas. La inclusión de escalas y cuestionarios psicológicos en los cuestionarios de las encuestas es cada vez más frecuente, especialmente, en encuestas de salud, calidad de vida, problemas sociales, etc. El objetivo de este trabajo es ilustrar la aplicación de las técnicas de detección del DIF con datos de encuesta. A modo de ejemplo, se presentarán los resultados obtenidos en la aplicación de la Regresión Logística a los ítems politómicos del Cuestionario General de Salud incluido en una encuesta nacional de salud. El trabajo ilustra como las evidencias sobre el posible DIF de los ítems pueden ayudar a construir argumentos de validez, cuando se comparan grupos de encuestados definidos por variables demográficas o culturales.

La utilización de escalas psicológicas dentro de los cuestionarios de las encuestas es cada vez más frecuente. Estos cuestionarios permiten medir variables psicológicas durante la administración de una encuesta, como es el caso de la Encuesta Nacional de Salud (ENS, 2007), que incluye cuestionarios psicológicos para medir aspectos relevantes en el concepto de salud, como pueden ser la salud mental, el apoyo social o el apoyo familiar.

La evaluación de las encuestas ha estado centrada tradicionalmente en la aparición de los errores relacionados con el muestreo. Sin embargo, en el error total en la encuesta también se incluyen los errores de no muestreo. Entre los errores no relacionados con el muestreo se encuentran los errores de medida que tienen que ver con: a) la especificación, o traducción inadecuada de los objetivos en las preguntas de la encuesta; b) los entrevistados, en relación a dificultades en la comprensión o recuerdo de la información o la emisión de una respuesta inadecuada; c) el cuestionario, por ejemplo un diseño deficiente, ambigüedad de las preguntas, flujos complejos, etc.

La evaluación de las escalas psicológicas pretende detectar, y en lo posible eliminar parte de los errores ajenos al muestro, incrementando así la calidad de la encuesta. Una forma de detectar los errores de medida son los estudios del Funcionamiento Diferencial de los Items (DIF por su nombre en inglés). Los estudios del DIF se han aplicado principalmente en contextos educativos y psicológicos, siendo muy limitada su aplicación en el contexto de encuestas. La importancia de detectar ítems con DIF en cuestionarios incluidos en encuestas se basa en el hecho de que la presencia de ítems con DIF altera las propiedades métricas de la escala e influye en sus puntuaciones (Li y Zumbo, 2009).

## *Concepto de DIF y técnicas para su detección*

El DIF se define como la diferencia en la probabilidad que tienen los participantes de dos o más grupos de dar una respuesta concreta a un ítem (Millsap y Meredith, 1992). Se produce DIF cuando sujetos con idéntico nivel en la característica medida –comparables– tienen distintas probabilidades de dar una determinada respuesta a un ítem, en función del grupo al que pertenezcan. Para estudiar el DIF se realizan comparaciones normalmente entre dos grupos sujetos, el grupo mayoritario (Grupo de

Referencia), y el grupo minoritario (Grupo Focal) (Gómez e Hidalgo, 1997). Entre estos dos grupos puede aparecer DIF de modo uniforme y no-uniforme (Mellenbergh, 1982).

Para identificar ambos tipos de DIF existen diversos procedimientos estadísticos que detectan ítems que se comportan diferencialmente (Gómez e Hidalgo, 1997). Una clasificación exhaustiva de las técnicas aplicadas para la detección del DIF puede encontrarse en Hidalgo y Gómez-Benito (2010), y en Teresi y Fleishman (2007). Uno de los procedimientos más utilizados es el Análisis de la Regresión Logística.

### *Regresión Logística*

De todas las técnicas disponibles, el Análisis de Regresión Logística propuesto por Swaminathan y Rogers (1990) para el análisis del DIF, es una de las técnicas más útiles y potentes y proporciona un marco general e integrado para analizar el DIF. La aplicación de la regresión logística puede realizarse con distintas estrategias (Hidalgo y Gómez-Benito, 2006a). La primera estrategia se basa en la comparación de modelos anidados. Se ajustan tres modelos en distintas etapas. En la primera etapa, se ajusta el modelo base de ausencia de DIF (Modelo 1). En la segunda etapa, se añade a la ecuación la variable de agrupamiento, ajustándose el modelo de DIF uniforme (Modelo 2). Por último, en la tercera etapa se introduce en la ecuación la interacción entre el grupo y la puntuación total en el test, valorándose el ajuste del modelo de DIF no-uniforme o modelo completo (Modelo 3). En esta estrategia de análisis un ítem muestra DIF uniforme si el efecto del grupo resulta estadísticamente significativo, mientras que la interacción habilidad por grupo no ejerce ningún efecto sobre el ítem. Por el contrario, si la interacción resulta estadísticamente significativa, el ítem presentaría DIF no-uniforme.

Por lo tanto, se clasifican con DIF aquellos ítems en el que uno de los modelos sea significativo. Sin embargo, la utilización de la puntuación total como criterio de agrupamiento implica incluir en dicha puntuación la aportación de los ítems con DIF y la posibilidad de que la precisión de la habilidad estimada decrezca. Para evitar este problema, Hidalgo y Gómez (2003) proponen utilizar procedimientos de purificación que consisten en realizar un nuevo análisis sólo con aquellos ítems que no presentan DIF en la evaluación inicial.

La detección del DIF mediante regresión logística se complica cuando se utilizan formatos de ítems politómicos. Una posible solución es utilizar Análisis Logístico Discriminante (ADL) (Miller y Spray, 1993), que se basa en un modelo de regresión logística binaria, donde la variable dependiente (criterio) es ahora el grupo de pertenencia y las variables independientes son  $X$  (puntuación total en el test) y  $U$  (respuesta al ítem). El análisis del DIF sigue la misma estrategia de comparación de modelos con el análisis de regresión logística dicotómico.

Una vez aplicado el procedimiento de purificación y detectado el DIF globalmente para un ítem, es necesario realizar un estudio de las categorías. Este estudio se puede llevar a cabo de forma visual o realizando un nuevo Análisis de Discriminación Logístico. Mellenbergh (1982) considera dos tipos de análisis: Modelos de Probabilidad Acumulativa y Modelos de Categorías Adyacentes.

### *Objetivo del estudio*

El objetivo de este estudio es realizar una aplicación de la regresión logística para la detección del DIF en los ítems politómicos de un cuestionario psicológico incluido en una encuesta de salud. Esta aplicación se centrará en las comparaciones entre grupos diferenciados por su país de nacimiento, siendo España el grupo de referencia y Rumanía el grupo focal.

## Método

### *Participantes*

Del total de participantes que respondieron al cuestionario GHQ incluido en la Encuesta Nacional de Salud se seleccionaron aquellos cuyo país de nacimiento era Rumanía, por ser uno de los países con una frecuencia superior a 200 participantes (209). A continuación, se seleccionó una muestra equivalente de participantes procedentes de España (300), contando así con un total de 509 participantes.

### *Instrumento*

El instrumento utilizado fue el Cuestionario de Salud General (General Health Questionnaire) GHQ incluido en el Cuestionario de Adultos de la Encuesta Nacional de Salud (INE, 2007). Existen varias versiones del cuestionario, siendo la más reducida la utilizada en este estudio. Esta versión contiene 12 ítems de respuesta múltiple con cuatro alternativas de respuesta adaptadas a cada una de las preguntas. Estudios de fiabilidad del GHQ realizados en población española muestran un alfa de .78 (Sánchez-López y Dresch, 2008).

### *Análisis*

Los análisis se dividieron en dos partes: por un lado, el análisis de ítem y de la escala los ítems, y por otro el estudio del DIF. En el estudio del DIF se realizó para cada ítem mediante el Análisis Discriminante Logístico a partir del cuál se detectaron los ítems con posible DIF mediante un proceso de purificación bietápica y, por último, se realizó un análisis de las categorías de respuesta.

## Resultados

### *Análisis de los ítems y de la escala*

La tabla 1 muestra por un lado el rango y el promedio de las medias de los ítems y por otro el rango y la media de los índices de discriminación.

*Tabla 1. Análisis de ítems por grupos*

	Estadísticos descriptivos de los ítems		Índices de discriminación	
	Rango	Media	Rango	Media
España	0.330-1.187	0.884	0.407-0.672	0.584
Rumania	0.249-1.043	0.788	0.157-0.611 *	0.418

\*Tres ítems con ID inferiores a 0.30 (3,4 y 8)

Como muestra la tabla 1 el rango de las medias de los ítems y el promedio de las medias es bajo, alcanzando en el grupo de Rumanía el menor valor. En cuanto a los índices de discriminación el promedio es adecuado en España en que todos los ítems presentan índices de discriminación superiores a 0.30 mientras que en Rumanía ya los ítems 3 y 4 recogen valores inferiores a 0.30. En el análisis de la fiabilidad, se obtuvo el coeficiente alfa para cada una de las submuestras: 0.86 para la muestra española y 0.77 para la muestra rumana.

*Estudio del DIF*

A continuación, se presentan los resultados obtenidos para los ítems detectados con DIF tras la purificación y el análisis de las categorías de respuesta realizado en dichos ítems. En cada modelo presentado (tanto en el de DIF Uniforme como en el de DIF No Uniforme) para ambas etapas del estudio del DIF se incluyen los valores de  $G^2$ , la probabilidad de éste y el incremento de  $R^2$ . La tabla 2 muestra los resultados obtenidos en la comparación entre España y Rumania.

Tabla 2. Estudio del DIF España- Rumania

Item	Modelo 1 (DIF Uniforme)			Modelo 2 (DIF No Uniforme)		
	$G^2$	$P$	$\Delta R^2$	$G^2$	$p$	$\Delta R^2$
1	9.179	.002	0.023	0.307	.579	0.001
9	6.840	.009	0.017	2.740	.098	0.007

Como muestra la tabla, fue detectado DIF en los ítems 1 y 9 en el modelo Uniforme. El incremento de  $R^2$  fue en ambos casos inferior a 0.030 y sus valores descendieron tras la purificación.

*Análisis de las categorías*

A continuación, se realizó el análisis de las categorías. En el caso del ítem 1 fue necesario colapsar las categorías 2 y 3 ya que la categoría 3 no fue elegida por ningún participante del grupo de Rumania. Por este motivo no existen valores para la OR3, ya que sólo fueron necesarias dos comparaciones entre las categorías. La tabla 3 muestra los resultados obtenidos.

Tabla 3. Análisis de las categorías de respuesta España – Rumania

Items	País	Frecuencias de elección				Análisis del DIF						
		0	1	2	3	Modelo 1			Modelo 2			
						$G^2$	$p$	OR1	OR2	OR3	$G^2$	$p$
1*	Esp.	6	240	46	8	7.600	.022	0.534	0.446		1.633	.442
	Rum.	10	184	15	0							
9	Esp.	146	98	49	7	6.847	.077	1.489	0.154	0.620	7.854	.049
	Rum.	104	66	35	4							

\* Las categorías 2 y 3 se han colapsado para realizar este análisis por la ausencia de elección de la categoría 3 en el grupo Rumania por lo tanto no existen datos para la OR3.

Como indica la tabla, aparecen altas frecuencias de elección de la categoría 0 en el ítem 9 para ambos países. En el ítem 1, no aparece elegida la categoría 3 en el grupo de Rumania y la frecuencia de elección de la categoría 0 es más alta en este grupo que en el grupo de España. En cuanto al estudio del DIF, la probabilidad de  $G^2$  no fue significativa en ninguno de los dos ítems para ninguno de los modelos, por lo que será necesario realizar análisis que permitan determinar si se trata de falsos positivos.

## Discusión

El estudio del DIF ha identificado diferentes ítems con DIF en la comparación entre España y Rumanía en los ítems 1 y 9. En ambos se detectó DIF sólo a nivel del ítem ya que no apareció DIF en el análisis de las categorías de respuesta.

Cuando estamos evaluando la equivalencia de encuestas aplicadas a varias culturas, tanto la equivalencia del constructo como la de los ítems necesitan ser asegurada. La evaluación del Funcionamiento Diferencial del Ítem (DIF) es un aspecto clave en el proceso. A su vez, dado que la mayor parte de las encuestas tratan con datos politómicos, las técnicas de DIF politómicas son las más apropiadas para identificar DIF.

Por último, futuras líneas de investigación deberán identificar las condiciones más adecuadas para la aplicación del Análisis Discriminante Logístico y las medidas de tamaño del efecto más adecuadas.

## Referencias

- Gómez, J. e Hidalgo, M.D. (1997). Evaluación del funcionamiento diferencial en ítems dicotómicos: una revisión metodológica. *Anuario de Psicología* 74, 3-32.
- Hidalgo, M.D. y Gómez-Benito, J. (2003). Test Purification and the Evaluation of Differential Item Functioning with Multinomial Logistic Regression. *European Journal of Psychological Assessment*, 19 (1), 1-11.
- Hidalgo, M. D. y Gómez-Benito, J. (2006a). Nonuniform DIF detection using Discriminant Logistic Analysis and Multinomial Logistic Regression: A comparison for polytomous items. *Quality & Quantity* 40, 805-823.
- Hidalgo, M. D. y Gómez-Benito, J. (2010). Education measurement: Differential item functioning. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd edition). USA: Elsevier - Science & Technology.
- Instituto Nacional de Estadística. (2007). Encuesta Nacional de salud.
- Li, Z. y Zumbo, B.D. (2009). Impact of Differential Item Functioning on Subsequent Statistical Conclusions Based on Observed Test Score Data. *Psicológica*, 30, 343-370.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics* 7, 105-118.
- Miller, T.R. y Spray, J.A. (1993). Logistic Discriminant Function Analysis for DIF Identification of polytomously scored items. *Journal of Educational Measurement* 30 (2), 107-122.
- Millsap, R. E. y Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement* 17, 297-334.
- Sánchez-López, M.P. y Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema* 20 (4), 839-843.
- Swaminathan, H. y Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement* 27, 361-370.
- Teresi, J.A. y Fleishman, J.A. (2007). Differential item functioning and health assessment. *Quality of Life Research* 16, 33-42.

# METODOLOGÍA PARA LA TRADUCCIÓN Y ADAPTACIÓN DE INSTRUMENTOS DE EVALUACIÓN ENTRE DISTINTOS GRUPOS LINGÜÍSTICOS Y CULTURALES

**Coordinadores: Nekane Balluerka<sup>1</sup> y José Luis Padilla<sup>2</sup>**

<sup>1</sup>Universidad del País Vasco

<sup>2</sup>Universidad de Granada

La metodología para la traducción y adaptación de instrumentos que evalúan características personales y sociales en individuos con diferentes antecedentes lingüísticos y culturales suscita cada vez mayor interés tanto en el ámbito de la investigación como por parte de organizaciones psicológicas y educativas. Este interés se halla estrechamente relacionado con el aumento de procesos de evaluación intercultural. Como ejemplos relevantes de tales procesos cabe destacar los siguientes: a) la adaptación a más de 30 lenguas de tests de rendimiento utilizados en procesos de evaluación de carácter internacional, tales como el *Third International Mathematical and Science Study* (TIMSS), o el *Programme for International Student Assessment* (PISA); b) la utilización internacional de exámenes para la acreditación profesional de grandes multinacionales (*Microsoft, Apple, etc.*) y c) el constante crecimiento de flujos migratorios que genera la necesidad de realizar evaluaciones educativas y sanitarias a grupos de personas con diferentes lenguas maternas dentro del mismo país. La trascendencia de los procesos de evaluación que acabamos de citar pone de manifiesto la necesidad de llevar a cabo evaluaciones interculturales que garanticen la equidad y la validez de las interpretaciones y de las decisiones que se deriven de ellas. Por tal razón, la traducción y adaptación de instrumentos de evaluación entre distintos grupos lingüísticos y culturales debe sustentarse en bases metodológicamente sólidas. Las tres comunicaciones que se incluyen en el presente simposio se centran en cuestiones referidas a la metodología para la adaptación de instrumentos de evaluación de unas culturas a otras. El primer trabajo, desarrollado por los profesores Gorostiaga, Balluerka e Isasi, se centra en el proceso de traducción y adaptación del cuestionario para medir la personalidad en niños y adolescentes EPQ-J (Eysenck y Eysenck, 1997) del castellano a una lengua minoritaria, como es el euskera. En el segundo trabajo, los profesores C. Arce, De Francisco, Andrade, Garrido, Torrado e I. Arce, exponen el proceso de adaptación del instrumento para la medida del *burnout* en deportistas ABQ (*Athlete Burnout Questionnaire*; Raedeke y Smith, 2001) del inglés al castellano. Por último, los profesores Benítez, Castillo, Padilla, Gómez-Benito, Flores-Galaz y Díaz-Loving comparan dos aproximaciones metodológicas, el escalamiento multidimensional y el análisis factorial confirmatorio, para examinar la equivalencia del constructo en contextos monolingües multiculturales.

# ADAPTACIÓN AL EUSKERA DEL EPQ-J (*CUESTIONARIO DE PERSONALIDAD PARA NIÑOS*)

Arantxa Gorostiaga, Nekane Balluerka y Xabier Isasi  
Universidad del País Vasco

El objetivo del presente estudio consistió en adaptar al euskera el Cuestionario para medir la personalidad en niños y adolescentes EPQ-J (*Eysenck Personality Questionnaire-Junior*). Una vez llevada a cabo la fase de traducción de los ítems y su administración a una muestra piloto, el instrumento fue aplicado, en la fase experimental, a una muestra de 886 participantes (432 varones y 454 mujeres) con un rango de edad de entre 8 y 16 años. Tras eliminar los sujetos que puntuaron por debajo del percentil 25 en la escala de Sinceridad así como los ítems de dicha escala, se realizó un análisis factorial confirmatorio a partir de la matriz de correlaciones tetracóricas, en el que no se confirmó la estructura de 3 factores (Neuroticismo, Psicoticismo y Extraversión) del instrumento original. El análisis de conglomerados llevado a cabo en una fase posterior puso de manifiesto la existencia de dos dimensiones, una de ellas formada por los ítems de los factores Neuroticismo y Psicoticismo, y otra que incluía todos los ítems de Extraversión. A su vez, tales dimensiones mostraron una aceptable consistencia interna y estabilidad temporal, así como evidencias de validez convergente y de relación con otras variables.

Desde siempre, los psicólogos han reconocido la importancia de describir los principales modelos de conducta y se han centrado en la búsqueda de las dimensiones fundamentales de la personalidad. Eysenck postuló un modelo de tres grandes dimensiones: Psicoticismo, Extraversión y Dureza (Eysenck, 1991; Eysenck, 1994; Eysenck, Barrett, Wilson, & Jackson, 1992; Eysenck & Eysenck, 1991).

El EPQ (*Eysenck Personality Questionnaire*) se elaboró a partir de diversos cuestionarios de personalidad anteriores. El primero de ellos, el *Maudsley Medical Questionnaire* (Eysenck, 1952), estaba compuesto por 40 elementos y medía el Neuroticismo o Emocionabilidad. A este le siguió el *Maudsley Personality Inventory* (MPI, Eysenck, 1959) que evaluaba el Neuroticismo y la dimensión Extraversión-Introversión. Posteriormente, se elaboró el *Eysenck Personality Inventory* (EPI, Eysenck y Eysenck, 1964) que añadía la escala L para medir la tendencia al disimulo. Finalmente, el EPQ-J (Cuestionario de personalidad para niños, Eysenck y Eysenck, 1975) se elaboró con el fin de evaluar tres dimensiones básicas de personalidad, a saber, el Neuroticismo o Emocionabilidad, la Extraversión y el Psicoticismo o Dureza. Asimismo, incluye una dimensión referida a la sinceridad de las personas evaluadas y otra relacionada con la conducta antisocial. Esta última dimensión está compuesta por ítems pertenecientes a las tres dimensiones básicas de la personalidad. El instrumento está destinado a niños/as y adolescentes de entre 8 y 16 años de edad y se compone de 81 ítems con una escala de respuesta dicotómica: SI/NO.

Partiendo de la importancia de la evaluación de la personalidad en la población infantil, así como de la carencia de instrumentos en euskera para medir tal constructo, el objetivo del presente estudio consistió en adaptar al euskera el Cuestionario para medir la personalidad en niños y adolescentes EPQ-J (*Eysenck Personality Questionnaire-Junior*).

## *Fase de traducción de los ítems*

En la fase de traducción de los ítems del EPQ-J se utilizó un diseño de traducción directa-inversa. En primer lugar, los ítems de la versión en castellano y 11 ítems adicionales utilizados en la versión original en inglés (para posibles sustituciones por originales no satisfactorios) fueron traducidos al euskera de

forma independiente por dos traductores con amplio dominio de ambas lenguas, familiarizados con las dos culturas e instruidos en aspectos psicométricos básicos relacionados con la construcción de ítems. Las dos traducciones fueron comparadas y sometidas a discusión hasta obtener una versión consensuada de cada uno de los ítems. Partiendo de esta versión, otros dos traductores de características similares a los que realizaron la traducción directa, tradujeron, de forma independiente, todos los ítems de la versión en euskera al castellano y llegaron a una versión consensuada del test. Finalmente, los cuatro integrantes del equipo de traductores compararon cada uno de los ítems de las versiones original e inversamente adaptada del instrumento, examinando las posibles no-equivalencias en cuanto al significado. Partiendo de este análisis se llevaron a cabo las modificaciones oportunas en la versión en euskera del EPQ-J.

### *Fase piloto*

A continuación, se llevó a cabo un estudio piloto a fin de obtener datos empíricos acerca del funcionamiento de la versión en euskera del EPQ-J. En el estudio participaron 102 niños/as y adolescentes (50 mujeres y 52 varones) con edades comprendidas entre los 9 y los 14 años. Los/as participantes pertenecían a diversos centros educativos de la Comunidad Autónoma del País Vasco y presentaban un alto dominio de la lengua vasca.

El análisis cuantitativo y cualitativo de los ítems nos llevó a eliminar ocho elementos originales y sustituirlos por aquellos adicionales que presentaban una mejor consistencia interna, así como a reformular algunos ítems que habían generado problemas de comprensión. El instrumento derivado de este proceso está compuesto por 81 ítems y es el que se utilizó en la fase experimental.

## Método

### *Participantes*

En una primera fase, la muestra experimental estuvo compuesta por 886 participantes (454 mujeres y 432 varones) con un rango de edad entre 8 y 16 años. En una segunda fase, se seleccionaron los/as participantes que habían puntuado por encima del percentil 25 en la dimensión de Sinceridad. Así, la muestra definitiva estuvo compuesta por 633 participantes (307 mujeres y 326 varones).

### *Instrumentos*

- Versión en euskera del EPQ-J (Cuestionario de Personalidad para niños de Eysenck y Eysenck, 1975).- Está compuesto por 81 ítems con formato de respuesta SI/NO que evalúan las dimensiones de Emocionabilidad, Extraversión y Dureza, así como la Sinceridad y la Conducta antisocial.
- Versión en euskera del CDS (Escala de depresión para niños de Lang y Tisher, 1978).- Está compuesta por 66 ítems con un formato de respuesta tipo Likert de 5 puntos que evalúan las dimensiones de Total positivo y Total depresivo.
- Cuestionario de variables sociodemográficas que incluye información referida al género, a la fecha de nacimiento, al contexto socio-familiar, etc.

### *Procedimiento*

Un psicólogo especializado en psicodiagnóstico infantil aplicó la versión en euskera del EPQ-J, junto con el CDS y el cuestionario de variables sociodemográficas, a los participantes en sus aulas correspondientes. Cuatro semanas más tarde, se repitió la aplicación del EPQ-J a una submuestra compuesta por 200 participantes.

Como ya se ha señalado en el apartado de Participantes, en una segunda fase se descartaron los/as participantes que mostraban una puntuación inferior al percentil 25 en la dimensión de sinceridad, es decir, aquellos/as que en el conjunto de los 20 ítems que componían esa dimensión habían obtenido 7 o menos puntos, dado que ello indica una elevada tendencia a falsificar las respuestas con el objetivo de presentar un “buen aspecto”.

### Resultados

En los análisis realizados con el fin de examinar las propiedades psicométricas del instrumento no se incluyeron los ítems que componen la escala de Sinceridad, puesto que tal escala no forma parte una de las dimensiones básicas de la personalidad.

### *Estructura factorial*

Con el fin de examinar la dimensionalidad del instrumento se realizó un análisis factorial confirmatorio a partir de la matriz de correlaciones tetracóricas mediante el software LISREL 8.8 (Jöreskog y Sörbom, 1999). Como puede observarse en la Tabla 1, el modelo de tres factores no presentó índices de ajuste adecuados.

*Tabla 1. Índices de bondad de ajuste para el modelo de 3 factores*

$\chi^2$	3662,13 (p=0,001)
g.l.	1766
RMSEA	0,044
GFI	0,83
AGFI	0,82
NNFI	0,78

Una vez comprobado que el modelo de tres factores no ajustaba de manera adecuada, se procedió a utilizar el software Factor (Lorenzo-Seva y Ferrando, 2007) para realizar un análisis de componentes principales con rotación Varimax a partir de la matriz de correlaciones tetracóricas de los 61 ítems. Sin embargo, dado que la matriz de correlaciones no cumplía las condiciones para realizar un análisis factorial exploratorio, se optó por un procedimiento que aunque menos utilizado, permite examinar las agrupaciones entre las variables de un instrumento, a saber, el análisis de conglomerados jerárquico. El método de conglomeración utilizado fue el Vecino más próximo y la medida de distancia seleccionada, la Distancia Euclídea al cuadrado. Dicho análisis puso de manifiesto la existencia de dos conglomerados de variables: el primer conglomerado, a excepción de 3 ítems, estuvo compuesto por los ítems de los factores de Emocionabilidad y Dureza; el segundo conglomerado, a excepción de 1 ítem, incluyó los ítems del factor de Extraversión. En la Figura 1 se puede observar el dendrograma correspondiente al análisis. Se han señalado en negrita los 4 ítems que no se ubicaron en el conglomerado correspondiente.

Consistencia interna

La consistencia interna de cada una de las dos dimensiones de la versión en euskera del EPQ-J se estimó mediante el coeficiente alpha de Cronbach. Dicho índice adoptó los valores de 0,77 y de 0,61 para las dimensiones de Emocionabilidad-Dureza (N-P) y de Extraversión (E), respectivamente.

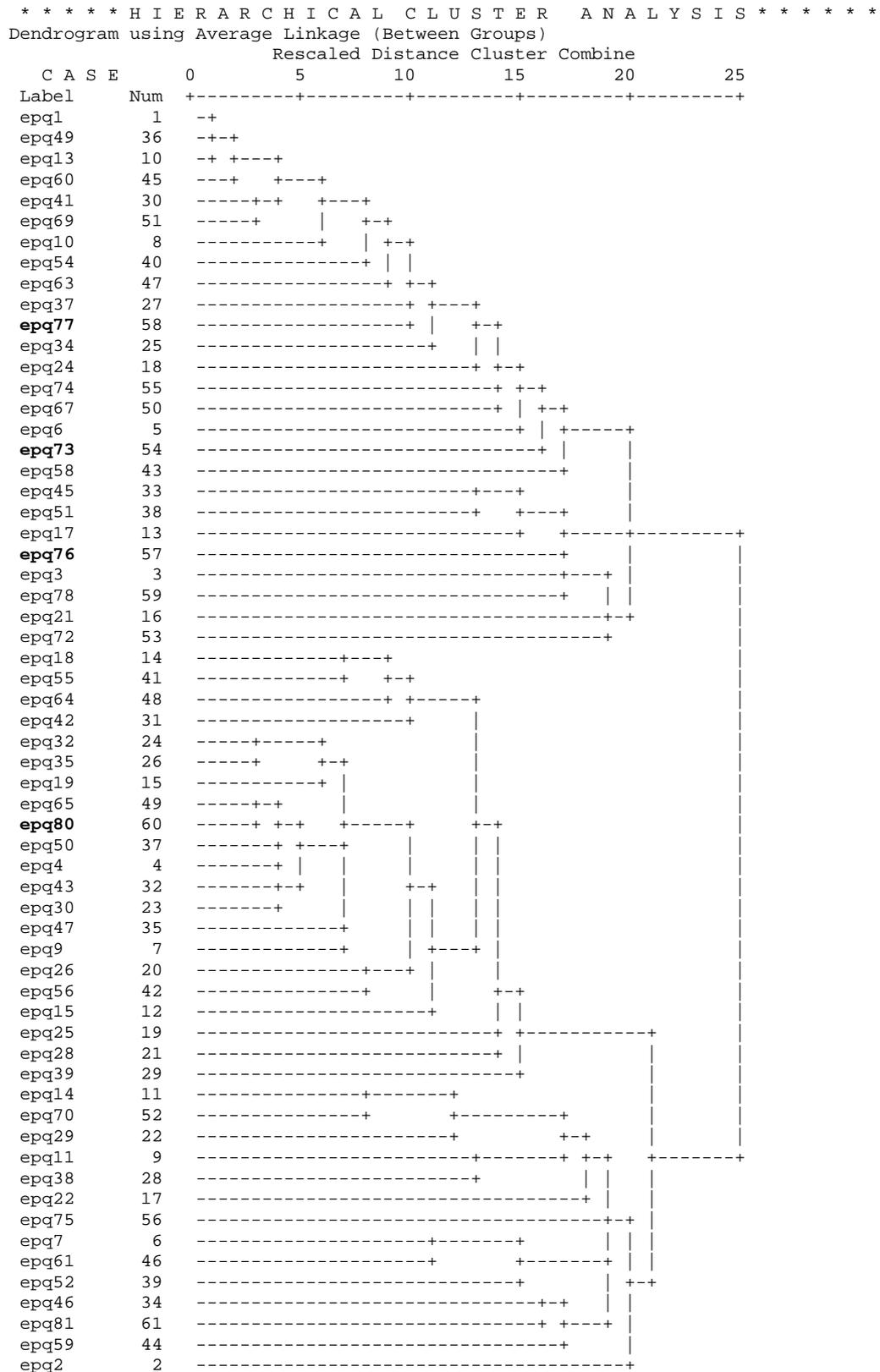


Figura 1. Dendrograma correspondiente al análisis de conglomerados

### *Estabilidad temporal*

La estabilidad temporal del instrumento se examinó estimando los índices de correlación entre las puntuaciones obtenidas por los/as participantes en cada una de sus dimensiones en dos aplicaciones separadas por un intervalo de cuatro semanas. Las correlaciones test-retest de las dos dimensiones del instrumento fueron moderadas, con un valor de 0,55 para la dimensión de Emocionabilidad-Dureza (N-P) y de 0,57 para la dimensión de Extraversión (E).

### *Relación con otras variables*

En primer lugar, se examinó la relación existente entre la dimensión Emocionabilidad-Dureza del EPQ y la dimensión Total depresivo del CDS. El resultado obtenido (Coeficiente de correlación=0,67) es coherente con los resultados encontrados en la adaptación española del EPQ-J (Eysenck y Eysenck, 1978) y en otros trabajos (Del Barrio, Moreno-Rosset, López-Martínez y Olmedo, 1997; Larsen, 1992).

En segundo lugar, a partir de los resultados de estudios previos (Anthony, 1977; Campbell y Hawley, 1982; Goh y Moore, 1978), se pronosticó que los/as niños/as que presentaran un mejor rendimiento académico, mostrarían un mayor nivel de extraversión que los/as niños/as con peor rendimiento. Los resultados confirmaron dicha hipótesis ( $t=-2,894$ ;  $p=0,001$ ;  $d$  de Cohen=0,49). Este resultado parece indicar que los/as niños/as extravertidos/as maduran antes que los/as introvertidos/as.

En tercer lugar, se examinó la diferencia existente en el nivel de Emocionabilidad-Dureza de los/as menores en función del sexo, dado que investigaciones anteriores (Rothen et al., 2008) sugieren que los varones presentan un mayor nivel de dureza que las niñas. Los resultados no confirmaron esta hipótesis ( $t=0,66$ ;  $p=0,509$ ;  $d$  de Cohen=0,05).

Finalmente, de acuerdo con estudios anteriores (Eysenck y Eysenck, 1978), la media en Emocionabilidad-Dureza de niños y niñas mayores de 12 años resultó significativamente más alta que la de los/as menores de esa edad, aunque el tamaño del efecto asociado a esa diferencia de medias presentó un valor entre pequeño y moderado ( $t=-4,71$ ;  $p=0,001$ ;  $d$  de Cohen=0,40).

## Conclusión

En primer lugar, cabe señalar que la versión en euskera del EPQ-J no replica la estructura de tres factores de personalidad de otras versiones. En la población vasca, las dimensiones de Emocionabilidad y Dureza tienden a agruparse.

Los índices de consistencia interna y de estabilidad temporal de la versión en euskera del EPQ-J resultan aceptables.

Por otra parte, el hecho de que algunas de las hipótesis sometidas a prueba con el fin de evaluar la validez del instrumento no se cumplan puede deberse a que las dimensiones evaluadas en estudios anteriores y en nuestro proceso de adaptación no son las mismas.

Aunque los resultados obtenidos sugieren realizar un estudio más exhaustivo de la validez factorial del instrumento con nuevas muestras de niños/as y adolescentes vascoparlantes, cabe afirmar que la versión en euskera del EPQ-J presenta unas características psicométricas aceptables. En consecuencia, creemos que los profesionales del ámbito clínico y educativo disponen de un instrumento que permite estudiar la estructura de personalidad de niños/as y adolescentes cuya lengua materna sea el euskera.

## Referencias

- Anthony, W. S. (1977). The development of extroversion and ability; an analysis of Rushton's longitudinal data. *British Journal of Educational Psychology*, 47, 193-196.
- Campbell, J. B. y Hawley, C. W. (1982). Study habits and Eysenck's theory of Extraversion-Introversion. *Journal of Research in Personality*, 16, 139-146.
- Del Barrio, V., Moreno-Rosset, C., López-Martínez, R. y Olmedo, M. (1997). Anxiety, depression and personality structure. *Personality and Individual Differences*, 23, 327-335.
- Eysenck, H. J. (1952). *The scientific study of personality*. London: Routledge & Kegan Paul.
- Eysenck, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London: University of London Press.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3? Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12, 773-790.
- Eysenck, H. J. (1994). The big five or giant three: criteria for a paradigm. In C. F. Halverson & G. A. Kohnstamm (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 37-51). Hillsdale: Lawrence Erlbaum Associates.
- Eysenck, H. J. y Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder & Stoughton Educational.
- Eysenck, H. J. y Eysenck, S. B. G. (1991). *Manual of the Eysenck personality scales*. London: Hodder & Stoughton Educational.
- Eysenck, H. J., Barrett, P. T., Wilson, G. y Jackson, C. J. (1992). Primary trait measurement of the 21 components of the P-E-N system. *European Journal of Psychological Assessment*, 8, 109-117.
- Eysenck, S. B. G. y Eysenck, H. J. (1964). An improved short questionnaire for the measurement of extraversion and neuroticism. *Life Science*, 305, 1103-1109.
- Eysenck, H. J. y Eysenck, S. B. G. (1978). *EPQ, Cuestionario de personalidad para niños (EPQ-J) y adultos (EPQ-A)*. Madrid: TEA Ediciones, S.A.
- Goh, D. S. y Moore, C. (1978). Personality and Academic Achievement in Three Educational Levels. *Psychological Reports*, 43, 71-79.
- Jöreskog, K. y Sörbom, D. (1999). *LISREL 8: User's Reference Guide*, 2nd ed. Lincolnwood: Scientific Software International.
- Lang, M. y Tisher, M. (1978). *Cuestionario de Depresión para Niños*. Madrid: TEA.
- Larsen, R. J. (1992). Neuroticism and selective encoding and recall of symptoms: Evidence from a combined concurrent retrospective study. *Journal of Personality and Social Psychology*, 62, 487-488.
- Lorenzo-Seva, U. y Ferrando, P. J. (2007). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers*, 38, 88-91.

Rothen, S., Vandeleur, C. L., Lustenberger, Y., Jeanprêtre, N., Ayer, E., Sisbane, F. et al. (2008). Validation of the French version on the EPQ-Junior. *Personality and Individual Differences*, 44, 464-474.

# ADAPTACIÓN DEL *ATHLETE BURNOUT QUESTIONNAIRE* (ABQ) DEL INGLÉS AL CASTELLANO

Constantino Arce, Cristina de Francisco, Elena Andrade, Javier Garrido, Julio Torrado e Iria Arce  
Universidad de Santiago de Compostela

El *Athlete Burnout Questionnaire* (ABQ) es un instrumento desarrollado por Raedeke y Smith (2001) para la medida del *burnout* en deportistas. Se compone de tres subescalas: agotamiento físico/emocional, reducida sensación de logro y devaluación del deporte. El objetivo de este estudio consistió en obtener una versión adaptada al español del ABQ. En una investigación preliminar (De Francisco, 2007), el ABQ fue traducido al castellano y aplicado a 140 jugadores de fútbol masculino con edades comprendidas entre los 16 y 19 años. Teniendo en cuenta las conclusiones de dicho trabajo, se elaboraron enunciados alternativos para algunos de los ítems. Como resultado, se obtuvo una versión preliminar de 21 ítems (15 originales y 6 alternativos) que se aplicó a una muestra de 326 jóvenes futbolistas. Mediante análisis de ítems y análisis factoriales exploratorios, el cuestionario fue reducido a 15 ítems para conservar la estructura del original (cinco ítems para cada una de las tres dimensiones). A continuación, se realizó un análisis factorial confirmatorio en el que se obtuvieron índices de ajuste aceptables con valores de ,97 para CFI, ,96 para NNFI, ,93 para GFI, ,059 para RMSEA y ,061 para RMR.

El *Athlete Burnout Questionnaire* (ABQ) es un cuestionario elaborado por Raedeke y Smith (2001) para evaluar el *burnout* en deportistas. Se compone de 15 ítems distribuidos equitativamente en tres subescalas: agotamiento físico/emocional, reducida sensación de logro y devaluación del deporte que se practica. Análisis factoriales confirmatorios muestran índices de ajuste satisfactorios y estudios como el de Lonsdale, Hodge y Jackson (2007) apoyan la estructural factorial del ABQ. Asimismo, los coeficientes de fiabilidad test-retest son altos situándose entre ,86 para reducida sensación de logro y ,92 para agotamiento físico/emocional y devaluación del deporte (Raedeke y Smith, 2001). Cresswell y Eklund (2006) también aportan datos favorables sobre la validez convergente y discriminante. En España, De Francisco (2007), tras un proceso de doble traducción, estudió las propiedades psicométricas del ABQ mediante análisis factoriales exploratorios y confirmatorios en una muestra de 140 jóvenes futbolistas varones. El trabajo que aquí se expone se realiza con el objetivo principal de mejorar esta versión española con datos procedentes de una muestra de 326 futbolistas de ambos sexos.

## Método

### *Participantes*

La muestra estaba formada por 326 futbolistas. El 72,7% eran hombres y el 27,3% mujeres, con un rango de edad que oscilaba entre los 14 y los 23 años, siendo la media de 16,82 y la desviación típica de 2,21. La media de entrenamientos semanales se situaba en 2,79, con una desviación típica de ,89, entrenando un promedio de 4,47 horas.

### *Instrumento*

El instrumento de medida fue una versión española preliminar del ABQ desarrollada por De Francisco (2007). En el presente trabajo, se optó por elaborar dos enunciados alternativos para tres ítems cuyo funcionamiento no era satisfactorio en dicha versión y añadirlos a los 15 originales. Se empleó, por

tanto, una versión de 21 ítems con formato de respuesta tipo Likert de cinco puntos: *Casi nunca* (1), *Pocas veces* (2), *Algunas veces* (3), *A menudo* (4) y *Casi siempre* (5).

### *Procedimiento*

La versión modificada del ABQ en español con 21 ítems fue aplicada a los participantes por tres psicólogos del deporte entre los meses de noviembre de 2007 y enero de 2008. Se siguió un protocolo estandarizado de forma que todos recibiesen las mismas instrucciones.

## Resultados

El análisis de los datos se realizó en dos etapas secuenciales. En la primera se llevaron a cabo, mediante el programa estadístico SPSS, múltiples análisis factoriales exploratorios (AFE), análisis de ítems y de consistencia interna de los factores. La segunda etapa estuvo orientada a la confirmación de las hipótesis generadas en la etapa anterior mediante análisis factoriales confirmatorios (AFC) realizados con el programa LISREL, que proporciona indicadores de ajuste del modelo.

### *Análisis factorial exploratorio*

Se comenzó el proceso de exploración realizando análisis factoriales con los 15 ítems originales del ABQ que hipotéticamente se agrupaban en tres factores. Dado que las soluciones obtenidas no lograron reproducir nítidamente la solución de tres factores hipotetizada, se procedió a factorizar los 21 ítems utilizados en la investigación. La selección de ítems se realizó atendiendo a su posicionamiento en la estructura factorial, a su contribución a la consistencia interna del factor y a los diferentes matices de significado que pudiera aportar cada enunciado al factor. De este modo, se elaboró una forma del cuestionario que quedó nuevamente conformada por 15 ítems (12 del cuestionario original y 3 alternativos), que se presenta en la Tabla 1.

### *Análisis de consistencia interna*

Los análisis de ítems y de consistencia interna apoyaron la composición de 5 ítems para cada una de las dimensiones teóricas. Para el primer factor, agotamiento físico/emocional, los valores de las correlaciones ítem-factor corregidas oscilaban entre ,54 (ítem 16) y ,66 (ítem 14), siendo el valor de alpha de Cronbach de ,82. Para el segundo factor, reducida sensación de logro, los valores de las correlaciones oscilaron entre ,28 (ítem 3) y ,43 (ítem 10). El coeficiente alpha fue de ,62. Finalmente, para el tercer factor, devaluación del deporte, las correlaciones ítem-factor se situaron entre ,53 (ítem 18) y ,68 (ítem 12), obteniéndose un coeficiente de consistencia interna de ,81. Para ningún factor, se encontraron mejoras si se eliminaba alguno de los ítems.

Tabla 1. Escala final de 15 ítems

Subescala	Ítem	Texto del ítem
AFE	5	El entrenamiento me deja tan cansado/a que me cuesta reunir la energía suficiente para hacer otras cosas.
	7	La práctica del fútbol me deja demasiado cansado/a.
	14	Me siento físicamente agotado/a por el fútbol.
	16	Me agotan las exigencias físicas y mentales del fútbol.
	19*	La práctica del fútbol me deja mentalmente agotado/a.
RSL	3	En el fútbol estoy logrando muchas cosas que valen la pena.
	8	Creo que no estoy logrando mucho en el fútbol.
	10	Pienso que no estoy rindiendo a mi nivel real en el fútbol.
	17	Parece que, haga lo que haga, no rindo como debería.
	20	Creo que tengo éxito en el fútbol.
DPD	9	Mi rendimiento en el fútbol me importa menos que antes.
	12	Creo que no me interesa tanto el fútbol como antes.
	13*	Estoy dejando de disfrutar del fútbol.
	15	Me preocupo menos que antes por triunfar en el fútbol.
	18*	Tengo dudas de si el fútbol merece todo el tiempo que le dedico.
AFE = agotamiento físico/emocional		
RSL = reducida sensación de logro		
DPD = devaluación de la práctica deportiva		
* Enunciados alternativos		

### *Análisis factorial confirmatorio*

En primer lugar, se realizó un AFC con el modelo de tres factores, con 5 ítems cada uno de ellos. El método de estimación empleado fue el de máxima verosimilitud. El modelo, que permitía las correlaciones entre factores, estaba sobreidentificado con 31 parámetros a estimar. Las cargas factoriales estimadas fueron todas ellas significativas, oscilando entre ,25 (ítem 3) y ,78 (ítem 9) con valores del término de error de entre ,94 y ,39. Los coeficientes de correlación estimados entre los tres factores fueron de ,48 para la relación entre agotamiento físico/emocional y devaluación del deporte, ,50 para la relación entre agotamiento físico/emocional y reducida sensación de logro y ,72 para la relación entre reducida sensación de logro y devaluación del deporte. Todos ellos fueron significativos (valores t de 8,81, 8,29 y 15,08, respectivamente).

Para evaluar el ajuste global del modelo se eligieron seis índices: la razón del valor del estadístico  $\chi^2$  entre sus grados de libertad, ya que éste por sí solo es especialmente sensible al tamaño muestral (Jöreskog y Sörbom, 1993), y otros índices que no dependen tanto del tamaño muestral como el índice de bondad de ajuste (GFI), el índice de ajuste comparativo (CFI), el índice de ajuste no normado (NNFI), el error cuadrático medio de aproximación (RMSEA) y la raíz media cuadrática residual (RMR). Se obtuvo un cociente entre  $\chi^2$  (234,00) y los grados de libertad (87) de 2,68 ( $p < ,001$ ). Los índices GFI, CFI y NNFI alcanzaron valores de ,91, ,95 y ,94, respectivamente, y los valores de RMSEA y RMR fueron de ,076 y ,066. Los valores observados evidenciaron un ajuste aceptable del modelo (Bentler, 1990; Bollen y Long, 1993; Browne y Cudeck, 1993; Jöreskog y Sörbom, 1993; Tucker y Lewis, 1973).

Se tuvieron en cuenta las medidas de ajuste individual y los índices de modificación proporcionados por el AFC para mejorar el ajuste del modelo. De este modo, se introdujeron secuencialmente dos cambios. Primero se permitió la correlación entre los términos de error de los ítems 5 y 7 y a continuación la de los ítems 10 y 17.

La Tabla 2 recoge los parámetros estimados después de efectuar dichas modificaciones. Las saturaciones factoriales mostraron valores entre ,36 (ítem 3) y ,78 (ítem 12 e ítem 14), siendo todas ellas significativas. Los coeficientes de correlación fueron de ,50 para la relación entre agotamiento físico/emocional y reducida sensación de logro, así como también para la relación entre agotamiento físico/emocional y devaluación del deporte. El valor de la correlación entre reducida sensación de logro y devaluación del deporte fue más elevado (,81). Los tres valores fueron significativos.

Con respecto a los índices de ajuste global, la razón del estadístico  $\chi^2$  (178,63) entre sus grados de libertad (87) fue de 2,10. El valor de GFI se igualó al anterior (,93). Se incrementaron los valores de CFI (,97) y NNFI (,96) y los valores de RMSEA (,059) y RMR (,061) disminuyeron. Por lo tanto, el ajuste global resultó más satisfactorio con este modelo reespecificado. Finalmente no se observaron otras modificaciones que fuesen sustentables teóricamente.

Tabla 2. Estimaciones del análisis factorial confirmatorio

Factor	Ítem	Carga factorial	Valores t	Términos de error	R <sup>2</sup>
AFE	14	,78	14,95	,40	,60
	19	,71	13,32	,50	,50
	5	,65	11,80	,58	,42
	7	,65	11,73	,58	,42
	16	,65	11,91	,58	,42
RSL	17	,57	9,09	,67	,33
	8	,51	8,11	,74	,26
	10	,48	7,39	,77	,23
	20	,40	6,40	,84	,16
	3	,36	5,65	,87	,13
DPD	12	,78	15,67	,39	,61
	13	,76	15,16	,42	,58
	9	,69	13,16	,53	,47
	18	,62	11,60	,61	,39
	15	,59	10,85	,65	,35

## Discusión y conclusión

El propósito de este estudio fue perfeccionar la versión española del ABQ propuesta por De Francisco (2007). En su investigación, se sugerían modificaciones en relación a tres de los quince ítems que componían el cuestionario. Por ello, en el presente trabajo se incluyeron dos enunciados alternativos para cada uno de ellos, empleando así una versión de 21 ítems.

Tras varios AFE, se seleccionaron los mejores enunciados en función de criterios estadísticos y sustantivos hasta reducir el cuestionario a 15 ítems (5 para cada dimensión), respetando, de este modo, la estructura del instrumento original. A continuación, tras emplear procedimientos de AFC, se realizaron dos modificaciones en el modelo, que permitieron mejorar su ajuste: se permitieron las correlaciones entre los términos de error del par 5 y 7 de agotamiento físico/emocional y del par 10 y

17 de reducida sensación de logro. Si acudimos a la Tabla 1, podemos apreciar, por un lado, que los enunciados 5 y 7 se refieren a cansancio físico provocado por la práctica del fútbol; y por otro, que los ítems 10 y 17 sugieren una falta de rendimiento. Estas modificaciones tienen una interpretación sustantiva bastante clara que, además, se ve apoyada por los comentarios sobre la redundancia de estos enunciados aportados por los participantes al cumplimentar el cuestionario.

Las correlaciones entre factores (entre .50 y .81) fueron similares a las referidas por Raedeke y Smith (2001) en el cuestionario original (.52-.71). El valor más elevado fue el hallado entre las dimensiones reducida sensación de logro y devaluación del deporte. Otras investigaciones apoyan este resultado (Cresswell, 2008; Hill, Hall, Appleton y Kozub, 2008; Lonsdale et al., 2007; Perreault, Gaudreau, Lapointe y Lacroix, 2007; Raedeke y Smith, 2004).

Estos resultados nos permiten llegar a una versión española del ABQ más depurada. Además, desde un punto de vista práctico, hemos de añadir una valoración positiva en relación a la facilidad de aplicación del cuestionario. Fue cumplimentado con extraordinaria rapidez por los deportistas y sin manifestar dificultades para comprender sus enunciados.

Como objetivo de una futura investigación debería contemplarse la generalización de esta versión a otros deportes distintos del fútbol, así como seguir investigando en la fiabilidad del cuestionario.

Agradecimientos. La presente investigación ha sido realizada con el apoyo económico del Ministerio de Educación y Ciencia y del Fondo Europeo de Desarrollo Regional-FEDER (SEJ2007-67278/PSIC).

### Referencias

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.
- Bollen, K. A. & Long, J. S. (Eds.). (1993). *Testing structural equations models*. Newbury Park, CA: Sage.
- Browne, M. W. y Cudeck, R. (1993). Alternative ways of assessing model fit. En Bollen, K. A. & Long, J. S. (Eds.). *Testing structural equations models*, (pp.136-162). Newbury Park, CA: Sage.
- Cresswell, S. L. (2008). Possible early signs of athlete burnout: A prospective study. *Journal of Science and Medicine in Sport* (2008), doi: 10.1016/j.jsams.2008.01.009.
- Cresswell, S. L. y Eklund, R. C. (2006). The convergent and discriminant validity of burnout measures in sport: A multi-trait/multi-method analysis. *Journal of Sports Sciences*, 24(2), 209-220.
- De Francisco, C. (2007). *Evaluación de un modelo para la medida de burnout en deportistas*. Tesis de licenciatura no publicada, Universidad de Santiago de Compostela, Santiago de Compostela.
- Hill, A. P., Hall, H. K., Appleton, P. R. y Kozub, S. A. (2008). Perfectionism and burnout in junior elite soccer players. The mediating influence of unconditional self-acceptance. *Psychology of Sport and Exercise*, 9(5), 620-629.
- Jöreskog, K. G. & Sörbom, D. (1993). LISREL 8. *Structural Equation Modeling with the SIMPLIS Command Language*. Hillsdale, NJ: LEA.

- Lonsdale, C., Hodge, K. J. & Jackson, S. A. (2007). Athlete engagement: II. Development and initial validation of the Athlete Engagement Questionnaire. *International Journal of Sport Psychology*, 38(4), 471-492.
- Perreault, S., Gaudreau, P., Lapointe, M. -C. y Lacroix, C. (2007). Does it take three to tango? Psychological need satisfaction and athlete burnout. *International Journal of Sport Psychology*, 38(4), 437-450.
- Raedeke, T. D. y Smith, A. L. (2001). Development and preliminary validation of an athlete burnout measure. *Journal of Sport and Exercise Psychology*, 23(4), 281-306.
- Raedeke, T. D. y Smith, A. L. (2004). Coping resources and athlete burnout: an examination of stress mediated and moderation hypotheses. *Journal of Sport and Exercise Psychology*, 26(4), 525-541.
- Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factors analysis. *Psychometrika*, 38, 1-10.

# ANÁLISIS DE LA EQUIVALENCIA DEL CONSTRUCTO “ASERTIVIDAD” EN MUESTRAS MEXICANAS Y ESPAÑOLAS

Isabel Benítez<sup>1</sup>, Miguel Castillo<sup>1</sup>, José Luis Padilla<sup>1</sup>, Juana Gómez-Benito<sup>2</sup>, Mirta Margarita Flores-Galaz<sup>3</sup> y Rolando Díaz-Loving<sup>3</sup>

<sup>1</sup>Universidad de Granada

<sup>2</sup>Universidad de Barcelona

<sup>3</sup>Universidad Autónoma de Yucatán

Existe un amplio consenso respecto a la necesidad de examinar la equivalencia del constructo objeto de la medición entre las poblaciones implicadas en las comparaciones transculturales. El examen de la equivalencia del constructo se debe realizar también cuando las poblaciones responden a la misma versión del cuestionario pero pertenecen a contextos culturales diferentes. Investigar la equivalencia en estas situaciones supone obtener evidencias sobre el “solapamiento” en las definiciones operacionales y la “red nomológica” de los constructos evaluados. El objetivo del estudio es examinar la equivalencia del constructo en contextos “monolingües” multi-culturales mediante el escalamiento multidimensional. A modo de ejemplo, se presentan los resultados obtenidos en el análisis de la equivalencia del constructo “asertividad” entre muestras mexicanas y españolas. 599 estudiantes de la Universidad Autónoma de Yucatán (México) y 646 estudiantes de la Universidad de Granada (España), respondieron a la Escala Multidimensional de Asertividad (Flores y Díaz-Loving, 2004). El escalamiento multidimensional proporcionó evidencias sobre el nivel de solapamiento del constructo y las comparaciones que pueden hacerse entre ambas poblaciones.

Uno de los objetivos de la psicología transcultural es la explicación de los fenómenos psicológicos por medio de procesos y constructos universales, *versus* las características ideosincráticas particulares y específicas de los diferentes grupos culturales (Van de Vijver y Poortinga, 1982). Para ello, desde la “Etnopsicología” se han conceptualizado y operacionalizado numerosos constructos de personalidad procedentes de diferentes culturas.

Dadas las relaciones históricas y culturales entre México y España, resulta interesante el análisis de las diferencias culturales en relación a la equivalencia conceptual del constructo “asertividad” entre ambas poblaciones. Flores y Díaz-Loving (2002) consideran la capacidad de emitir una conducta asertiva como un tipo de habilidad social, que depende del contexto cultural en el que se desarrolla el individuo. Incluso dentro de una misma cultura, se destaca la influencia de la educación, el estatus social, el sexo y la edad, de manera que existen muchas variables personales y situacionales que hacen que una conducta apropiada en una situación, no lo sea en otra. Estos aspectos implican una gran ambigüedad a la hora de definir el constructo y determinar qué es una conducta asertiva, por ello Flores y Díaz-Loving (2002, 2004) proponen un marco teórico-conceptual que permite el estudio de la asertividad en un contexto cultural determinado. Dentro de este marco, la expresión de una respuesta asertiva está influida por la cultura en la que se desarrolla el individuo, por el contexto situacional y psicológico, por las personas involucradas y/o percibidas en la interacción, por las variables sociodemográficas tales como el sexo, edad, escolaridad, nivel socioeconómico y por variables de personalidad.

La influencia de la cultura sobre el constructo asertividad se observa tanto en la emisión de determinadas conductas por parte de un actor como en la consideración de dichas conductas como asertivas o no asertivas por parte de un espectador. En este sentido, los autores Flores y Díaz-Loving (2004) desarrollaron la Escala Multidimensional de Asertividad (EMA) para evaluar la asertividad a partir de la consideración de estos aspectos culturales y personales. El interés principal de este estudio

subyace en la medición del constructo a través de la EMA en muestras de diferentes culturas (México y España), comparando la equivalencia conceptual de éste en ambas muestras. En este tipo de estudios se busca determinar si la definición de un constructo elaborada pensando en una cultura determinada, en este caso México, es igualmente adecuada en otra cultura distinta, en este caso España. En esta situación, compartir un mismo idioma permite atribuir las posibles diferencias en los indicadores o en la relación con otros constructos, a las diferencias culturales en la expresión del constructo asertividad.

La equivalencia conceptual o psicológica no es inherente a la equivalencia lingüística existente entre dos versiones de un mismo cuestionario. En este caso una sola versión del cuestionario administrada a dos muestras diferentes, puede implicar diferentes significados para ellas. Es importante por ello, tener presente la definición considerada por Van de Vijver y Poortinga (2005) en la que se define el sesgo como cualquier factor de ruido que pone en duda la interpretación de diferencias en los resultados entre los grupos.

La posible falta de equivalencia puede manifestarse a diferentes niveles. Este estudio se centra en la equivalencia estructural con el objetivo de asegurar la comparabilidad de las puntuaciones a nivel grupal. La equivalencia estructural permitirá realizar comparaciones entre las medias de los grupos descartando la posibilidad de que dichas diferencias se deban al sesgo (Van de Vijver y Poortinga 2005).

Para evaluar la equivalencia del constructo se utilizó el Escalamiento Multidimensional, uno de los procedimientos recomendados por Sireci, Patsula, y Hambleton (2005). El Escalamiento Multidimensional (EM) es una aproximación atractiva cuando se trata de cuestionarios con una dimensionalidad compleja o durante fases iniciales de investigaciones transculturales con diferentes grupos lingüísticos o culturales. Mediante el EM las diferencias entre los grupos son analizadas mediante los pesos grupales los cuáles indican la importancia para cada grupo o la capacidad de cada dimensión para explicar las diferencias encontradas en las respuestas dadas por cada grupo de ítems agrupados en una dimensión. Es decir, que cuanto más similares sean los pesos de los grupos en cada una de las dimensiones ajustadas, más confianza se puede tener al proponer la equivalencia estructural.

El objetivo de este estudio es realizar una comparación intercultural que permita determinar la equivalencia del constructo “asertividad” entre muestras mexicanas y españolas. Este objetivo pretende dar seguridad a los estudios en los que se comparen muestras mexicanas y muestras españolas. La finalidad última de este análisis de la equivalencia es eliminar el sesgo como explicación alternativa en aquellos casos en que aparezcan diferencias entre las mediciones obtenidas con diferentes muestras.

## Método

### *Participantes*

1245 participantes respondieron a la EMA, 599 estudiantes de la Universidad Autónoma de Yucatán (México) y 646 estudiantes de la Universidad de Granada (España). En la muestra española, la edad promedio fue de 21,29 años ( $DT = 5,23$ ), incluyendo 470 mujeres, y 171 hombres. En la muestra mexicana, hubo 317 hombres y 273 mujeres, con una edad promedio de 19,58 años ( $DT = 3,23$ ).

### *Instrumento*

#### *Escala Multidimensional de Asertividad (EMA)*

Elaborada por Flores Galaz y Díaz-Loving (2004), la EMA es un instrumento que consta de 45 ítems con formato de respuesta tipo Likert con cinco opciones de respuesta. La EMA está constituida por tres

dimensiones de 15 ítems cada una de ellas: 1) *Asertividad indirecta*, entendiéndose como la inhabilidad del individuo de poder tener enfrentamientos directos cara a cara con otras personas en situaciones cotidianas o de trabajo llevándolo por lo tanto a expresar sus opiniones, deseos, sentimientos, peticiones, limitaciones personales, realizar peticiones, decir no, dar y recibir alabanzas y manejar la crítica por medio de cartas, teléfono 2) *No asertividad*, es decir, la inhabilidad del individuo de poder expresar sus deseos, opiniones, sentimientos, limitaciones, dar y recibir alabanzas, iniciar la interacción con otras personas y manejar la crítica y 3) *Asertividad*, definida como la habilidad del individuo de poder expresar sus limitaciones, sentimientos, opiniones, deseos, derechos, dar y recibir alabanzas, hacer peticiones y manejar la crítica (Flores y Díaz-Loving, 2004). La EMA proporciona una medida de global de asertividad y una puntuación para cada una de las subescalas.

### *Procedimiento*

Los participantes del estudio respondieron al EMA en sesiones colectivas dentro del horario de clase tras solicitar el permiso a los responsables académicos. La participación fue voluntaria. Las instrucciones y las condiciones de aplicación garantizaban la confidencialidad de las respuestas. El tiempo para responder a la escala osciló entre 10-15 minutos.

### *Análisis de datos*

Los análisis se realizaron mediante el SPSS v.15. En primer lugar, se analizaron las propiedades psicométricas de la EMA, incluyendo el análisis de las respuestas a los ítems del cuestionario y de las propiedades de la escala. A continuación, se analizó la fiabilidad de las subescalas y, por último, se aplicó el EM.

## Resultados

### *Propiedades métricas de los ítems y de la escala*

Durante el análisis de las características psicométricas de la escala en su conjunto, se observó que la consistencia interna de la escala estimada con el coeficiente alfa de Cronbach fue de 0,718 en la muestra española y de 0,816 en la muestra mexicana. Estos valores aumentaron cuando se realizaron análisis separados para cada una de las subescalas, encontrándose lo siguiente: los ítems del componente de asertividad mostraron un coeficiente alfa de Cronbach de 0,867 para la muestra mexicana y de 0,771 para la muestra española; los ítems del componente de no asertividad ofrecieron un coeficiente alfa de Cronbach de 0,871 para la muestra mexicana y de 0,853 para la muestra española; y por último, los ítems del componente de asertividad indirecta mostraron un coeficiente alfa de Cronbach de 0,916 para la muestra mexicana y de 0,902 para la muestra española. Por otro lado, los valores del índice de discriminación fueron adecuados para todos los ítems, así como el coeficiente alfa si se elimina el elemento. La existencia de correlaciones altas entre los ítems de cada subescala y dicha subescala, frente a la aparición de correlaciones bajas entre los ítems y la puntuación total corrobora la multidimensionalidad del constructo.

### Escalamiento Multidimensional: Selección del número de dimensiones.

En primer lugar se seleccionó el número de dimensiones a partir del índice de stress y los valores de  $R^2$ . Los investigadores suelen recurrir a consideraciones teóricas y a los valores del índice de Stress para determinar el número más adecuado de dimensiones. El índice de Stress puede entenderse como los “auto-valores” en el contexto del análisis factorial exploratorio. La tabla 1 muestra los índices de stress y los valores de  $R^2$  obtenidos.

Tabla 1. Índices de stress y valores de  $R^2$

D	Stress	$R^2$
4	0,08	0,98
3	0,10	0,96
2	0,09	0,97
1	0,35	0,67

Como muestra la Tabla 1, la “caída” en los valores del índice de Stress se produce entre las dimensiones 1 y 2. Sin embargo, la estructura teórica del cuestionario y los objetivos del estudio aconsejaron optar por 3 dimensiones como el número más idóneo. Habiendo determinado el número de dimensiones, el siguiente paso es darles una interpretación teórica revisando el mapa perceptual y la tabla de coordenadas aportadas por el EM. La figura 1 muestra la agrupación de los ítems considerando tres dimensiones. Cada uno de los ítems es identificado mediante su número y la dimensión a la que corresponden teóricamente, siendo: A= Asertividad; N= No asertividad; I= Asertividad Indirecta.

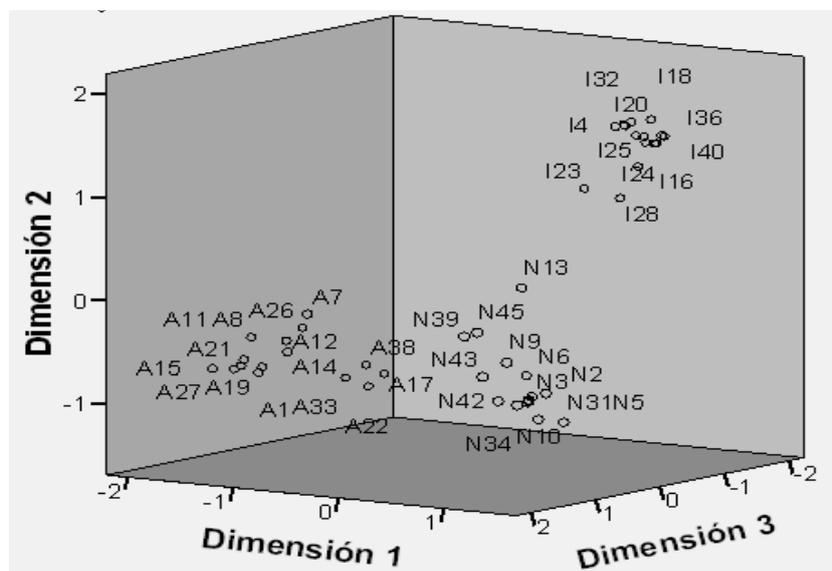


Figura 1. Localización de los ítems en tres dimensiones

Las tres dimensiones agrupan los ítems en una forma consistente con la teoría seguida en la elaboración de la EMA. Cada una de las dimensiones agrupa los ítems que pertenecen a cada una de las 3 subescalas de la EMA.

Habiendo dado “significado” a las dimensiones, el siguiente paso es evaluar si las dimensiones son igualmente relevantes para explicar las respuestas a los ítems dadas por los diferentes grupos. Se trata de determinar si los resultados del análisis apoyan las inferencias sobre la equivalencia de constructo de las mediciones obtenidas con la EMA a través de los grupos. La tabla 2 muestra los pesos de cada dimensión para los dos grupos.

Tabla 2. Pesos de las dimensiones

Grupos	D1	D2	D3
España	0,12	0,97	0,07
México	0,98	0,03	0,07

Analizando los pesos de cada dimensión para los dos grupos, un patrón de resultados, como muestra la Tabla 2, aparece con claridad. Las dimensiones 1 y 2 tienen una importancia “opuesta” para cada grupo. Mientras que la dimensión 3 no parece especialmente relevante para ninguno de los dos grupos.

### Discusión

El principal objetivo de este estudio era realizar una comparación intercultural para determinar la equivalencia del constructo asertividad entre la población mexicana y la española. Dicho objetivo, se ha cubierto con la realización de un análisis de la equivalencia entre las mediciones aportadas por el EMA en distintas poblaciones, lo que permitirá dar seguridad a los estudios en los que se comparen muestras mexicanas y muestras españolas. El análisis de la equivalencia conceptual realizado apunta a la existencia de posibles diferencias respecto de la definición operacional del constructo asertividad entre muestras mexicanas y españolas.

Este estudio concreto, es un estudio metodológico previo al estudio comparativo. En este tipo de investigaciones, se pretende asegurar un nivel de equivalencia que permita realizar, en futuros estudios, comparaciones entre las muestras implicadas; entendiéndose por comparaciones la interpretación de diferencias entre las mediciones aportadas por la escala. Dadas las diferencias encontradas respecto de la importancia diferencial de las dimensiones en cada muestra, no sería posible realizar comparaciones grupales entre las mediciones del constructo asertividad entre muestras mexicanas y españolas.

A partir del análisis estadístico de los ítems y de las evidencias sobre la estructura factorial del EMA, se identificaron las diferencias en el funcionamiento de los ítems en ambas muestras. Los análisis mostraron propiedades psicométricas más adecuadas para las subescalas *Asertividad*, *No asertividad* y *Asertividad indirecta* que para la escala general. Los análisis estadísticos de los ítems muestran, tal y como se esperaba, un mejor funcionamiento de los ítems a través de las subescalas que a través de la escala general. Se encontraron correlaciones altas entre los ítems de cada subescala y dicha subescala, mientras que las correlaciones entre los ítems y la puntuación total fueron bajas, lo que corrobora la multidimensionalidad del constructo.

Por otra parte, el EM aportó evidencias de la falta de equivalencia conceptual del constructo “asertividad” en las muestras mexicana y española. Esta ausencia de equivalencia puede interpretarse como la diferencia en la relevancia de los indicadores medidos por los ítems para medir la asertividad en ambas muestras. Como se ha observado, el EMD es un procedimiento útil para orientar la comparación transcultural.

Futuras investigaciones, podrían centrarse en un análisis de equivalencia métrica a partir de la identificación de ítems con un funcionamiento diferencial al comparar las dos muestras. La aparición de diferencias en el funcionamiento de los ítems a través de las muestras, podrían indicar la existencia de diferencias culturales, que se entenderían como una desigualdad en la relevancia de la conducta medida entre las dos muestras.

## Referencias

- Flores, G. M. y Díaz-Loving, R. (2002). *Asertividad: Una alternativa para el óptimo manejo de las relaciones interpersonales*. México: Porrúa-UADY.
- Flores, G. M. y Díaz-Loving, R. (2004). *Escala Multidimensional de Asertividad*. México: Manual Moderno.
- Sireci, S.G., Patsula, L. y Hambleton, R.K. (2005). Statistical methods for identifying flaws in the test adaptation process. En R.K. Hambleton, P.F. Merenda y S.D. Spielberger (eds.): *Adapting educational and psychological tests for cross-cultural assessment* (pp. 93-115). New Jersey: Lawrence Erlbaum Associates.
- Van de Vijver, F. J. R. y Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology*, 13, 387-408.
- Van de Vijver F. J. R. y Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. En R. K. Hambleton, P. F. Merenda, & C. D. Spielberger, C. D. *Adapting educational and psychological tests for cross-cultural assessment*. LEA, NJ.

# TEORÍA DE LA GENERALIZABILIDAD: INVESTIGACIONES APLICADAS

**Coordinadores: Ángel Blanco<sup>1</sup> y Verónica Morales<sup>2</sup>**

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Universidad de Málaga

Cronbach, Gleser, Nanda y Rajaratnam (1972), han desarrollado la “*teoría de la generalizabilidad*” (TG), asumiendo que hay otras fuentes de variación además de las diferencias individuales e integrando cada una de estas fuentes de variación en una estructura global, que permite aplicaciones particulares de la teoría estadística del muestreo. La teoría de la generalizabilidad reconoce explícitamente las múltiples fuentes de error de medida (participantes, contextos, tratamientos, observaciones, sesiones, ocasiones de medida...). Podemos estimar cada una de estas fuentes de error así como las diferentes interacciones entre ellas. El error de medida no es más que el efecto de las fluctuaciones debidas a la elección aleatoria de individuos, contextos, tratamientos, observaciones, sesiones, ocasiones de medida..., es decir al muestreo de niveles particulares en cada una de las facetas (variables) del universo de observaciones posibles. Optimizar dicha medida es adaptar nuestro diseño para reducir al máximo la variancia del muestreo debida a estas facetas. Aunque las aplicaciones de la TG en el ámbito de la metodología observacional tienen que ver habitualmente con el cálculo de la fiabilidad (intra-observador e inter-observadores) mediante el coeficiente de correlación intraclase (CCI) y de la validez en la construcción de sistemas de categorías para la observación directa del comportamiento, también puede utilizarse para la optimización de los tamaños de muestra ideales que nos permiten inferir de las mismas a las poblaciones pertinentes, constituyendo un estudio del coste-beneficio que nos permite mayor flexibilidad y parsimonia en las diferentes ocasiones de medida. Simultáneamente, la TG nos permite llevar a cabo un análisis de la variabilidad estimada a través de los componentes de variancia del grado de influencia en el diseño de cada una de las facetas, individualmente o en interacción. Estas tres direcciones de estudio de la TG van a estar presentes en las comunicaciones que se van a mostrar en este simposio, realizadas en diferentes ámbitos, ya sea en el ámbito de la actividad física y del deporte, ya sea en el ámbito clínico de detección del TDAH, ya sea en la construcción de sistemas para evaluar la calidad en diversos ámbitos de intervención.

# DISEÑO DE MEDIDA EN LA CONSTRUCCIÓN DE UNA HERRAMIENTA PARA EVALUAR LA CALIDAD DE LOS PROGRAMAS DE VOLUNTARIADO AMBIENTAL

Encarnación Chica<sup>1</sup>, Antonio Hernández Mendo<sup>2</sup> y Verónica Morales<sup>2</sup>

<sup>1</sup>Universidad de Cádiz

<sup>2</sup>Universidad de Málaga

El objetivo de esta investigación es elaborar una herramienta para evaluar la calidad de los programas de voluntariado ambiental, a través de la satisfacción de los voluntarios. Dada la significatividad social del voluntariado, es importante para las organizaciones establecer estrategias que ayuden a incrementar la satisfacción con la actividad, favorecer así la permanencia de los voluntarios/as y su compromiso con la entidad. Desde este planteamiento se ofrece una herramienta útil en la gestión de programas, que evalúe la calidad de los mismos, incorporando una metodología formativa y participativa. Una parte del análisis se ha realizado con el paquete estadístico SAS® con el objetivo de realizar un análisis de componentes de variancia utilizando el procedimiento GLM (*General Linear Model*) con el fin de obtener la suma de cuadrados necesaria para un análisis de generalizabilidad a través del software GT. Este análisis nos va a permitir asegurar que: (a) el número de ítems son suficientes para conseguir resultados fiables y generalizables, (b) la muestra utilizada permite considerar que el diseño es óptimo en términos de fiabilidad y generalización. Con ello calcular un adecuado *Plan de Optimización de la Calidad* en los programas de voluntariado ambiental mediante la satisfacción de los voluntarios.

En los últimos años ha aumentado la preocupación por el medio ambiente. Las encuestas de opinión de valores, así como las agendas políticas y las movilizaciones sociales, expresan dicha preocupación. El medio ambiente se convierte en meta y valor, consistente en *alcanzar una calidad de vida ambiental*.

Aparece el concepto de *desarrollo sostenible*. La definición más aceptada se recoge en el Informe Brundtland (1987) como *la única vía posible para continuar creciendo sin agotar los recursos del planeta*. También surgen disciplinas implicadas en la búsqueda de soluciones, como la Psicología ambiental y la Educación Ambiental. Al mismo tiempo, grupos y personas intervienen de forma activa y crítica, constituyéndose el *voluntariado ambiental*. Éste se constituye en organizaciones, que buscan dar respuesta desde diferentes sectores, y lo hacen mediante la participación ciudadana en programas y proyectos.

Constatamos la importancia social del voluntariado, pero también la dificultad para evaluar las acciones que se realizan, por lo que es importante incorporar procesos de mejora en la gestión de las entidades voluntarias y estrategias que favorezcan la satisfacción de los voluntarios con la actividad, para favorecer la captación y el mantenimiento de voluntarios.

En este tema, Dávila y Chacón (1991) proponen el Modelo Básico. Con él explican mediante una estructura de relaciones que el compromiso con la organización predice la permanencia en diferentes tipos de voluntarios a través de la intención de continuar, e incluye la satisfacción.

Dada la significatividad social del voluntariado, si queremos trabajar la permanencia en una organización, este *modelo básico* nos desvela la importancia de establecer estrategias que ayuden a incrementar la satisfacción con la actividad, que en última instancia, predice la permanencia.

Actualmente, medir la satisfacción del consumidor, cliente o voluntario en nuestro caso, es un objetivo importante para empresas y organizaciones. La satisfacción es una medida adicional, relativa a la calidad del producto o servicio (percibida a través de la evaluación de la experiencia). Al mismo tiempo, la satisfacción está relacionada con la fidelidad, por lo que influirá en la permanencia o compromiso según confirman los modelos teóricos del voluntariado.

Desde este planteamiento queremos ofrecer una herramienta o cuestionario *Inventario de Calidad en Programas de Voluntariado Ambiental* (ICPVA) útil en la gestión de programas, que evalúa la calidad de los mismos, incorporando una metodología participativa de los voluntarios, desde un compromiso e implicación con la organización.

En una investigación más amplia, nos planteábamos como objetivo general *elaborar una herramienta para evaluar la calidad de los programas de voluntariado ambiental, a partir de la satisfacción de las personas implicadas* y como objetivo específico que centraría el trabajo que presentamos, *sería estimar los diseños de medida óptimos para una adecuada evaluación de la calidad de dichos programas*.

## Método

### *Participantes*

El primer cuestionario se administró a 102 participantes, jóvenes universitarios con edades comprendidas entre los 18 y 23 años, que colaboraban de forma habitual y/o puntual en programas de voluntariado. El segundo cuestionario se administró a 190 participantes, con un rango de edad más amplio, entre los 15 y 53 años.

### *Diseño*

Hemos seleccionado un diseño multifaceta, dado que se han utilizado distintas variables de clasificación; parcialmente anidado, ya que hemos contemplado dentro del cuestionario diferentes escalas; y no ortogonal teniendo en cuenta que hemos realizado este estudio en un momento y evaluado a diferentes personas.

### *Material*

Se elaboró un primer cuestionario *Inventario de Calidad en Programas de Voluntariado Ambiental* (ICPVA-v.1.0) y tras los resultados y análisis del primer estudio empírico, vimos necesario construir nuevos ítems en algunas de las escalas y aumentar la muestra. Se planteó entonces llevar a cabo un segundo estudio con el nuevo cuestionario (ICPVA-v-2-0), elaborado a partir del primero, y administrarlo a un número mayor de voluntarios.

El primer cuestionario consta de un total de 46 ítems distribuidos en 5 escalas e incorpora la recogida de datos relativos a la edad y género. El segundo cuestionario, que se elaboró a partir del anterior, consta de 71 ítems, distribuidos también en 5 escalas y se le añadió la recogida de datos sobre profesión y nivel de estudios.

## Programas

- Paquete estadístico SPSS v.15.0
- Programa estadístico de Análisis de Teoría de la Respuesta al Ítem: Acer ConQuest (Wu, Adams & Wilson, 1998)
- Programa LISREL 8.30 y PRELIS 2.30 (Kart Jöreskog and Dag Sörbom, 1999)
- Paquete estadístico SAS (Schlotzhauer & Littell, 1997) para el análisis de componentes de la varianza
- Programa estadístico GT (Ysewijn, 1996) para el análisis de generalizabilidad.

## Procedimiento

Respecto al procedimiento, además de la elaboración de los cuestionarios, se realizó un *análisis factorial exploratorio*, un *análisis de fiabilidad* y *análisis de respuesta al ítem* en las dos versiones del cuestionario.

Tras la obtención de los resultados y el análisis de los datos, confirmamos mejores resultados en el primer estudio realizado con la primera versión del cuestionario (ICPVA v.1.0) y a éste se le realizó un *análisis factorial confirmatorio*, utilizando como procedimiento de cálculo para el ajuste del modelo el de máxima verosimilitud.

Según los resultados hemos obtenido una herramienta de fiabilidad satisfactoria y con una estructura factorial parsimoniosa.

Posteriormente llevamos a cabo un *análisis de generalizabilidad* (TG). Pero antes de aplicar la TG, se realiza un Análisis de Variabilidad con los datos obtenidos. El análisis de varianza incluye:

1. Un análisis de componentes de la varianza, mediante el Modelo Lineal General (GLM), con el paquete estadístico SAS
2. El análisis de generalizabilidad, desarrollada a través del programa estadístico GT.

Según Blanco (1989) y Anguera y Blanco (1993), las etapas esenciales de este análisis son tres: análisis de la varianza, análisis de generalizabilidad y plan de optimización.

En primer lugar realizamos un análisis de la varianza con el fin de obtener la suma de cuadrados necesaria para un Análisis de Generalizabilidad. Aplicando las técnicas del análisis de la varianza se puede cuantificar la importancia de cada fuente de variación (variable de clasificación), y en nuestro estudio hemos utilizado las *variables de clasificación*: Usuarios (U), Años (A), Género (G), Escalas (E) e Ítems (I). Las puntuaciones sobre cada ítems se consideraron la variable de medida.

## Resultados

El análisis del modelo general nos permite analizar las diferentes estructuras de diseño independientes, donde no importa el peso que tiene cada variable sobre el modelo general sino el tipo de interacciones particulares que se establecen así como su significación estadística independientemente del resto de variables y la posible precisión que se establecen a la hora de generalizar.

A partir de los resultados obtenidos en el análisis de variabilidad, comprobamos el ajuste de distintos modelos, con el objetivo de estimar la varianza explicada y su significación (Morales, Blanco y Hernández Mendo, 2004).

Entre los diseños de medida hemos estimado:

Un modelo de cinco facetas (UAGEI), en el que U: corresponde a usuarios (102), A: años (3 intervalos entre 18 y 23 años), G: género (2), E: escalas (5 escalas), I: ítems (6 a 13 ítems en las distintas escalas).

Para un adecuado *Plan de Optimización*, es necesario realizar distintas aproximaciones en los diferentes diseños de medida, estimando cada una de las facetas, anteriormente mencionadas y cada una de sus posibles interacciones.

Los diseños de medida propuestos y analizados han sido

- Cinco diseños en los que se estima una faceta a partir de las otras cuatro (UAGE/I, UAGI/E, UA EI/G, AGEI/U, UGEI/A)
- Diez diseños en los que se estima la interacción de dos facetas a partir de las otras tres (UAG/EI, UAI/GE, UGI/AE, UGI/UE, AGE/UI, UGE/AI, UAE/GI, AIE/UG, UIE/AG, GIE/UA)
- Diez diseños con la interacción de tres facetas a partir de dos (UA/GEI, UG/AIE, UI/AGE, UE/AGI, AG/UEI, AI/UGE, AE/UGI, GI/UAE, GE/UAI, IE/UAG)
- Cinco diseños en los que interaccionan cuatro facetas a partir de una (U/AGEI, A/UGEI, I/UAGE, E/UAGI, G/UA EI)

Del Análisis de la varianza del modelo de 5 facetas (UAGEI) con el paquete estadístico SAS, los resultados nos devuelven que este modelo es fiable ( $<.0001$ ), pero sólo explica el 27.83% de la varianza. También las facetas A (Años) y G (Género) no tienen peso, por lo que no aportan varianza al modelo.

Esto nos lleva a tomar la decisión de desestimar dichas facetas (A y G), y analizar un nuevo modelo de tres facetas: Usuarios x Escalas x Ítems (UEI). En el análisis de la varianza, los resultados nos devuelven que explica un 100% de la misma.

Los diseños de medida fueron

- Tres diseños en los que se estima la interacción de dos facetas a partir de otra (UE/I, UI/E, IE/U)
- Y otros tres en los que se estima la interacción de una faceta a partir de las otras dos (U/EI, E/UI, I/EU)

Con estos datos pasamos a realizar el análisis de generalizabilidad con el programa GT. La teoría de la Generalizabilidad se asienta en el postulado de que en cualquier situación de medida existen múltiples fuentes de variación (facetas). Se trata de analizar esas fuentes de variación que pueden afectar a un dato observacional. La aplicación de esta teoría permite estimar el grado de generalización de un diseño de medida con unas condiciones particulares a un valor teórico buscado.

En el análisis de generalizabilidad obtenemos índices de fiabilidad y de generalizabilidad para la optimización de los diseños de medida, con un análisis independiente de cada modelo. Esta optimización se resuelve en términos costo-beneficio que indudablemente inciden en la utilización y reasignación de recursos en la investigación.

En la tabla 1 presentamos los resultados de las aproximaciones realizadas a los distintos diseños de medida del modelo de tres facetas (Usuarios x Escalas x Ítems)

*Tabla 1. Resultados de los diseños de medida en el modelo de tres facetas (Usuarios x Escalas x Ítems)*

DISEÑOS DE MEDIDA (Modelo: Usuarios x Escalas x Ítems)						
	UE/I	UI/E	EI/U	U/EI	E/UI	I/EU
Índice de fiabilidad $e^2$	0.885	0.582	0.963	0.790	0,880	0.000
Índice de generalizabilidad $\Phi$	0.885	0.559	0.955	0.731	0.863	0,000
Nº observaciones	6630	6630	6630	6630	6630	6630

Vemos que en el diseño UE/I, en el que estimamos la faceta “ítems” (13) a partir de “usuarios” (102) y “escalas” (5), los valores de los índices de fiabilidad y generalizabilidad son buenos (0,885). Si realizamos un plan de optimización aumentando el número de ítems, aumenta el número de observaciones, por lo que consideramos adecuados los ítems que componen las escalas de la herramienta (entre 6 y 13).

Así mismo, en el diseño en el que estimamos la faceta “escalas” sobre “usuarios” e “ítems” (UI/E), el índice de fiabilidad es de 0.582 y el de generalizabilidad de 0,559. Estos resultados son coherentes con los resultados obtenidos en el análisis de fiabilidad y factorial confirmatorio realizados en la investigación más amplia. Consideramos por tanto, óptimo el cuestionario con 5 escalas. Una fiabilidad baja, muestra que las escalas miden constructos distintos.

También estimamos la faceta “usuarios” a partir de las otras dos: “ítems” y “escalas” (EI/U). El total de observaciones obtenidas es de 6630, el índice de fiabilidad es 0.963 y el de generalizabilidad 0,995.

En este modelo (UEI), continuamos con una adecuada optimización del análisis de medida en términos de costo-beneficio con el objetivo de una mejora continua en la gestión de entidades de voluntariado. Tras realizar distintas aproximaciones con diferentes niveles en la faceta Usuarios, podemos observar que si reducimos el número de ellos a 50, disminuimos el número de observaciones a 3250 y obtenemos valores altos de fiabilidad y generalizabilidad.

Así mismo, la interacción de Usuarios e Ítems a partir de la faceta Escala, presenta un índice de fiabilidad de 0.880 el de generalizabilidad de 0.863, manteniendo unos valores óptimos al bajar el número de usuarios a 50 y manteniendo el número de ítems

Tabla 2. Resultados de la optimización (modelo usuarios x escalas x ítems)

Modelo UEI		Nivel Actual	Aproximación óptima
EI/U	Nº de usuarios	102	50
	Índ. Fiab.	0.963	0.927
	Índ. General	0.955	0.912
	Nºobservaciones	6630	3250
E/UI	Nº de usuarios - Ítems	102 / 13	50/ 13
	Índ. Fiab.	0.880	0.859
	Índ. General.	0.863	0.827
	Nºobservaciones	6630	3250

### Conclusión

Consideramos pues que el número de Escalas y de Ítems es adecuado en términos de fiabilidad y generalizabilidad. Una estrategia de evaluación que recoja nuestro plan de optimación de los modelos que han sido sometidos a consideración (la situación ideal) sería una evaluación realizada con 50 usuarios, confirmando que el número de ítems y las escalas que configuran la herramienta es óptimo (entre 6 y 13 ítems y 5 escalas).

### Referencias

- Blanco Villaseñor, A. (1991). La Teoría de la Generalizabilidad aplicada a diseños observacionales. *Revista Mexicana de Análisis de la Conducta/Mexican Journal of Behavior Analysis*, 14(3), 23-64.
- Blanco Villaseñor, A., Losada, J.L. y Anguera, M.T. (1991). Estimación de la precisión en diseños de evaluación ambiental. *Evaluación Psicológica / Psychological Assessment*, 7 (2), 223-257.
- Blanco Villaseñor, A., Losada López, J.L. (2004). Generalización en Diseños Observacionales: alternativas de estimación y modelización. *Metodología de las Ciencias del Comportamiento*. Suplemento, 79-84.
- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- Hernández Mendo, A. (2001). Un cuestionario para evaluar la calidad en programas de actividad física. *Revista de Psicología del Deporte*, 10, 179-196.
- Morales Sánchez, V., Blanco Villaseñor, A. y Hernández Mendo, A. (2004). Optimización de modelos de medida en la evaluación de programas de actividad física. *Metodología de las Ciencias del Comportamiento. Vol. Especial*. 437-443.
- Ysewijn, P. (1996). *GT software for generalizability studies*. Mimeografía

# ESTIMACIÓN DE DISEÑOS PARA EVALUAR LA CALIDAD EN LOS CENTROS DE ATENCIÓN INFANTIL TEMPRANA (CAIT)

Rita P. Romero<sup>1</sup> y Verónica Morales<sup>2</sup>

<sup>1</sup>Centro de Atención Infantil Temprana (CAIT) del Ayuntamiento de Antequera (Málaga)

<sup>2</sup>Universidad de Málaga

Transcurridas tres décadas desde que se iniciaron en nuestro país los primeros tratamientos de *atención temprana* consideramos primordial determinar no solo la calidad de estos servicios sino también su eficiencia. La *atención temprana* no es ajena a la creciente preocupación por la calidad que impregna la sociedad actual. Nuestra labor profesional ha de estar orientada hacia la *calidad total* ya que, tener esta meta, nos permitirá incrementar nuestra competitividad y la satisfacción de nuestros clientes. Estos argumentos nos llevan a considerar que el marco metodológico más adecuado para conseguir parte de estos objetivos es la Teoría de la Generalizabilidad (TG). La TG es una teoría de los errores multifaceta de una medición conductual. El objetivo del análisis llevado a cabo con esta teoría es desglosar, en cualquier tipo de medición, la variabilidad real de la variabilidad del error. En nuestra investigación, el cumplimiento de este objetivo pasa por estimar los componentes de varianza de diversas facetas, tales como centros, usuarios, género y escalas. La cuestión central de la TG se encuentra en los componentes de variancia, dado que su magnitud nos aporta información sobre las fuentes de error que están afectando una medición conductual. Asimismo, utilizamos la TG para optimizar los tamaños de muestra ideales que nos permiten inferir de las mismas a las poblaciones pertinentes, constituyendo un estudio del coste-beneficio que nos permite mayor flexibilidad y parsimonia en las diferentes ocasiones de medida.

Se entiende por *atención temprana* el conjunto de intervenciones, dirigidas a la población infantil de cero a seis años, a la familia y al entorno, que tiene por objetivo dar respuesta lo más pronto posible a las necesidades transitorias o permanentes que presentan los niños con trastornos en su desarrollo o que tienen el riesgo de padecerlos. Estas intervenciones, que deben considerar la globalidad del niño, han de ser planificadas por un equipo de profesionales de orientación *interdisciplinar o transdisciplinar* (Grupo de Atención Temprana, 2000).

En los últimos años, cualquier sector de actividad productiva ofrece la calidad como un valor intrínseco de sus productos o servicios. Desde esta perspectiva, la calidad de servicio es una de las principales áreas de estudio del comportamiento de los consumidores y usuarios, dado que el rendimiento de las organizaciones de servicios es valorado por las propias personas que adquieren y/o utilizan estos bienes de consumo y/o servicios. Por ello los estudios de la calidad del servicio y de la satisfacción de los consumidores y usuarios tienen en común la importancia del punto de vista de los clientes al valorar el producto y el servicio que proporciona la propia organización. Para realizar la gestión adecuada de la calidad es necesario conocer qué buscan las personas en sus actividades de consumo (Quintanilla, 2002).

Cronbach, Glesser, Nanda y Rajaratnam (1972), han desarrollado la *teoría de la generalizabilidad* (TG), asumiendo que hay otras fuentes de variación además de las diferencias individuales e integrando cada una de estas fuentes de variación en una estructura global, que permite aplicaciones particulares de la teoría estadística del muestreo. La *teoría de la generalizabilidad* reconoce explícitamente las múltiples fuentes de error de medida (participantes, contextos, tratamientos, observaciones, sesiones, ocasiones de medida, etc.). Podemos estimar cada una de estas fuentes de error así como las diferentes interacciones entre ellas. El error de medida no es más que el efecto de las fluctuaciones debidas a la elección aleatoria de individuos, contextos, tratamientos, observaciones, sesiones, ocasiones de

medida, etc., es decir, al muestreo de niveles particulares en cada una de las facetas (variables) del universo de observaciones posibles. Optimizar dicha medida es adaptar nuestro diseño para reducir al máximo la variancia del muestreo debida a estas facetas (Morales, 2009).

Aunque las aplicaciones de la TG en el ámbito de la Metodología Observacional tienen que ver habitualmente con el cálculo de la fiabilidad (intraobservador e inter-observadores) mediante el coeficiente de correlación intraclass (CCI) y de la validez en la construcción de sistemas de categorías para la observación directa del comportamiento, también puede utilizarse para la optimización de los tamaños de muestra ideales que nos permiten inferir de las mismas a las poblaciones pertinentes, constituyendo un estudio del coste-beneficio que nos permite mayor flexibilidad y parsimonia en las diferentes ocasiones de medida. Simultáneamente, la TG nos permite llevar a cabo un análisis de la variabilidad estimada a través de los componentes de variancia del grado de influencia en el diseño de cada una de las facetas, individualmente o en interacción (Morales, 2003).

Concretamente en nuestro estudio, este análisis de la variabilidad estimada a través de los componentes de variancia del grado de influencia en el diseño de cada una de las facetas, individualmente o en interacción, nos ha permitido calcular un adecuado *Plan de optimización de la Calidad en Atención Temprana*. Realizando distintas aproximaciones de los diferentes diseños de medida, estimando cada una de las facetas y sus interacciones.

## Método

### *Participantes*

Los participantes en el estudio se dividen en 2 bloques: tres *encuestadores*, que ayudaron en la recogida de información, y 102 *usuarios*, en este caso los padres y madres de los niños-as que reciben tratamiento de atención temprana en tres CAIT de la ciudad de Málaga.

### *Material*

Para el estudio global se utilizaron:

- Inventario de Calidad en los Centros de Atención Infantil Temprana (ICCAIT) (Romero Galisteo y Morales, 2009). Instrumento que evalúa la calidad en los centros de atención infantil temprana.
- Paquete estadístico SAS (Schlotzhauer, S. D. & Littell, R. C., 1997) para realizar el análisis de componentes de la varianza, siguiendo el Modelo Lineal General (GLM).
- El programa estadístico GT (Ysewijn, 1996) para el análisis de generalizabilidad.

### *Procedimiento*

El estudio suponía la estimación de diferentes diseños para evaluar la calidad en los centros de atención infantil temprana. El cuestionario utilizado (ICCAIT) consta de seis escalas: relativa a las instalaciones (ítem del 1 al 17), relativa a las salas de tratamiento y material (ítem del 18 al 26), relativa a la atención al usuario (ítem del 27 al 36), relativa al personal especializado (ítem del 37 al 47), respecto a la información general (ítem del 48 al 55) y relativa a la información técnica (ítem del 56 al 63).

Antes de aplicar la TG, realizamos un Análisis de la Variabilidad con los datos obtenidos, donde hemos utilizado como variables de clasificación: usuarios (U), centros (C), Género (G) y escalas (E).

Mediante la TG se llevó a cabo la estimación de distintos modelos teniendo como objetivo la minimización de la varianza error. Con cada una de las facetas: usuarios, centros, género y escalas, elaboramos catorce diseños de medida teniendo en cuenta la interacción de sus distintas facetas. Cada uno de los diseños propuestos se analizó de forma independiente, concluyendo en cada diseño la estimación de cada una de las facetas individualmente o en interacción, de tal forma, que nos indique el índice de fiabilidad y generalizabilidad para que nos permita generalizar con precisión los resultados evaluados y, a su vez, poder establecer un adecuado *Plan de optimización*.

## Resultados

Un análisis pormenorizado del modelo UCGE permite analizar detalladamente estructuras de diseño independientes, donde no importa el peso que tiene cada variable sobre el modelo general sino el tipo de interacciones particulares que se establecen así como su significación estadística independientemente del resto de variables y la posible precisión que se establece a la hora de generalizar.

Se obtuvieron excelentes resultados en el estudio de tres diseños de medida en los que se estima una sola faceta: CGE/U ¿es necesario aumentar el número de usuarios para generalizar con mayor precisión?, UGE/C ¿han sido suficientes los centros evaluados para generalizar con mayor precisión? y UCE/G. Del mismo modo, obtuvimos muy buenos resultados en otros tres diseños en los que se estimaba la interacción de dos facetas: EC/UG, EU/CG y EG/UC ¿han sido suficientes los usuarios y el número de centros para generalizar con mayor precisión?, así como en un séptimo diseño de medida en el que se estimaba la interacción de tres facetas: E/UCG (Tabla 1). En los siete diseños de medida restantes: UCG/E, UC/EG, CG/UE, UG/EC, C/EGU, G/ECU y U/CGE, no se obtuvieron resultados relevantes.

*Tabla 1: Resultados del análisis de generalizabilidad*

	CGE/U	UGE/C	UCE/G	EC/UG	EG/UC	EU/CG	E/UCG
Índice de fiabilidad	0.995	1.000	1.000	0.995	0.995	1.000	0.995
Índice de generalizabilidad	0.993	1.000	1.000	0.993	0.993	1.000	0.993
Nº de observaciones	3336	3336	3336	3336	3336	3336	3336

Para una adecuada optimización del análisis de medida en términos de costo-beneficio, hemos realizado distintas aproximaciones con diferentes niveles en las facetas usuarios y centros. A modo de ejemplo, elegiremos las interacciones de las facetas usuarios-centro en la optimización del modelo UCGE. En el diseño de medida EG/UC, observamos que si reducimos los usuarios a 50 y los centros a 1, reducimos el número de observaciones de 3636 a 600 y disminuyendo tan solo el índice de fiabilidad a 0.991 y el generalizabilidad a 0.985. En la estimación de la faceta usuario en el diseño CGE/U, el resultado es que reduciendo el número de usuarios a 10, los índices se mantienen excelentes y el número de observaciones se reduciría a 360 (Tabla 2).

Tabla 2: Optimización del modelo UCGE

Modelo UCGE		Nivel actual	Aprox. óptima
EG/UC	Nº usuarios/centros	101/3	50/1
	Índice Fiabilidad	0.995	0.991
	Índice Generaliza.	0.993	0.985
	Nº observaciones	3636	600
CGE/U	Nº de usuarios	101	10
	Índice Fiabilidad	0.995	0.956
	Índice Generaliza.	0.993	0.930
	Nº observaciones	3636	360

### Conclusión

Aunque las aplicaciones de la TG en el ámbito de la Metodología Observacional tienen que ver habitualmente con el cálculo de la fiabilidad (intraobservador e inter-observadores) mediante el coeficiente de correlación intraclass (CCI) y de la validez en la construcción de sistemas de categorías para la observación directa del comportamiento, también puede utilizarse para la optimización de los tamaños de muestra ideales que nos permiten inferir de las mismas a las poblaciones pertinentes, constituyendo un estudio del coste-beneficio que nos permite mayor flexibilidad y parsimonia en las diferentes ocasiones de medida.

Es importante destacar que el número de usuarios y centros es adecuado y permite considerar que el diseño es óptimo en términos de fiabilidad y generalización.

### Referencias

- Blanco-Villaseñor, A. (1993). Fiabilidad, precisión, validez y generalización de los diseños observacionales. En M. T. Anguera (Ed.), *Metodología observacional en la investigación psicológica*. Vol 2 Fundamentación, 151-261. Barcelona: PPU.
- Blanco-Villaseñor, A., Losada, J. L. y Anguera, M. T. (1991). Estimación de la precisión en diseños de evaluación ambiental. *Evaluación Psicológica. Psychological Assessment*, 7 (2), 223-257.
- Cronbach, L.J., Gleser, G.C., Nanda, H. y Rjaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- Grupo de Atención Temprana (G.A.T.) (2000): *Libro Blanco de la Atención Temprana*. Real Patronato de Prevención y de Atención a Personas con Minusvalía. Ministerio de Trabajo y Asuntos Sociales. Madrid: Autor.
- Máiquez, M. L., Blanco-Villaseñor, A., Rodrigo, M.J. y Vermaes, I.P. (2000). La evaluación de la eficiencia en la intervención familiar: generalizabilidad y optimización del programa experiencial para padres. *Psicothema*, 12 (4). 533-542.

- Morales, V. (2003). *Evaluación psicosocial de la calidad en los servicios municipales deportivos: aportaciones desde el análisis de variabilidad*. Universidad de Málaga: Tesis doctoral.
- Morales Sánchez, V. (en prensa). Evaluación de la Calidad en Organizaciones Deportivas: Análisis de Generalizabilidad. *Revista de Psicología General y Aplicada*.
- Quintanilla, I. (2002). *Psicología social del consumidor*. Valencia: Promolibro.
- Romero Galisteo, R.P. y Morales, V. (2008). *Evaluación de la calidad en los centros de desarrollo infantil y atención temprana: proyecto piloto de investigación en los centros de la provincia de Málaga*. En I Jornadas Interdisciplinarias de Atención Temprana. El profesional de Atención Temprana: calidad en la intervención. Zaragoza, 27-28 noviembre (paper).
- Schlotzhauer, S. D. y Littell, R. C. (1997). *SAS Sistem for Elementary Statical Analysis*. Cary, NC: SAS Institute Inc.
- Ysewijn, P. (1996). *Software for generalizability studies*. Memeographed.

# DISEÑO DE PRECISIÓN DE UNA HERRAMIENTA PARA LA EVALUACIÓN DE LA CALIDAD DE UNA INSTALACIÓN MUNICIPAL DEPORTIVA

Verónica Morales y Pablo Gálvez  
Universidad de Málaga

El incremento de la práctica deportiva ha aumentado el interés de las organizaciones en establecer estrategias de diferenciación, basadas en la mejora de la calidad a través de la satisfacción del usuario como elemento clave de la estrategia de beneficios y la competitividad. El objetivo del presente trabajo consiste en aplicar la Teoría de la Generalizabilidad (TG) en el *Cuestionario de Evaluación de la Calidad de los Servicios Deportivos* (CECASDEP), aplicado a una muestra de 110 usuarios de los servicios deportivos ofertados por el Patronato Municipal de Deportes de Vélez – Málaga y de Ronda (ambos de la provincia de Málaga). El cuestionario pretende ser una herramienta valiosa y de fácil utilización para profesionales de la gestión deportiva, proporcionando unos resultados que cumplan con las condiciones de fiabilidad y validez. Los resultados muestran que el diseño de la herramienta es óptimo en términos de fiabilidad y de generalizabilidad, obteniéndose resultados superiores a .90 en ambos casos.

El deporte se ha convertido en una actividad muy accesible, significando un elemento muy importante para la salud de la colectividad, propia del llamado estado del bienestar, constituyendo un adecuado medio para la promoción de valores sociales positivos.

Las transformaciones que ha sufrido la sociedad en las últimas décadas han desembocado, en materia de deporte, en un cambio de los hábitos deportivos. Los nuevos modelos deportivos recogen ya tendencias de salud, diversión, entretenimiento, bienestar, cuidado de la imagen, etc., lo que hacen concebir al deporte como un sistema abierto. Asimismo, las numerosas connotaciones económicas que están presentes en cualquiera de las manifestaciones deportivas lo convierten en una auténtica industria de enorme peso socio-económico, lo que ha supuesto grandes cambios en los hábitos de consumo en materia de deporte, entendiendo que el deporte supone un producto de consumo muy atractivo en el mercado de bienes.

Todos estos cambios están afectando a la gestión de servicios e instalaciones deportivas. Las cambiantes demandas de los usuarios así como la aparición de nuevas modalidades deportivas han modificado la tipología de los espacios deportivos tradicionales, donde la funcionalidad se convierte en el criterio básico sin renunciar a criterios de calidad, confort y seguridad. Estos aspectos nos llevan a elaborar un cuestionario para evaluar la calidad de los servicios deportivos, debido a las cambiantes pautas de consumo y por la necesidad de actualización de las herramientas existentes.

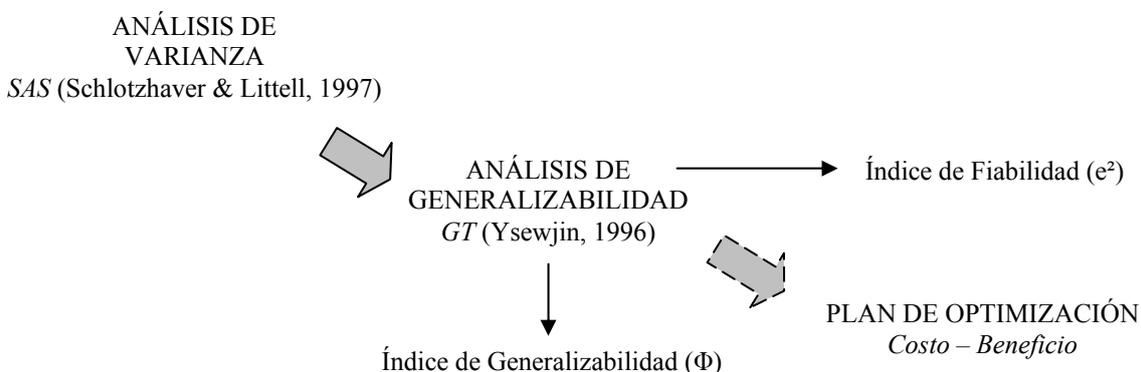
Este estudio pertenece a una investigación de mayor alcance, cuyo objetivo es elaborar una herramienta que cumpliendo con las condiciones de fiabilidad y validez, nos permita evaluar la calidad de los servicios deportivos. Los objetivos específicos que nos planteamos en esta investigación son: (a) demostrar que el número de programas de actividad física son suficientes para conseguir resultados fiables y generalizables, (b) asegurar que la muestra utilizada permite considerar que el diseño es óptimo en términos de fiabilidad ( $e^2$ ) y generalizabilidad ( $\Phi$ ). De esta forma, podremos calcular con gran precisión un adecuado *plan de optimización* (Cronbach, Gleser, Nanda y Rajaratnam, 1972) de la calidad de los servicios deportivos en términos de *costo – beneficio*. En definitiva, se trata de aportar una herramienta de fácil utilización en el ámbito de la gestión deportiva, que permita conocer de forma

fiable y precisa la calidad de los servicios ofrecidos por una organización, otorgándoles a los gestores deportivos el máximo beneficio y rentabilidad con el mínimo de recursos, con el consiguiente ahorro en términos de tiempo y dinero.

### Método

Para este estudio empleamos la metodología selectiva, que según Anguera (2003) es la más utilizada dentro de la investigación social y resulta muy adecuada en el ámbito del deporte por sus innumerables aplicaciones posibles. Entre las diversas aplicaciones que ofrece esta metodología utilizamos el *cuestionario*, que es un procedimiento sistemático y flexible que aporta información estandarizada con gran ahorro de tiempo. Hemos seleccionado un diseño no ortogonal, multifaceta y parcialmente anidado.

Para cumplimentar los objetivos específicos de esta investigación, hemos utilizado el paquete estadístico SAS® (Schlotzhauer & Littell, 1997) con el fin de realizar un análisis de componentes de varianza a través del procedimiento GLM (*General Linear Model*), obteniendo así la suma de cuadrados necesaria para poder realizar posteriormente un análisis de generalizabilidad mediante el programa estadístico GT (Ysewijn, 1996). La teoría de la Generalizabilidad (TG) (Cronbach, Gleser, Nanda y Rajaratnam, 1972), nos permite comprobar la precisión de generalización de los resultados de cualquier investigación. De esta forma, mediante este análisis pretendemos generalizar con precisión los resultados evaluados a través de distintos diseños de medida para un adecuado plan de optimización en términos de costo – beneficio.

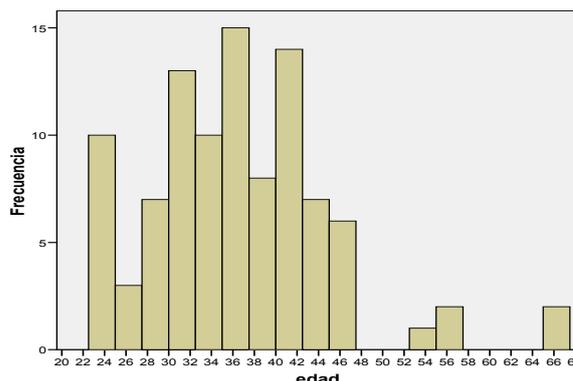


### Participantes

La muestra está compuesta por 110 usuarios, de los cuales 48 (43.6%) son de género femenino mientras que 62 (56.4%) son de género masculino, situándose la media de edad en 36.13 años. En cuanto al rango de edad, obtenemos que la edad mínima es de 23 años y la máxima de 66 años. La selección de la muestra corresponde a un muestreo por conglomerados, utilizando los municipios de Vélez-Málaga y Ronda, ambos de la provincia de Málaga.

*Tabla 1. Resultados estadísticos relativos a la edad*

N	Válidos	110
	Perdidos	12
Media		36.13
Moda		37
Desv. típica		8.41
Mínimo		23
Máximo		66



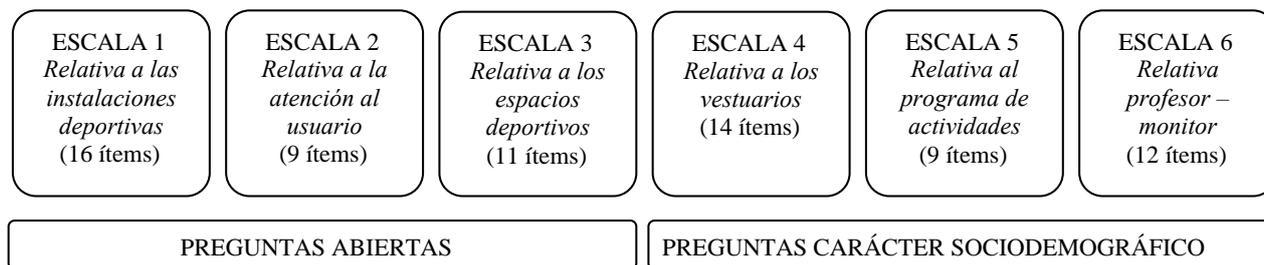
*Figura 1. Rango de edad de los participantes en el estudio*

*Procedimiento*

Realizamos una recopilación de información sobre las herramientas de evaluación de la calidad de los servicios deportivos, basándonos tanto en la herramienta *SERVQUAL* (Parasuraman, Zeithaml y Berry, 1988, 1993) como en el cuestionario *ICPAF* (Hernández Mendo, 2001), así como en los trabajos de Morales (2003) y Morales, Hernández Mendo y Blanco (2005, 2009). A partir de aquí hemos elaborado el cuestionario en diversas fases para realizar la recogida de información durante los meses de diciembre de 2008 y enero de 2009.

*Material*

Hemos utilizado la herramienta *CECASDEP* (Cuestionario de Evaluación de la Calidad de los Servicios Deportivos), compuesto por 71 ítems estructurados en 6 escalas, finalizando con una pregunta abierta por cada escala, y una serie de preguntas de carácter sociodemográfico: nombre de la instalación deportiva, actividad practicada, horario de la actividad, género, fecha y lugar de nacimiento y, por último, nivel de estudios. Las respuestas se producen en un continuo del 1 al 5, donde 1 significa “nada de acuerdo” y 5 “muy de acuerdo” con el enunciado de la pregunta.



*Figura 2. Estructura del cuestionario CECASDEP*

**Resultados**

En primer lugar se ha comprobado la fiabilidad del cuestionario, sometiendo los datos a un análisis para determinar el coeficiente Alfa de Cronbach para cada una de las escalas. Los resultados obtenidos muestran una elevada consistencia interna, correspondiendo el valor mínimo a la escala 2 (.711) mientras que el valor máximo pertenece a la escala 6 (.951), por lo que teniendo en cuenta el rango establecido

por Nunnally (1976), obtenemos una consistencia interna aceptable para la escala 2, mientras que los resultados son satisfactorios para el resto de las escalas (tabla 2).

A continuación se llevó a cabo un análisis preliminar para comprobar la pertinencia del análisis factorial exploratorio teniendo en cuenta los siguientes indicadores: determinante de la matriz, la medida de adecuación muestral Kaiser-Meyer-Olkin (KMO), el índice de esfericidad de Bartlett y el % de la varianza explicada. En este caso, examinando la matriz de correlaciones, el valor de los determinantes es adecuado, siendo próximos a 0 en todas las escalas; en cuanto a la medida de adecuación muestral KMO, los resultados obtenidos también son adecuados, presentando valores que oscilan entre .671 de la escala 2 como valor mínimo y .906 de la escala 6 como valor máximo; para el test de esfericidad de Bartlett, presenta significatividad en todas las escalas ( $p < 0.05$ ); por último, los resultados de la varianza explicada, obtenemos que todas las escalas presentan valores superiores al 60% (tabla 2).

Tabla 2. Resultados del análisis de fiabilidad y del análisis factorial exploratorio

Escala	Nº ítems	Alpha de Cronbach	Determinante	KMO	Varianza explicada
1	16	.817	.002	.779	65.748 %
2	9	.711	.059	.671	64.942 %
3	11	.863	.007	.835	66.319 %
4	14	.907	.001	.869	63.441 %
5	9	.886	.007	.855	67.688 %
6	12	.951	.000003	.906	66.360 %

Para el análisis de componentes de varianza, hemos utilizado 5 variables de clasificación que son: (U) usuarios: 110, (I) instalaciones: 2, (E) espacios deportivos: 6, (A) actividades: 6, y (K) escalas: 6, obteniéndose un total de 30 diseños de medida. Los resultados del análisis de generalizabilidad muestran altos índices de fiabilidad ( $e^2$ ) y generalizabilidad ( $\Phi$ ) en todos los diseños de medida, salvo en aquellos en los que la faceta *escala* interacciona con otras facetas, donde observamos que los resultados obtenidos muestran índices bajos, por lo que se confirma que cada escala mide constructos distintos (tabla 3).

Tabla 3. Resultados del análisis de generalizabilidad en distintos diseños de medida

Diseños de medida	Fiabilidad ( $e^2$ )	Generalizabilidad ( $\Phi$ )
U I E K / A	1.000	1.000
I E A K / U	.975	.956
U E K / I A	1.000	1.000
I E K / U A	.975	.956
I E A / U K	.793	.000
U K / I E A	1.000	1.000
E K / U I A	.975	.956
I E / U A K	.634	.000
K / U I E A	.975	.956
I / U E A K	.507	.000

Para una adecuada optimización del análisis de medida en términos de costo – beneficio, mostramos a continuación distintas aproximaciones con diferentes niveles en cada una de las distintas facetas (tabla 4).

Tabla 4. Plan de optimización de distintos diseños de medida

Diseños de medida	Número	e <sup>2</sup>	Φ	Nº observaciones	
I E A K / U	110	.975	.956	47520	Nivel actual
	40	.935	.887	17280	Aprox. óptima
I E K / U A	110 / 6	.975	.956	47520	Nivel actual
	40 / 1	.935	.887	2880	Aprox. óptima
K / U I E A	110 / 2 / 6 / 6	1.000	1.000	47520	Nivel actual
	40 / 1 / 1 / 1	.935	.887	240	Aprox. óptima

En el primer caso, comprobamos la interacción de la faceta *usuarios* sobre las facetas *instalación*, *espacios deportivos*, *actividades* y *escalas*, obteniendo índices de fiabilidad y de generalizabilidad de .935 y .887 respectivamente con un número de 40 usuarios; en el segundo ejemplo, interaccionamos las facetas *usuarios* y *actividades* a partir de las facetas *instalaciones*, *espacios deportivos* y *escalas*, utilizando 40 usuarios y pasando de 6 a 1 actividad deportiva, obteniendo igualmente altos índices de fiabilidad (.935) y de generalizabilidad (.887); en el último ejemplo, planteamos el plan de optimización que de todos los modelos que han sido sometidos a consideración, supondría la situación ideal, pues observamos la interacción de las facetas *usuarios*, *instalaciones*, *espacios deportivos* y *actividades* a partir de la faceta *escalas*, donde con 40 usuarios, 1 instalación, 1 espacio deportivo y 1 actividad, obtenemos un índice de fiabilidad de .935 y de generalizabilidad de .887, reduciendo el número de observaciones a 240. Por tanto, podemos afirmar que tanto el número de programas de actividad física evaluados, como también el tamaño de la muestra utilizada, nos permite considerar que el diseño de la herramienta es óptima en términos de fiabilidad y de generalizabilidad.

#### Referencias

- Anguera, M.T. (2003). La metodología selectiva en el deporte. En A. Hernández Mendo (Coord.), *Psicología del Deporte (Vol. II) – Metodología* (74-96). Buenos Aires: Tulio Guterman (<http://www.efdeportes.com>).
- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- Morales, V. (2003). *Evaluación psicosocial de la calidad en servicios municipales deportivos: aportaciones desde el análisis de variabilidad*. Universidad de Málaga: Tesis doctoral.
- Morales, V., Hernández Mendo, A. y Blanco, Á. (2005). Evaluación de la calidad en los programas de actividad física. *Psicothema*, 17(2), 311-317.
- Morales, V., Hernández Mendo, A. y Blanco, Á. (2009). Evaluación de la calidad en organizaciones deportivas: adaptación del modelo SERVQUAL. *Revista de Psicología del Deporte*, 18(2), 137-150.
- Nunnally, J.C. (1976). *Psychometric theory*. New York: McGraw-Hill.

- Parasuraman, A., Zeithaml, V. y Berry, L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.
- Parasuraman, A. Zeithaml, V. y Berry, L. (1993). More on improving service quality measurement. *Journal of Retailing*, 69, 140-147.
- Schlotzhauer, S. D. & Littell, R. C. (1997). *SAS System for Elementary Statistical Analysis*. Cary, NC: SAS Institute Inc.

# ESTIMACIÓN DE PARÁMETROS BAJO EL SUPUESTO DEL INCUMPLIMIENTO DE LAS CONDICIONES DE APLICACIÓN

**Coordinador: Joan Guàrdia**  
Universidad de Barcelona

Está claro que el objetivo fundamental de las técnicas estadísticas se centra en la obtención de estimaciones cada vez más eficientes y suficientes, en los términos propios de la estimación estadística, de los parámetros comprometidos en el modelo estadístico de interés. Se trata, pues, de una fase de la modelización estadística que no debe ser tratada de una forma ligera, lo cual es más habitual de lo que sería deseable, puesto que de ella depende buena parte del resultado de ese proceso. En aquellas aproximaciones empíricas basadas en sistemas de simulación, la fase de estimación suele ser el propio objeto de estudio y, por tanto, la anterior aseveración adquiere una importancia en sí misma. Ahora bien, cuando se trata de analizar datos aplicados derivados de procesos de muestreo, a veces poco rigurosos, la fase de estimación adquiere una relevancia excepcional puesto que no hay más opción que analizar los datos disponibles. Una correcta tipificación de las variables y una correcta elección de la técnica de estimación de parámetros son cruciales para obtener resultados sólidos y rigurosos. Al margen de que las estimaciones de parámetros a partir de datos aplicados se complementen con simulaciones para ofrecer estimaciones más rigurosas (tanto en la estimación de los parámetros estructurales como de sus errores estándar), es esta una cuestión que sigue generando amplia bibliografía al respecto. En las siguientes comunicaciones se pretende aportar algunas soluciones básicas para la fase de estimación de parámetros en situaciones muy comunes en las que están comprometidas las condiciones de aplicación, especialmente aquellas que hacen referencia al tamaño muestral y a las distribuciones observadas. En el primer caso, el tamaño de muestra, se hace muy habitual el uso de estimaciones no especialmente adaptados a muestras pequeñas o, en varios casos, con tamaños muestrales que de hecho impedirían una correcta estimación. Lo mismo sucede con las distribuciones observadas en las que es fácil apreciar valores de asimetría muy distintos de cero (ya sean positivos o negativos). Tales circunstancias han sido analizadas bajo diversos supuestos y situaciones, intentándolas resolver con diferentes aproximaciones y posibilidades estadísticas. Ello se ha efectuado fundamentalmente en técnicas de modelización muy usadas en la investigación psicológica aplicada como son los Modelos de Ecuaciones Estructurales (SEM) y la Regresión Logística (LR). La primera aproximación (SEM) implica sistemas de estimación de parámetros “*full estimation*” es decir, sistemas globales de estimación de parámetros, lo cual siempre ha sido valorado como una de sus principales ventajas frente a sistemas de estimación parciales; pero, a su vez, hace mucho más complicada la fase de estimación y mucho más frágil toda esa arquitectura frente a la vulneración de supuestos básicos. Por lo que se refiere a la segunda aproximación (LR), se trata de una opción cada vez más oportuna a la vista del uso de variables categóricas y a la cada vez más usual tratamiento estadístico desde una óptica epidemiológica (estimaciones de riesgo, de *odds ratio*, etc.) de datos psicológicos.

Estas ideas son las que prevalecen de forma reiterada en los distintos trabajos que se incorporan en este grupo. De este modo, en el trabajo de Pero, Guàrdia, Freixa y Blanxart se aborda el grado de acuerdo en la decisión tomada a partir de las diferentes estrategias de decisión basadas en el estadístico *t de Student* de grupos independientes, prueba no paramétrica U de *Mann-Whitney*, comparación de intervalos de confianza alrededor de la media y comparación de intervalos de confianza alrededor de la mediana, contruidos éstos últimos a partir del error estándar o de la distribución binomial. Se trata, pues, del análisis de la posible alternativa que el estudio de la Mediana pueda suponer como alternativa a las pruebas más clásicas. Siguiendo con el esquema centrado en la aplicación de procedimientos estadísticos en situaciones de incumplimiento de los supuestos, el trabajo de Benítez, Guàrdia, Arch

y Jarne trata, mediante simulación y datos reales, de analizar la distribución observada de cada índice de ajuste de los Modelos de Ecuaciones Estructurales (SEM) según los datos vinculados a un modelo complejo, y analizar a la vez si esos procesos de simulación pudieran ser una buena opción al estudio del propio ajuste de los modelos propuestos. En idéntica posición se centra el trabajo de Carrera, Feixas y Guàrdia que, en este caso centrado en la existencia de datos no registrados (*missing data*) revisa las diversas opciones en función de las propiedades que tengan la distribución de esos datos no registrados y analiza la bondad de las diversas alternativas de simulación, mostrando el efecto de concentración de algunas de esas técnicas y, más importante, la necesidad del estudio sistemático de la distribución de los propios datos no registrados. Finalmente, en este grupo de trabajos en torno al estudio de estadísticos en situaciones con anomalías en sus datos, el trabajo de Gordóvil y Guàrdia trata de analizar el proceso de estimación de parámetros en modelos estadísticos con muestra pequeña. Para ello, se generan tres situaciones en las que la variable dependiente es tratada como variable dicotómica, politómica y continua. Se comparan diversas técnicas de estimación de parámetros, unas basadas en los presupuestos más clásicos y otras basadas en criterios bayesianos en modelos de regresión logística binaria, multinomial y lineal mixta bajo la condición, como decíamos, de tamaño muestral pequeño.

De todas las opciones que se presentan, quizás deberemos aumentar los trabajos y debates sobre ellas, añadiendo a ello, y con mayor intensidad, las posibilidades de la estadística bayesiana. Esta especial óptica ya se ha hecho habitual en muchos ámbitos de la ciencia y la idea de la valoración probabilística de cada una de las posibles causas, una vez observado el efecto, ya no es una novedad, a pesar de que en Psicología estamos lejos, en este tema, de otras disciplinas como la Biología o la propia Medicina. Quizás sea interesante recordar aquí las palabras de Albert cuando dice que quizás el problema de la decisión y estimación estadística implica, como mínimo, un proceso previo de debate y “ruido” para saber, como mínimo, lo poco que sabemos sobre muchas cosas.

# APLICACIÓN DE MODELOS DE REGRESIÓN BAJO CONDICIONES DE MUESTRA PEQUEÑA

Amalia Gordóvil<sup>1</sup> y Joan Guàrdia<sup>1,2</sup>

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Instituto de Investigación en Cognición, Cerebro y Conducta (IR3C)

Los procedimientos clásicos de estimación en modelos de regresión presentan una serie de limitaciones bien conocidas. Una de las más frecuentes, derivadas de trabajos clínicos, es la dificultad de acceso a muestras suficientes. Así, se generan trabajos con tamaños muestrales pequeños en los que la estimación de coeficientes se ve comprometida. Además, el escaso tamaño de muestra en variables categoriales determina la presencia de ceros muestrales que aun compromete más la estimación de coeficientes. Esto es detectable a partir de valores elevados en el error estándar de la estimación así como la tendencia a infinito de los intervalos de confianza (Field, 2005; Hosmer y Lemeshow, 1989). A fin de solventar este problema, se ha propuesto la corrección de Haldane, adecuada ante variables discretas pero no categóricas ni continuas. Además, el trabajo con muestras pequeñas incrementa el hecho de que las observaciones influyentes (*outliers* o con gran influencia en la recta de regresión) puedan sesgar las estimaciones. (Academic Technology Service [UCLA], 2008). El presente trabajo muestra algunas aplicaciones de la estimación de la regresión con variables categóricas y muestras pequeñas, a partir de datos derivados de la administración de la *Escala de Calidad de Vida en Epilepsia Infantil (CCVEI)*.

Las investigaciones derivadas de planteamientos aplicados sitúan frecuentemente al investigador ante situaciones que producen déficit en la estimación de parámetros. Tres ejemplos prototípicos son: 1) presencia de ceros muestrales, lo cual producirá problemas en la estimación de coeficientes (Greenland, Schwartzbaum y Finkle, 2000); 2) presencia de *outliers*, contribuyendo a un sesgo de resultados (UCLA, 2008) y 3) grupos de diferentes tamaños; las clasificaciones en modelos de predicción son sensibles al grupo de mayor tamaño, independientemente del ajuste del modelo (Hosmer y Lemeshow, 1989). Todo ello deriva en la obtención de valores elevados en coeficientes y errores estándar así como en la presencia de intervalos de confianza de considerable amplitud. En consecuencia, la estimación clásica de parámetros en modelos de regresión se ve considerablemente comprometida.

La elevada ocurrencia de tales situaciones ha generado algunas propuestas para mitigar tales efectos. En el caso de la regresión logística, la *Corrección de Haldane* propone solventar el problema de los ceros muestrales incrementando .5 el valor de cada casilla. Sin embargo es estrictamente adecuada para variables discretas y no categóricas o continuas (Field, 2005). Ante el manejo de datos en pruebas de independencia con tamaño muestral pequeño, se plantea la construcción de tests exactos de muestra pequeña utilizando la distribución hipergeométrica generalizada; opción aplicable a tablas de contingencia 2x2 (Agresti, 1999). Si extendemos tales disertaciones a modelos de regresión en general, encontramos algunas preferencias como la elección de la ratio de Máxima Verosimilitud (MV) frente a tests específicos en los tests de hipótesis (Kleinbaum, 1992) y la elección de regresión jerárquica frente a no jerárquica, lo cual permite reducir efectos derivados del tamaño muestral según Sculz y Betebenner (2004). Estos autores presentan la generación de estimaciones bayesianas como propuesta alternativa. En comparación a estimaciones por MV basadas en el total de la muestra, éstas presentan valores similares en términos de validez predictiva basándose en el 25% de la muestra.

El presente estudio pretende analizar un modelo clínico derivado de datos sobre calidad de vida en epilepsia infantil así como la estructura de medida utilizada. Para ello, se generaron tres situaciones en

las que la variable dependiente (calidad de vida general [CVG]) fue tratada como variable dicotómica, politómica y continua. Por tanto, el objetivo de estudio consistió en la comparación de estimaciones clásicas y alternativas en modelos de regresión logística binaria, multinomial y lineal mixta bajo la condición de tamaño muestral pequeño ( $n=24$ ).

## Método

### Participantes

Se trabajó con cuidadores principales de pacientes diagnosticados de epilepsia. Los criterios de inclusión para dichos pacientes fueron: 1) edad entre 4 y 18 años y 2) presencia de mínimo una crisis durante los dos años previos.

La distribución por género de la muestra final ( $n=24$ ) fue de 62.5% niñas y 37.5% niños con una media de edad (recortada al 5%) de 10.17 (DT=4.02; IC al 95%=8.6-12). El 40% presentó un grado escolar ajustado a la edad cronológica.

Los datos clínicos indicaron presencia de comorbilidad para el 29.2%. La mediana de edad en que se produjo la primera crisis epiléptica fue de 1.5 años ( $\bar{x}_{\text{recortada al 5\%}}=2.75$  años; DT=3.15; IC al 95%=1.65-4.31). En la actualidad el 16.7% consume 1 fármaco antiepiléptico (FAE), el 50% consume 2, el 29.1% presenta una prescripción de 3 FAEs y el 4.2% consume 4. El valor de la mediana de tiempo de consumo de FAEs es de 4.79 años ( $\bar{x}_{\text{recortada al 5\%}}=6.2$ ; DT=4.96; IC al 95% = 4.36-8.35). La frecuencia de crisis epilépticas durante el mes previo a la evaluación ( $n=17$ ) presentó un valor mediano de 6.5 ( $\bar{x}_{\text{recortada al 5\%}}=23.64$ ; DT=61.250; IC al 95%=6.93-61.25). Finalmente, la figura 1 resume el grado de severidad de crisis.

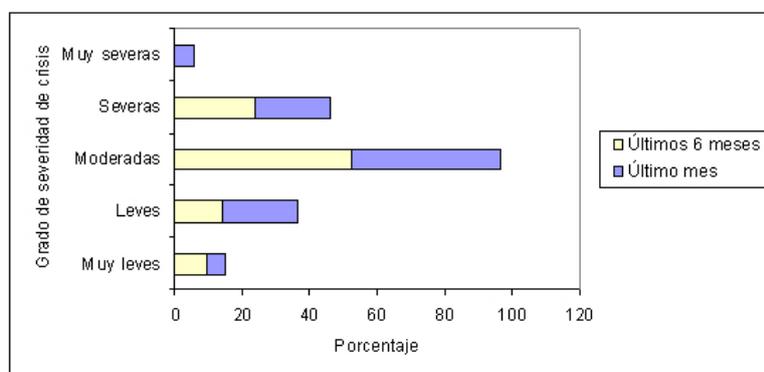


Figura 1. Severidad de crisis epilépticas de acuerdo a la percepción del cuidador principal, durante 6 los meses ( $n=21$ ) y 1 mes ( $n=18$ ) previos

### Instrumentos

Se utilizaron de dos instrumentos: un cuestionario de datos clínicos (CDC) y la versión adaptada del *Quality of Life in Childhood Epilepsy Questionnaire* ([QOLCE] Sabaz et al., 2000).

El CDC se generó “ex profeso” para la investigación y consistió en preguntas sobre las siguientes cuestiones: datos demográficos (edad, género, grado escolar), edad en que apareció la primera crisis epiléptica y signos de detección, frecuencia y severidad de crisis durante los 6 y 1 meses previos,

farmacología (consumo de antiepilépticos, edad de inicio de consumo, tipo, número y dosis de FAEs consumidos en la actualidad), sometimiento a intervención quirúrgica, tipo de epilepsia) y posibles trastornos comórbidos (presencia y tipo).

El *CCVEI* es la adaptación del cuestionario originariamente australiano *QOLCE* (Sabaz et al., 2000), un instrumento multi-faceta específico para epilepsia, de 79 ítems, aplicable a niños de entre 4 y 18 años. Los dominios y subescalas constitutivos del instrumento se muestran en la tabla 1. Además de la obtención de valores para cada subescala, proporciona el cálculo de un índice de CVG. Tanto las subescalas como el índice de CVG oscilan entre 0 y 1, siendo 1 el ajuste adecuado para el paciente según su grupo de edad.

*Tabla 1. Estructura del CCVEI por ámbitos, subescalas y número de ítems*

Ámbitos	Subescalas	Nº de ítems
Aspectos Físicos	Restricciones Físicas	10
	Energía/Fatiga	2
Bienestar	Depresión	4
	Ansiedad	5
	Control/Impotencia	4
	Autoestima	5
Cognición	Atención/Concentración	5
	Memoria	6
	Lenguaje	8
	Otros Aspectos Cognitivos	3
Actividades Sociales	Interacciones Sociales	5
	Actividades Sociales	3
	Estigma	1
Comportamiento	Comportamiento	16
	Salud General	1
	Calidad de Vida	1

La respuesta a los ítems del *CCVEI* consiste en la valoración mediante escala tipo *Likert* (de 1 a 5). Además, existe la opción de respuesta “ítem no aplicable”. Finalmente, para cada ámbito examinado se incluye una pregunta abierta. Todas las preguntas se refieren a aspectos ocurridos durante el mes previo a la administración.

### *Procedimiento*

Una vez confeccionados los protocolos de administración y definido el contacto con los centros colaboradores, se procedió a la aplicación de los instrumentos entre enero y julio de 2009 en Barcelona (España). Las características logísticas de las diferentes instituciones definieron tres modos de administración: presencial individual (45.9%), presencial colectiva (33.3%) y vía correo postal (20.8%).

## Análisis de Datos

Se ejecutó una fase preliminar de análisis descriptivos univariados para la detección de posibles anomalías en la distribución de variables. A continuación se procedió a la descripción de las variables implicadas en el estudio. Posteriormente, se establecieron tres modelos de análisis (regresión logística binaria, multinomial y regresión lineal mixta) en función del tratamiento otorgado a la variable dependiente (dicotómica, politómica y continua) tal y como indica la tabla 2.

Tabla 2. Análisis de datos: modelos de regresión y estimaciones ejecutadas para los tres tratamientos de la variable dependiente

	VD	Modelo de regresión	Estimación
Dicotómica	Baja: $x_i < \bar{x}$ (49%) Alta: $x_i > \bar{x}$ (51%)	Logística binaria	Clásica Bayes
Politómica	Baja: $x_i = \bar{x} - 2DT$ (8,33%) Alta: $x_i = \bar{x} + 2DT$ (4,16%) Media: Resto valores (87,51%)	Logística multinomial	Clásica MV Bayes
Continua	$\bar{x} = .62$ ; $DT = .105$	Lineal mixta	OLS corregidos Bayes

Los análisis fueron realizados mediante SPSS 16.0 y TETRAD 4.3.9.9.

## Resultados

De acuerdo a los objetivos planteados, la exposición de los resultados tratará en primer lugar sobre el análisis del modelo clínico en relación a la calidad de vida (CV) en epilepsia infantil y acerca de la estructura del CCVEI. Recordemos que, en ambos casos el motivo de estudio es el análisis del funcionamiento de las diversas estimaciones empleadas en modelos de regresión.

En el análisis de tipo clínico, se generaron modelos de predicción a partir de las variables que la literatura científica reporta como significativamente relacionadas con la CV en epilepsia infantil. Se trata de las siguientes: edad, grado escolar, edad de inicio de crisis y tiempo de consumo de FAEs (Yong, Chengye, y Jiong, 2006), número de FAEs (Campos y Carreño, 2009), frecuencia de crisis (Adeuwuya y Osen, 2005; Ronen, Streiner, y Rosenbaum, 2003; Yong et al., 2006), severidad de crisis (Conolly, Sabaz, Lawson, y Bye 2005; Ronen et al., 2003) y comorbilidad (Felder et al., 2004; Ronen et al., 2003).

Tal y como presentamos en la tabla 3 las estimaciones bayesianas detectan la significación estadística de la mayor parte de variables, en contraposición a los métodos de estimación clásicos. La comparación de las estimaciones bayesianas en los tres modelos generados, indica significación para las mismas variables: ajuste en grado escolar, edad de primera crisis, frecuencia y severidad de crisis durante el mes previo, tiempo de consumo y número de FAEs y número de trastornos comórbidos.

Tabla 3. Valores de significación estadística para las variables introducidas en el modelo.  
Diferenciación por tipo de modelo de regresión y estimación

Variable	CVG Dicotómica		CVG politómica			CVG Cuantitativa	
	Estimación Clásica	Estimación Bayesiana	Estimación Clásica	Estimación MV	Estimación Bayesiana	Estimación OLS corregidos	Estimación Bayesiana
Edad	.066	.101	.091	.097	.120	.098	.121
GEajust. <sup>#1</sup>	.023	-.235 *	.089	.011	-.341 *	.092	-.347 *
Edad1 <sup>a</sup> cr <sup>2</sup>	-.117	-.366*	-.202	-.199	-.400 *	-.217	-.412 *
Sev.6m <sup>#3</sup>	.061	.021	.021	.044	.056	.032	.076
Frec.1m <sup>#4</sup>	-.119	-.298 *	-.201	-.177	-.334 *	-.247	-.376 *
Sev 1m <sup>#5</sup>	-.167	-.377 *	-.201	-.189	-.402 *	-.244	-.412 *
T. FAE <sup>6</sup>	-.200	-.521 *	-.211	-.261	-.599 *	-.361	-.612 *
Nº FAE <sup>7</sup>	-.099	-.401 *	-.101	-.189	-.416 *	-.190	-.487 *
Nº Trast <sup>8</sup>	-.094	-.267 *	-.118	-.132	-.366 *	-.150	-.312 *

<sup>1</sup>GEajust: ajuste en grado escolar; <sup>2</sup>Edad1<sup>a</sup>cr: edad primera crisis; <sup>3</sup>Sev.6m: severidad de crisis durante 6 meses previos; <sup>4</sup>Frec.1m: frecuencia de crisis durante 1 mes previo; <sup>5</sup>Sev 1m: severidad de crisis durante 1 mes previo; <sup>6</sup>T. FAE: tiempo de consumo de FAEs; <sup>7</sup>Nº FAEs: número de FAEs consumidos actualmente; <sup>8</sup>Nº Trast: número de trastornos comórbidos.

Estimación Bayesiana: En términos de la estimación bayesiana propia de los SEM (TETRAD). \* p < .05 # Variables categóricas

En el estudio de la estructura interna del CCVEI (ver tabla 4) observamos que, nuevamente las estimaciones bayesianas reflejan mayor cantidad de significaciones estadísticas. Las subescalas detectadas a partir de este tipo de estimación son: Energía, Control, Autoestima, Atención, Lenguaje, Otros Aspectos Cognitivos, Actividades sociales y Estigma.

Tabla 4. Estudio de la estructura del CCVEI. Valores de significación estadística para cada subescala. Diferenciación por tipo de modelo de regresión y estimación

Variable	CVG Dicotómica		CVG politómica			CVG Cuantitativa	
	Estimación Clásica	Estimación Bayesiana	Estimación Clásica	Estimación MV	Estimación Bayesiana	Estimación OLS corregidos	Estimación Bayesiana
R.Físicas <sup>1</sup>	-.088	-.083	-.084	-.089	-.093	-.085	-.081
En/Fat <sup>2</sup>	.211 *	.227 *	.221 *	.277 *	.277 *	.237*	.276 *
Depresión	-.042	-.049	-.063	-.077	-.060	-.053	-.062
Ansiedad	.082	.069	.060	.091	.082	.062	.088
Cntr/Imp <sup>3</sup>	.161	.201 *	.177	.221 *	.211 *	.182	.214 *
Autoestima	.212 *	.221 *	.201 *	.278 *	.260 *	.204*	.267 *
Atención	.199 *	.211 *	-.101	.249 *	.221 *	-.121	.225 *
Memoria	-.005	-.006	-.009	-.010	-.009	-.008	-.010
Lenguaje	.327 *	.399 *	.288 *	.349 *	.410 *	.298 *	.412 *
OAC. <sup>4</sup>	.161 *	.201 *	.187 *	.218 *	.231 *	.191 *	.234 *
IS <sup>5</sup>	.121	.121	.101	.161	.141	.104	.144
AS <sup>6</sup>	.412 *	.526 *	.377 *	.482 *	.540 *	.382 *	.543 *
Estigma	.299 *	.341 *	.218 *	.320 *	.374 *	.220 *	.378 *
Comport. <sup>7</sup>	.008	.009	.007	.008	.012	.009	.018

<sup>1</sup>R.Físicas: Restricciones Físicas; <sup>2</sup> En/Fat: Energía/Fatiga; <sup>3</sup> Cntr/Imp: Control/Impotencia; <sup>4</sup> OAC: Otros Aspectos Cognitivos;

<sup>5</sup>IS: Interacciones Sociales; <sup>6</sup>AS: Actividades Sociales; <sup>7</sup>Comport: Comportamiento.

Estimación Bayesiana: En términos de la estimación bayesiana propia de los SEM (TETRAD). \* p < .05

## Discusión

El establecimiento de modelos de regresión en base a estimaciones clásicas de parámetros compromete la estabilidad de coeficientes y por tanto la fiabilidad de resultados. Por ello, se están generando líneas de trabajo en base a propuestas alternativas. Una de las más prometedoras son las estimaciones bayesianas. El presente estudio demuestra cómo precisamente este tipo de planteamiento logra detectar significación estadística en variables que la literatura científica reporta como ciertamente relacionadas con la variable criterio introducida en el modelo, todo ello, trabajando con muestra pequeña (n=24).

Los resultados obtenidos indican además, que el planteamiento más adecuado para la variable CVG es de tipo continuo, tal y como trabajan otros autores del campo sustantivo (Sabaz et al., 2000). Finalmente, la elección de estimaciones bayesianas y modelo de regresión lineal mixto, nos permiten establecer conclusiones con un mayor grado de fiabilidad. Van en el sentido que, las variables con mayor grado de predicción sobre la CVG de pacientes con epilepsia son: grado de severidad de crisis durante el mes previo y aquéllas referidas a aspectos farmacológicos (número de FAEs consumidos en la actualidad y tiempo de consumo de FAEs).

Complementariamente, las subescalas del *CCVEI* con mayor potencia predictora son Lenguaje y Actividades Sociales.

Las conclusiones metodológicas planteadas nos sitúan ante un panorama en el que resulta necesaria la investigación sobre métodos de estimación alternativos que permitan solventar algunas de las limitaciones derivadas de estimaciones clásicas de parámetros. Por ello, paralelamente al estudio con datos procedentes de campos aplicados, se hacen necesarios estudios de simulación en los que puedan establecerse diversidad de condiciones (por ejemplo, distintos tamaños muestrales) y en consecuencia, estudiar diferentes alternativas de estimación para valorar hasta qué punto y bajo qué supuestos y condiciones es preferible optar con modalidades clásicas o alternativas.

Agradecimientos. El estudio fue subvencionado por *El Comissionat per a Universitats i Recerca del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya* y el *Fondo Social Europeo*.

## Referencias

- Academic Technology Services. (2008). Logistic regression diagnostics. Recuperado 26 Febrero, 2008, de Word Wide Web <http://www.ats.ucla.edu/stat/stata/webbooks/logistic/chapter3/statalog3.htm>.
- Adeuwuya, A. O. y Osen, S. B. (2005). Impact of psychiatric morbidity on parent-rated quality of life in Nigerian adolescents with epilepsy. *Epilepsy & Behavior*, 7, 497-501.
- Agresti A. (1999). Modelling ordered categorical data. Recent advances and future challenges. *Statistics in Medicine*, 18, 2191-2207.
- Campos, M. R. y Carreño, M. (2009). Utilización de carbamazepina y oxcarbazepina en pacientes pediátricos con epilepsia parcial en España. Estudio observacional. *Neurología*, 24 (1), 30-39.
- Conolly, A. M., Sabaz, M., Lawson, J. A. y Bye, A. (2005). Quality of life in childhood epilepsy: validating the QOLCE. *Journal of Paediatrics and Child Health*, 41, 156-158.

- Felder, R., Frey, E., Proksch, K., Varni, J. W., Gadner, H. y Topf, R. (2004). Validation of the German version of the Pediatric Quality of Life Inventory (PedsQL) in childhood cancer patients off treatment and children with epilepsy. *Quality of Life Research* 13, 223–234.
- Field, A. (2005). *Discovering statistics using SPSS*. London: SAGE publications.
- Greenland, S., Schwartzbaum, J. A. y Finkle, W. D. (2000). Problems due to small samples and sparse data in conditional logistic regression analysis. *American Journal of Epidemiology*, 151 (5), 531-539.
- Hosmer, D. W. y Lemeshow, S. (1989). *Applied logistic regression*. USA: John Wiley & Sons, Inc.
- Kleinbaum, D. G. (1992). *Logistic regression. A self-learnign text*. New York: Springer-Verlag.
- Ronen, G. M., Streiner, D. L. y Rosenbaum, P. (2003). Health-related quality of life in childhood epilepsy: Moving beyond ‘seizure control with minimal adverse effects. *Health and Quality of Life Outcomes*, 1, 1-10.
- Sabaz, M., Cairns, D.R., Lawson, J.A., Nheu, N., Bleasel, A.F. y Bye, A.M.E. (2000). Validation of a new quality of life measure for children with epilepsy. *Epilepsia*, 41, 765-74.
- Sculz, E. M. y Betebenner, D. (2004). Hierarchical logistic regression in course placement. *Journal of Educational Measurement*, 41 (3), 271-286.
- Yong, L., Chengye, J. y Jiong, Q. (2006). Factors affecting the quality of life in childhood epilepsy in China. *Acta Neusologica Scandinavica*, 113, 3, 167-173.

# EL USO DE LA MEDIANA COMO ESTADÍSTICO RESISTENTE. COINCIDENCIAS EN LA DECISIÓN A PARTIR DE DIFERENTES ESTADÍSTICOS

Maribel Però<sup>1,2</sup>, Joan Guàrdia<sup>1,2</sup>, Montserrat Freixa<sup>1</sup> y Jaume Turbany<sup>1</sup>

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Instituto de Investigación en Cognición, Cerebro y Conducta (IR3C)

En diferentes trabajos realizados vía simulación hemos estudiado la potencia de la comparación de los intervalos de confianza basados en las medianas como prueba de decisión en el caso de trabajar en un diseño de grupos independientes. Los resultados hallados hasta el momento no muestran una adecuada potencia de la comparación de estos intervalos de confianza en la decisión estadística, aunque cabe destacar, que en determinadas situaciones a pesar de que la potencia no es muy buena, es mejor que la potencia observada en las pruebas utilizadas clásicamente como pueden ser el estadístico t de *Student* de grupos independientes o la prueba no paramétrica U de Mann-Whitney. Con el fin de poder diseñar nuevas condiciones de simulación en este trabajo se muestran los resultados de la comparación de las coincidencias en la decisión de las diferentes estrategias estudiadas. Los resultados muestran un patrón diferente, en función de las condiciones de simulación y del criterio de decisión establecido en la comparación de los intervalos de confianza.

En la mayoría de los trabajos aplicados en el ámbito de las ciencias sociales se hace un uso masivo de la estadística clásica en detrimento del uso de pruebas más resistentes o robustas ante el incumplimiento de las condiciones de aplicación. Ya en 1977, Tukey propuso que para un adecuado análisis de datos estos se deben mirar, entender y pensar sobre ellos, pilares básicos del análisis exploratorio de datos. Pero esta estrategia que si se ha difundido a nivel descriptivo no ha tenido tanto éxito cuando se aplica a la estadística inferencial, que en el mejor de los casos, ante el incumplimiento de las condiciones de aplicación, el investigador utiliza pruebas no paramétricas. Ahora bien, son diferentes los trabajos que defienden el uso de los intervalos de confianza en la decisión estadística conjuntamente con los intervalos de las medidas del tamaño del efecto (Cohen, 1994; Wilkinson y the Task Force on Statistics Inference, 1999; Cumming y Finch, 2001 o Cumming y Finch, 2005). De hecho, Hagen, en 1997, ya apuntaba la idea de que los intervalos de confianza aportan la misma información que el contraste de hipótesis. De todos modos no existe la misma tradición en el uso de indicadores resistentes, a excepción de algunos trabajos que se centran en la obtención de intervalos de confianza alrededor de la mediana (Woodruff, 1952; Sheather y McKean, 1987; Bonett y Price, 2002 o Dubnicka, 2007). Así pues, desde hace algunos años, hemos centrado parte de nuestra investigación en el estudio de la comparación de los intervalos de confianza alrededor de la media y de la mediana en la decisión estadística (Freixa, Turbany y Però, 2005; Guàrdia, Però, Freixa y Turbany, 2007 o Però, Guàrdia, Freixa y Turbany, 2008). Los resultados encontrados muestran una elevada especificidad de la decisión tomada a partir de los intervalos de confianza, ya sean centrados en la media o en la mediana, pero no así una elevada sensibilidad. Es por este motivo, que en el presente trabajo se pretende estudiar el grado de acuerdo en la decisión tomada a partir de las diferentes estrategias de decisión utilizadas en los trabajos anteriormente citados, a saber, estadístico t de *Student* de grupos independientes, prueba no paramétrica U de Mann-Whitney, comparación de intervalos de confianza alrededor de la media, comparación de intervalos de confianza alrededor de la mediana, construidos a partir del error estándar o de la distribución binomial (para un mayor desarrollo de la obtención de estos intervalos de confianza consultar Però, et al. 2008).

## Método

### Procedimiento

Los datos utilizados en el presente trabajo se han generado vía simulación utilizando para ello el programa EXCEL. Los datos se han simulado bajo el supuesto de trabajar con poblaciones que se distribuyen según el modelo de la ley normal en la población, en concreto se han creado cinco situaciones diferentes en función del valor de la media parámetro (100, 100, 110, 120 y 130), lo que ha permitido disponer de 4 situaciones de comparación, una con igualdad poblacional de medias (100 – 100) y tres con diferencia poblacional de medias en diferente grado (100 – 110, 100 – 120 y 100 -130). Además se ha trabajado con diferentes valores de desviación típica (10, 20 y 30) y tamaños de muestra (10, 15, 20, 25 y 30). De todos modos, comentar que en todas las comparaciones realizadas se ha trabajado bajo igualdad de desviaciones típicas y bajo un diseño de muestras equilibradas. Así pues, cada una de las situaciones simuladas implica 15 condiciones (3 niveles de desviación típica x 5 niveles de tamaño de muestra). Para cada una de las 15 condiciones se han generado 5000 parejas de muestras.

Dado que en todas las condiciones generadas se trabajó con muestra pequeña en los dos grupos a comparar, los intervalos de confianza se obtuvieron a partir de la distribución t de *Student*.

## Resultados

En la figura 1 se muestra este porcentaje de acuerdo entre los cinco criterios de decisión en la solución correcta. En la situación de igualdad de medias este porcentaje oscila entre un 80 y un 88% para las 15 condiciones estudiadas si el criterio de decisión es el no estricto, siendo superior este valor si el criterio de decisión es el estricto (alrededor del 93%). Cuando existe diferencia de medias poblacional el acuerdo es más bajo.

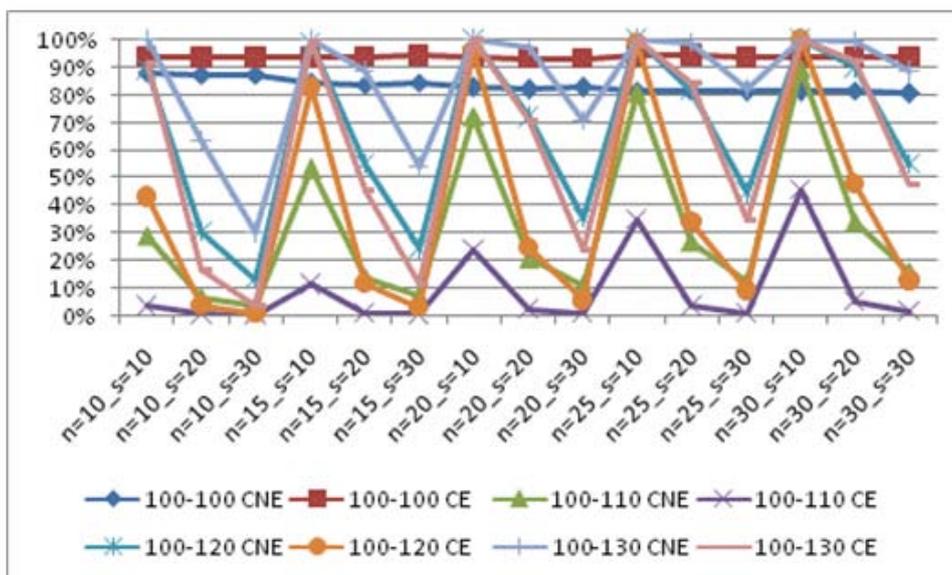


Figura 1. Porcentaje de acuerdo en los cinco criterios según condición (n: tamaño de muestra, s: desviación típica, CNE: criterio de decisión no estricto en la comparación de los intervalos de confianza, CE: criterio de decisión estricto en la comparación de los intervalos de confianza)

Dada la disparidad en el porcentaje de acuerdo entre los cinco criterios en las cuatro situaciones estudiadas se decidió realizar un análisis más pormenorizado de este grado de acuerdo, comparando para ello el acuerdo en la decisión correcta entre el estadístico *t* de *Student* de grupos independientes y la prueba no paramétrica *U* de Mann-Whitney (figura 2), y por otra parte la decisión tomada a partir de la comparación de los intervalos de confianza (media, mediana error estándar y mediana binomial) con la decisión tomada a partir del estadístico *t* de *Student* por una parte y con la prueba no paramétrica *U* de Mann-Whitney por otra parte (figuras 3 a 6). La ausencia de valores en algunas condiciones es debido a que en ninguna de las dos pruebas comparadas se había tomado la decisión de no rechazar la hipótesis nula.

En la figura 2 se puede observar que el grado de acuerdo es elevado en la decisión tomada a partir del estadístico *t* de *Student* de grupos independientes y la prueba no paramétrica *U* de Mann-Whitney en las cuatro situaciones estudiadas.

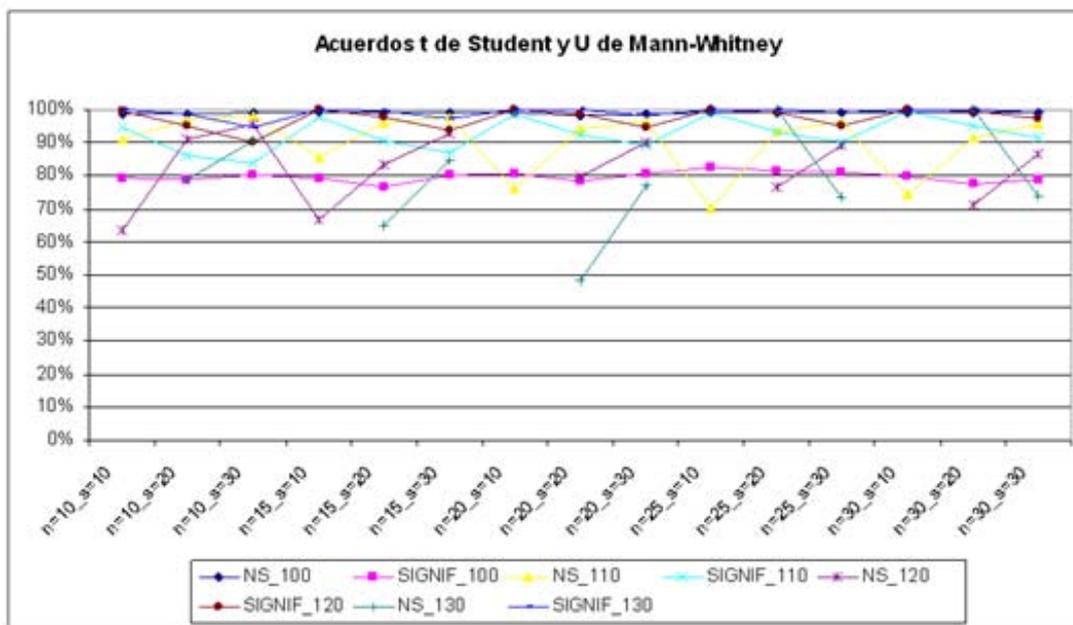


Figura 2. Porcentaje de acuerdo en la decisión para la prueba *t* de *Student* y la prueba *U* de Mann-Whitney para las diferentes situaciones y condiciones estudiadas (*n*: tamaño de muestra, *s*: desviación típica, *NS*: diferencias estadísticamente no significativas y *SIGNIF*: diferencias estadísticamente significativas)

En las figuras 3 a 6 se muestran los porcentajes de acuerdo en la decisión estadística para las cuatro situaciones estudiadas entre la decisión tomada a partir de la comparación de los intervalos (tanto si el criterio de decisión es el no estricto como si es el estricto) y la prueba *t* de *Student* de grupos independientes o la prueba no paramétrica *U* de Mann-Whitney. En la situación de igualdad de medias poblacional el acuerdo es muy alto en el caso de la comparación de los intervalos de confianza a partir del criterio estricto, en tanto que para las tres situaciones de diferencia de medias poblacionales (figuras 4 a 6) el acuerdo es mayor si el criterio de decisión es el no estricto, a pesar de que en determinadas condiciones el acuerdo es muy bajo.

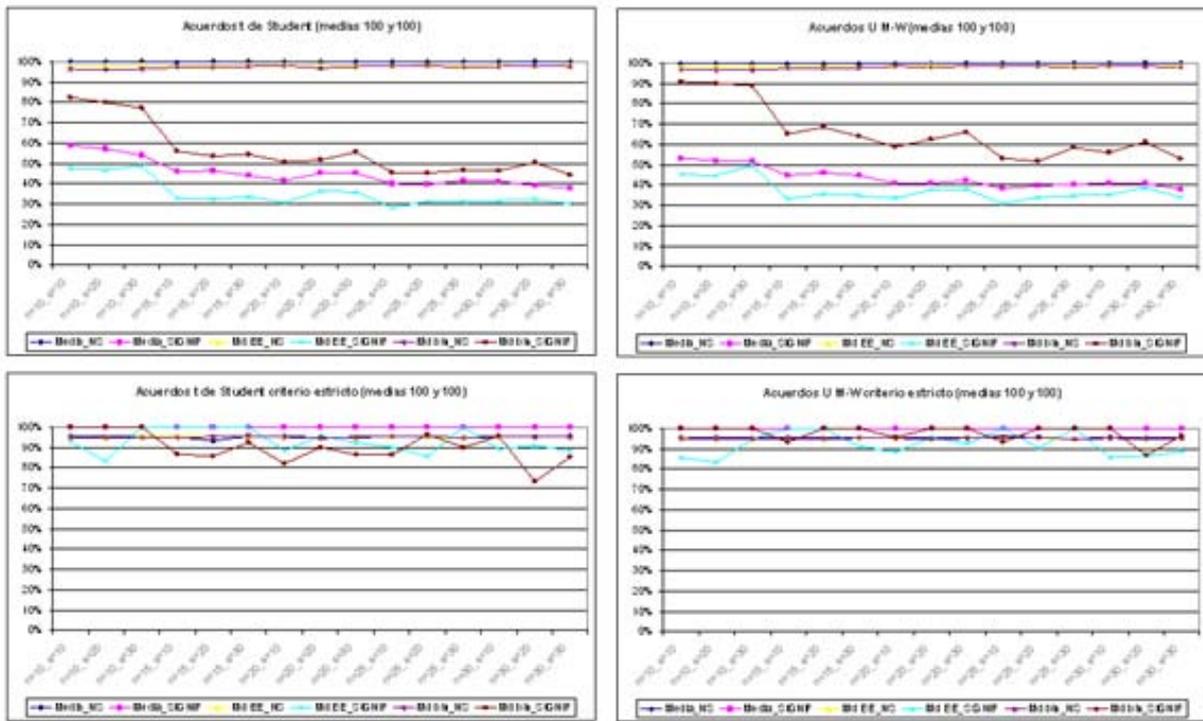


Figura 3. Porcentaje de acuerdo en la decisión para la prueba t de Student de grupos independientes y la prueba no paramétrica U de Mann-Whitney y la comparación de los intervalos de confianza en la situación de no diferencias entre las dos poblaciones (medias de 100) (n: tamaño de muestra, s: desviación típica, NS: diferencias estadísticamente no significativas y SIGNIF: diferencias estadísticamente significativas)

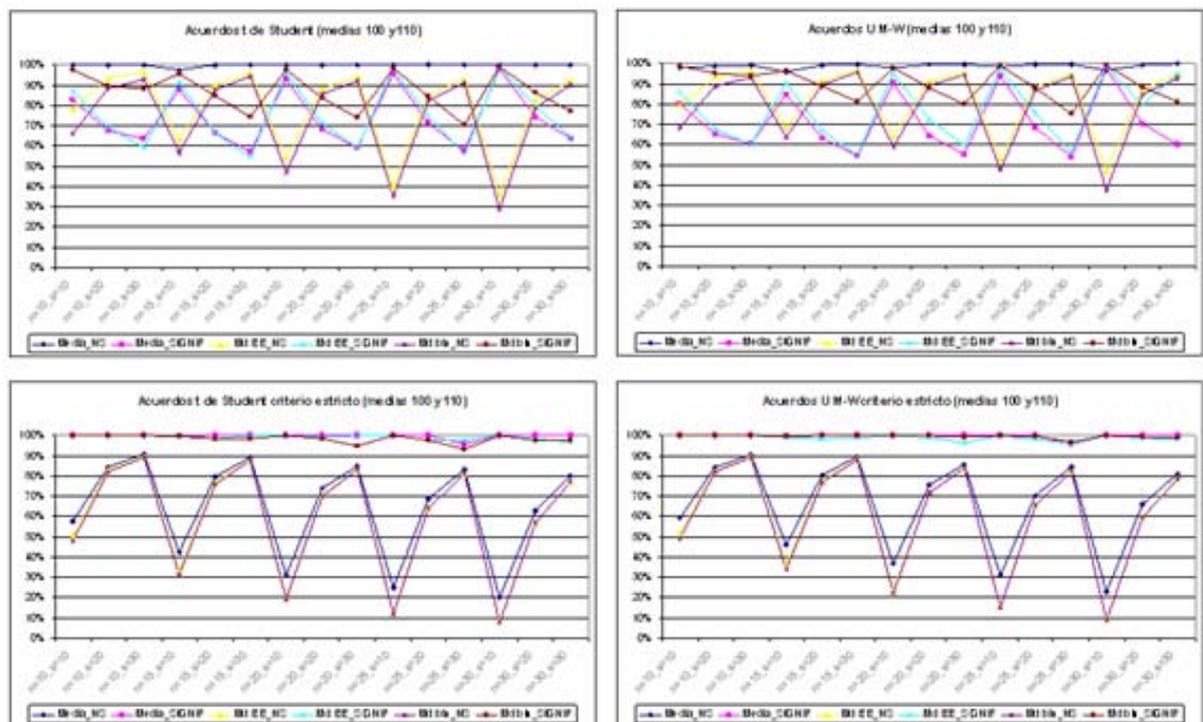


Figura 4. Porcentaje de acuerdo en la decisión para la prueba t de Student de grupos independientes y la prueba no paramétrica U de Mann-Whitney y la comparación de los intervalos de confianza en la situación de diferencias entre las dos poblaciones (medias de 100 y 110) (n: tamaño de muestra, s: desviación típica, NS: diferencias estadísticamente no significativas y SIGNIF: diferencias estadísticamente significativas)

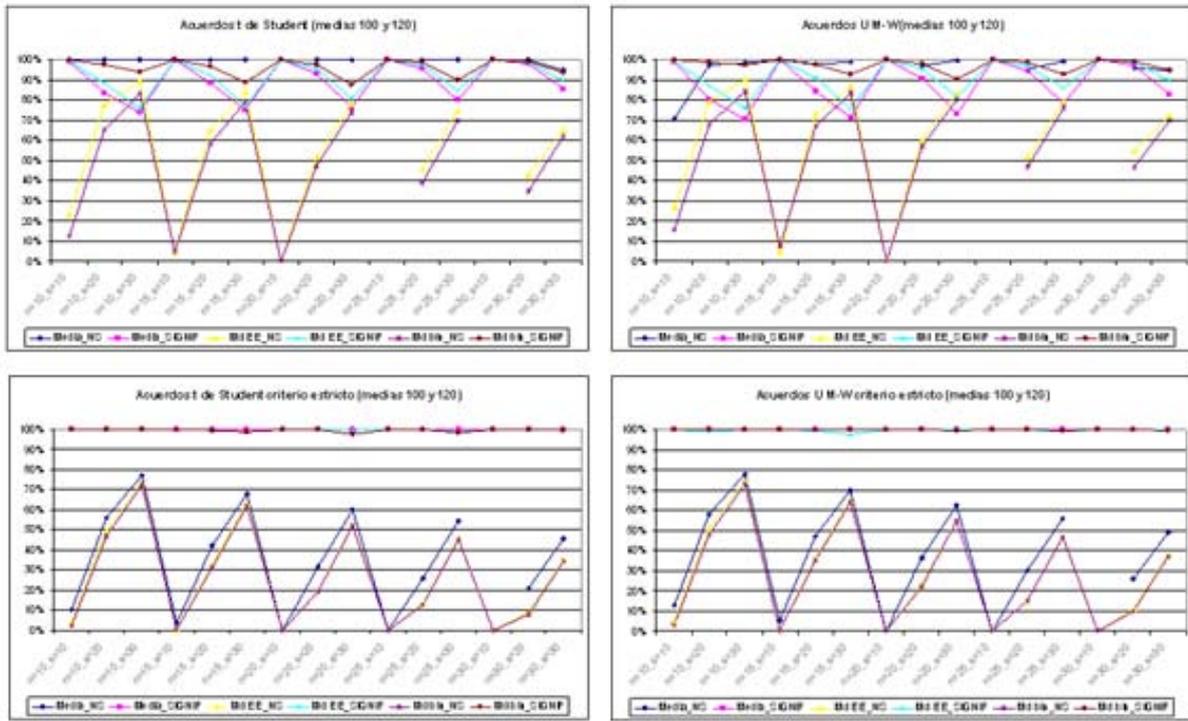


Figura 5. Porcentaje de acuerdo en la decisión para la prueba *t* de Student de grupos independientes y la prueba no paramétrica *U* de Mann-Whitney y la comparación de los intervalos de confianza en la situación de diferencias entre las dos poblaciones (medias de 100 y 120) (*n*: tamaño de muestra, *s*: desviación típica, NS: diferencias estadísticamente no significativas y SIGNIF: diferencias estadísticamente significativas)

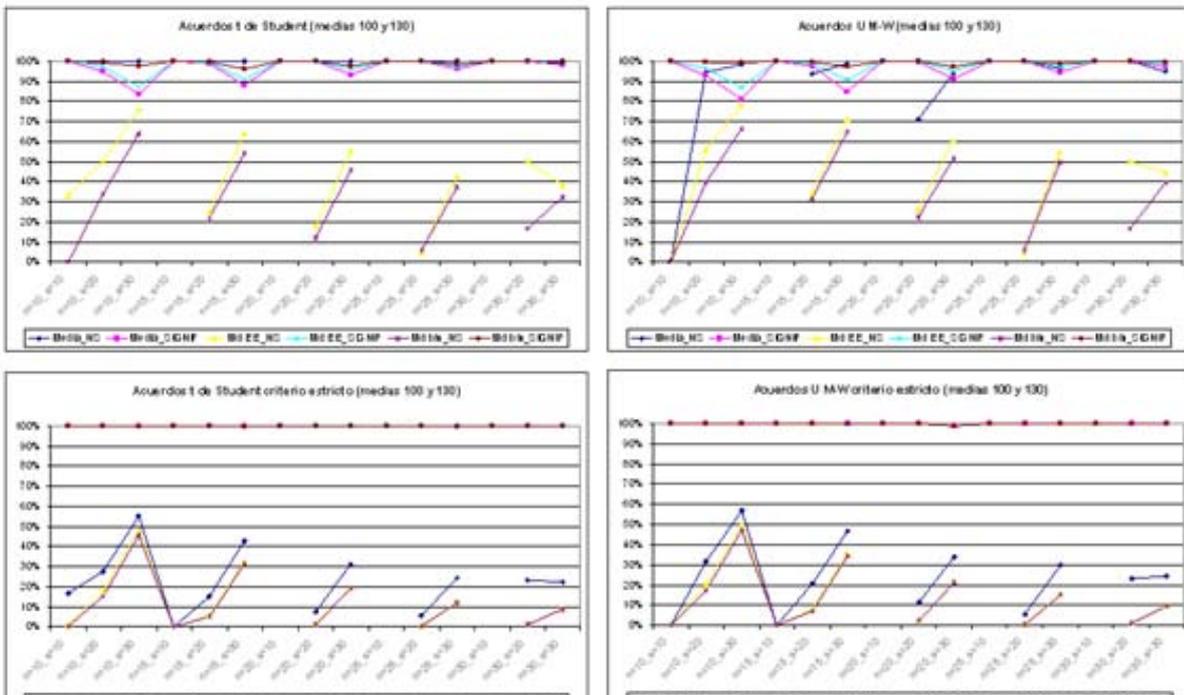


Figura 6. Porcentaje de acuerdo en la decisión para la prueba *t* de Student de grupos independientes y la prueba no paramétrica *U* de Mann-Whitney y la comparación de los intervalos de confianza en la situación de diferencias entre las dos poblaciones (medias de 100 y 130) (*n*: tamaño de muestra, *s*: desviación típica, NS: diferencias estadísticamente no significativas y SIGNIF: diferencias estadísticamente significativas)

## Conclusión

Al estudiar el grado de acuerdo en la decisión correcta entre las cinco pruebas analizadas, se ha mostrado que en la situación de no diferencias entre las dos poblaciones estudiadas cuando se utiliza el criterio estricto en la comparación de los intervalos de confianza el acuerdo está alrededor del 93%, por tanto bastante elevado, y queda reducido a un 80%-88% cuando el criterio de decisión es el no estricto. Una situación diferente se produce en las tres situaciones en las que existen diferencias entre las medias de las dos poblaciones, en estas tres situaciones se ha mostrado que el acuerdo entre las cinco pruebas analizadas es mejor si se ha tomado la decisión a partir del criterio no estricto que si la decisión se ha tomado a partir del criterio estricto. Además, el porcentaje de acuerdo aumenta al aumentar el tamaño de muestra y disminuir la variabilidad y obviamente, este patrón presenta un mayor acuerdo al aumentar la distancia poblacional de las medias.

Dada esta situación de resultados dispares en el acuerdo en la decisión, se optó por estudiar el acuerdo entre la decisión tomada a partir de los intervalos de confianza y la tomada a partir de las pruebas utilizadas clásicamente en la decisión estadística. En este punto cabe remarcar el hecho que el patrón observado para el estadístico *t* de *Student* se reproduce prácticamente igual en el caso de la prueba no paramétrica *U* de Mann-Whitney, lo cual resulta lógico dado que el acuerdo entre estas dos pruebas es muy elevado, aunque es mayor para las decisiones correctas que para las decisiones incorrectas en las cuatro situaciones estudiadas. Así pues, al comparar la decisión a partir de los intervalos de confianza con las pruebas clásicas en la situación de no diferencias de medias en la población, el acuerdo es mayor si se sigue el criterio estricto de decisión que si se sigue el criterio no estricto de decisión y este acuerdo es mayor en la decisión correcta que en los casos en que se toma la decisión incorrecta. Para las tres situaciones en las que a nivel poblacional existen diferencias de medias, en aquellos pares de muestras que se ha tomado la decisión correcta; el acuerdo es mayor si se sigue el criterio estricto en la comparación de los intervalos que si se sigue el criterio no estricto. Ello se invierte si la decisión que se toma es la incorrecta, dado que en este caso el acuerdo es superior si la decisión se ha tomado siguiendo el criterio no estricto que el criterio estricto.

Además, comentar, que si la decisión tomada es la correcta, rechazar la hipótesis nula, el intervalo de confianza de la mediana basado en la ley binomial es el que proporciona mayor acuerdo, en tanto que si la decisión tomada es incorrecta, existen más oscilaciones en el grado de acuerdo a través de las condiciones estudiadas, aumentando a medida que aumenta la desviación típica, aunque el grado de acuerdo en la decisión incorrecta ha disminuido al aumentar la diferencia poblacional entre las medias. Una posible explicación de ello podría estar en el hecho de que las decisiones incorrectas eran menores que cuando la diferencia poblacional de las medias era menor. Estos resultados no confirmarían del todo la afirmación de Hagen (1997) en el sentido que los intervalos de confianza aportan la misma información que el contraste de hipótesis.

## Referencias

- Bonett, D.G. y Price, R.M. (2002). Statistical inference for a linear function of medians: confidence intervals, hypothesis testing, and sample size requirements. *Psychological Methods*, 7 (3), 370-383.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49 (12), 997-1003.
- Cumming, G. y Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61 (4), 532-574.

- Cumming, G. y Finch, S. (2005). Inference by eye. Confidence intervals and how to read pictures of data. *American Psychologist*, 60 (2), 170-180.
- Dubnicka, S.R. (2007). A confidence interval for the median of a finite population under unequal probability sampling: a model-assisted approach. *Journal of Statistical Planning and Inference*, 137 (7), 2429-2438.
- Freixa, M., Turbany, J. y Però, M. (2005). El Uso y el abuso de la distancia con respecto a la media aritmética como la mejor medida de sensibilidad en la investigación aplicada en psicología. Comunicación presentada en la X Conferencia Española de Biometría, Oviedo (Spain).
- Guàrdia, J., Però, M., Freixa, M. y Turbany, J. (2007). Generación mediante simulación de intervalos de confianza en torno a estadísticos resistentes: el caso de la mediana. Comunicación presentada en el X Congreso de Metodología de las Ciencias Sociales y de la Salud, Barcelona (Spain).
- Hagen, R.L. (1997). In praise of the null hypothesis statistical test. *American Psychologist*, 52 (1), 15-24.
- Peró, M., Guàrdia, J., Freixa, M. y Turbany, J. (2008). Técnicas basadas en la mediana como alternativa a las pruebas clásicas de inferencia estadística. *Psicothema*, 20 (4), 857-862.
- Sheater, S.J. y McKean, J.W. (1987). A comparison fo testing and confidence interval methods for the median. *Statistics & Probabilitiy Letters*, 6 (1), 31-36.
- Tukey, J.W. (1977). *Exploratory data analysis*. Reading, Massachussets: Addison-Wesley.
- Wilkinson, L. y the Task Force on Statistical Inference (1999). Statistical methods in psychology journals. Guidelines and explanations. *American Psychologist*, 54 (8), 594-604.
- Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, 47, 635-646.

# LA FASE DE ESTIMACIÓN DE PARÁMETROS EN LOS MODELOS DE ECUACIONES ESTRUCTURALES CON VARIABLES CATEGÓRICAS

Sonia Benítez<sup>1</sup>, Joan Guàrdia<sup>1,2</sup>, Mila Arch<sup>1</sup> y Adolfo Jarne<sup>1</sup>

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Instituto de Investigación en Cognición, Cerebro y Conducta (IR3C)

En los Modelos Estadísticos Multivariantes, una de las situaciones que es frecuente se relaciona con el uso de variables categóricas y el efecto que ello tiene en la fase de estimación. Ello justifica sobradamente la gran cantidad de bibliografía que, en la última década, se ha generado sobre el uso de este tipo de variables en la estimación de parámetros y posterior ajuste de modelos estadísticos. Los Modelos de Ecuaciones Estructurales (SEM) no son una excepción, aunque en este caso parece que existe cierta evidencia de cómo superar la limitación que implica la utilización de variables categóricas, centrando su foco de atención en modificaciones importantes de algoritmos ampliamente conocidos como las estimaciones elípticas o el uso de estimadores robustos derivados de la familia de los mínimos cuadrados (LS). En el presente trabajo, se pretende presentar algunas opciones y propuestas para tratar de solventar esas limitaciones, siguiendo en la línea de lo que otros autores ya han propuesto (Lee & Poon, 1987; Poon, Wang & Lee, 1999; Flora & Curran, 2004; Zhang & Browne, 2006), de modificar el algoritmo en la fase de estimación para ajustarse a las características de los datos originales o adaptar la correlación o covarianza de partida para adaptarla a las propiedades del algoritmo de estimación, sea éste de base lineal o no lineal.

En un estudio previo realizado (Benítez, Guàrdia, Arch y Jarne, 2008) ya se abordaron algunos de los incumplimientos de las condiciones de aplicación durante el proceso de estimación de parámetros, ya sea el uso de variables categóricas, trabajar con muestras pequeñas, la identificación de procesos complejos (relaciones recíprocas entre variables) o el ajuste del modelo en el caso de analizar efectos tanto recursivos como no recursivos. El objetivo de dicho estudio fue presentar los efectos del uso de variables categóricas y muestras pequeñas en la estimación de parámetros en Modelos de Ecuaciones Estructurales, con especificación de los efectos recíprocos.

Se realizó a partir de datos propios de la Psicología Forense, concretamente del ámbito de familia, puesto que en procedimientos de divorcio es habitual que las condiciones de uno u otro cónyuge ( $\eta_1$ ) determinen el tipo de recomendación forense respecto a qué progenitor otorgar la custodia exclusivo o sobre la viabilidad de la custodia compartida ( $Y_2$ ) así como la recomendación de custodia exclusiva o custodia compartida determina la importancia que se le asigna a determinadas situaciones de los cónyuges (Arch, 2008); siendo el modelo propuesto el que aparece en la Figura 1.

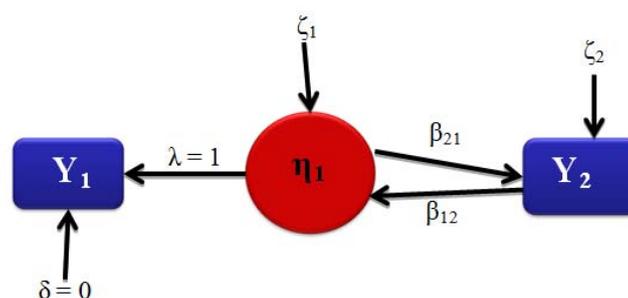


Figura 1. Path diagrama del modelo propuesto en el estudio previo realizado en 2008

Los resultados reflejaron que el uso de coeficientes de correlación no adecuados al tipo de escala, como es el caso de los basados en la linealidad directa no son eficientes para la estimación de parámetros cuando trabajamos con variables categóricas y tamaño de muestra pequeño. Como en trabajos parecidos los resultados mostraron que, en ese caso, es preferible el uso de correlaciones policóricas.

A partir de los resultados de ese estudio previo, planteamos analizar de nuevo esos datos especificando un modelo más complejo y generando modelos de medida latentes con el objetivo de analizar la importancia de los efectos recíprocos entre variables latentes (y no entre una variable latente y una variable observada como en el estudio previo) manteniendo las condiciones de medir variables categóricas y muestras pequeñas. Para ello, planteamos un estudio con los mismos datos surgidos de la Psicología Forense en el ámbito de familia y ya utilizados para el estudio previo. Se deseaba seguir el análisis de los criterios que utilizan los psicólogos forenses en la asignación de la guarda y custodia de menores.

Para el análisis de los efectos estadísticos, planteamos también la realización de un estudio de simulación con los siguientes objetivos: a) para tratar de analizar la distribución observada de cada índice de ajuste de los datos al modelo complejo propuesto; b) para obtener mejores estimaciones de los cuatro parámetros y, por último, c) evaluar si los procedimientos de simulación son una alternativa para solventar los problemas de trabajar con tamaños de muestra pequeños.

## Método

### *Participantes*

La muestra estuvo formada por un total de 66 psicólogos forenses que ejercen en España (Arch, 2008), expertos en evaluación de custodia. Durante el proceso de selección se contactó con psicólogos forenses: a) inscritos en las listas de los Turnos de Intervención Profesional de peritos en las diferentes sedes del Colegio Oficial de Psicólogos; b) ejercieran en nuestro país desde los diferentes servicios de asesoramiento técnico dependientes del Ministerio de Justicia y c) inscritos en bases de datos públicas en Internet que ofrecen sus servicios para este tipo de evaluaciones. Además, como criterios de inclusión se consideraron: a) que hubieran realizado un mínimo de cuatro evaluaciones de custodia y b) que dispusieran de un mínimo de dos años de experiencia en el ámbito de la psicología clínica.

Del total de la muestra, el 76.9% son mujeres y un 23.1% son hombres; siendo la media de edad de 40.42 años (con un rango de edad de 27 a 64 años). La mitad de ellos, el 51.5%, ejercían en Cataluña mientras que el 48.5% restante lo hacían en el resto de Comunidades Autónomas españolas tal como refleja la Figura 2.

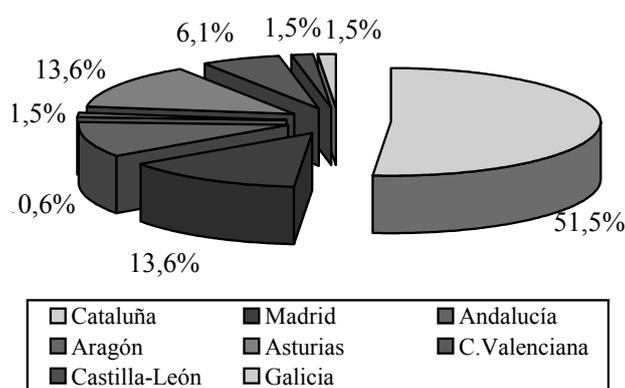


Figura 2. Distribución de la muestra en función de la variable Comunidad Autónoma en la que ejercían

Si tenemos en cuenta el ámbito en el que ejercían, el 78.8% de los psicólogos forenses lo hacían en la práctica privada y únicamente el 9.1% lo hacía en el servicio público, mientras que el 12.1% lo hacía en los dos ámbitos. Finalmente, la figura 3 muestra la distribución de la muestra en función de variables relacionadas con el ejercicio profesional.

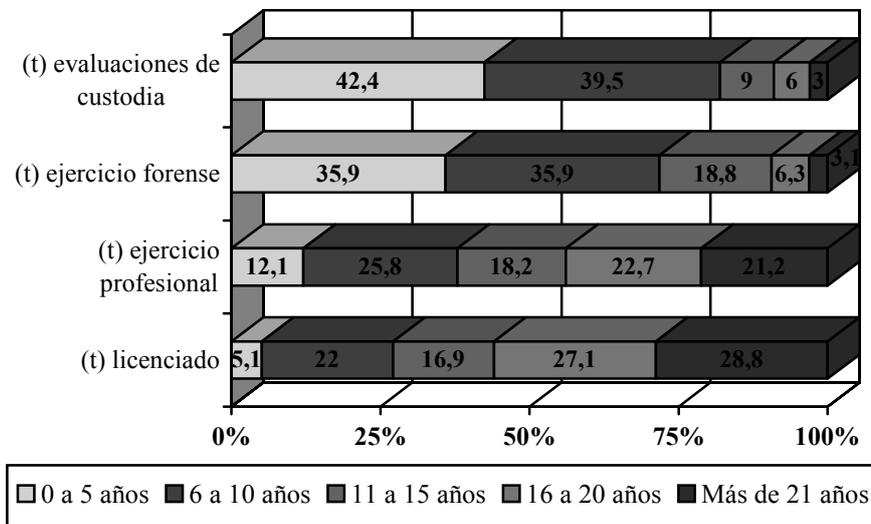


Figura 3. Distribución de la muestra en función de variables relacionadas con el ejercicio profesional

La media de años que llevan licenciados fue de 16.63 (s = 2.24) aunque más de la mitad de los participantes llevaban licenciados entre 16 y 20 años o más de 21 años. La media de años de ejercicio profesional se situó en 14.27 (s = 3.22), destacando que más de la mitad de ellos llevan en el ejercicio forense y, concretamente en la evaluación de custodia, más de 6 años.

*Instrumentos*

El instrumento utilizado fue el cuestionario *ad-hoc* elaborado por Arch (2008). Consta de 93 ítems que exploran las experiencias, actividades y creencias de los profesionales en 3 áreas: 1) (11 ítems) explora datos demográficos (personales y profesionales) de los participantes que forman parte de la muestra; 2) en la segunda (10 ítems), se solicita información relacionada con una variedad de aspectos, técnicas y procedimientos de evaluación de custodia utilizados por los profesionales y 3) (72 ítems) que exploran los criterios en una doble vía, 42 ítems exploran criterios relacionados con la toma de decisiones en recomendaciones de custodia exclusiva y 30 ítems exploran criterios relacionados con la toma de decisiones en las recomendaciones de custodia compartida. El análisis que aquí se presenta solo utilizó los ítems de esta última parte de los cuales se muestra alguno en la tabla 1.

Tabla 1. Muestra de ítems de la escala usada en esta investigación

	¿Cómo de importante?									Custodia a:		
	1	2	3	4	5	6	7	8	9	A	B	Indistintamente
El progenitor B es alcohólico activo.												
El progenitor A muestra mejores habilidades parentales que el progenitor B.												

	¿Cómo de importante?									Custodia compartida:		
	1	2	3	4	5	6	7	8	9	Viable	Inviabile	Indeterminado
Ambos progenitores presentan estabilidad emocional												

Los participantes debían dar una respuesta mediante una escala tipo Likert de 1 a 9 respecto a la importancia que otorgaran a cada uno de los ítems durante el proceso de decisión. Además, también debían indicar, en base a ese único criterio, si recomendarían la custodia a un progenitor determinado (A o B) en el caso de la custodia exclusiva, o indicar la viabilidad o inviabilidad en el caso de la custodia compartida.

A partir de las respuestas obtenidas por los 66 psicólogos forenses se planteó el modelo estructural que aparece en la Figura 4 de acuerdo con los propuesto teóricamente por Arch (2008):

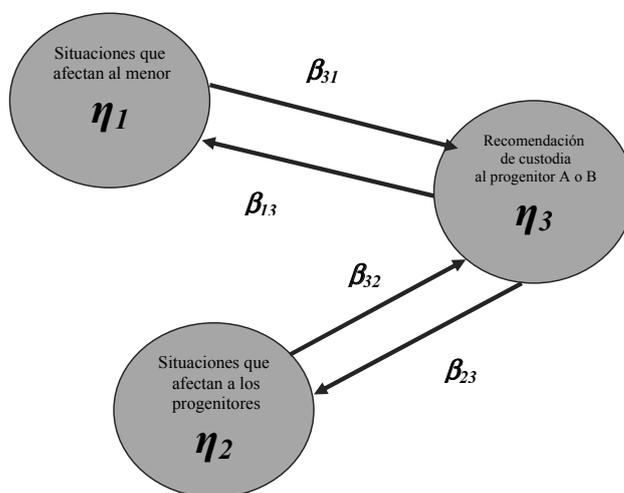


Figura 4. Path diagrama del modelo propuesto en el presente estudio

Hipotetizamos que situaciones concretas tanto de los menores ( $\eta_1$ ) como de los progenitores ( $\eta_2$ ) y la importancia que le otorgan los psicólogos forenses a las mismas ( $\eta_3$ ) determinan la decisión de recomendar custodia exclusiva o custodia compartida, y esa decisión define al mismo tiempo la importancia que le otorgan a las situaciones de los menores y progenitores; por tanto, que existe una relación recíproca entre ambas variables latentes.

*Procedimiento y Análisis de Datos*

Con los datos obtenidos se calcularon los valores de ajuste global del modelo propuesto y de cada parámetro estructural. Para ello, se estimó la matriz de correlaciones policóricas y, como técnica de estimación de parámetros, se utilizó el método de mínimos cuadrados ordinarios (OLS) corregido para variables categóricas. Para la realización del estudio de simulación posterior, se generaron 10.000 muestras mediante el procedimiento de remuestreo (*leave-one-out*), y se utilizó los componentes de la matriz de correlaciones policóricas inicial como parámetros poblaciones. Con estos datos para cada muestra, se calcularon los valores de ajuste global del modelo y la estimación de cada parámetro estructural.

El análisis de los datos se realizó con el software EQS versión 6.1 para Windows.

## Resultados

Los resultados obtenidos del análisis del ajuste de los datos aplicados así como de los datos obtenidos por simulación al modelo estructural propuesto anteriormente se muestran a continuación.

Tabla 1. Índices de bondad de ajuste de la estimación de parámetros global

Índices de ajuste global	Valor
Jhi cuadrado	$\chi^2 = 47.78$ g.l. = 12 $p < .001$
Goodness of Fit Index (GFI)	.712
Adjusted Goodness of Fit Index (AGFI)	.692
Bentler-Bonnet Normat Fit Index (BBNFI)	.699
Root Mean Square Error of Approximation (RSMSE)	.082

*GFI*: Índice de bondad de ajuste. *BBNFI*: Índice de ajuste normalizado de Bentler-Bonnet. *AGFI*: Índice de bondad de ajuste corregido. *RMSEA*: Residual de la media cuadrática de aproximación.

A partir de la tabla 1 se observa que el estadístico  $\chi^2$  no indica un buen ajuste de los datos para el modelo propuesto puesto que resulta estadísticamente significativo, aunque cabe mencionar que se trata de un índice muy sensible al tamaño de muestra (Bentler y Bonnet, 1980) y que el criterio de la razón de grados de libertad ( $\chi^2/g.l. = 3.98$ ) se muestra un valor compatible con ajustes aceptables (entre 2 y 5). Los índices GFI y AGFI son insensibles al tamaño de muestra, y en este caso, ambos índices de ajuste no indican un buen ajuste del modelo porque sus valores no tienden a 1; lo mismo sucede con el índice BBNFI. Por último, el estadístico RMSEA como estimador del error de aproximación relativo del modelo tampoco confirmó el ajuste del modelo ya que su valor fue excesivamente alto ( $> 0.04$ ).

Dado que los índices globales del modelo no indicaron un buen ajuste del modelo, creímos adecuado valorar la significación de la estimación de cada parámetro (tabla 2).

Tabla 2. Significación de la estimación de los parámetros

Parámetro	Estimación	Error estándar	Significación
$\beta_{13}$	3.124	2.789	.458
$\beta_{31}$	6.143	5.551	.845
$\beta_{23}$	12.134	3.126	< .001
$\beta_{32}$	15.721	3.945	< .001

En la tabla 2 queda reflejado que únicamente ha resultado significativa la estimación de los parámetros  $\beta_{23}$  y  $\beta_{32}$ , lo que parece indicar que solo una de las variables latentes ( $\eta_2$ ), la que está relacionada con situaciones que afectan a los progenitores, tiene más efecto en el modelo propuesto. En el estudio de simulación realizado posteriormente, los resultados del ajuste del modelo fueron más favorables.

Tabla 3. Distribución observada de los índices de bondad de ajuste en el estudio de simulación

Índice de ajuste	$\bar{X}$	SD	IC	p
GFI	.799	.061	.771 - .827	
AGFI	.732	.055	.719 - .745	
BBNFI	.782	.041	.778 - .786	
RSMSE	.042	.002	.039 - .045	
B <sub>13</sub>	7.184	1.01	7.11 – 7.25	< .001
B <sub>31</sub>	8.124	1.45	8.02 – 8.22	< .001
B <sub>23</sub>	9.743	.99	9.65 – 9.83	< .001
B <sub>32</sub>	9.996	1.27	9.85 – 10.13	< .001
Ajuste $\chi^2$	63% modelos ajustados			

$\bar{X}$ : Media de la estimación. *SD*: Desviación estándar observada. *IC*: Intervalo de confianza. *p*: Significación. *GFI*: Índice de bondad de ajuste. *BBNFI*: Índice de ajuste normalizado de Bentler-Bonnet. *AGFI*: Índice de bondad de ajuste corregido. *RMSEA*: Residual de la media cuadrática de aproximación.

En la tabla 3 se puede observar que los índices globales de ajuste (GFI, AGFI, BBNFI y RSMSE) no indican un buen ajuste del modelo con datos simulados. Por otro lado, en este caso sí que han resultado significativas las estimaciones de todos los parámetros, lo que estaría indicando que en el estudio de simulación ambas variables latentes ( $\eta_1$  y  $\eta_2$ ) tienen efecto en el modelo propuesto. Por último, cabe comentar que el 63% de modelos ajustados en el estudio de simulación realizado puede interpretarse como un mejor buen ajuste global del modelo que el obtenido con los datos originales.

### Discusión

Después del análisis de los resultados obtenidos, podemos establecer las siguientes conclusiones:

- Con el estudio de simulación no se mejoran los resultados de los índices de ajuste global obtenidos previamente en el estudio aplicado.
- Con los datos del estudio aplicado solo han resultado significativas las estimaciones de dos de los cuatro parámetros, los relacionados con una de las variables latentes (la relacionada con situaciones de los progenitores); mientras que los datos simulados permiten significar las estimaciones de todos los parámetros. Por tanto, a través de los datos simulados queda reflejado el efecto de las dos variables latentes en el modelo propuesto (mientras que el estudio aplicado sólo refleja el efecto de una de las variables).
- El estudio de simulación permite incrementar el porcentaje de ajuste del modelo (hasta un 63%), lo que indica un mejor ajuste global del modelo.
- En este sentido, parece razonable pensar que, desde el punto de vista forense, los resultados obtenidos a partir del estudio de simulación reflejan que tanto las situaciones que afectan a los progenitores como las situaciones que afectan a los menores afectan y al mismo tiempo determinan las recomendaciones de los psicólogos forenses en el sentido de valorar la viabilidad de la custodia compartida y de recomendar la custodia exclusiva a un progenitor o a otro.

Después de todo el análisis realizado hasta el momento, parece que los estudios de simulación pueden ser una buena alternativa para analizar el ajuste de modelos estructurales cuando trabajamos con variables categóricas y tamaños de muestra pequeños, dada la complejidad de realizar estos análisis con datos aplicados.

## Referencias

- Arch, M. (2008). *La intervención de los psicólogos forenses en las evaluaciones periciales de guarda y custodia de los niños*. Tesis Doctoral. Barcelona: U.B.
- Benítez, S., Guàrdia, J., Arch, M. y Jarne, A. (2009). La fase de estimación de parámetros en los modelos de ecuaciones estructurales con variables categóricas. Comunicación presentada al *XI Congreso de Metodología de las Ciencias Sociales y de la Salud*, celebrado en Málaga del 16 al 18 de septiembre de 2009.
- Bentler, P.M. y Bonnet, D.G. (1980). Significance test and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Flora, D.B. y Curran, P.J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods*, 9 (4), pp. 466-491.
- Lee, S.Y. y Poon, W.Y. (1987). Maximum likelihood estimation of multiple correlations and canonical correlations with categorical data. *Applied Psychological Measurement*, 11 (3), pp. 317-323.
- Poon, W.Y., Wang, S.J. y Lee, S.Y. (1999). Influence analysis of structural equation models with polytomous variables. *Psychometrika*, 64 (4), pp. 461-473.
- Zhang, G. y Browne, M.W. (2006). Bootstrap fit testing, confidence intervals, and standard error estimation in the factor analysis of polychoric correlation matrices. *Behaviormetrika*, 33 (1), pp. 61-74.

# EFFECTO DE LOS *MISSING DATA* EN UN DISEÑO MULTIVARIANTE DE GRUPOS INDEPENDIENTES CON MUESTRAS PEQUEÑAS

María Jesús Carrera<sup>1</sup>, Guillem Feixas<sup>1</sup> y Joan Guàrdia<sup>1,2</sup>

<sup>1</sup>Universidad de Barcelona

<sup>2</sup>Instituto de Investigación en Cognición, Cerebro y Conducta (IR3C)

Los datos perdidos o no registrados (*missing data*) en diseños multivariantes tienen diversos efectos. Por una parte la pérdida de validez debida al riesgo de que los sujetos con datos completos tengan características distintas en relación con los que presentan datos incompletos y por otra, implica un aumento en la probabilidad de cometer un error de tipo I puesto que los criterios de decisión estadística pueden variar. De ahí la importancia de asegurar un buen manejo de los mismos. La asignación de valores simulados a los *missings* está en función del mecanismo de generación. Al trabajar con grupos independientes debe, además, considerarse si la técnica de asignación elegida funcionará igual ante un distinto número de *missings*. Si la respuesta fuera negativa podríamos requerir distintas técnicas de asignación. En muestras grandes el método de máxima verosimilitud (ML) parece ser el idóneo, pero es necesario comprobar si su efecto es el mismo con muestras pequeñas o si al efecto de la técnica de asignación debemos añadir el efecto del tamaño de la muestra. Así pues, en este trabajo se compararon distintas técnicas de asignación de valores simulados a los *missing data*, encontrándose que los métodos de máxima verosimilitud resultan los más adecuados cuando se aplican a muestras pequeñas en diseños multivariantes.

En ocasiones parece que la mejor solución para manipular los *missing data* es que no existan. Al menos esta suele ser una de las metas de los investigadores al realizar sus estudios: la gran mayoría actúa simplemente como si no existieran *missings data* en sus investigaciones. Sin embargo, sabemos que los *missings* son como una pandemia que es imposible evitar totalmente, lo que podemos hacer es tomar las medidas preventivas necesarias y, en caso de contagio, emplear todos los recursos disponibles.

Desde la década de los años 70 se han hecho notables contribuciones al manejo de los *missing data*. Rubin (1976) realizó una temprana aportación al conceptualizar los tres procesos generales por los que se producen *missing data*, estos son los mecanismos MCAR (*Missing Completely At Random*), MAR (*Missing At Random*) y MNAR (*Missing Not At Random*). Cuando se está ante un mecanismo MCAR los *missings* son ocasionados por un proceso aleatorio y no están relacionados con otras variables observadas o no, por lo tanto las estimaciones de los parámetros de la población se conservan insesgadas. Los *missings* MAR dependen únicamente de los valores de otras variables observadas que son medidas en todos los participantes. Si se estiman los parámetros de la población basándose sólo en los casos con datos completos estos serán típicamente sesgados. Finalmente, los MNAR dependen de la puntuación de un individuo en la variable de interés, sea medida o no. Nuevamente las estimaciones de parámetros poblacionales serán sesgadas. (West, 2001). En otras palabras los MNAR dependerían de los datos no observados (Schafer & Graham, 2002).

Algunos autores como Graham (2009) consideran que en 1987 comenzó una revolución en el manejo de los *missings* con la publicación de dos libros: *Statistical Analysis with Missing Data* de Little & Rubin (1987) y *Multiple Imputation for Nonresponse in Surveys* de Rubin (1987). Estas publicaciones y algunos otros artículos, aunados al avance tecnológico, significaron un punto y aparte en el tratamiento del *missing*.

McKnight, McKnight, Sidani & Figueredo (2007) analizaron los tipos de soluciones estadísticas para los *missing data*, señalando que la mejor solución es en primer lugar planearlos y prevenirlos, pero que cuando esto falla deben emplearse procedimientos que sean valorados según sus fortalezas y debilidades en términos de los objetivos y diseño de cada estudio en particular. Estos autores clasificaron las diversas opciones como: *data deletion methods* (*listwise deletion*, *pairwise deletion*, *available item analysis*, *individual growth curve analysis* y *multisample analysis*); *data augmentation procedures* (*maximum likelihood*, *expectation-maximization*, *Markov chain Monte Carlo*, *weighting* y *dummy variable*); *single imputation procedures* (*mean substitution*, *ML estimated mean substitution*, *median substitution*, *zero imputation*, *hot deck imputation*, *cold deck imputation*, *conditional mean imputation*, *last value carried forward*, *next value carried backward*, *mean previous observations*, *last and next average*, *regression*, *regression with error*) y *multiple imputation (MI process)*.

Por su parte Graham (2009) realizó un interesante resumen sobre las antiguas aproximaciones a los *missing data* en comparación con los métodos modernos. Los criterios que considera para evaluar los distintos métodos son:

1. El método debe producir estimaciones insesgadas en un amplio rango de parámetros.
2. Debería incluir un procedimiento para evaluar el grado de incertidumbre de las estimaciones de parámetros, esto es, estimaciones razonables del error estándar y de los intervalos de confianza.
3. Una vez han sido tratados los sesgos y los errores estándar, el método debe tener un buen poder estadístico.

Graham considera dentro de los métodos antiguos los siguientes: *complete cases analysis* (AKA, *listwise deletion*), *pairwise deletion*, *mean substitution*, *dummy variable* y *regression based single imputation*; mientras que en los modernos sitúa a: *EM algorithm*, *multiple imputation under the normal model (MI)* y *full information maximum likelihood methods (FIML)*.

En general, ninguno de los métodos antiguos es recomendable. La sugerencia de emplear alguno de los métodos modernos está en función del objetivo de la investigación y las características del diseño. Por ejemplo, el algoritmo EM no es adecuado para emplearse en pruebas de hipótesis, entre otras cosas porque aunque provee buenas estimaciones de parámetros no proporciona el error estándar como parte automática del proceso; pero la matriz de covarianzas EM es una excelente base para un análisis factorial exploratorio (Graham, 2009). Para pruebas de hipótesis deben emplearse procedimientos MI o FIML. Los métodos modernos funcionan bien y producen resultados insesgados incluso cuando los datos son MAR (West, 2001).

Una condición a tomar en cuenta es cuando se trabaja con muestras pequeñas. La dificultad en muestras pequeñas no es el *missing per se*, sino que no hay muchos datos con los cuales trabajar sobre esos *missings*. MI funciona bien en muestras pequeñas (N=50), incluso con modelos de regresión múltiple muy largos (18 predictores) y con el 50% de *missings* (Graham & Schafer, 1999). MI también ha demostrado en simulaciones que funciona bien con datos no normales. McKnight, et al. (2007) advierten que aunque MI generalmente ofrece grandes ventajas sobre otros métodos, hay situaciones donde puede ser problemático. MI puede proveer estimaciones robustas de parámetros, pero requiere tamaños de muestra que no siempre están disponibles en la investigación en ciencias sociales (Schafer & Graham, 2002).

El objetivo que se planteó en este trabajo fue comparar distintas técnicas de asignación de valores simulados a los *missing data* con la finalidad de constatar cuál es el procedimiento que funciona mejor cuando se aplica a muestras pequeñas en diseños multivariantes.

## Método

### *Participantes*

Se seleccionaron dos muestras, una procedente de la población de estudiantes universitarios y otra a partir de la población general. Del total de personas evaluadas, el 53.7% forma parte del grupo denominado “estudiantes” (58 personas), el 43.5% pertenecen al grupo llamado “muestra comunitaria” (47) y el 2.8% restante (3) fueron descartadas por tener una edad muy dispar al resto, quedando un total de 105 sujetos clasificados (55.2% en el grupo de estudiantes y 44.8% en la muestra comunitaria). La distribución por género resultante fue de 74.28% (78) mujeres y 25.71% (27) hombres.

La muestra de estudiantes estuvo integrada por un 86.20% (50) de mujeres y 13.79% (8) de hombres, todos ellos estudiantes universitarios de licenciatura y de posgrado del área de psicología, con edades que oscilaron entre 18 y 45 años, una media de edad de 25.17 años ( $sd= 5.299$ ; intervalo de confianza al 95% entre 23.78 y 26.57). Los años de escolarización se ubicaron entre 13 y 17, con una media de 15.67 ( $sd= 1.161$ ; intervalo de confianza al 95% entre 15.37 y 15.98).

Por otra parte, la muestra comunitaria se conformó por un 59.57% (28) de mujeres y 40.42% (19) de hombres, con edades entre 21 y 53 años, media de edad de 30.70 ( $sd=7.363$ ; intervalo de confianza al 95% entre 28.54 y 32.86). Este grupo abarcó entre 10 y 18 años de escolarización, con media de 15.68 ( $sd=2.033$ ; intervalo de confianza al 95% entre 15.08 y 16.28).

Finalmente, señalar que la técnica de muestreo en ambas muestras fue de carácter no probabilístico y accidental, siguiendo para la muestra comunitaria la técnica de bola de nieve.

### *Instrumentos y medidas*

Los instrumentos utilizados en este estudio fueron los siguientes:

- *Escala de Autoestima de Rosenberg (RSE)* (Rosenberg, 1965, adaptación española de Martín-Albó, Nuñez, Navarro & Grijalvo, 2007). Escala autoadministrada de 10 ítems que da como resultado una medida global de la autoestima de la persona, entendida como los sentimientos de valía personal y de respeto a sí mismo.
- *NEO Five Factor Inventory (NEO-FFI)* (Costa & McCrae, 1985, adaptación española de Cordero, Pamos y Seisdedos, 1999). Versión reducida de NEO-PI-R, que recoge los 60 mejores ítems del cuestionario original y permite evaluar en menor tiempo los cinco grandes factores de personalidad: Neuroticismo (Neo-N), Extraversión (Neo-E), Apertura (Neo-O), Amabilidad (Neo-A) y Responsabilidad (Neo-C).
- *Beck Depression Inventory (BDI-II)* (Beck, A. T., Steer, R. A., & Brown, G. K., 1996, adaptación española de Sanz, J., Vázquez, C. & Navarro, M. E., 2003). El BDI-II contiene 21 ítems, que se evalúan en una escala del 0 al 3, indicando las puntuaciones más altas mayores síntomas depresivos.
- *Cuestionario de 90 Síntomas Revisado (SCL-90-R)* (Derogatis, L. R., 1983, adaptación española de González de Rivera, J. L., De las Cuevas, C., Rodríguez, M. & Rodríguez, F., 2002). Cuestionario autoaplicado para la detección y medida de nueve dimensiones sintomáticas de psicopatología y tres índices globales de malestar. Las dimensiones sintomáticas son Somatización, Obsesión-compulsión, Sensibilidad interpersonal, Depresión, Ansiedad, Hostilidad, Ansiedad fóbica, Ideación paranoide, Psicoticismo y Síntomas misceláneos (ítems adicionales). Los índices globales de malestar son el Índice Sintomático General (GSI, *Global Severity Index*), el Total de Síntomas Positivos (PST, *Positive Symptom Total*) y el Índice de Distrés de Síntomas Positivos (PSDI, *Positive Symptom Distress Index*).

Todas las pruebas han mostrado evidencias empíricas suficientes para que sus adaptaciones sean válidas y fiables en términos psicométricos.

### *Procedimiento*

Los participantes de la muestra de estudiantes se reclutaron en distintas clases de la Facultad de Psicología de la Universidad de Barcelona, solicitándoles su colaboración voluntaria en la investigación. Los cuestionarios fueron autoadministrados.

Para la constitución de la muestra comunitaria se optó por un muestreo accidental y de bola de nieve. Los criterios de exclusión fueron ser menor de edad, ser psicólogo/a o estudiante de psicología y estar actualmente en algún proceso terapéutico (psicológico o psiquiátrico). La aplicación se realizaba en una sola sesión. En todos los casos los evaluadores fueron psicólogos o bien estudiantes del último curso de psicología entrenados a tal efecto. Estas dos muestras forman parte de una investigación más amplia que se está realizando actualmente.

Los análisis se efectuaron mediante el programa *SPSS* 16.0, realizando su corrección en este mismo programa.

### *Análisis de datos*

De cada prueba se tomaron los resultados globales y el análisis se efectuó a partir de las submuestras constituidas por los sujetos sin ningún valor perdido. Algunas pruebas consignan con el valor “0” tanto la no respuesta como el valor menor del rasgo y, a veces, la no posibilidad de aplicación de un determinado ítem también se valora como “0”. Además y como consecuencia de lo anterior, se seleccionaron los sujetos de la muestra que no tenían ningún valor “0”, para asegurar que no hubiera confusión entre unos y otros valores “0”, obteniéndose las siguientes medias, desviaciones estándar y valores de *t* para cada grupo:

*Tabla 1. Media, desviación estándar y valor de t de Student para las variables analizadas*

VARIABLES	Estudiantes (n=42)	Comunitaria (n=31)	Valor de <i>t</i>
BDI	7.25 (6.94)	5.83 (5.28)	1.191
Autoestima RSE	32.93 (4.92)	33.81 (3.86)	.928
Neuroticismo-NEO	21.50 (8.94)	17.53 (8.93)	2.269*
Extraversión-NEO	32.77 (7.08)	32.74 (7.81)	.005
IGS-SCL	.5571 (.4370)	.3969 (.3389)	2.001
TSP-SCL	32.41 (17.44)	25.42 (15.68)	2.078*
IDSP-SCL	1.42 (.409)	1.29 (.325)	1.673*

\*  $p < .05$

IGS: Índice General de Sintomatología

TSP: Total de Síntomas Positivos

IDSP: Índice de Distrés de Síntomas Positivos

Se asumieron los valores de los estadísticos de las dos muestras iniciales como valores de referencia poblacionales y, a partir de los mismos, se simularon 500 muestras de cada uno de los grupos y en cada muestra se incorporaron aleatoriamente un determinado porcentaje de *missing data*. En concreto, se simularon los siguientes porcentajes de *missings* en cada muestra: 5, 10, 20 y 30%. No se simularon valores de presencia de *missings* cruzados, es decir, que en un grupo hubiera, por ejemplo,

un 10% de *missings* y en el otro grupo un 20%, en todos los casos se simularon el mismo porcentaje de datos perdidos en cada grupo. Obviamente, esto impide analizar los efectos más complejos (diferencia entre presencia de *missings* por grupo o distancia entre porcentajes que determina soluciones distintas). Para cada par de muestras, una vez simulada la presencia de los distintos porcentajes de *missings*, se aplicaron las diferentes técnicas de simulación para la recuperación de los datos perdidos que se consideran más adecuadas y potentes. En concreto, se usaron las siguientes técnicas de imputación: Imputación mediante la Media, Estimación mediante Regresión Lineal, estimaciones e imputaciones a partir de EM, MI, ML y FIML. Una vez efectuadas dichas imputaciones se procedió a comparar las medias obtenidas tras ese proceso y estimar el estadístico de contraste *t* de *Student*. Los procedimientos de simulación que se aplicaron fueron los propios de los planteamientos de Montecarlo operativizados en el programa EQS en su versión 6.1. y a partir del criterio “*leave-one-out*”.

## Resultados

Para cada una de las situaciones de presencia de *missings*, se obtuvo el porcentaje de resultados iguales a los obtenidos en la tabla 1. En las tablas 2 a 5 se presentan, pues, esos datos de ajuste que se obtienen con cada simulación y porcentaje de *missing* con relación a la solución original.

*Tabla 2. Resultados de ajuste con una simulación simétrica de missing del 5%  
Porcentaje de muestras cuya comparación de medias coincide con la situación de partida*

	Mean	Reg.	EM	MI	ML	FIML
BDI	89.72	85.12	81.66	91.12	89.12	91.12
Autoestima Rosenberg	77.65	79.12	84.66	88.33	91.13	89.99
Neuroticismo	84.48	82.66	87.12	87.45	90.04	91.14
Extraversión	91.78	87.45	86.47	90.05	92.14	89.12
IGS	87.49	86.77	85.33	86.33	93.12	92.67
TSP	89.12	81.12	88.59	81.12	90.40	91.55
IDSP	91.45	89.33	91.05	86.17	87.66	92.13

*Tabla 3. Resultados de ajuste con una simulación simétrica de missings del 10%  
Porcentaje de muestras cuya comparación de medias coinciden con la situación de partida*

	Mean	Reg.	EM	MI	ML	FIML
BDI	55	65	81	86	88	89
Autoestima Rosenberg	52	67	80	87	87	87
Neuroticismo	51	68	81	82	86	86
Extraversión	54	62	79	84	87	87
IGS	52	65	81	85	84	88
TSP	52	62	82	86	88	89
IDSP	51	61	83	88	81	91

Tabla 4. Resultados de ajuste con una simulación simétrica de missings del 20%  
Porcentaje de muestras cuya comparación de medias coinciden con la situación de partida

	Mean	Reg.	EM	MI	ML	FIML
BDI	32	45	65	78	78	82
Autoestima Rosenberg	31	43	66	77	79	81
Neuroticismo	33	44	68	80	77	83
Extraversión	34	42	65	81	76	84
IGS	35	43	61	81	78	82
TSP	31	42	67	77	80	81
IDSP	33	41	68	78	79	80

Tabla 5. Resultados de ajuste con una simulación simétrica de missings del 30%  
Porcentaje de muestras cuya comparación de medias coinciden con la situación de partida

	Mean	Reg.	EM	MI	ML	FIML
BDI	30	38	60	68	69	75
Autoestima Rosenberg	29	35	61	69	71	74
Neuroticismo	31	34	61	71	69	76
Extraversión	28	37	59	72	68	73
IGS	31	38	58	71	72	74
TSP	30	35	62	66	68	72
IDSP	29	37	57	67	71	71

### Conclusión

En las tablas anteriores puede observarse que con un porcentaje de *missings* del 5% prácticamente todos los métodos probados dieron buenos resultados, siendo mínima la mejoría en el ajuste que presentan los métodos modernos. Sin embargo, con el 10% de *missings* la pérdida de ajuste de los métodos antiguos es considerable y se diferenció claramente de los resultados obtenidos por métodos modernos. La tendencia a perder ajuste de los métodos antiguos se mantuvo y sus resultados fueron muy pobres al llegar al 20% de *missings*, aunque en ese momento también los métodos modernos perdieron ajuste, especialmente EM. Con el 30% de *missings* todos los métodos continuaron perdiendo ajuste.

Puede concluirse que con muestras pequeñas y a medida que los *missings* aumentan, la imputación mediante la sustitución por la media y la imputación individual basada en la regresión no dieron buenos resultados. Para estos casos fueron mejores los métodos modernos basados en estimaciones de máxima verosimilitud. Esto parece coincidir en parte con lo encontrado por Graham & Schafer (1999) quienes sostienen que MI funciona bien con muestras pequeñas incluso con el 50% de los *missings*, sin embargo la tendencia a un peor ajuste lleva a pensar que probablemente al llegar al 50% de *missings* los resultados obtenidos con MI fueran ya deficientes. Los mismos autores plantean que MI requiere tamaños de muestra que no siempre están disponibles para proveer estimaciones robustas de parámetros (Schafer & Graham, 2002), lo que podría ser una explicación para la deficiencia en el ajuste en este caso.

Debido a lo anterior quisiéramos señalar por último que dentro de los métodos de máxima verosimilitud el que presentó mejores resultados en estas condiciones fue el método FIML, conservando un buen ajuste con el 30% de *missings*.

### Referencias

- Costa, P.T. y McCrae, R.R. (1985). The NEO Personality Inventory. Odessa, FI: Psychological Assessment Resources. Adaptación española de Cordero, A., Pamos, A. & Seisdedos, N. (1999). *Inventario de Personalidad NEO reducido*. Madrid: TEA Ediciones.
- Derogatis, L. R. (1994). SCL-90-R Symptom Checklist 90 Revised. Minnesota: NCS Pearson. Adaptación española de González de Rivera, J. L., De las Cuevas, C., Rodríguez, M. & Rodríguez, F. (2002) *SCL-90-R Cuestionario de síntomas*. Madrid: TEA Ediciones.
- Graham, J.W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60, 549-576.
- Graham, J.W. y Schafer, J.L. (1999) On the performance of multiple imputation for multivariate data with small sample size. In *Statistical Strategies for Small Sample Research*, ed. R. Hoyle, 1:1-29. Thousand Oaks, CA: Sage.
- Little, R. J. A. y Rubin, D.B. (1987) *Statistical analysis with missing data*. New York: Wiley.
- Martín Albo, L., Nuñez, J. L., Navarro, J. G. y Grijalvo, F. (2007). The Rosenberg Self-Esteem Scale: Translation and Validation in University Students. *Spanish Journal of Psychology*, 10, 458-468.
- McKnight, P. E., McKnight, K. M., Sidani, S. y Figueredo, A. J. (2007) *Missing data. A gentle introduction*. New York: Guilford.
- Rubin, D.B. (1976) Inference and missing data. *Biometrika* 63, 581-92.
- Rubin, D.B. (1987) *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Sanz, J., Navarro, M. E. y Vázquez, C. (2003) Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): propiedades psicométricas en estudiantes universitarios. *Análisis y modificación de conducta*, 29, (124), 239-288.
- Schafer, J.L. y Graham, J.W. (2002) Missing data: our view of the state of the art. *Psychological Methods*, 7, 147-177.
- West, S. G. (2001) New approaches to missing data in psychological research: introduction to the special section. *Psychological Methods*, vol. 6, 4, 315-316.



# SESIONES PARALELAS



# CARACTERÍSTICAS Y ANÁLISIS DE LOS DISEÑOS DE MEDIDAS REPETIDAS EN LA INVESTIGACIÓN EXPERIMENTAL EN ESPAÑA EN LOS ÚLTIMOS 10 AÑOS

Paula Fernández García<sup>1</sup>, Guillermo Vallejo<sup>1</sup>, Pablo Livacic-Rojas<sup>2</sup> y Ellián Tuero<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Santiago de Chile

En la presente investigación se lleva a cabo un estudio de contenido que tiene por objetivo conocer cuáles son los diseños más utilizados y sus características principales en tres revistas científicas de Psicología editadas en castellano y en el período 1999-2008. Las revistas se eligieron arbitrariamente de entre las que cumplían determinados criterios. Se ha puesto de relieve que son los diseños experimentales más utilizados que los cuasiexperimentales, y de ellos, los diseños de medidas repetidas. También que la investigación básica es superior a la aplicada y que los estudiantes de Psicología son el tipo de muestra más abundante.

La evaluación de la producción científica es un tema que cotiza al alza en la actualidad. En nuestro país empezó a alcanzar fuerza este debate en la década de 1990 (véase Pelechano, 2002a, 2002b, 2002c). Abundan los estudios sobre evaluación de artículos, revistas, investigadores, universidades, etc desde las más diversas perspectivas. En lo que respecta a la evaluación de las revistas científicas podríamos clasificarla desde tres ópticas distintas: Análisis Bibliométrico (v.g., Buela-Casal, 2003; Buela-Casal, Carretero-Dios y Santos-Roig, 2002); Análisis de contenido (v.g., García- Jiménez y Rubio Jerónimo, 1990, 1991; Moreno y Sánchez, 1998; Montero y León, 2001) y Análisis de Calidad de las investigaciones en ellas contenidas. Las primeras abordaría la producción y las dos últimas el producto.

En la presente investigación se lleva a cabo un estudio de contenido, y tiene como objetivo conocer cuáles son las preferencias en los últimos 10 años (1999-2008) de los investigadores que publican en revistas españolas de Psicología sobre el modo de recoger los datos, esto es, sobre la estructura de los diseños de investigación, y qué características les definen. Este estudio se reduce exclusivamente a las investigaciones experimentales (con excepción de los diseños de replicación intra-sujeto o N=1) y cuasiexperimentales (con excepción de aquellos que sean de series temporales). Para ello, decidimos elegir arbitrariamente tres revistas científicas españolas de Psicología que cumplieren tres criterios: a) ser revistas de temática general; b) aparecer en el *Journal Citation Reports (JCR)* con factor de impacto (en el año 2008); c) aparecer en *IN-RECS* con índice de impacto en 2008 dentro del primer cuartil.

## Método

Este trabajo constituye un estudio descriptivo mediante análisis de documentos.

### *Materiales*

Se eligieron arbitrariamente tres revistas científicas que cumplían los criterios anteriormente referidos:

1. *Psicothema*: Editada por la Facultad de Psicología de la Universidad de Oviedo y el Colegio Oficial de Psicólogos. Fundada en 1989 con la finalidad de publicar trabajos de carácter empírico o teórico, sobre cualquier ámbito de la psicología. En un principio semestral, ha experimentado un notable crecimiento y actualmente edita cuatro números al año.

2. *Internacional Journal and Health Psychology* (IJCHP): Editada por la Asociación Española de Psicología Conductual y cuenta con la colaboración de ABA-Colombia. Fundada en el año 2001 con la finalidad de publicar trabajos de carácter aplicado, tanto teóricos como experimentales, que contribuyan al avance de cualquier ámbito de la Psicología Clínica y de la Salud. Edita tres números al año.
3. *Psicológica*: Editada por el Departamento de Metodología de las Ciencias del Comportamiento de la Universidad de Valencia. Fundada en 1980 con la finalidad de publicar trabajos de Metodología y de Psicología experimental. Edita dos números al año.

Se han exceptuado de la revista *Psicothema* los números que son suplementos. Sin embargo se han considerado los números especiales y monográficos de la revista *Psicológica*.

### *Unidad de análisis*

La unidad de análisis ha sido el estudio, considerando unidades independientes cada uno de los estudios experimentales y cuasiexperimentales publicados en un mismo artículo. Consideramos utilizar como unidad superior de análisis los artículos publicados por las tres revistas anteriormente referidas durante los años 1999-2008.

### *Diseño y Procedimiento*

Se recogieron datos de todos los estudios contenidos en los artículos de las revistas para el período de tiempo (1999-2008). La información acerca de los diseños y sus características considerada fueron los apartados introducción, método, procedimiento y conclusiones. Dos grupos de expertos (profesores de Psicología) compuestos por dos profesionales cada grupo analizaron aleatoriamente la mitad de los artículos. En cada grupo cada artículo era analizado por cada experto en las variables de interés. Posteriormente se comparaban las clasificaciones obtenidas por ambos. El porcentaje medio de acuerdo entre los grupos fue de un 95%. En aquellos casos en los que no coincidían se revisaban otra vez con la colaboración de un tercer experto y se analizaba el contenido del artículo hasta que los tres expertos llegasen a una conclusión compartida.

### *Variables registradas:*

1. *Temática*: Se han clasificado en cuatro grupos: 1) Psicofisiología; 2) Aprendizaje, Memoria; 3) Psicología Cognitiva y 4) Psicología Clínica (Aplicación de técnicas de intervención conductual, programas de tratamiento o rehabilitación). Las tres primeras constituyen investigación básica y la última investigación aplicada.
2. *Diseño*: Se han clasificado en 11 grupos: 1) Simple de MR; 2) Factorial de MR; 3) Split-Plot; 4) Simple de MR (2 registros); 5) Factorial de MR (2 registros); 6) Split-Plot (2 registros); 7) Trasversal de grupos; 8) Trasversal factorial; 9) Trasversal bloques al azar; 10) Trasversal jerárquico y 11) Otros.
3. *Muestra*: Se ha clasificado en 8 grupos: 1) Estudiantes de Psicología o afín (créditos o compromiso); 2) Estudiantes de otros centros universitarios (en este caso también intervienen con frecuencia estudiantes de Psicología); 3) Personas voluntarias: Con alguna excepción testimonial, son estudiantes de Psicología que participan de modo voluntario en una investigación sin verse comprometidos a hacerlo y sin recibir créditos por ello; 4) Estudiantes de colegios o institutos; 5) Personas que acuden o pertenecen a centros de salud; 6) Muestra específica (son muestras de pacientes, v.g. padecen esquizofrenia, depresión, dislexia..., siempre con una finalidad aplicada. Todas aquellas investigaciones que se han incluido en la temática de Psicología Clínica tienen esta clasificación

de muestra dado que la finalidad siempre ha sido aplicada; 7) Animales; 8) Selección aleatoria de los sujetos; 10) Estudiantes de Psicología y muestra específica y 11) Estudiantes de Psicología y estudiantes de colegios.

4. *Constitución de los grupos de sujetos que se someten a estudio:* 1) Grupo único de sujetos; si el diseño cuenta con más de un grupo de sujetos será, 2) Diseño equilibrado o 3) Diseño no equilibrado.
5. *Tamaño de grupos:* Si sólo hay un grupo de sujetos se recoge el tamaño del grupo; si existe más de un grupo de sujetos, se recoge su tamaño si es equilibrado; si el diseño no es equilibrado se recogen sólo los tamaños de los grupos con mayor y menor número de casos.

De todos los diseños, son diseños cuasiexperimentales básicos (no tienen grupo de control o no tiene medida pretratamiento):

- Los diseños del grupo 4 (Simple de MR, 2 registros), donde sólo existe un grupo de sujetos y los registros constituyen una variable intra-sujeto de tiempo (pre-post).
- Los diseños del grupo 7 (Trasversal de grupos) donde los grupos están constituidos por pacientes y no han sido asignados aleatoriamente a los niveles de la variable de tratamiento, y tampoco se ha registrado una medida pre-tratamiento.

Son diseños propiamente cuasiexperimentales:

- Los diseños del grupo 6, Split-Plot, donde la variable entre-sujetos es de clasificación (un grupo ha sido grupo control) y la variable intra-sujeto es el tiempo (pre-post). Estos diseños son, o bien Diseños de Grupo control no equivalente, o bien diseños de discontinuidad en la regresión.
- El resto de diseños son diseños experimentales.

## Resultados

Se ha realizado un análisis exclusivamente descriptivo de las variables estudiadas, totales, frecuencias y porcentajes.

- Se han revisado un total de 1534 artículos científicos, 1048, 304 y 162 en las revistas *Psicothema*, *IJCHP* y *Psicológica* respectivamente.
- En los 1534 artículos se encuentran 493 estudios y en cada uno de ellos un diseño de investigación experimental o cuasiexperimental.
- De los 493 diseños de investigación considerados, 16 (3,24%) son diseños propiamente cuasiexperimentales (D. de Grupo Control no Equivalente y/o Discontinuidad en la regresión); 12 (2,43%) suponen diseños cuasiexperimentales básicos. Por lo tanto, los diseños experimentales suponen el volumen mayor (94,33%) de las investigaciones realizadas.
- Los estudios que son investigación básica suponen el 60%, 100% y el 97% de las investigaciones realizadas en las revistas *Psicothema*, *IJCHO* y *Psicológica* respectivamente. La investigación aplicada suponen un porcentaje importante en las publicaciones de la revista *IJCHP* (40%), y testimonial en la revista *Psicothema* (2%).
- En la Tabla 1 observamos que los diseños más utilizados en las investigaciones publicadas en las tres revistas científicas consideradas son los diseños Split-Plot o mixtos (tienen variables intra-sujeto y entre-sujetos) 329 (66.73%). Los diseños menos utilizados son los diseños experimentales de bloques al azar, factoriales jerárquicos y otros, que en este caso sólo existe 1 y es un diseño de cuadrado latino Youden.
- De los 493 diseños de investigación encontrados 429 son diseños de medidas repetidas (simples, factoriales y Split-Plot), esto es, el 87.01% del total. De ellos, los diseños de medidas repetidas con más de dos niveles intra-sujeto son 254 (59.20%).

*Tabla 1. Frecuencias y porcentajes de los diseños de investigación utilizados y temática investigada en las tres Revistas Científicas*

Revista	Diseños		Temática			
	n	p	n	p		
Psicothema	Simple de medidas repetidas	17	5.9	Psicofisiología	72	25
	Factorial de medidas repetidas	16	5.6	Aprendizaje, memoria	116	40.3
	Split-Plot	90	31.3	P. Cognitiva	93	32.3
	Simple de medidas repetidas (2 registros)	7	2.4	P. Clínica	7	2.4
	Factorial de medidas repetidas	5	1.7	Total	288	100
	Split-Plot (2 registros)	101	35.1			
	Trasversal de grupos	27	9.4			
	Trasversal factorial	19	6.6			
	Trasversal Bloques al azar	4	1.4			
	Trasversal jerárquico	1	.3			
	Otros	1	.3			
Total	288	100				
IJCHP	Simple de medidas repetidas	5	14,3	Psicofisiología	10	28.6
	Split-Plot	15	42,9	Aprendizaje, memoria	11	31.4
	Simple de medidas repetidas (2 registros)	4	11,4	P. Cognitiva	0	0
	Split-Plot (2 registros)	9	25,7	P. Clínica	14	40
	Trasversal de grupos	2	5,7	Total	35	100
	Total	35	100			
Psicológica	Simple de medidas repetidas	2	1.2	Psicofisiología	38	22.4
	Factorial de medidas repetidas	28	16.5	Aprendizaje, memoria	87	51.2
	Split-Plot	81	47.6	P. Cognitiva	45	26.5
	Simple de medidas repetidas (2 registros)	3	1.8	P. Clínica	0	0
	Factorial de medidas repetidas	13	7.6	Total	170	100
	Split-Plot (2 registros)	33	19.4			
	Trasversal de grupos	5	2.9			
	Trasversal factorial	5	2.9			
	Total	170	100			

Leyenda: n=Frecuencia absoluta; p=Porcentaje

En cuanto a la muestra, 189 (38.3%) de los estudios se realizan con estudiantes de Psicología a cambio de créditos o por compromiso con el profesor; 81(16,4%) son sujetos voluntarios a cambio de nada, que exceptuando muy pocos casos, también son estudiantes de Psicología; 34 (6,9%) son estudiantes de colegios; 68 (13.8%) son pacientes; 99 (20,1%) animales; sólo hay 4 (0.8%) en el que la muestra sea extraída aleatoriamente de la población de interés. El porcentaje restante, 3,6% se distribuye entre estudiantes de otros centros universitarios (1,8%), sujetos de centros de salud (1%) y muestras mixtas formadas por estudiantes de Psicología y Colegios (0.6%) y Estudiantes de Psicología y pacientes (0.2).

*Tabla 2. Frecuencias y porcentajes de los diseños en cuanto al número de grupos y su formación*

	Características de tamaños de grupo					
	n	p	r	T.g	nc	p
Solo un grupo de sujetos	101	20.5	3-193	8	12	11.9
				10	10	9.9
Varios Grupos: Diseño Equilibrado	329	66.7	2-325	6	16	4.9
				8	20	6.1
				9	10	3
				10	39	11.9
				12	33	10
				15	18	5,5
				16	23	7
				20	28	8.5
				25	11	3.3
30	15	4.6				
Varios Grupos: Diseño no equilibrado	63	12.8		*5-10		
				22-238		
Total	493	100				

Leyenda: r= rango de tamaños de grupo; T.g.= Tamaño de grupo nc=número de estudios; \*= en diseños no equilibrados los tamaños de grupos en una misma investigación van 5 el menor y 10 sujetos el mayor, hasta 238 sujetos el mayor de los grupos y 22 en el menor de ellos. Resto, ver Tabla 1

En la Tabla 2 advertimos que el volumen de diseños que tiene un solo grupo de sujetos, diseños simples de medidas repetidas o factoriales de medidas repetidas son 101(20.5%). El resto de diseños contienen más de un grupo, siendo los diseños balanceados los que constituyen el mayor porcentaje 329(66.7%). Los tamaños de los grupos más abundantes son de 8 y 10 sujetos cuando sólo hay un grupo, y de 10, 12 y 20 sujetos cuando hay más de un grupo. En los diseños no equilibrados existen diseños donde el grupo con menor tamaño es de 5 y el de mayor de 10, hasta alguno donde los grupos de menor y mayor tamaño son respectivamente 22 y 38.

### Discusión y conclusión

En este estudio descriptivo se ha puesto de relieve que en el conjunto de tres revistas científicas importantes de Psicología editadas en castellano, *Psicothema*, *IJCHP* y *Psicológica* se publica un mayor volumen de investigación experimental que cuasiexperimental, y, aunque con diferencias entre las tres revistas, la cantidad de investigación básica es muy superior a la investigación aplicada. También se ha puesto de relieve la mayor preferencia por utilizar diseños de medidas repetidas en lugar de diseños trasversales. García-Jiménez y Rubio-Jerónimo (1991) también encontraron este resultado en otro conjunto de revistas españolas de Psicología (sólo coincide con ésta investigación la revista *Psicológica*) en el período 1981-1988. Se pone de manifiesto que los alumnos de Psicología constituyen el tipo de muestra más utilizado para llevarlos a cabo, este mismo dato también lo destacaron García-Jiménez y Rubio-Jerónimo (1990).

Lo expuesto en este texto constituye una pequeña parte de la comunicación presentada en el XI Congreso de Metodología celebrado en Málaga (septiembre de 2009), que, a su vez, es parte de una investigación que se continúa realizando.

Este trabajo ha sido realizado con la ayuda concedida por el MCI (Ref.: PS-I2008-03624).

## Referencias

- Buela-Casal, G., Carretero-Dios, H. y Santos-Roig, M. (2002). Estudio comparativo de las revistas de Psicología en castellano con factor de impacto. *Psicothema*, 14, 837-852.
- Buela-Casal, G. (2003). Evaluación de la calidad de los artículos y de las revistas científicas: Propuesta del factor de impacto ponderado y de un índice de calidad. *Psicothema*, 15(1), 23-35
- García- Jiménez, M.V. y Rubio-Jerónimo, A. (1990). Análisis de las muestras empleadas en la investigación experimental. Comunicación. *VIII Congreso Nacional de Psicología*. Barcelona.
- García- Jiménez, M.V. y Rubio-Jerónimo, A. (1991). Análisis de variables y diseños empleados en la investigación experimental. *Currículum*, 1, 29-36.
- Montero, I. y León, O.G. (2001). Usos y costumbres metodológicas en la Psicología española: Un análisis a través de la revista *Psicothema* (1990-1999). *Psicothema*, 13, 671-677.
- Moreno, S. y Sánchez, A. (1998). Análisis bibliométrico de la revista "Psicothema" (1989-1997). *Psicothema*, 10, 23-27.
- Pelechano, V. (2002a). Presentación del monográfico del Monográfico del 2002: La valoración de la actividad científica. *Análisis y Modificación de Conducta*, 28, 311-316.
- Pelechano, V. (2002b). Valoración de la actividad científica en Psicología ¿pseudoproblema, sociologismo e idealismo? *Análisis y Modificación de Conducta*, 28, 323-362.
- Pelechano, V. (2002c). Epílogo del monográfico del Monográfico del 2002: La valoración de la actividad científica. *Análisis y Modificación de Conducta*, 28, 477.

# TENDENCIAS DE LA POTENCIA DE PRUEBA EN UN DISEÑO DE MEDIDAS REPETIDAS CARENTE DE ESFERICIDAD CON Y SIN AUTOCORRELACIÓN SERIAL

Paula Fernández García<sup>1</sup>, Guillermo Vallejo<sup>1</sup>, Pablo Livacic-Rojas<sup>2</sup> y Ellián Tuero<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Santiago de Chile

En este trabajo examinamos el comportamiento de cuatro estadísticos univariados para analizar los datos de un diseño de medidas parcialmente repetidas. Los cuatro asumen que la matriz de desviación subyacente es no esférica, pero dos de los procedimientos presuponen que la correlación entre los datos es arbitraria y otros dos que existe autocorrelación serial de primer orden. Mediante investigación Montecarlo los cuatro son comparados con respecto a su potencia en ausencia de esfericidad y en ambas situaciones, bajo correlación serial y bajo correlación arbitraria. Cuatro han sido las variables manipuladas: el tamaño de muestra total, la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión, la estructura de la matriz de covarianza de la población y la forma de distribución de los datos.

La ausencia de esfericidad es el principal problema al que se enfrentan los investigadores cuando analizan un diseño de medidas repetidas. Para resolverlo, y en función de si las matrices de dispersión son o no homogéneas, pueden corregir los valores críticos de la  $F$  univariada (v.g. Greenhouse & Geisser, 1959; Huynh & Feldt, 1976; Quintana & Maxwell, 1994). Sin embargo, estos procedimientos asumen que la correlación entre las observaciones en distintos puntos del tiempo no es una función de la distancia temporal entre ellas.

Otro enfoque univariado, más flexible y versátil que los anteriores es el Modelo Lineal Mixto (MLM). El MLM extiende el modelo clásico a situaciones donde los supuestos de independencia y homogeneidad no son requeridos y las variables son fijas y/o aleatorias. Sin embargo, si bien es verdad que tiene múltiples ventajas (Littel, Milliken, Stroup, Wolfinger & Schabenberger, 2006), no está exento de inconvenientes (Keselman, Algina, Kowalchuck & Wolfinger, 1999; Vallejo, Fernández & Ato, 2003; Vallejo, Arnau, Bono, Cuesta, Fernández & Herrero, 2002).

Existen otros procedimientos univariados orientados a corregir la ausencia de esfericidad atendiendo a los efectos de la dependencia serial y que han sido escasamente estudiados. Recientemente dos de ellos, los procedimientos de Hearne, Clark & Hatch, (1983) y Jones (1985), han sido estudiados con respecto al error de Tipo I junto con otros procedimientos univariados por Fernández, Vallejo, Livacic-Rojas, Herrero & Cuesta (2008) y también con respecto a la potencia de prueba (Fernández, Vallejo y Livacic-Rojas, 2009). Estas investigaciones se realizaron asumiendo un diseño balanceado y homogeneidad de las matrices de dispersión entre los grupos.

El objetivo de esta investigación ha sido evaluar la Potencia empírica de estos dos procedimientos junto con los propuestos por Greenhouse y Geisser (1959) y la corrección de Lecoutre (1991) cuando en los datos subyace ausencia de homogeneidad entre las matrices de dispersión de los grupos, tanto cuando el diseño es balanceado como cuando no lo es, y cuando en los datos subyace distribución normal y no normal. La potencia se examinó en ausencia de esfericidad bajo correlación serial y bajo correlación arbitraria.

## Método

Se realizó un estudio de simulación Montecarlo para un diseño de medidas repetidas (3x4) (modelo aditivo) donde las matrices de covarianza grupales fueron heterogéneas en la relación  $\Sigma_1=1/3\Sigma_2$  y  $\Sigma_3=5/3\Sigma_2$ .

*Procedimientos que se someten a investigación:* Greenhouse & Geisser (1959), corrección de Lecoutre (1991) del estadístico de Huynh & Feldt (1976), Hearne et al., (1983) y Jones (1985), que denominaremos GG, LEC, HCH y JN respectivamente. La formulación de las dos primeras se puede consultar en (v.g., Vallejo, 1991; Fernández et al. 2007) y las de las dos últimas en Hearne et al. (1983) y Jones (1985) respectivamente.

### *Variables manipuladas:*

1. *Tamaño de muestra total.* Se utilizaron dos tamaños (N): N=15 y 30. El coeficiente de variación muestral fue 0 cuando el diseño estaba equilibrado ( $n_j=5$  y 10 respectivamente en los tamaños N anteriores). Cuando fue 0.33 el diseño estaba no equilibrado del siguiente modo (N=15:  $n_j=3, 5$  y 7; N=30:  $n_j=6, 10$  y 14).

2. *Patrones de covarianza empleados para generar los datos:* Nuestro principal interés ha sido observar qué influye más en la estimación de la potencia de prueba empírica de los procedimientos, si la magnitud de la autocorrelación serial de primer orden ( $\rho$ ) o la cuantía de la desviación de la esfericidad ( $\varepsilon$ ). Para ello generamos los datos utilizando dos estructuras de covarianza no estacionarias desestructuradas, ARAH[1] y NE, ambas con dos desviaciones de la esfericidad ( $\varepsilon=.56$  y  $\varepsilon=.75$ ). En las matrices ARAH[1] existe autocorrelación serial de primer orden y por lo tanto la correlación entre la  $k$ th y la  $k'$ th observación es  $\rho^{|k-k'|}$ ,  $\rho= [0.2:0.8: (0.2)]$ . Las Matrices NE presentan una estructura de correlación arbitraria, y, como consecuencia, la autocorrelación serial es cero ( $\rho=0$ ).

3. *Emparejamiento de las matrices de covarianza y el tamaño de los grupos.* Razón C/H. H0 (diseño equilibrado y heterogeneidad entre grupos), H+ y H- (diseño no equilibrado y relación entre los tamaños de los grupos y el de la matriz de dispersión positiva y negativa respectivamente).

4. *Forma de distribución de la variable de medida.* Dos han sido las formas de distribución estudiadas: Distribución normal ( $\gamma_1=0$  y  $\gamma_2=0$ ) y Distribución Exponencial ( $\gamma_1=2$  y  $\gamma_2=6$ ). Los datos han sido generados mediante las distribuciones  $g$  y  $h$  introducidas por Tukey (1977).

*Cálculo de las Potencia de Prueba Teórica:* Calculamos la Potencia Teórica siguiendo las recomendaciones de Muller y Barton (1989) y tomando como referencia de las mismas el valor medio de la potencia prueba teórica en todas las magnitudes de correlación para los procedimientos GG y LEC en  $n_j=5$  y  $q=4$ . Así, las potencias teóricas son: ARAH[1] y  $\varepsilon=.50$ ,  $(1-\beta)=.75$ ; ARAH[1] y  $\varepsilon=.75$ ,  $(1-\beta)=.77$ ; NE y  $\varepsilon=.50$ ,  $(1-\beta)=.64$  y NE y  $\varepsilon=.75$ ,  $(1-\beta)=.73$ .

## Resultados

La tasa de potencia empírica se calculó dividiendo el número de veces que la Hipótesis nula es rechazada correctamente para el nivel  $\alpha=.05$  entre el número de ejecuciones efectuadas ( $10^4$ ). En la Tabla 1 se muestran los resultados destacándose en negrita las potencias empíricas inferiores a las teóricas anteriormente referidas.

*Matriz subyacente* ARAH[1]: Observamos, de una parte, que una desviación de la normalidad reduce la potencia de prueba en todos los procedimientos, y de otra, que el procedimiento HCH es el que mayor potencia empírica alcanza seguido de LEC y JN, siendo GG el que experimenta la menor potencia en ambas magnitudes de  $\epsilon$ .

Tabla 1. Tasas de Potencia empírica asociadas al Tratamiento Intra-sujeto para los cuatro procedimientos en un diseño Split-Plot (3x4).  
Matrices de desviación ARAH [1] YNE

Distribución normal ( $\gamma_1=0$ y $\gamma_2=0$ )												
GG	LEC	GG	LEC	GG	LEC	$\rho$	$\eta_1$	$\rho$	GG	LEC	GG	
.404	.486	.426	.523	.340	.408	.362	.449	.510	.581	.527	.615	.20
.451	.551	.468	.589	.371	.471	.391	.510	.550	.655	.569	.685	.40
.534	.690	.556	.702	.448	.611	.467	.641	.634	.786	.653	.807	.60
.658	.888	.671	.902	.577	.829	.594	.850	.763	.944	.776	.952	.80
.227	.364	.244	.395	.180	.314	.197	.344	.291	.460	.307	.488	.00
.769	.845	.778	.854	.680	.772	.689	.786	.854	.918	.964	.925	.20
.806	.896	.814	.902	.726	.843	.726	.853	.895	.953	.899	.957	.40
.877	.958	.882	.962	.816	.918	.823	.924	.937	.984	.949	.986	.60
.947	.998	.949	.998	.902	.993	.905	.994	.975	.999	.976	.999	.80
.447	.703	.456	.716	.364	.620	.371	.632	.535	.803	.542	.813	.00
Distribución Exponencial ( $\gamma_1=2$ y $\gamma_2=6$ )												
GG	LEC	GG	LEC	GG	LEC	$\rho$	$\eta_1$	$\rho$	GG	LEC	GG	
.20	.231	.289	.240	.327	.147	.217	.167	.255	.309	.389	.339	.430
.40	.243	.348	.271	.392	.177	.273	.199	.317	.348	.450	.375	.491
.60	.321	.519	.346	.557	.243	.422	.267	.459	.443	.624	.466	.651
.80	.453	.724	.474	.744	.357	.648	.379	.674	.590	.804	.610	.817
.00	.144	.236	.159	.262	.120	.197	.132	.220	.193	.320	.206	.348
.20	.599	.694	.612	.707	.469	.588	.484	.608	.746	.804	.757	.817
.40	.642	.761	.653	.774	.531	.665	.543	.684	.790	.855	.798	.864
.60	.743	.857	.750	.863	.638	.792	.648	.803	.863	.918	.867	.922
.80	.854	.944	.863	.947	.775	.917	.782	.921	.937	.971	.940	.973
.00	.298	.515	.305	.527	.246	.438	.253	.452	.368	.624	.376	.635
Distribución Exponencial ( $\gamma_1=2$ y $\gamma_2=6$ )												
H0			H+			H-			H			
$\epsilon$	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	
.20	.359	.390	.335	.356	.286	.307	.249	.270	.483	.517	.454	
.40	.385	.442	.330	.364	.296	.355	.249	.284	.511	.570	.443	
.60	.449	.586	.369	.485	.349	.487	.283	.389	.586	.698	.494	
.80	.568	.759	.501	.667	.467	.681	.401	.574	.707	.738	.645	
.00	.241	.342	.227	.329	.204	.295	.193	.282	.302	.439	.286	
.20	.750	.753	.699	.705	.630	.659	.575	.602	.874	.857	.835	
.40	.774	.809	.688	.724	.675	.718	.580	.619	.895	.894	.831	
.60	.838	.874	.750	.804	.641	.816	.641	.721	.928	.935	.870	
.80	.907	.954	.867	.922	.841	.926	.786	.877	.965	.978	.943	
.00	.414	.616	.393	.693	.355	.537	.333	.518	.498	.718	.472	
Distribución Exponencial ( $\gamma_1=2$ y $\gamma_2=6$ )												
H0			H+			H-			H			
$\epsilon$	.50	.75	.50	.75	.50	.75	.50	.75	.50	.75	.50	
.20	.359	.390	.335	.356	.286	.307	.249	.270	.483	.517	.454	
.40	.385	.442	.330	.364	.296	.355	.249	.284	.511	.570	.443	
.60	.449	.586	.369	.485	.349	.487	.283	.389	.586	.698	.494	
.80	.568	.759	.501	.667	.467	.681	.401	.574	.707	.738	.645	
.00	.241	.342	.227	.329	.204	.295	.193	.282	.302	.439	.286	
.20	.750	.753	.699	.705	.630	.659	.575	.602	.874	.857	.835	
.40	.774	.809	.688	.724	.675	.718	.580	.619	.895	.894	.831	
.60	.838	.874	.750	.804	.641	.816	.641	.721	.928	.935	.870	
.80	.907	.954	.867	.922	.841	.926	.786	.877	.965	.978	.943	
.00	.414	.616	.393	.693	.355	.537	.333	.518	.498	.718	.472	

Leyenda:  $\rho=20$ ,  $.40$ ,  $.60$ ,  $.80$ = autocorrelación serial positiva de primer orden en ARAH[1];  $\rho=00$ = ausencia de autocorrelación serial (correlación arbitraria), matrizNE.

El resto de etiquetas no conducen a error dado que están igualmente definidas en el texto.

*Matriz subyacente* NE: En la misma Tabla podemos advertir que HCH muestra el mejor comportamiento y GG el peor comportamiento para todo valor  $n_j$  y  $\varepsilon$ . Un incremento en el valor de  $\varepsilon$  provoca un incremento de la potencia.

Independientemente de la matriz subyacente a los datos (ARAH [1] o NE) advertimos que un incremento en el tamaño de la muestra provoca un incremento de la potencia y se igualan los resultados de los cuatro procedimientos.

### Discusión y conclusión

De los resultados anteriores se extraen cinco tendencias comunes a los cuatro procedimientos:

- 1.- En función de la razón C/H, la potencia empírica es mayor cuando la razón es H-, le sigue H0 y, se halla la potencia empírica menor bajo la razón H+.
- 2.- La potencia empírica incrementa conforme incrementa N y disminuye conforme nos alejamos de la normalidad.
- 3.- Los cuatro procedimientos experimentan mayor potencia empírica cuando  $\varepsilon=.75$  que cuando  $\varepsilon=.50$ .
- 4.- Los procedimientos GG, LEC, HCH y JN experimentan una potencia empírica de mayor a menor cuando  $\rho+>\rho=0$ .
- 5.- Cuando la autocorrelación serial es de primer orden, un incremento en su valor incrementa la potencia en los procedimientos, dejándose notar claramente cuando  $n_j=5$ .

¿Qué influye más sobre la potencia empírica, el tamaño de  $\varepsilon$ , o la magnitud de  $\rho$ ? En la investigación realizada por Fernández, Vallejo, Livacic-Rojas y Tuero (2010) se puso de manifiesto que con respecto al error de Tipo I, claramente era el tamaño de  $\varepsilon$ . Sin embargo, para la potencia de prueba, si bien es cierto que la potencia es mayor cuanto mayor es  $\varepsilon$ , también es cierto que la autocorrelación serial ejerce un efecto sobre ella mayor que la desviación de la esfericidad, dado que cuando no existe autocorrelación serial (matriz NE) la potencia empírica es muy inferior a cuando  $\rho$  es positiva (ARAH[1]).

Estos resultados están en sintonía con los obtenidos por Fernández et al. (2008) y Fernández et al. (2009).

Agradecimientos. Este trabajo ha sido realizado con la ayuda concedida por el MCI (Ref.: PS-I2008-03624).

### Referencias

- Fernández, P., Livacic-Rojas, P. & Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. *International Journal of Clinical and Health Psychology*, 7(1), 153-175.
- Fernández, P., Vallejo, G., Livacic-Rojas, Herrero, J. & Cuesta, M. (2008). Comparative robustness of six tests in repeated measures designs with specified departures from sphericity. *Quality & Quantity*, DOI: 10.1007/S11135-008-9198-3.

- Fernández, P., Vallejo, G., Livacic-Rojas, P. & Tuero, E. (2009). Comparison of the power of four statistics in repeated measures design in the absence of sphericity with and without serial autocorrelation. *Review of Psychology*, 16.
- Fernández, P., Vallejo, J., Livacic-Rojas, P. & Tuero, E. (2009). Análisis de la robustez de cuatro estadísticos univariados en un diseño de medidas repetidas carente de esfericidad con, y sin autocorrelación serial. Comunicación presentada en el *XI Congreso de Metodología de las Ciencias Sociales y de la Salud*, Málaga (España). 15-18 de Septiembre, 2009.
- Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Hearne, E.M., Clark, G.M. & Hatch, J.P. (1983). A test for serial correlation in univariate repeated-measures analysis. *Biometrics*, 39, 237-243.
- Huynh, H. & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69-82.
- Jones, R.H. (1985). Repeated measures, interventions, and time series analysis. *Psychoneuroendocrinology*, 10(1), 5-14.
- Keselman, H.J., Algina, J., Kowalchuck, R.K. & Wolfinger, R.D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, 52, 63-78.
- Lecoutre, B. (1991). A correction for the  $\epsilon$  approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics*, 16, 371-372.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D. & Schabenberger, O. (2006). *SAS System for Mixed Models*, Cary, NC: SAS Institute Inc.
- Muller, K.E. & Barton, C.N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association* 84 (406), 549-555.
- Quintana, S. & Maxwell, S. E. (1994). A Monte Carlo comparison of seven e-adjustment procedures in repeated measures designs with small sample sizes. *Journal of Educational Statistics*, 19, 57-71.
- Tukey, J.W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southern Massachusetts University (North Dartmouth, MA).
- Vallejo, G., Arnau, J., Bono, R., Cuesta, M., Fernández, P. & Herrero, J. (2002). [Analysis of short time series designs]. Análisis de diseños de series temporales cortas. *Metodología de las Ciencias del Comportamiento*, 4(2), 301-323.
- Vallejo, G., Fernández, P. & Ato, M. (2003). Tasas de potencia de dos enfoques robustos para analizar datos longitudinales. *Psicológica*, 24, 109-122.

# POTENCIA DEL ESTADÍSTICO $f$ VS. $F$ CUANDO NO SE ASUME LA NORMALIDAD

José Fernando García Pérez<sup>1</sup>, Enrique Gracia<sup>1</sup>, Sergio Murgui<sup>2</sup> y M. del Castillo Fuentes<sup>1</sup>

<sup>1</sup> Universidad de Valencia

<sup>2</sup> Universidad Católica de Valencia

El estadístico ' $f$ ' se computa como el ' $F$ ' de Snedecor si bien suponiendo que la distribución de la población padre es rectangular y discreta. Este estadístico ha sido recomendado para las situaciones en que la distribución de la variable dependiente es asimétrica. Las colas largas pueden incrementar considerablemente la varianza del error y, consecuentemente, la probabilidad del error del Tipo II. En este trabajo se presentan varias simulaciones para determinar la potencia y el error del Tipo I de los dos estadísticos, ' $f$ ' y ' $F$ ', variando: a) la asimetría de la variable dependiente, b) el tamaño de la muestra, y c) el tamaño del efecto.

El estadístico  $f$  (García, Pascual y Frías, 2002; Pérez, Navarro y Llobell, 2000; Pérez, Llobell, Navarro, 2004) se define formalmente a partir de una población padre, la de medida, que sea una variedad de la distribución rectangular (uniforme) discreta  $R$ , con un recorrido entre los valores 1, el mínimo, y  $R$ , el máximo, de tal manera que  $1 \leq r \leq R$ , siendo  $r$  cualquier número entero 1, 2, ...,  $R$ . La función de probabilidad en cualquier punto  $r$  será  $1/R$ , y el de distribución,  $r/R$ . Los momentos centrales de esta familia de distribuciones también se conocen, siendo la  $\mu = (R + 1)/2$  y la  $\sigma^2 = (R + 1)(R - 1)/12$ . Son siempre distribuciones simétricas con una curtosis de  $(3 - (4/(R + 1)(R + 3)))(3/5)$  (Evans, Hastings y Peacock, 1993). Una función de probabilidad rectangular discreta implica una forma acumulada de la población padre de  $R$  incrementos constantes con igual amplitud, un concepto de escala lineal que es monótonicamente creciente respecto de una unidad de  $1/R$  proporciones de la distribución (Pérez, 2008).

A diferencia del estadístico  $F$  que asume una distribución padre normal, la distribución  $f$  implica una transformación de las puntuaciones directas para que se ajusten a la distribución rectangular uniforme discreta (Pérez et al., 2004). La unidad de la medida directa del estudio, como sean los microsegundos para el tiempo de reacción, el acierto para el número de aciertos en una tarea experimental, cada punto de las escalas Likert para la suma de las respuestas de una factor de personalidad, ya no pueden seguir una relación lineal con las puntuaciones  $R_{tiles}$ , conservando únicamente la relación de orden con las unidades de medida directa de la conducta (García et al., 1999; Pérez, 2008). Se impone sobre el proceso de transformación de los datos una restricción de *muchos a uno* de manera que una parte variable de valores directos pasan a tener una misma puntuación  $R_{til}$  conservando los valores así transformados únicamente una relación de orden con los directos. Perdiendo la relación lineal los valores de la escala directa de medida de la conducta y los de la  $R_{til}$  (Pérez, 2008).

Considerando el paralelismo entre la forma de los valores distribuidos con la escala de medida directa cuando la forma original sea normal el estadístico  $F$  será más potente y limitará el error del Tipo I a los valores nominales (Bono y Arnau, 1995; García, Pascual, Frías, Van Krunckelsven y Murgui, 2008). Sin embargo, cuando la forma de los valores distribuidos en la escala de medida directa presente una forma que no se ajuste a la distribución normal (e.g., el alargamiento de las colas) el estadístico  $F$  será menos potente que el  $f$  (García et al., 2002; Pérez et al., 2000; Pérez et al., 2004). Los valores extremos de la distribución implican un incremento considerable en el componente de varianza de error que ocasiona la consecuente pérdida de potencia (García, Musitu y Veiga, 2006). Fijando el tamaño del efecto, el tamaño muestral y el margen nominal del error del Tipo I,  $F$  será más potente para distribuciones padre normales y  $f$  para las que no lo sean (García et al., 2008).

## Método

Se simulan datos procedentes de la distribución padre normal y de otra asimétrica ( $\chi^2$  con muy pocos grados de libertad) con 100 réplicas de cada situación. Se varía el tamaño de efecto y de la muestra, manteniendo el error del *Tipo I* en el 5% y los valores de  $R$  en 2, 3, 4, 5 y 6. En todos los casos se aplica la prueba de la hipótesis con las distribuciones muestrales de  $F$ , asumiendo la situación más desfavorable para  $f$ : no se dispone de sus distribuciones exactas. Las simulaciones se realizaron con la versión 17 del SPSS.

## Resultados

*Tamaño del efecto pequeño.* Cuando la distribución fue normal se fijó el tamaño del efecto para que fuera  $d = 0,2$ , con las medias en 1,0 y 1,2 y las dos desviaciones típicas de 1,0. En  $\chi^2$  los datos de la primera condición se distribuían con 1,0 grado de libertad y los de la segunda con 1,3, siendo las desviaciones típicas de 1,41 y 1,61 (Evans et al., 1993), y el valor de  $d = 0,198$  (García et al., 2008). Los tamaños de cada muestra se mantuvieron en 100, 200, 300 y 800 (Tabla 1).

Tabla 1. Potencia y error del Tipo I cuando el tamaño del efecto es pequeño

$d = 0,2$		Potencia (1 - $\beta$ )				$d = 0,0$		Error del tipo I ( $\alpha = 0,05$ )			
$N(1,0, 1)$	$n_1 =$	100	200	300	800	$N(1, 1)$	$n_1 =$	100	200	300	800
$N(1,2, 1)$	$n_2 =$	100	200	300	800	$N(1, 1)$	$n_2 =$	100	200	300	800
	1 $F$	32	53	69	99		1 $F$	4	7	10	7
	2 $f(2)$	22	39	46	92		2 $f(2)$	7	6	9	6
	3 $f(3)$	27	48	58	96		3 $f(3)$	5	4	11	9
	4 $f(4)$	26	49	66	96		4 $f(4)$	5	4	8	8
	5 $f(5)$	29	50	65	97		5 $f(5)$	7	4	10	7
	6 $f(6)$	27	51	64	98		6 $f(6)$	3	4	8	8
$\chi(1,0)$	$n_1 =$	100	200	300	800	$\chi(1,0)$	$n_1 =$	100	200	300	800
$\chi(1,3)$	$n_2 =$	100	200	300	800	$\chi(1,0)$	$n_2 =$	100	200	300	800
	1 $F$	21	53	66	97		1 $F$	8	7	1	4
	2 $f(2)$	31	62	74	99		2 $f(2)$	6	7	2	4
	3 $f(3)$	29	63	78	100		3 $f(3)$	10	4	3	4
	4 $f(4)$	30	70	87	100		4 $f(4)$	6	6	4	2
	5 $f(5)$	32	76	91	100		5 $f(5)$	6	3	2	3
	6 $f(6)$	31	69	89	100		6 $f(6)$	8	4	3	5

Se observa en los resultados de las simulaciones cuando el tamaño del efecto es pequeño (Tabla 1) que la potencia es sistemáticamente mayor con la prueba  $F$  cuando la distribución padre es normal. Cuando la distribución padre es  $\chi^2$ , entonces el patrón de potencia es el opuesto: la prueba  $f$  supera a la  $F$ . Todos estos patrones se confirman cuando se incrementa el tamaño del efecto, y en general, al incrementar el valor de  $R$  los datos mejoran la potencia. No hay diferencias significativas ( $\alpha = 0,05$ ) entre los errores del Tipo I en la distribución normal,  $F(1, 22) = 0,17$ , entre la prueba  $F$  y la  $f$  ( $M = 7,00$ ,  $DT = 2,12$  vs.  $M = 6,65$ ;  $DT = 2,20$ ), tampoco en  $\chi^2$  hay diferencias,  $F(1, 22) = 1,42$  ( $M = 7,00$ ,  $DT = 2,12$  vs.  $M = 6,65$ ;  $DT = 2,20$ ).

*Tamaño del efecto mediano.* Cuando la distribución fue normal se fijó el tamaño del efecto para que fuera  $d = 0,5$ , con las medias en 1,0 y 1,5 y las dos desviaciones típicas de 1,0. En  $\chi^2$  los datos de la primera condición se distribuían con 1,0 grado de libertad y los de la segunda con 1,9, siendo las desviaciones típicas de 1,41 y 1,95, y el valor de  $d = 0,529$ . Los tamaños de cada muestra: 20, 30, 50 y 120 (Tabla 2).

Tabla 2. Potencia y error del Tipo I cuando el tamaño del efecto es mediano

$d = 0,5$		Potencia (1 - $\beta$ )				$d = 0,0$		Error del tipo I ( $\alpha = 0,05$ )			
$N(1,0, 1)$	$n_1 =$	20	30	50	120	$N(1, 1)$	$n_1 =$	20	30	50	120
$N(1,5, 1)$	$n_2 =$	20	30	50	120	$N(1, 1)$	$n_2 =$	20	30	50	120
	1 $F$	36	48	70	97		1 $F$	6	7	7	4
	2 $f(2)$	15	36	55	85		2 $f(2)$	2	7	7	5
	3 $f(3)$	24	31	63	92		3 $f(3)$	7	3	10	4
	4 $f(4)$	35	39	65	95		4 $f(4)$	13	6	8	4
	5 $f(5)$	<b>38</b>	44	67	95		5 $f(5)$	8	7	7	3
	6 $f(6)$	32	37	66	96		6 $f(6)$	7	4	9	5
$\chi(1,0)$	$n_1 =$	20	30	50	120	$\chi(1,0)$	$n_1 =$	20	30	50	120
$\chi(1,9)$	$n_2 =$	20	30	50	120	$\chi(1,0)$	$n_2 =$	20	30	50	120
	1 $F$	52	61	82	99		1 $F$	2	4	2	6
	2 $f(2)$	<b>40</b>	71	91	99		2 $f(2)$	0	6	5	5
	3 $f(3)$	57	71	96	100		3 $f(3)$	2	4	8	5
	4 $f(4)$	69	77	93	100		4 $f(4)$	4	5	8	4
	5 $f(5)$	67	81	94	100		5 $f(5)$	5	4	8	7
	6 $f(6)$	66	78	95	100		6 $f(6)$	1	5	7	6

Cuando el tamaño del efecto es mediano (Tabla 2) se aprecian dos excepciones que se han marcado en negrita. No hay diferencias significativas ( $\alpha = 0,05$ ) entre los errores del Tipo I en la distribución normal,  $F(1, 22) = 1,01$ , entre la prueba  $F$  y la  $f$  ( $M = 6,00, DT = 1,22$  vs.  $M = 6,30; DT = 2,59$ ), tampoco en  $\chi^2$  hay diferencias,  $F(1, 22) < 0,01$  ( $M = 3,50, DT = 1,66$  vs.  $M = 4,95; DT = 2,13$ ).

Tamaño del efecto grande. Cuando la distribución fue normal se fijó el tamaño del efecto para que fuera  $d = 0,8$ , con las medias en 1,0 y 1,8 y las dos desviaciones típicas de 1,0. En  $\chi^2$  los datos de la primera condición se distribuían con 1,0 grado de libertad y los de la segunda con 2,5, siendo las desviaciones típicas de 1,41 y 2,24, y el valor de  $d = 0,802$ . Los tamaños de cada muestra: 5, 10, 15 y 40 (Tabla 3).

Tabla 3. Potencia y error del Tipo I cuando el tamaño del efecto es grande

$d = 0,8$		Potencia (1 - $\beta$ )				$d = 0,0$		Error del tipo I ( $\alpha = 0,05$ )			
$N(1,0, 1)$	$n_1 =$	5	10	15	40	$N(1, 1)$	$n_1 =$	5	10	15	40
$N(1,8, 1)$	$n_2 =$	5	10	15	40	$N(1, 1)$	$n_2 =$	5	10	15	40
	1 $F$	21	58	67	99		1 $F$	5	6	6	4
	2 $f(2)$	0	26	44	83		2 $f(2)$	0	3	6	3
	3 $f(3)$	12	31	51	92		3 $f(3)$	5	1	6	4
	4 $f(4)$	<b>26</b>	53	52	97		4 $f(4)$	7	12	6	6
	5 $f(5)$	19	44	62	96		5 $f(5)$	5	5	6	3
	6 $f(6)$	7	54	<b>68</b>	97		6 $f(6)$	5	6	6	4
$\chi(1,0)$	$n_1 =$	5	10	15	40	$\chi(1,0)$	$n_1 =$	5	10	15	40
$\chi(2,5)$	$n_2 =$	5	10	15	40	$\chi(1,0)$	$n_2 =$	5	10	15	40
	1 $F$	21	52	65	96		1 $F$	1	3	2	4
	2 $f(2)$	<b>0</b>	<b>34</b>	<b>50</b>	<b>94</b>		2 $f(2)$	0	2	4	2
	3 $f(3)$	26	<b>44</b>	<b>65</b>	99		3 $f(3)$	2	3	6	2
	4 $f(4)$	41	69	69	98		4 $f(4)$	7	8	5	2
	5 $f(5)$	28	57	76	98		5 $f(5)$	3	2	7	1
	6 $f(6)$	23	64	75	98		6 $f(6)$	2	6	7	2

Cuando el tamaño del efecto es grande (Tabla 3) se aprecian ocho excepciones que se han marcado en negrita. No hay diferencias significativas ( $\alpha = 0,05$ ) entre los errores del Tipo I en la distribución normal,  $F(1, 22) = 2,28$ , entre la prueba  $F$  y la  $f$  ( $M = 5,25$ ,  $DT = 0,83$  vs.  $M = 4,95$ ;  $DT = 2,40$ ), tampoco en  $\chi^2$  hay diferencias,  $F(1, 22) < 0,96$  ( $M = 2,50$ ,  $DT = 1,12$  vs.  $M = 3,65$ ;  $DT = 2,33$ ).

### Discusión

Los resultados del trabajo confirman de manera bastante satisfactoria las previsiones de que la prueba  $F$  es más potente para distribuciones con la forma normal y la  $f$  para cuando no lo son (una distribución asimétrica) (García et al., 2002; Pérez et al., 2000; Pérez et al., 2004). También es importante destacar que los resultados confirman que el error del Tipo I se mantiene en márgenes semejantes para las dos pruebas independientemente de la forma de la distribución. Las limitaciones de este trabajo proceden del reducido número de réplicas que se han realizado, aspecto que se deberá tener en cuenta en distintos trabajos. Si bien los resultados son extraordinariamente buenos con las muestras grandes que requieren un tamaño del efecto pequeño, también es cierto que al disminuir el tamaño del efecto y reducir el tamaño de las muestras, algunos resultados (muestras pequeñas, menos de 20 observaciones, y con  $R$  muy pequeño, 2 y 3) no fueron tan claros (García et al., 2002). Sucesivos trabajos tienen que determinar si se trata de que las réplicas fueron pocas o más bien se trata de que el estadístico  $f$  requiere muestras mayores cuando  $R$  es muy pequeño (menor de 4) aunque el tamaño del efecto sea grande.

### Referencias

- Bono, R. & Arnau, J. (1995). General considerations about statistical power studies. *Anales de Psicología*, 11, 193-202.
- Evans, M., Hastings, N. A. J. & Peacock, J. B. (1993). *Statistical distributions* (2nd ed.). New York: J. Wiley.
- García, J. F., Frías, M. D. & Pascual, J. (1999). *Los diseños de la investigación experimental: Comprobación de las hipótesis*. Valencia, Spain: Cristóbal Serrano Villalba.
- García, J. F., Musitu, G. & Veiga, F. H. (2006). Autoconcepto en adultos de España y Portugal [Self-concept in adults from Spain and Portugal]. *Psicothema*, 18, 551-556.
- García, J. F., Pascual, J. & Frías, M. D. (2002). F-sample distribution -f- when the measurement scale is discrete and rectangular. *Metodología de las Ciencias del Comportamiento*, 4, 219-223.
- García, J. F., Pascual, J., Frías, M. D., Van Krunkelsven, D. & Murgui, S. (2008). Design and power analysis: n and confidence intervals of means. *Psicothema*, 20, 933-938.
- Pérez, F. G., Llobell, J. P. & Navarro, M. D. F. (2004). Tables of f statistic with one degree of freedom in the numerator. *Metodología de las Ciencias del Comportamiento*, 6, 237-243.
- Pérez, J. F. G. (2008). *Métodos de investigación, diseño y técnicas en las ciencias del comportamiento*. Valencia, Spain: Palmero Ediciones.
- Pérez, J. F. G., Navarro, M. D. F. & Llobell, J. P. (2000). Randomness tests versus F-distribution when the measurement scale is discrete. *Psicothema*, 12, 253-256.

# EVALUACIÓN DEL COMPORTAMIENTO DE SELECTORES DE ESTRUCTURAS DE COVARIANZA EN DISEÑOS DE MEDIDAS REPETIDAS

Pablo Livacic-Rojas<sup>1</sup>, Guillermo Vallejo<sup>2</sup> y Paula Fernández García<sup>2</sup>

<sup>1</sup>Universidad de Santiago de Chile

<sup>2</sup>Universidad de Oviedo

Se evaluó en dos fases el comportamiento del Criterio de Akaike (AIC) para seleccionar estructuras de covarianza y la potencia respecto a otros selectores. La primera fase, analizó la frecuencia de selección de la estructura de covarianza y los niveles de error entre AIC y el Modelo Correctamente Identificado (MCI). Mediante el método de simulación Montecarlo e implementado con el paquete estadístico SAS 9.1 se utilizó un diseño split-plot. Las variables manipuladas fueron: el tamaño muestral, la relación entre el tamaño de los grupos y el de las matrices de dispersión, el tipo de matrices de dispersión y la forma de la distribución. En la segunda fase se analizó la potencia para los efectos principales entre AIC, el procedimiento de Brown-Forsythe (BF) y MCI. Las variables manipuladas fueron: el tamaño muestral, la relación entre el tamaño de los grupos y el de las matrices de dispersión, el tipo de matrices de dispersión y los patrones de medias. Los resultados muestran que AIC seleccionó mayormente estructuras de covarianza heterogéneas y exhibió tasas de error mayor que MCI. Respecto a la potencia, BF tiene niveles de más altos que AIC y MCI.

Una de las dificultades más relevantes cuando se analizan datos que utilizan diseños de medidas repetidas ocurre cuando los datos no se ajustan a una matriz esférica y cuando las matrices de covarianza son heterogéneas. Entre los diferentes procedimientos que se han desarrollado para tratar el impacto negativo sobre las tasas de error de tipo I para evaluar los efectos principales y de la interacción, Vallejo y Livacic-Rojas (2005), Vallejo y Ato (2006), Vallejo, Moris y Conejo (2006) señalan que es posible aplicar un test robusto mediante el procedimiento modificado de Brown Forsythe (BF), el cual no requiere asumir la homogeneidad de las matrices de covarianza. En tal sentido, es necesario que el investigador acuda a procedimientos de búsqueda semiautomática (previo al contraste de las hipótesis referidas al diseño) a fin de conocer la estructura que subyace a los datos y, con ello, incrementar la sensibilidad para detectar los efectos de los tratamientos (Kowalchuk, Keselman, Algina y Wolfinger, 2004) y evaluarlos con mayor precisión (mediante estadísticos que se ajusten mejor a las características reales de los datos). Este asunto es de particular importancia debido a que, a su vez, los selectores de estructuras de covarianza muestran una baja de elección correcta de la estructura de covarianza conforme a la distribución que presentan los datos del estudio (ver también en Akaike, 1974 y Wolfinger R. 1996). Al realizar este tipo de análisis, se pretende analizar la tasa empírica del error que cometen los diferentes procedimientos de selección automática, así como también, analizar los niveles de la potencia para estimar parámetros y realizar inferencias con más precisión. A su vez, es importante señalar la importancia que tiene el estudio de los efectos de la interacción en estos diseños, los cuales, a su vez presentan entre varios de los eventuales riesgos es la presencia de la dependencia entre las observaciones a través del tiempo (Livacic-Rojas, Vallejo y Fernández, 2010).

El objetivo del presente trabajo fue realizar en la primera fase un análisis de la frecuencia de selección de estructuras de covarianza, a saber, Simetría Compuesta (CS), No Estructurada (NE), Autorregresiva de orden uno [AR(1)], Huynh-Feldt (HF), Simetría Compuesta Heterogénea (CSH), Autorregresiva de Orden Uno Heterogénea [ARH(1)], Coeficientes Aleatorios Lineal (CAL), No Estructurada Heterogénea (NEJ), Huynh-Feldt Heterogénea (HFJ), Autorregresiva Heterogénea ARHJ y Coeficientes Aleatorios Heterogénea (CAJ)) y las tasas de error de tipo I entre AIC y MCI para los efectos principales. En la segunda fase, se hizo una evaluación de los niveles de la potencia para los efectos principales entre AIC, BF y MCI.

### *Procedimientos del estudio*

Según Vallejo *et. al.* (2005), el modelo mixto lineal general para analizar medidas repetidas extiende el modelo lineal clásico a situaciones en que no se requieren los supuestos de independencia y homogeneidad y, a otras donde las variables son tanto fijas como aleatorias. En su forma estándar, dicho modelo puede ser escrito como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (1)$$

donde  $\mathbf{y}$  es un vector de datos observados,  $\boldsymbol{\beta}$  contiene el efecto fijo desconocido a ser estimado con una matriz de diseño conocida  $\mathbf{X}$ ,  $\mathbf{u}$  es un vector desconocido de efectos aleatorios con una matriz de diseño conocida  $\mathbf{Z}$ , y  $\mathbf{e}$ , representa el vector de datos del error desconocido cuyos elementos no requieren ser independientes y homogéneos.

Por su parte, el procedimiento BF aborda el problema de encontrar una aproximación a la distribución de las matrices sumas de cuadrados y productos cruzados (SC&PC) debidas a la hipótesis y al error, respectivamente, cuando las covarianzas son heterogéneas. El enfoque BF utiliza el criterio de los mínimos cuadrados ordinarios y se basa en un modelo donde los efectos aleatorios son tratados como fijos y los parámetros de la matriz de dispersión como arbitrarios. La matriz SC&PC correspondiente al error se estima usando un enfoque similar al estimador Sandwich, utilizado comúnmente con el método de ecuaciones de estimación generalizada propuesto por Liang y Zeger (1986), para acomodar la dependencia serial propia de los estudios longitudinales. Esta forma de construir la matriz SS&CP debida al error, asegura que bajo hipótesis nula su valor esperado coincida con el de la matriz SS&CP referida a hipótesis. Esta solución es sustancialmente menos vulnerable a la violación de los supuestos distribucionales de homogeneidad y normalidad multivariada que el tradicional enfoque MANOVA. Aunque el enfoque BF solventa alguno de los problemas que plantea el enfoque MANOVA cuando las matrices de dispersión son heterogéneas, sin embargo, otros permanecen y son básicamente cuatro en contextos donde el control experimental es limitado: (1) Desgaste de la muestra; (2) Número de ocasiones en que se realizan las mediciones sea menor que el número de participantes; (3) No permite manejar covariadas cambiantes; (4) Puede resultar ineficiente cuando la estructura de la matriz de dispersión requiera estimar un número reducido de parámetros (Livacic-Rojas, Vallejo y Fernández 2006).

### Método

El estudio se desarrolló en dos fases mediante el método de simulación Montecarlo implementado con el paquete estadístico SAS 9.1 y con un diseño split-plot con un factor entre sujetos ( $p=3$ ) y un factor intra sujetos ( $k=4$ ). La primera fase fue un análisis comparativo de la frecuencia de selección de las estructuras de covarianza y los niveles de error entre Criterio Selección de Akaike (AIC) y Modelo Correctamente Identificado (MCI) para los efectos principales y los efectos de la interacción. Las variables manipuladas fueron cuatro: el tamaño de muestra total, la relación entre el tamaño de los grupos y el de las matrices de dispersión, el tipo de matrices de dispersión y la forma de la distribución. En la segunda fase se evaluaron comparativamente los niveles de la potencia para los efectos principales entre AIC, el procedimiento de Brown-Forsythe (BF) y el MCI. Se manipularon cuatro variables: el tamaño de muestra total, la relación entre el tamaño de los grupos y el de las matrices de dispersión, el tipo de matrices de dispersión y los patrones de medias.

### Resultados

La tabla 1 (Fase I del estudio) muestra que AIC escoge en un mayor porcentaje estructuras de covarianza heterogéneas en vez de la estructura de covarianza original y de la relación de éstas con el tamaño grupal. Para la estructura de covarianza que realiza una selección más ajustada es para la

ARH (1), particularmente, cuando la distribución es normal, leve y moderadamente sesgada. Cuando la distribución es fuertemente sesgada, AIC escoge en un mayor porcentaje la estructura de covarianza UNJ. La mayoría de los cambios observados en la selección de estructuras de covarianza se observan asociados a la distribución moderadamente desviada de la normalidad.

Tabla 1. Porcentaje de selección de estructuras de covarianza original (CA, ARH (1) y NE)

Distribución de Datos	Estructura de Covarianza		
	NE	ARH(1)	CA
Normal	ARHJ (41.10)	ARHJ (55.90)	CAJ (86.97)
	NEJ (22.26)	CAJ (29.66)	
	CAJ (20.81)		
Levemente Sesgada	ARHJ (34.57)	ARHJ (53.51)	CAJ (74.31)
	NEJ (24.39)	CAJ (23.32)	NEJ (13.93)
	CAJ (16.20)	NEJ (13.42)	
Moderadamente Sesgada	NEJ (53.97)	ARHJ (43.32)	NEJ (43.90)
	CAJ (27.66)	NEJ (13.03)	CAJ (39.92)
		CAJ (13.68)	
Fuertemente Sesgada	NEJ (68.71)	NEJ (55.54)	NEJ (78.93)
	CAJ (24.44)	ARHJ (38.80)	

Leyenda: Estructura Covarianza Autorregresiva Heterogénea (ARH1); Estructura Covarianza Autorregresiva de Orden Uno Heterogénea (ARHJ); Estructura de Covarianza No Estructurada (NE); Estructura de Covarianza No Estructura Heterogénea (NEJ); Estructura de Covarianza de Coeficientes Aleatorios (CA); Estructura de Covarianza Coeficientes Aleatorios Heterogénea (CAJ).

Respecto a las Tasas de error de Tipo I (Ver Tabla nº2), tanto AIC como MCI sobrepasan los niveles nominales en una mayor proporción para los efectos principales que para los efectos de la interacción. Sin embargo, entre ambos, el AIC lo hace en una mayor proporción para las diferentes estructuras de covarianza y con independencia de su relación con el tamaño de los grupos como del tipo de distribución de los datos. La mayor proporción de errores cometidos se encuentra asociado a la estructura de covarianza de Coeficientes Aleatorios (CA).

Tabla 2. Porcentajes comisión de Tasas de Error de Tipo I en función de las estructuras de covarianza original (CA, ARH (1) y NE) y el criterio de selección de Akaike

Tasas de Error Tipo I	Estructura de Covarianza		
	NE	ARH(1)	CA
<u>AIC</u>			
Efectos Principales	61.11	61.11	66.67
Efectos Interacción	47.22	36.11	41.66
<u>MCI</u>			
Efectos Principales	58.33	66.67	63.89
Efectos Interacción	25.00	22.22	38.89

Leyenda: Estructura Covarianza Autorregresiva Heterogénea (ARH1); Estructura Covarianza Autorregresiva de Orden Uno Heterogénea (ARHJ); Estructura de Covarianza No Estructurada (NE); Estructura de Covarianza No Estructura Heterogénea (NEJ); Estructura de Covarianza de Coeficientes Aleatorios (CA); Estructura de Covarianza Coeficientes Aleatorios Heterogénea (CAJ); Criterio de Selección de Akaike (AIC); Modelo Correctamente Identificado (MCI).

Las tablas 3 y 4 (Fase II del estudio) muestran los niveles de potencia promedio que los selectores de estructuras de covarianza BF y AIC tienen cuando la estructura de covarianza es la correcta. En términos generales, los selectores de estructuras de covarianza muestran rendimiento similar (BF levemente superior a AIC) cuando la distribución se desvía de la normalidad para los efectos principales (potencia moderada; entre 0.40-0.60). Esto, se encuentra asociado a las matrices de covarianza NE y ARH(1). La situación cambia cuando la matriz de covarianza es CA dado que los niveles de potencia son bajos (potencia baja; entre 0.20-0.30). De igual modo, los niveles de potencia decrecen en función de la relación entre el tamaño de los grupos y las matrices de dispersión del siguiente modo: Neutro, Negativo y Positivo.

En términos específicos (ver Tabla nº3), cuando el selector de estructura de covarianza es BF los niveles de potencia superiores respecto de AIC y MCI en casi todas las estructuras de covarianza (excepto para ARH (1)). A su vez, bajo NE los niveles de potencia son superiores a dos puntos promedio respecto de ARH (1) y en una medida claramente superior para CA (superior a veinte puntos promedio) para los diferentes tipos de sesgo de los datos.

*Tabla 3. Niveles de Potencia Promedio para selectores de estructuras de covarianza BF cuando la estructura de covarianza es la correcta*

Estructuras Covarianza	Sesgo de los Datos		
	Leve	Moderado	Fuerte
	BF	AIC	MCI
NE	58.26	58.11	49.26
ARH (1)	56.62	57.72	48.35
CA	32.03	29.83	21.08

Leyenda: Estructura de Covarianza No Estructurada (NE); Estructura Covarianza Autorregresiva Heterogénea (ARH1); Estructura de Covarianza de Coeficientes Aleatorios (CA); Procedimiento de Brown y Forsythe (BF); Criterio de Selección de Akaike (AIC); Modelo Correctamente Identificado (MCI).

A su vez, cuando el selector de estructura de covarianza es AIC (Ver Tabla nº4), BF nuevamente exhibe niveles de potencia superiores respecto de AIC y MCI en casi todas las estructuras de covarianza (excepto para ARH (1)). De igual modo, bajo NE los niveles de potencia son superiores a tres puntos promedio respecto de ARH (1) y en una medida claramente superior para CA (superior a veinticinco puntos promedio) para los diferentes tipos de sesgo de los datos.

*Tabla 4. Niveles de Potencia Promedio para selectores de estructuras de covarianza AIC cuando la estructura de covarianza es la correcta*

Estructuras Covarianza	Sesgo de los Datos		
	Leve	Moderado	Fuerte
	BF	AIC	MCI
NE	58.59	57.72	47.79
ARH (1)	52.77	53.51	44.13
CA	24.72	23.53	16.43

Leyenda: Estructura de Covarianza No Estructurada (NE); Estructura Covarianza Autorregresiva Heterogénea (ARH1); Estructura de Covarianza de Coeficientes Aleatorios (CA); Procedimiento de Brown y Forsythe (BF); Criterio de Selección de Akaike (AIC); Modelo Correctamente Identificado (MCI).

## Discusión

El presente estudio evaluó en dos fases el comportamiento de AIC para seleccionar estructuras de covarianza y analizó la potencia de éste respecto de BF y del MCI. Los resultados para la fase uno del estudio muestran que AIC seleccionó mayormente estructuras de covarianza heterogéneas en vez de la original conforme la distribución se alejaba de la normalidad. Respecto a las tasas de error, el AIC sólo exhibió una mayor frecuencia de error respecto al MCI cuando la distribución de los datos se aleja de la normalidad para los diferentes tipos de efecto. Es oportuno señalar que los niveles de error promedio se observan mayormente para los efectos principales que para los de la interacción, situación particularmente relevante en los diseños con datos longitudinales. Para la fase dos, se observa que BF y AIC exhiben un rendimiento similar (niveles de potencia) para los efectos principales cuando la distribución era normal y alejada de la normalidad como cuando los patrones de medias eran cercanos y lejanos entre sí y asociados a NE.

## Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on automatic Control*, *AC-19*, 716-723.
- Kowalchuck, R.K., Keselman, H.J., Alginan, J. y Wolfinger. (2004). The analysis of repeated measurements with mixed-model adjusted F test. *Educational and Psychological Measurements*, *64*, 224-242.
- Livacic-Rojas, P., Vallejo, G. y Fernández, P. (2006). Diferentes procedimientos alternativos para evaluar la robustez mediante diseños de medidas repetidas. *Revista Latinoamericana de Psicología*, *38*, 3, 579-598.
- Livacic-Rojas, Vallejo, G. y Fernández, P. (2010). Analysis of Type I Error rates of Univariate and Multivariate Procedures in Repeated Measures Designs. *Communications in Statistics - Simulation and Computation*, *39*: 3, 624-640.
- Vallejo, G. & Livacic-Rojas, P. (2005). A comparison of two procedures for analyzing small sets of repeated measures data. *Multivariate Behavioral Research*, *40*, 179-205.
- Vallejo, G. & Ato, M. (2006). Modified Brown-Forsythe procedure for testing interaction effects in split-plot designs. *Multivariate Behavioral Research*, *41*, 549-578.
- Vallejo, G., Moris, J. & Conejo, N. M. (2006). A SAS/IML program for implementing the modified Brown-Forsythe procedure in repeated measures designs. *Computer Methods & Programs in Biomedicine*, *83*, 169-177.
- Wolfinger R. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environment Statistics*, *Volume 1, Number 2*, (205-230.).

# APLICACIÓN DE PROCEDIMIENTOS HETEROSCEDÁSTICOS TRANSFORMADOS AL ANÁLISIS DE DISEÑOS FACTORIALES DESEQUILIBRADOS

Guillermo Vallejo<sup>1</sup>, Manuel Ato<sup>2</sup>, Paula Fernández García<sup>1</sup> y Ellián Tuero<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Murcia

El presente trabajo usa la transformación de Hall para combatir el sesgo y dos pruebas heteroscedásticas para combatir la heterogeneidad, a saber: el estadístico Tipo Box desarrollado por Brunner, Dette y Munk (BDM) y el estadístico Welch-James desarrollado por Johansen (WJ). Las características operantes de ambos enfoques fueron evaluadas en un diseño factorial cuando se violaba separada y conjuntamente los supuestos de normalidad y homogeneidad. Cuando las varianzas fueron heterogéneas y los datos severamente sesgados, ningún enfoque se comportó de manera completamente satisfactoriamente. No obstante, el enfoque BDM controló las tasas de error sustancialmente mejor que el estadístico WJ.

Los diseños factoriales hacen referencia a disposiciones experimentales en las que se estudia simultáneamente la acción de dos o más variables independientes, con los niveles de cada una de ellas combinado con los niveles de las restantes. Su empleo reduce los costos, incrementa la información y potencia la base inductiva de las conclusiones. Si las fuentes de variación (FV) son independientes, la simplicidad analítica e interpretativa de estos diseños es equivalente a los del análisis de la varianza de un factor (ANOVA). Desafortunadamente, la situación cambia radicalmente cuando las FV no son ortogonales.

Durante las últimas décadas ha surgido un extenso debate acerca de cuál es el mejor método para analizar diseños factoriales carentes de balanceo. Tras decantarse por soluciones mínimo cuadráticas ordinarias (MCO) ejecutadas comparando modelos, gran parte del debate se ha centrado en la elección de la suma cuadrática (SC) más adecuada para contrastar las hipótesis de interés; elección fácil de abordar cuando se conocen las causas del desgaste de muestra. Aquí hemos asumido que el mecanismo responsable de los datos faltantes es independiente de todas las variables y covariables presentes en el estudio.

Cuando se prueban los efectos del diseño factorial adoptando la solución MCO basada en la comparación de modelos, se asumen errores normales y homogéneos. El usual ANOVA es demasiado liberal (conservador) cuando el tamaño de muestra está correlacionado negativamente (positivamente) con el de la varianza. Diversas soluciones están disponibles para vencer el impacto negativo de la heterogeneidad. Aunque tanto el esfuerzo computacional, como la calidad de las inferencias revela importantes diferencias entre ellas. Sirva de botón de muestras las cinco que siguen: (a) Las centradas en estabilizar las varianzas cambiando la escala de los datos. (b) Las basadas en corregir los valores críticos ajustando los grados de libertad a partir de la muestra de datos disponible. (c) Las basadas en ponderar las medias cuadráticas de la razón F por la desigualdad de las varianzas. (d) Las caracterizadas por utilizar métodos de estimación basados en el principio de máxima verosimilitud. (e) Las centradas en calcular los valores críticos usando técnicas de remuestreo.

De una u otra manera, todas las soluciones han sido evaluadas por nuestro grupo. Por ejemplo, Vallejo, Ato, Fernández y Livacic-Rojas (2008a) y Vallejo, Fernández y Livacic-Rojas (2008b) desarrollaron una solución basada en generalizar el enfoque Brown-Forsythe. Los resultados pusieron

de relieve que cuando los datos eran normales el procedimiento controlaba adecuadamente las tasas de error, inclusive bajo condiciones elevadas de heterogeneidad, aunque resultó bastante sensible a la falta de normalidad. Recientemente, Vallejo, Fernández y Livacic-Rojas (2009a) han investigado como afecta al desempeño de los enfoques de Brunner, Dette y Munk (BDM, 1997), Welch-James (WJ) desarrollado por Johansen (1980), Mixto, BDM con la transformación Box-Cox (BC), WJ-BC y Mixto-BC la violación separada y conjunta de los supuestos de normalidad y homogeneidad. Cuando la distribución era simétrica, todos los enfoques controlaban las tasas de error. Cuando la distribución era asimétrica, ningún enfoque controlaba las tasas de error. Este patrón de resultados fue consistente a través de las SC Tipo I, Tipo II y Tipo III empleadas para contrastar las hipótesis. Más recientemente, Vallejo, Fernández y Livacic-Rojas (2009b) han examinado el comportamiento de los enfoques WJ y BDM con los valores críticos derivados empíricamente, en vez de hacerlo analíticamente desde la teoría normal. Los procedimientos se comportaban de un modo liberal cuando los datos eran extraídos a partir de distribuciones sesgadas y el grado de heterogeneidad de las varianzas era elevado (1:36).

La meta actual se ha centrado en transformar las pruebas heteroscedásticas WJ y BDM usando un procedimiento desarrollado por Hall (1992) para corregir el sesgo de la prueba  $t$ . Así pues, este trabajo usa la transformación de Hall para combatir el sesgo y dos tests heteroscedásticos para combatir la heterogeneidad. Las características operantes de los métodos propuestos son evaluadas numéricamente cuando se viola separada y conjunta de los supuestos de normalidad y homogeneidad.

### *Definición de las pruebas estadísticas*

Considérese un diseño en el cual los  $n_{jk}$  ( $\sum_j \sum_k n_{jk} = N; j=1, \dots, J; k=1, \dots, K$ ) participantes de cada una de las  $JK$  celdas sean medidos en una única ocasión. Usando el modelo de medias, la respuesta dada por el  $i$ -ésimo participante en el  $j$ -ésimo nivel de  $A$  y en el  $k$ -ésimo nivel de  $B$  es representada mediante la ecuación  $Y_{ijk} = \mu_{jk} + \epsilon_{ijk}$  con  $\epsilon_{ijk} \sim N(0, \sigma^2)$ . Bajo este modelo, el investigador suele estar interesado en determinar los efectos de  $A$ ,  $B$  y  $AB$ . Como hemos mencionado anteriormente, cuando el número de participantes varía de celda a celda existen, al menos, tres tipos diferentes de sumas cuadrados. Dado que generalmente cada una de ellas contrasta hipótesis nulas ( $H_0$ ) distintas, excepto para el caso de la interacción, la interpretación de los resultados puede resultar confusa; sobre todo, cuando algún efecto resulta significativo utilizando un tipo de suma de cuadrados y no significativo utilizando otro (Ato & Vallejo, 2007).

Por tanto, antes de describir los enfoques procede definir las hipótesis a probar. De acuerdo con Searle (1987), las  $H_0$  asociadas con el método basado en medias no ponderadas usadas en el presente trabajo son:

$$\begin{aligned}
 H_0(A) &= \sum_{j=1}^J \frac{\mu_{jk}}{K} - \sum_{j=1}^J \frac{\mu_{j'k}}{K} = 0, \forall j \text{ y } j', \\
 H_0(B) &= \sum_{k=1}^K \frac{\mu_{jk}}{J} - \sum_{k=1}^K \frac{\mu_{jk'}}{J} = 0, \forall k \text{ y } k', \\
 H_0(AB) &= \mu_{jk} - \mu_{j'k} - \mu_{jk'} + \mu_{j'k'} = 0, \forall j, j' \text{ y } k, k'.
 \end{aligned} \tag{1}$$

Desde la perspectiva de comparación de modelos, con este enfoque cada efecto principal se prueba comparando la suma de cuadrados del error correspondiente al modelo no aditivo con la suma de cuadrados residual obtenida tras eliminar del modelo completo el efecto referido a la hipótesis nula de interés. En concreto, los modelos a comparar para contrastar  $H_0(A)$  serían:  $y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$  versus  $y_{ijk} = \mu + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$ . Este procedimiento de estimación se corresponde con la suma de

cuadrados Tipo III ofrecida por defecto en los principales paquetes estadísticos, y es similar al análisis de medias no ponderadas descrito por Horst y Edwards (1982); de hecho, los resultados obtenidos con ambos enfoques son idénticos si los factores tienen sólo dos niveles. Siempre que el investigador esté interesado en probar aquellas hipótesis que surgen de su ámbito de trabajo y no aquellas otras que dependan del número de sujetos que aparecen en las celdas, este enfoque representa la mejor opción.

*Transformación de Hall para remover el sesgo de una distribución asimétrica.*

Para eliminar los efectos del sesgo de un estadístico asimétrico, Hall (1992) propuso la transformación siguiente:

$$T_H = T_H(S) = S + n^{-1}b\hat{\gamma} + a\hat{\gamma}S^2 + \frac{1}{3}a^2\hat{\gamma}^2S^3, \tag{2}$$

donde  $S = (\hat{\theta} - \theta) / \hat{\sigma}$ ,  $\hat{\theta} = \bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ ,  $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ ,  $\hat{\gamma}$  es el sesgo definido por  $\hat{\gamma} = n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^3 / \hat{\sigma}^3$ ,  $a = 1/3$ ,  $b = 1/6$  y  $Y_i (i = 1, 2, \dots, n)$  es una muestra aleatoria independiente de un solo grupo.

Para aplicar el método descrito en el trabajo de Hall para manejar el problema de la asimetría en un ANOVA factorial, reemplazamos el de tamaño muestra  $n$ , la media muestral  $\hat{\theta}$ , la varianza muestral  $\hat{\sigma}^2$ , y el sesgo muestral  $\hat{\gamma}$  con  $n_{jk}$ ,  $(\bar{Y}_{jk} - \bar{Y}) = (\hat{\alpha}_j + \hat{\beta}_k + \hat{\gamma}_{jk})$ ,  $\hat{\sigma}_{jk}^2$  y  $n_{jk}^{-1} \sum_{i=1}^{n_{jk}} (Y_{ijk} - \bar{Y}_{jk})^3 / \hat{\sigma}_{jk}^3$ . Substituyendo estos valores en la ecuación (2), se sigue que la media muestral para la  $jk$ -ésima combinación de tratamiento con la transformación de Hall es dado por la expresión

$$\tilde{Y}_{jk} = \tilde{\sigma}_{jk}^{-1} \left[ (\bar{Y}_{jk} - \bar{Y}) + \frac{\hat{Y}_{jk}}{6\tilde{\sigma}_{jk}^2} + \frac{\hat{Y}_{jk}}{3\tilde{\sigma}_{jk}^4} (\bar{Y}_{jk} - \bar{Y})^2 + \frac{\hat{Y}_{jk}^2}{27\tilde{\sigma}_{jk}^8} (\bar{Y}_{jk} - \bar{Y})^3 \right], \tag{3}$$

donde  $\hat{Y}_{jk} = n_{jk}^{-2} \sum_{i=1}^{n_{jk}} (Y_{ijk} - \bar{Y}_{jk})^3$ ,  $\tilde{\sigma}_{jk}^2 = \hat{\sigma}_{jk}^2 / n_{jk}$  y  $\bar{Y} = \sum_{j=1}^a \sum_{k=1}^b \bar{Y}_{jk} / ab$ .

Después, para calcular los procedimientos WJ y BDM descritos en Vallejo et al. (2009, 2010) basados en la transformación de Hall, sustituimos el vector de medias de las muestras por el vector de medias transformado por el método de Hall,  $\tilde{\boldsymbol{\mu}} = (\tilde{Y}'_1, \dots, \tilde{Y}'_a)'$ , con  $\tilde{Y}_j = (\tilde{Y}_{j1}, \dots, \tilde{Y}_{jb})'$ . De acuerdo con lo dicho, los estadísticos heteroscedásticos transformados se pueden reescribir como sigue:

$$T_{HWJ} = (\mathbf{C}\tilde{\boldsymbol{\mu}})' (\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}')^{-1} (\mathbf{C}\tilde{\boldsymbol{\mu}}), \tag{4}$$

$$F_{HBDM} = N(\tilde{\boldsymbol{\mu}}'\mathbf{H}\tilde{\boldsymbol{\mu}}) / \text{tr}(\mathbf{H}_D\hat{\mathbf{S}}). \tag{5}$$

Bajo  $H_0$ :  $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$  es razonable aproximar la distribución nula de las pruebas  $T_{HWJ}$  y  $F_{HBDM}$  con la distribución  $F$  central con los grados de libertad especificados por Vallejo et al. (2010a) para los estadísticos  $T_{WJ}$  y  $F_{BDM}$

Finalmente, se puede construir un intervalo de confianza para los efectos del diseño adoptando la prueba estadística de  $t$ , pero sin agrupar las varianzas del error (véase Lix y Keselman, 1995). Por ejemplo, siguiendo a Guo y Luh (2000), el intervalo de confianza para los efectos diferenciales del factor fila al nivel de confianza  $(1-\alpha)$  sería

$$\sum_{k=1}^b (\bar{Y}_{jk} - \bar{Y}_{j'k}) \pm \sum_{k=1}^b (\tilde{\sigma}_{jk}^2 + \tilde{\sigma}_{j'k}^2)^{1/2} \{3[(1 + \hat{\gamma}(t_{\alpha/2, v} - \hat{\gamma}/6))^{1/3} - 1] / \hat{\gamma}\}, \tag{6}$$

donde  $\sum_{k=1}^b (\tilde{\sigma}_{jk}^2 + \tilde{\sigma}_{j'k}^2)^{1/2}$  es el estimador del error estándar de  $\sum_{k=1}^b (\bar{Y}_{jk} - \bar{Y}_{j'k})$ ,  $\{3[(1 + \hat{\gamma}(t_{\alpha/2, v} - \hat{\gamma}/6))^{1/3} - 1] / \hat{\gamma}\}$  es el valor crítico, cuyo valor ha sido determinado usando la función inversa de  $T_H(S)$  y  $t_{\alpha/2, v}$  es el cuantil  $(1 - \alpha/2)$  de la distribución  $t$  de Student con grados de libertad

$$\hat{v} = \left[ \sum_{k=1}^b (\tilde{\sigma}_{jk}^2 + \tilde{\sigma}_{j'k}^2) \right]^2 / \sum_{k=1}^b \left[ \frac{\tilde{\sigma}_{jk}^4}{(n_{jk} - 1)} + \frac{\tilde{\sigma}_{j'k}^4}{(n_{j'k} - 1)} \right]. \quad (7)$$

## Método

En orden a evaluar la robustez y sensibilidad de los tests citados, llevamos a cabo sendos estudios de simulación usando un diseño factorial de dos factores con  $J = 3$  y  $K = 5$ . El primer estudio se centró en comparar la robustez de los enfoques propuestos cuando se incumplían los supuestos de normalidad y homogeneidad. Para tal fin fueron manipuladas las cuatro variables siguientes:

- *Tamaño de muestra de las celdas.* El desempeño fue investigado usando dos tamaños:  $N = 104$  ( $n_{1k} = 2, 3, 4, 5, 7$ ,  $n_{2k} = 3, 6, 8, 10, 12$  y  $n_{3k} = 4, 6, 9, 12, 13$ ) y  $N = 208$  (el tamaño era doblado). El grado de desequilibrio era severo ( $c = 5$ )
- *Relación entre el tamaño de las celdas y el tamaño de las varianzas.* El desempeño se investigó bajo relaciones positivas y negativas. El tamaño de las varianzas fue:  $\sigma_{1k}^2 = 2, 4, 6, 9, 12$ ;  $\sigma_{2k}^2 = 4, 9, 16, 25, 36$  y  $\sigma_{3k}^2 = 6, 16, 30, 49, 73$ .
- *Forma de la distribución de la población.* En orden a investigar los efectos que la forma de la distribución tiene en la robustez de los enfoques, generamos datos desde distribuciones normales y no normales. Específicamente, además de la distribución normal ( $\gamma_1 = 0$ ;  $\gamma_2 = 0$ ), también fueron investigadas una distribución simétrica con un moderado grado de curtosis ( $\gamma_1 = 0$ ;  $\gamma_2 = 3$ ) y dos asimétricas. Una moderadamente sesgada ( $\gamma_1 = 2$ ;  $\gamma_2 = 6$ ) y otra fuertemente sesgada ( $\gamma_1 = 6$ ;  $\gamma_2 = 114$ ).

El segundo estudio se centró en comparar la sensibilidad de los enfoques cuando las tasas de error eran razonablemente controladas. Para evitar efectos de techo, seleccionamos valores de  $\mu_{jk}$  que ofreciesen potencias nominales de 0.60. Las potencias teóricas fueron calculadas usando el procedimiento WJ y la SC Tipo III. Para realizar los cálculos se utilizó un MACRO escrito en lenguaje SAS/IML y las condiciones del diseño fueron replicadas 10000 veces.

## Resultados

Como se pone relieve en la Figura 1, Cuando los datos eran extraídos desde distribuciones simétricas, tanto el enfoque HBDM como enfoque HWJ controlaban aceptablemente las tasas de error Tipo I. Aunque no se muestra en las figuras, bajo esta situación existían escasas diferencias entre los procedimientos examinados y los originales.

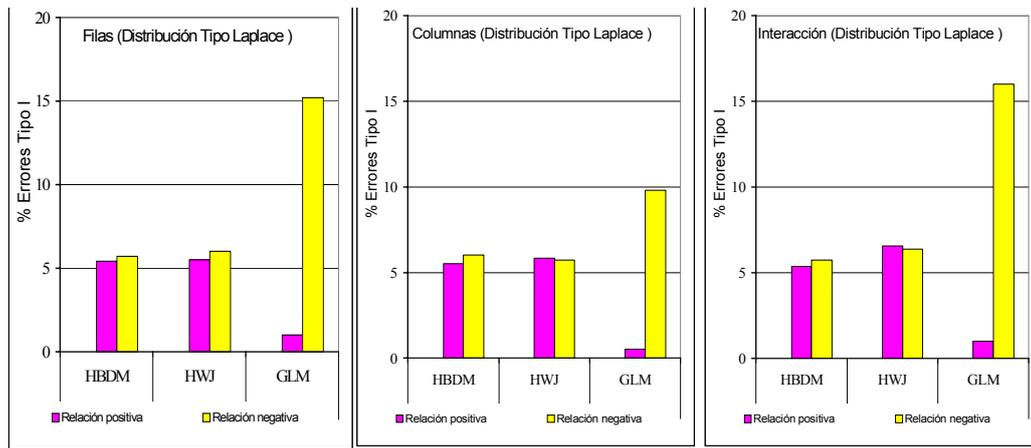


Figura 1. Resultados del estudio de robustez bajo una distribución tipo Laplace

Cuando los datos fueron generados desde distribuciones moderadamente sesgadas, el enfoque HBDM controló las tasas de error, sin embargo, el enfoque HWJ se volvía liberal cuando la relación entre el tamaño de las celdas y el tamaño de la varianzas era negativa. Cuando los datos fueron generados desde distribuciones severamente sesgadas ningún enfoque controló las tasas de error. El HBDM mostró una tendencia liberal en una de las veinticuatro condiciones examinadas, mientras que el enfoque HWJ lo hizo en diez. Las ejecuciones del GLM se vieron muy afectadas por la heterogeneidad y escasamente por la falta de normalidad. En la figura 2 aparecen recogidos los resultados obtenidos bajo la última condición reseñada.

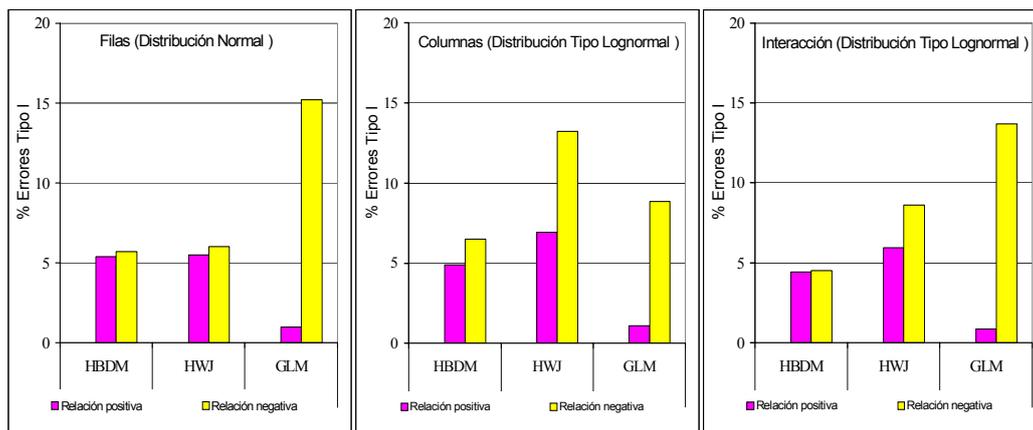


Figura 2. Resultados del estudio de robustez bajo una distribución tipo Lognormal

Con respecto a la potencia, ningún enfoque era uniformemente más potente que otro, no obstante, la HWJ fue generalmente más potente que la HBDM y muchas veces sustancialmente más potente. El patrón de diferencias entre ambas pruebas para los efectos principales de las filas y las columnas depende de la configuración de medias manipulada. Con otras configuraciones el patrón de resultados sería distinto. No obstante, las diferencias de potencia que favorecen al enfoque HBDM son siempre más pequeñas que las que favorecen al enfoque HWJ.

### Conclusión

Finalmente, resaltar que resulta difícil encontrar una solución completamente válida cuando las varianzas son heterogéneas y la distribución está severamente sesgada. En una futura investigación sería informativo examinar el desempeño del modelo lineal utilizando técnicas donde la distribución del término de error no se restrinja a ser normal y se permita relajar el supuesto de varianza constante. Estas técnicas, conocidas como modelos lineales generalizados, son una extensión del modelo lineal

general en el que la distribución del error no se limita a ser normal, sino que se amplía para dar cabida a muchas de las distribuciones encontradas cuando se trabaja con datos reales, incluidas las examinadas en el presente trabajo. Una información detallada de lo dicho se puede encontrar en McCulloch, Searle, y Neuhaus (2008). El procedimiento GLIMMIX de SAS permite ajustar modelos estadísticos cuando los supuestos de normalidad y homogeneidad de varianza no son necesariamente satisfechos (ver Schabenberger, 2007, para detalles y sugerencias). No obstante, esta solución requiere conocer la forma de la distribución y la relación entre el predictor lineal y la media de la función de distribución.

Reconocimientos. Este trabajo ha sido financiado mediante sendos proyectos de investigación concedidos por el MCI (Ref.: PSI-2008-03624-Ref.: PSI2005-09002).

### Referencias

- Ato, M. & Vallejo, G. (2007). *Diseños Experimentales en Psicología*. Madrid: Pirámide.
- Brunner, E., Dette, H. & Munk, A. (1997). Box-type approximations in heteroscedastic factorial designs. *Journal of the American Statistical Association*, 92, 1494-1503.
- Guo, J. H. & Luh, W. M. (2000). An invertible transformation two-sample trimmed t-statistic under heterogeneity and nonnormality. *Statistics and Probability Letters*, 49, 1-7.
- Hall, P. (1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B*, 54, 221-228.
- Horst, P. & Edwards, A. (1982). The k factorial experiment. *Psychological Bulletin*, 91, 190-192.
- Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, 67, 85-92.
- Lix, L. M. & Keselman, H. J. (1995). Approximate degrees of freedom tests: a unified perspective on testing for mean equality. *Psychological Bulletin*, 117, 547-560.
- McCulloch, C. E., Searle, S. R. & Neuhaus, J. W. (2008). *Generalized, Linear, and Mixed Models* (2nd ed.). New York: Wiley.
- Searle, S. R. (1987). *Linear Models for Unbalanced Data*. New York: Wiley.
- Schabenberger, O. (2007). Growing Up Fast: SAS® 9.2 Enhancements to the GLIMMIX Procedure. *Proceedings of the 2007 SAS Global Forum*. Cary, NC: SAS Institute Inc.
- Vallejo, G., Ato, M., Fernández, P. & Livacic-Rojas, P. E. (2008a). A practical method for analyzing factorial designs with heteroscedastic data. *Psychological Reports*, 102, 643-653.
- Vallejo, G., Fernández, P. & Livacic-Rojas, P. E. (2008b). Generalización del enfoque Brown-Forsythe a diseños factoriales. *Psicothema*, 20, 969-973.
- Vallejo, G., Fernández, P. & Livacic-Rojas, P. E. (2010a). Analysis of unbalanced factorial designs with heteroscedastic data. *Journal of Statistical Computation and Simulation*. DOI: 10.1080/0949650802482386.
- Vallejo, G., Fernández, P. & Livacic-Rojas, P. E. (2010b). Pruebas robustas para modelos ANOVA de dos factores con varianzas heterogéneas. *Psicológica: International Journal of Methodology and Experimental Psychology*, 31, 121-129.

# ANÁLISIS DE DISEÑOS LONGITUDINALES CON DATOS INCOMPLETOS

Guillermo Vallejo<sup>1</sup>, Paula Fernández García<sup>1</sup>, Ellián Tuero<sup>1</sup> y Pablo Livacic-Rojas<sup>2</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Universidad de Santiago de Chile

Este trabajo examinó el desempeño de la técnica de imputación múltiple (MI) seguida del enfoque MBF (Vallejo y Ato, 2006) cuando existía pérdida monótona de datos. Con propósitos de comparación también se incluyeron otros dos métodos diseñados para analizar datos longitudinales carentes de balanceo, a saber: el modelo lineal mixto (MLM) y el de las ecuaciones de estimación generalizadas ponderado (WGGE). Los resultados mostraron que ningún enfoque fue capaz de controlar las tasas de error en todas las condiciones investigadas. El enfoque MI-MBF se comportó adecuadamente cuando el interés se centró en el efecto entre grupos e interacción de éstos con las ocasiones de medida, pero resultó inaceptable para contrastar el efecto principal de las ocasiones. En este último caso, los enfoques MLM y WGGE exhibieron un desempeño superior.

La pérdida arbitraria o monótona de datos resulta bastante habitual en los estudios longitudinales, y soslayar esta cuestión acarrea graves riesgos para la gran mayoría de las metodologías estadísticas existentes. Como ha sido puesto de relieve por Fitzmaurice, Laird y Ware (2004), el desgaste de muestra tiene tres implicaciones importantes para el análisis longitudinal. Primero, cuando existen datos faltantes el diseño está desequilibrado, como resultado, la aplicación de los métodos que requieren datos completos se complica por causa de su rigidez. Segundo, en presencia de datos perdidos se reduce el tamaño de muestra, lo que acarrea un aumento en los errores estándar de los estimadores de los parámetros. Finalmente, bajo ciertas circunstancias los datos perdidos pueden introducir sesgos e inconsistencias en la estimación de los parámetros del modelo, en cuyo caso las inferencias realizadas pueden resultar erróneas.

El comportamiento de los diferentes métodos usados para analizar datos faltantes depende en gran medida del mecanismo de no respuesta, lo cual atañe directamente a las causas que originan la pérdida de datos y, en particular, a si la pérdida depende de las respuestas, observadas o no. Basándose la distribución del mecanismo de pérdida, Rubin (1976) define una jerarquía de tres procesos diferentes de no respuesta, a saber: Datos perdidos completamente al azar (MCAR), datos perdidos al azar (MAR) y datos perdidos no ignorables (NMAR). La literatura empírica pone de relieve que algunos métodos analíticos se hallan escasamente afectados por el desgaste de muestra cuando la pérdida es independiente de las respuestas. Sin embargo, la validez de los mismos puede verse comprometida cuando el mecanismo responsable de la pérdida y la conducta de interés estén relacionados (Dermitas, 2004).

Además de los tradicionales métodos basados en eliminar casos y en imputar un solo valor para cada dato perdido, los investigadores disponen de toda una gama de métodos para tratar los datos faltantes, en especial los basados en estimar directamente los parámetros del modelo a partir de los valores observados vía máxima verosimilitud, en imputar múltiples valores para cada dato perdido, seguido de los procedimientos utilizados para analizar datos completos, tales como el enfoque Brown-Forsythe modificado por Vallejo y Ato (2006), en relajar los supuestos distribucionales ponderando el método de las ecuaciones de estimación generalizada (GEE) por la inversa de la probabilidad de responder (WGEE; Robins, Rotnitzky & Zhao, 1995) y en seleccionar modelos. Schafer y Graham (2002) presentan una discusión detallada con los principales defectos que presentan las soluciones reseñadas.

En la actualidad no está claro que procedimiento - MBF vía SAS Proc MI, modelo lineal mixto (MLM) vía SAS Proc Mixed y GEE ponderado (WGEE) vía SAS Proc Genmod,- debe ser adoptado para reemplazar los enfoques convencionales a fin de ofrecer análisis fiables cuando existe pérdida de datos. Así, el propósito de este trabajo es examinar la validez de estos enfoques cuando se violan simultáneamente las hipótesis de normalidad de las observaciones y de homogeneidad de las matrices de dispersión. Aunque criterios de información están disponibles en PROC MIXED para ayudar a seleccionar la estructura de covarianza que mejor se aproxima al verdadero proceso generador de los datos, en este trabajo las inferencias sobre los parámetros que definen la estructura medias se hicieron siempre en base al modelo correcto, con los parámetros de covarianza estimada por el método REML. La aplicación del método WGEE requiere modelar adecuadamente el mecanismo de pérdida, hallando la probabilidad que tiene cada sujeto de responder en cada ocasión. En este trabajo, los pesos estimados se basan en una regresión logística donde la variable de respuesta es binaria y las covariables son el grupo, el tiempo, la interacción tiempo  $\times$  grupo y la medida realizada en el tiempo anterior (WEE1) o en el tiempo actual (WEE2). El verdadero mecanismo de pérdida es representado por WEE1. El enfoque de WGEE se aplicó asumiendo una matriz de trabajo no estructurada.

### *Definición de los procedimientos estadísticos que motivaron la investigación*

Las pruebas MBF y MLM son precisamente descritas en el texto de Vallejo y Ato (2006).

### *Método GEE ponderado (WGEE)*

Cuando existan datos faltantes y la estructura de correlación de trabajo no coincida con la verdadera, las inferencias realizadas vía GEE pueden resultar incorrectas, a no ser que el proceso de no respuesta sea MCAR. Para resolver este problema, Robins et al. (1995) propusieron modificar el enfoque GEE utilizando información relacionada con la pérdida. Sucintamente, la aplicación del enfoque resultante, denominado GEE ponderado, requiere llevar a cabo dos operaciones. Por un lado, modelar el mecanismo de pérdida hallando la probabilidad que tiene cada sujeto de responder en cada ocasión, dada la información que se considere de interés  $\pi_{ik} = Pr(R_{ik} = 1 | z_{ik})$ . Para ello se puede utilizar el modelo logístico u otro similar. Por otro lado, ponderar la función de estimación por la inversa de la probabilidad de ser observado y conducir un GEE estándar.

Las ecuaciones de estimación modificadas que ponderan las GEE por la inversa de la probabilidad de responder son dadas por

$$U_i^o(\beta_W) = \sum_{i=1}^N \mathbf{D}_i^{o'} [V_i^o(\hat{\alpha})]^{-1} \mathbf{W}_i^o (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o) = \mathbf{0}, \quad (1)$$

donde  $\mathbf{W}_i^o$  es una matriz diagonal con pesos  $R_{ik} \pi_{ik}^{-1}$  a lo largo de la diagonal principal. El estimador robusto de la covarianza de  $\beta_W$  se obtiene como

$$\text{Cov}(\hat{\beta}_W) = \left( \sum_{i=1}^N \mathbf{D}_i^{o'} [V_i^o(\hat{\alpha})]^{-1} \mathbf{W}_i^o \mathbf{D}_i^o \right)^{-1} \left( \sum_{i=1}^N \mathbf{E}_i^o \mathbf{E}_i^{o'} \right) \left( \sum_{i=1}^N \mathbf{D}_i^{o'} [V_i^o(\hat{\alpha})]^{-1} \mathbf{W}_i^o \mathbf{D}_i^o \right)^{-1}, \quad (2)$$

donde  $\mathbf{E}_i^o = U_i^o(\beta_W) - \left( \sum_{i=1}^N U_i^o(\beta_W) S_i^{o'} \right) \left( \sum_{i=1}^N S_i^o S_i^{o'} \right)^{-1} S_i^o$ , siendo  $S_i^o = (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$ .

Resaltar, finalmente, que si bien el enfoque WGEE resulta fácil de realizar utilizando la expresión SCWGT del procedimiento GENMOD del SAS (para detalles véase Hogan, Roy & Korkontzelou, 2004), su validez depende de especificar correctamente un modelo para el mecanismo de pérdida (Dermitas, 2004; Schafer & Graham, 2002).

## Método

Los procedimientos fueron investigados, en términos de robustez, con un diseño univariado de medidas parcialmente repetidas. Las variables manipuladas en el estudio fueron las siguientes: (a) Porcentaje de datos faltantes (10% y 20%), (b) número de mediciones repetidas ( $K = 4$  y  $K = 8$ ), (c) relación entre el tamaño de los grupos y el de las matrices de dispersión (nula, positiva y negativa) y (d) forma de distribución de la población (normal y exponencial). Las constantes fueron: (a) patrón de datos faltantes (monótono), (b) número de grupos ( $J = 3$ ), (c) tamaño de la muestra total ( $N = 60$ ), (d) el grado de igualdad / desigualdad de covarianza de grupo matrices, (e) patrón de la estructura de covarianza (ARH), y (f) tipo de mecanismo de datos faltantes (MAR). Para realizar los cálculos se utilizó un MACRO escrito en lenguaje SAS/IML. Como el tiempo de cálculo requerido por el programa era muy alta, tan sólo realizamos 2000 réplicas de cada condición estudiada con un nivel de significación del 5%.

## Resultados

La Tabla 1 contiene las tasas de error empíricas correspondientes a los efectos del diseño. Por lo que respecta al control de las tasas de error correspondientes al efecto principal de los grupos, los resultados mostraron que el desempeño de la técnica MI seguida del enfoque MBF (MI-MBF) resultó ligeramente superior al desempeño del enfoque PROC MIXED con la corrección Kenward y Roger (1997; 2009) basado en la verdadera estructura de covarianza, y también al desempeño de los enfoques WGEE.

En relación con el control de las tasas de error correspondientes al efecto de las ocasiones, los cuatro enfoques evidenciaron discrepancias entre las tasas de error nominal y estimada, particularmente conforme el número de ocasiones aumentaba. Sin embargo, nuestro estudio mostró que PROC MIXED era preferible al resto de los enfoques.

Con respecto a la interacción, los resultados obtenidos con los enfoques MI-MBF y PROC MIXED revelan un control aceptable de las tasas de error. Las tasas de WGEE también fueron en general bien controladas, excepto cuando las pérdidas eran del 20%. Como se ha mencionado, el mecanismo verdadero era representado por WEE1. No obstante, en términos de control de las tasas de error, WEE1 no funcionaba mejor que WEE2.

Tabla 1. Porcentaje de error Tipo I para el efecto principal de los grupos, de las ocasiones e interacción grupos x ocasiones

%	PDS	APA	MI (MBF)		MIXED		WGEE		MI (MBF)		MIXED		WGEE		MI (MBF)		MIXED		WGEE	
			COM	MON	COM	MON	WEE1	WEE2	COM	MON	COM	MON	WEE1	WEE2	COM	MON	COM	MON	WEE1	WEE2
K = 4																				
10	MVN	=	4.65	4.40	6.15	5.55	3.80	4.25	4.85	8.60	5.35	4.85	5.00	5.65	4.90	4.95	4.90	4.95	4.10	4.15
10	MVN	+	4.25	4.95	4.95	4.65	3.55	3.30	4.85	9.55	5.45	4.05	2.85	3.10	4.45	4.45	6.35	5.90	2.35	2.40
10	MVN	-	5.00	5.00	4.60	5.65	4.30	4.65	6.00	8.60	4.50	4.75	6.20	6.55	5.30	5.30	5.40	5.40	4.30	7.75
10	EXD	=	4.60	3.70	7.10	6.00	5.45	5.50	7.00	10.70	4.70	5.20	6.25	6.70	4.00	4.00	5.10	5.10	4.15	4.40
10	EXD	+	3.45	3.25	6.35	4.50	4.15	4.45	6.45	10.65	6.05	4.80	4.25	4.85	3.25	3.25	3.85	4.70	2.75	2.75
10	EXD	-	5.60	5.00	7.30	8.25	7.45	7.95	6.70	10.20	6.15	5.85	6.95	7.65	4.30	4.30	6.00	6.25	3.85	3.95
20	MVN	=	4.65	4.70	6.15	5.85	4.60	5.15	4.85	13.70	5.35	5.25	6.00	6.65	4.90	5.95	4.90	4.35	4.20	4.35
20	MVN	+	4.25	4.45	4.95	4.15	3.90	4.15	4.85	12.95	5.45	4.75	3.80	3.95	4.45	4.45	6.30	4.25	3.10	3.95
20	MVN	-	5.00	4.90	4.60	5.45	4.20	4.70	6.00	13.60	4.50	6.15	6.70	6.30	5.30	5.30	5.40	6.50	5.80	5.50
20	EXD	=	4.60	4.25	7.10	6.90	5.75	6.05	7.00	16.85	4.70	6.60	6.75	7.90	4.00	4.00	5.10	6.10	4.85	4.05
20	EXD	+	3.45	3.45	6.35	5.60	4.40	4.75	6.45	15.55	6.05	5.10	4.30	5.65	3.25	3.25	3.85	4.05	2.70	3.05
20	EXD	-	5.60	5.25	7.30	7.45	7.35	7.50	6.70	14.65	6.15	6.25	7.65	8.55	4.30	4.30	6.00	6.80	5.75	5.95
K = 8																				
10	MVN	=	5.45	4.70	4.95	5.15	4.10	4.05	4.85	10.10	4.80	6.15	8.20	9.05	4.85	4.70	4.85	4.20	3.80	3.95
10	MVN	+	5.25	3.75	5.25	4.75	4.75	4.90	4.90	10.85	4.95	5.85	2.85	3.00	5.10	4.25	5.10	4.05	1.45	1.50
10	MVN	-	5.70	5.00	5.30	4.35	4.80	5.05	3.55	7.55	4.50	4.65	9.10	9.75	4.75	6.20	4.75	4.25	6.20	6.15
10	EXD	=	5.20	5.10	5.85	6.50	6.15	6.20	9.00	15.50	7.45	8.50	13.65	14.80	4.25	4.50	4.25	4.85	4.75	4.80
10	EXD	+	4.75	4.15	4.80	4.90	4.05	4.25	8.55	15.75	7.55	7.70	5.15	5.55	2.75	4.05	2.70	2.55	1.95	2.05
10	EXD	-	5.20	7.30	6.80	6.65	8.95	9.05	7.60	12.90	7.50	7.75	13.35	14.95	5.60	5.25	5.60	7.15	4.85	4.95
20	MVN	=	5.45	5.50	4.95	4.25	4.80	4.95	4.85	18.70	4.80	4.75	9.20	9.15	4.85	4.70	4.85	4.30	4.10	4.25
20	MVN	+	5.25	5.05	5.25	4.65	4.80	4.55	4.90	17.95	4.95	4.95	2.70	2.75	5.10	4.25	5.10	4.55	1.50	1.65
20	MVN	-	5.70	4.70	5.30	4.75	6.80	6.90	3.55	10.10	4.50	5.15	12.20	12.30	4.75	6.20	4.75	5.25	4.80	5.10
20	EXD	=	5.20	4.75	5.85	6.00	7.85	8.05	9.00	16.85	7.45	8.80	15.05	15.35	4.25	4.50	4.25	5.25	4.95	5.55
20	EXD	+	4.75	4.65	4.80	4.50	5.40	5.45	8.55	14.65	7.55	8.60	5.60	6.35	2.75	4.05	2.70	2.65	2.20	2.25
20	EXD	-	5.50	6.45	6.80	7.35	9.15	9.75	7.60	18.35	7.50	8.35	12.45	15.65	5.60	5.25	5.60	7.20	7.15	6.95

Legenda: MI (MBF) = imputación múltiple segura del enfoque MBF; MIXED = SAS Proc Mixed; WGEE = método GEE ponderado; WEE1 = WGEE con los pesos basados en la medida previa. WEE2 = WGEE con los pesos basados en la medida actual. % = Porcentaje de datos faltantes; PDS = forma de distribución de la población; APA = relación entre el tamaño de los grupos y el de las matrices de dispersión; COM = vectores de respuestas completos; MON = patrón de pérdida de datos monótono; MVN = distribución normal multivariada ( $\gamma_1 = 0, \gamma_2 = 0$ ); EXD = distribución exponencial ( $\gamma_1 = 2, \gamma_2 = 6$ ); = relación nula ( $n_1 = 20, n_2 = 20; n_3 = 20$ ); + relación positiva ( $n_1 = 12, n_2 = 20; n_3 = 28$ ); - relación negativa ( $n_1 = 28, n_2 = 20; n_3 = 12$ ).

## Conclusión

En este trabajo, examinamos la robustez del enfoque MBF cuando existían pérdida de datos al azar. Con propósitos de comparación, también se incluyeron otros métodos especialmente diseñados para analizar datos longitudinales carentes de balanceo, a saber: los enfoques MLM y GEE ponderado (WGGE) implementados en el programa SAS.

Los resultados mostraron que ningún enfoque fue capaz de controlar las tasas de error en todas las condiciones investigadas. El enfoque MI-MBF controló aceptablemente las tasas de error referidas al efecto principal de los grupos e interacción de éstos con las ocasiones, pero resultó inaceptable para contrastar el efecto de las ocasiones de medida. En este último caso, los enfoques PROC MIXED y WGGE eran más recomendables.

Reconocimientos. Este trabajo ha sido financiado mediante el proyecto de investigación concedido por el MCI (Ref.: PSI-2008-03624) y galardonado con el premio AEMCCO en la categoría de Póster.

## Referencias

- Demirtas, H. (2004). Assessment of relative improvement due to due to weights within generalized estimating framework for Incomplete clinical trials data. *Journal of Biopharmaceutical Statistics, 14*, 1085-98.
- Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley.
- Hogan, J. W., Roy, J. & Korkontzelou, C. (2004). Handling drop-out in longitudinal studies. *Statistics in Medicine, 23*, 1455-1497.
- Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997.
- Kenward, M. G. & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis, 53*, 2583-2595.
- Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.
- Robins, J. M., Rotzinsky, A. & Zhao L.P. (1995). Analysis of semiparametric regression models for repeated measures in the presence of missing data. *Journal of the American Statistical Association, 90*, 106-121.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.
- Schafer, J. L. & Graham, J. W. (2002): Missing data: Our view of the state of the art, *Psychological Methods, 7*, 147-177.
- Vallejo, G. & Ato, M. (2006). Modified Brown-Forsythe procedure for testing interaction effects in split-plot designs. *Multivariate Behavioral Research, 41*. 549-578.

# COMPARACIÓN DE CAQDAS PARA LA INVESTIGACIÓN PSICOSOCIAL: ATLAS-TI *VERSUS* NVIVO

M. Teresa Anguera y Olatz López Fernández  
Universidad de Barcelona

El CAQDAS (“Computer Assisted Qualitative Data Analysis Systems”) hace referencia al uso de software específico para realizar el análisis de datos cualitativos, que se obtienen frecuentemente en investigaciones psicosociales en un amplio espectro de ámbitos. Recientemente ha adquirido relevancia en el área de la metodología cualitativa, aunque todavía no existe un acuerdo común respecto a las potencialidades de los programas, como indican Lewins y Silver (2007). El propósito de este trabajo es realizar una comparación de dos CAQDAS ampliamente reconocidos: Atlas-ti y NVivo. En concreto, se pretende analizar conceptualmente el objetivo metodológico de cada programa, así como las ventajas e inconvenientes del procedimiento analítico que proporcionan al investigador en ciencias sociales y de la salud. Con este análisis previo, se extraerán las principales aportaciones de cada uno de los programas mediante la comparación de sus características técnicas, con el fin de obtener aquellos aspectos del análisis de datos cualitativos que pueden ser más adecuados en una investigación psicosocial, cuyos datos tengan un formato textual. Finalmente, se realizarán recomendaciones respecto a los hallazgos técnico-metodológicos resultantes del presente análisis comparativo.

El CAQDAS es el acrónimo de “Computer Assisted Qualitative Data Analysis Systems” y hace referencia al uso de software específico para realizar el análisis de datos cualitativos. Un tipo de datos muy usuales en investigaciones psicosociales y otros ámbitos de las ciencias sociales. También son conocidos como “QDA / qualitative software” y su desarrollo se inicia en 1980, por lo que son más recientes que los tradicionales paquetes estadísticos, propios del análisis de datos de carácter cuantitativo.

Por datos cualitativos se entiende básicamente los textuales, aunque también habría que incluir lo visual, multimedia e información no numérica en general. Estos datos han de ser interpretados y, para ello, primero han de ser identificados, para poder ser codificados en temas, conceptos, procesos, contextos, etc. Para poder construir explicaciones o tratar de formular teorías o comprobarlas. Las técnicas de investigación cualitativa suelen estar enmarcadas en aproximaciones propias de la metodología cualitativa, como la Teoría Fundamentada, el Análisis Conversacional o Narrativo, entre otras. En cualquier caso, suelen ser básicamente aquellas técnicas en las que el investigador interacciona con el investigado en un contexto determinado, es decir, las entrevistas, los *focus groups* y la observación participante.

Sus principales características y similitudes consisten en que los CAQDAS permiten el análisis de datos cualitativos textuales, la codificación, la especificación de los memorándum y la obtención del resumen de citas codificadas por código o conjunto de códigos. Sin embargo, tienen sus ventajas e inconvenientes (ver figura 1).

<b>VENTAJAS</b>	<b>INCONVENIENTES</b>
<ul style="list-style-type: none"> <li>- Software específico para el análisis de datos cualitativos, esencialmente permite codificar texto, recuperarlo y hacer búsquedas sofisticadas (entre otras opciones).</li> <li>- Es un programa a modo de base de datos cualitativos, que garantiza la fiabilidad por el detalle y transparencia del análisis (Flick, 2007)</li> <li>- Permite importar datos analizados cualitativamente a software cuantitativo para poder realizar investigación híbrida ("mixed methods")</li> </ul>	<ul style="list-style-type: none"> <li>- Ningún CAQDAS puede tomar decisiones acerca la codificación de los materiales textuales o la interpretación de resultados (Weitzman &amp; Miles, 1995)</li> <li>- La mayoría son softwares comerciales, excepto alguno reciente de código abierto (por ejemplo: <i>WeftQDA</i>)</li> <li>- Cada programa tiene sus propias características gráficas y funcionalidades comunes y propias, por lo que hay que conocerlos para escoger cual utilizar.</li> </ul>

Figura 1. Ventajas e inconvenientes de los CAQDAS

En general, existe un debate o falta de acuerdo por su utilidad respecto a los siguientes elementos:

- Cuantificación
- Desfragmentación del discurso
- Descontextualización
- No interacción
- Similitud con el procesador de textos
- Fiabilidad entre observadores
- El sistema de análisis es más propio de la *Grounded Theory*

A pesar de ello, los CAQDAS parecen utilizarse cada vez más por las siguientes virtudes:

- Rapidez
- Nuevas oportunidades de análisis
- Garantía del análisis
- La estructuración de los datos
- Frecuencias

Recientemente los CAQDAS han adquirido relevancia en el área de la metodología cualitativa, aunque todavía no existe un acuerdo común respecto a las potencialidades de los programas, como indican Lewins y Silver (2007). Estas autoras indican que las diferentes herramientas existentes hasta el momento soportan determinadas tareas comunes y cada software tiene sus peculiaridades en cuanto a posibilidades para manejar los datos y realizar el análisis cualitativo.

El propósito de este trabajo es precisamente apuntar hacia esta cuestión. Por ello se propone una comparación de dos CAQDAS ampliamente reconocidos: Atlas-ti y NVivo. En concreto, se pretende introducir ambos softwares y analizarlos conceptualmente con el objetivo metodológico de cada programa, así como valorar sus ventajas e inconvenientes del análisis, que proporcionan al investigador en ciencias sociales y de la salud unos procedimientos y resultados aplicables a sus investigaciones de carácter cualitativo o híbrido (metodologías cualitativas y cuantitativas).

## NVivo vs Atlas-ti

Comparar Nvivo vs. Atlas-ti, fue una elección basada en el reconocimiento internacional de ambos CAQDAS, puesto que son los más utilizados en investigación psicosocial, y aunque ninguno es líder de la industria, QSR NUD\*IST Vivo (“*Non-numerical Unstructured Data Indexing Searching and Theorizing*”) o NVivo es el más extendido.

Los elementos en común que tienen ambos CAQDAS son:

- El ser comerciales (a través de Internet existen demos y precios para estudiantes y/o académicos)
- El ser utilizados en investigación psicosocial, educativa, empresarial, etc.
- El tener un interfaz inicial similar:
  - en el que se espera la creación de una base de datos del análisis cualitativo (N: proyecto; A: unidad hermenéutica)
  - la introducción de registros normalmente textuales (N: documentos; A: documentos primarios)
  - el análisis de datos cualitativos:
    - La codificación (N: nodos; A: códigos)
    - La creación de códigos más abstractos (N: conjuntos; A: familias)
    - Las operaciones *booleanas* entre éstos (N: preguntar a la codificación; A: herramienta de pregunta)
    - La creación de representaciones gráficas de los códigos (N: modelos; A: redes semánticas)
    - Entre otras operaciones complementarias (uso de frecuencias, tablas de contingencia, etc.)

En cuanto a los elementos comunes en el nivel más técnico, permiten:

- Importan y muestran texto enriquecido
- La construcción de listas de códigos de forma
- La recuperación de texto previamente codificado
- La evaluación de texto codificado en el contexto original de los documentos
- La escritura de memos que pueden ser enlazadas a códigos y/o a documentos
- La representación gráfica del análisis cualitativo

El NVivo, es un CAQDAS que ayuda a analizar, gestionar y dar forma a los datos cualitativos de forma jerarquizada. Proporciona seguridad al guardar la base de datos y archivos juntos en un solo fichero, aunque ello requiere el uso de memoria del ordenador. Es de valorar el que permite el uso de múltiples lenguas, de trabajo en equipo (“*merge function*”) y permite al investigador manipular fácilmente los datos y conducir búsquedas. Así como también puede mostrar gráficamente códigos y categorías. El software es australiano (<http://www.qsrinternational.com>) y su última versión proviene del internacional QSR y combina las características del N6 (Nud.ist) y Nvivo 2.0, estando sólo disponible para Windows PC. En cambio, Atlas-ti permite organizar los archivos de texto, gráficos, audio y visual, así como codificarlos, añadirles memos y encuentros dentro de un proyecto. Tiene la ventaja que es más rápido de utilizar, pues permite el arrastrar y dejar códigos en una pantalla interactiva y rápidamente buscar, recuperar y seleccionar todos los segmentos de datos y notas en un mapa conceptual. Los datos pueden ser exportados a SPSS, HTML, XML, CSV y requieren de poca memoria del ordenador porque enlaza directamente los datos al proyecto. En cambio, sólo permite a un grupo de investigadores trabajar en equipo en un proyecto y hacer comparaciones acerca de la codificación de cada uno. El software es alemán (<http://www.atlasti.com>) y también está sólo disponible para Windows PC.

Sus principales diferencias radican en como se realizan los procedimientos para manejar los datos y realizar el análisis cualitativo, según Gibbs (2007) (véase tabla 1).

*Tabla 1. Comparación general de 2 CAQDAS: NVivo vs. Atlas-ti*

NVivo (versión 8)	Atlas-ti (versión 5)
<ul style="list-style-type: none"> <li>• Tiene el soporte más simple de codificación jerarquizada, es decir, al inicio del análisis</li> <li>• Tiene la herramienta de búsqueda más potente al incluir la búsqueda de matrices (incluye tablas de comparación)</li> <li>• Permite enlazar imágenes, sonidos y video, pero no se pueden codificar</li> <li>• Los elementos de la gráfica final están directamente enlazados con los códigos y los datos</li> <li>• Acepta archivos RTF y reconoce secciones y partes del archivo utilizando estilos en los párrafos (ello permite la auto-codificación)</li> </ul>	<ul style="list-style-type: none"> <li>• La jerarquía la logra a través de los servicios de su red de trabajo, al final del análisis</li> <li>• Tiene una herramienta de búsqueda básica (operadores booleanos)</li> <li>• Además de permitir enlaces audiovisuales también se pueden codificar</li> <li>• A parte de enlazar gráfica con códigos, también tiene un conjunto de relaciones lógicas a añadir que se enlazan directamente al análisis</li> <li>• Acepta archivos RTF y los párrafos pueden ser agrupados, de modo que cada uno acabe con dos retornos (ello permite la auto-codificación)</li> </ul>

### Conclusión

Los dos CAQDAS comparten facilidades básicas para manipular las transcripciones, codificarlas, recuperar los fragmentos textuales resultantes, mostrar los códigos contextualizados, escribir memos, hacer búsquedas diversas y representaciones gráficas.

Las principales diferencias radican en el interfaz, que en ambos se estructura en función de las opciones de trabajo. Parece que NVivo es más completo y/o permite más opciones de análisis, con una estructura jerarquizada desde un inicio y unas rutinas similares para distintas fases del análisis. Sin embargo, parece que Atlas-ti es más claro, permite un análisis más libre (o no tan estructurado inicialmente) y es más simple de utilizar, con acciones más directas y procedimientos más lógicos.

Existen criterios para seleccionar CAQDAS (Creswell & Maietta, 2002):

1. Facilidad de integración al usar el programa: ¿es compatible con PC o MAC?, ¿qué tipo de datos el programa aceptará?, ¿se puede trabajar fácilmente a través del documento?
2. Tipos de datos que el programa aceptará: ¿cogerá datos textuales?, ¿y datos en multimedia?...
3. Leer y revisar texto: ¿puedo subrayar y conectar fragmentos?, ¿puedo buscar un fragmento textual específico?
4. Escritura de memos: ¿tiene la capacidad de añadir notas o memos?, ¿puedes fácilmente acceder a las memos que escribes?
5. Categorización: ¿puedes crear códigos?, ¿puedes aplicar fácilmente códigos al texto o imágenes?, ¿puedes mostrar fácilmente códigos?, ¿puedes fácilmente revisar y hacer cambios en los códigos?

6. Análisis de inventario y evaluación: ¿puedes clasificar códigos específicos?, ¿puedes combinar códigos en la búsqueda?, puedes desarrollar un mapa conceptual con los códigos?, ¿puedes hacer comparaciones demográficas con los códigos?
7. Datos cuantitativos: ¿puedes importar bases de datos cuantitativas (p.e. de SPSS)?, ¿puedes exportar una palabra o imagen de una base de datos cualitativa a una base de datos cuantitativa?
8. Fusión del proyecto: ¿pueden dos o más investigadores analizar los datos y pueden estos análisis fusionarse?

Sin embargo, no parece que ningún CAQDAS sea mejor que otro, pues hay herramientas en común y características diferenciales (Lewins & Silver, 2007). Entonces, su elección dependerá de varios factores:

- Paradigma de investigación cualitativa
- Estado de la cuestión (según el tema y métodos empleados)
- Preferencia de la interfaz y/o sistema de análisis del investigador, así como su experiencia previa con algún CAQDAS u otros programas de análisis de datos
- Disponibilidad del software (por parte de la institución)
- Etc.

Se recomienda también probar los software con un estudio a modo de piloto para aprender a utilizarlos (obtener tutorial de la web oficial) y comprobar el funcionamiento y resultados que nos puede proporcionar. Por todo ello, en conclusión, la elección de un CAQDAS va a depender tanto de los aspectos técnicos como sobretodo los metodológicos y analíticos, puesto que va a facilitar determinados procedimientos para sintetizar los datos, extraer resultados e interpretarlos en el contexto de la investigación psico-social de carácter cualitativa o híbrida que se desarrolle con el soporte de algunos de estos dos CAQDAS.

### Referencias

- Creswell, J. W, & Maietta, R. C. (2002). Qualitative research. In D. Miller & N.J. Salkind (Eds.), *Handbook of research design and social measurement*. Thousand Oaks, CA: Sage.
- Gibbs, G.R. (2007). *Analyzing qualitative data*. London: Sage.
- Lewins, A. & Silver, C. (2007). *Using software in qualitative research. A step-by-step guide*. London: Sage.

# ANÁLISIS OBSERVACIONAL DE LA FASE OFENSIVA DEL FÚTBOL DE ALTO NIVEL

Claudio A. Casal y Toni Ardá  
Universidad de La Coruña

En el presente estudio nos hemos planteado determinar qué conductas estratégicas nos permiten predecir el resultado final de las transiciones ofensivas. Para ello hemos analizado los encuentros disputados entre las mejores selecciones nacionales europeas, durante la Eurocopa disputada en Austria-Suiza en el 2008. En el fútbol existe un claro predominio de la defensa sobre el ataque, finalizando sólo en gol el 1% de las acciones ofensivas (Dufour, 1993; Silva, 2005; Sledziewski y Kisonda, 1982). La principal preocupación para aquellos entrenadores que buscan una mayor rentabilidad de sus equipos, y para aquellos investigadores que tratan de hacer más espectacular este deporte, se basa en la obtención de más situaciones de finalización. Las transiciones ofensivas en el fútbol, debido a la estructura del mismo, con un gran número de jugadores y un amplio espacio de juego, suponen un momento propicio para crear situaciones ventajosas sobre el equipo adversario. La metodología que hemos empleado para realizar el estudio ha sido una metodología observacional directa, sistematizada, no participativa y en un ambiente natural. Con los datos registrados hemos llevado a cabo, en un primer momento un análisis descriptivo y posteriormente un análisis predictivo, empleando para ello la técnica estadística de la regresión logística.

## *Objetivos*

- Identificar y cuantificar los elementos observables de las transiciones ofensivas.
- Especificar las variables significativas que aparecen de forma repetida en el juego, que nos permitan conocer el resultado final.
- Conocer la capacidad predictiva de las variables seleccionadas, relativamente a la consecuencia de finalización.

## *Hipótesis*

- Creemos que podemos predecir el tipo de éxito que puede llegar a tener una secuencia ofensiva en función de las variables implicadas.
- Es de esperar que la zona de inicio de la secuencia ofensiva (ZI), la configuración espacial de interacción de inicio (CEI), la organización defensiva del equipo rival (ORD), y la intención táctica del equipo (INT), determinen, en gran medida, el éxito de la transición ofensiva.
- Suponemos que las transiciones ofensivas que se producen con una configuración espacial de interacción de inicio MR, AR o AØ, ante una organización defensiva del equipo rival CIR, y con una intención táctica del equipo PR, tienen la mayor probabilidad de finalizar con éxito.

## *Diseño*

Para dar respuesta a los objetivos que nos hemos planteado y para tratar de confirmar las hipótesis del estudio, hemos utilizado una metodología observacional directa, sistematizada, no participativa y en un ambiente natural. Dentro de los posibles diseños que presenta esta metodología, el trabajo corresponde a un diseño nomotético, de seguimiento y multidimensional.

## Muestra

Hemos codificado 7 partidos, los disputados a partir de los cuartos de final de la Eurocopa Austria-Suiza 08. En total han sido analizadas 743 transiciones ofensivas.

## Instrumentos de observación

Hemos elaborado un instrumento de observación “*ad hoc*” constituido por la combinación de formatos de campo y sistemas de categorías (tabla 1).

Tabla 1. Relación de criterios y categorías empleados en la investigación

MACROCRITERIOS	CRITERIOS	CATEGORÍAS
INICIO DE LA POSESIÓN	1.- Zona de inicio. (ZI)	1.1.- Defensiva (DF)
		1.2.- Medio defensiva (MD)
		1.3.- Central (CE)
		1.4.- Medio ofensiva (MO)
		1.5.- Ofensiva (OF)
2.- Configuración espacial de interacción de inicio. (CEI)	2.- Configuración espacial de interacción de inicio. (CEI)	2.1.- PA
		2.2.- RA
		2.3.- RM
		2.4.- MR
		2.5.- MM
		2.6.- MA
		2.8.- AR
		2.9.- AM
		2.10.- AØ
		3.- Organización defensiva del equipo adversario. (ORD)
3.2.- Defensa circunstancial (CIR)		
4.- Tiempo del encuentro. (TP)	4.- Tiempo del encuentro. (TP)	4.1.- Entre el minuto 0 y el minuto 15. (15)
		4.2.- Entre el minuto 16 y el minuto 30. (30)
		4.3.- Entre el minuto 31 y el final del primer tiempo. (45)
		4.4.- Entre el inicio del segundo tiempo y el minuto 60. (60)
		4.5.- Entre el minuto 61 y el minuto 75. (75)
		4.6.- Entre el minuto 76 y el final del segundo tiempo.(90)
		4.7.- Entre el inicio de la prórroga hasta el final del primer tiempo de la misma. (105)
		4.8.- Entre el inicio de la segunda parte de la prórroga hasta el final de la misma. (120)
5.- Intención táctica del equipo observado. (INT)	5.- Intención táctica del equipo observado. (INT)	5.1.- Progresar. (PR)
		5.2.- Conservar la posesión del balón. (CONS)
DESARROLLO DE LA POSESIÓN	1.- Número de jugadores. (NJ)	
	2.- Número de pases. (NP)	

FINAL DE LA POSESIÓN	1.- Configuración espacial de interacción de finalización. (CEIF)	1.1.- PAF 1.2.- RAF 1.3.- RMF 1.4.- MRF 1.5.- MMF 1.6.- MAF 1.7.- ARF 1.8.- AMF 1.9.- AØF
	2.- Resultado de la jugada. (RJ)	2.1.- Éxito. (EX) 2.2.- No éxito. (NEX)
RESULTADO	1.- Resultado del partido. (R)	1.1.- Ganando. (G) 1.2.- Empatando.(E) 1.3.- Perdiendo. (P)

### Resultados del análisis descriptivo

Tabla 2. Frecuencias absolutas (FA) y frecuencias relativas (FR) correspondientes al criterio zona de inicio de la posesión (ZI)

Macrocriterio inicio de la posesión			
Criterio	Categoría	FA	FR
Zona de inicio (ZI)	DF	230	31,0
	MD	332	45,0
	CE	120	16,0
	MO	56	7,0
	OF	5	1,0
	TOTAL	743	100%

El 45% de las transiciones ofensivas se inician en la zona medio defensiva.

Tabla 3. Frecuencias absolutas (FA) y frecuencias relativas (FR) de las diferentes configuraciones espaciales de interacción de inicio (CEI)

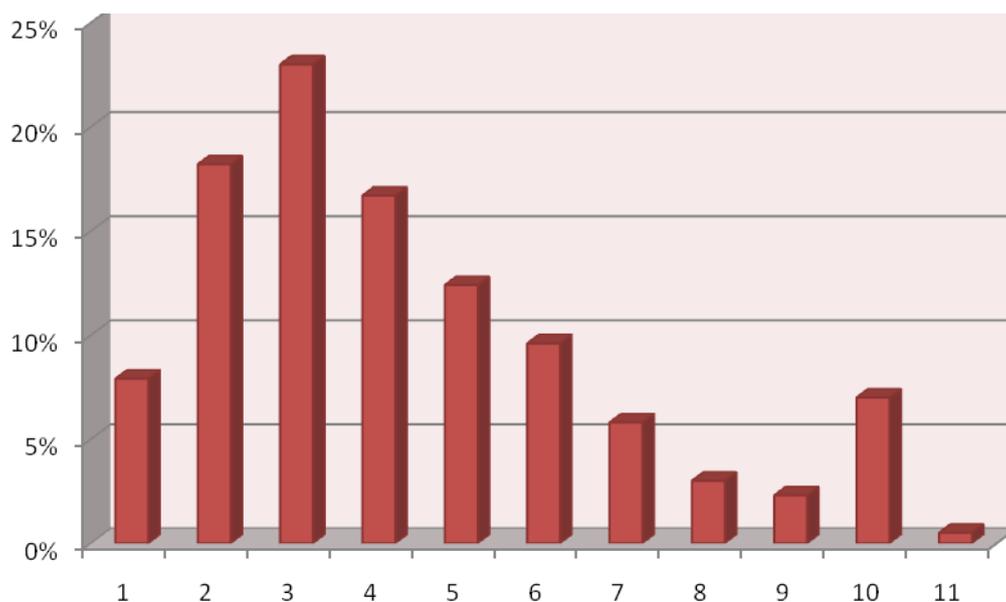
Macrocriterio inicio de la posesión			
Criterio	Categoría	FA	FR
Configuración espacial de interacción de inicio (CEI)	PA	132	17,8
	RA	263	35,4
	RM	53	7,1
	MR	7	0,9
	MM	250	33,6
	MA	12	1,6
	AR	21	2,8
	AM	5	0,7
	AØ	0	0,0
	TOTAL	743	100%

Los equipos recuperan la posesión del balón en un 35,3% de los casos con la línea de jugadores más retrasada.

*Tabla 4. Frecuencias absolutas (FA) y frecuencias relativas (FR) obtenidas del agrupamiento del tiempo del encuentro (TP)*

Macrocriterio inicio de la posesión			
Criterio	Categoría	FA	FR
Tiempo del encuentro (TP)	15	149	20,1
	30	99	13,3
	45	102	13,7
	60	117	15,7
	75	93	12,5
	90	109	14,7
	105	43	5,8
	120	31	4,2
TOTAL		743	100%

La mayor parte de las transiciones ofensivas se realizan en los primeros 15 minutos correspondientes a los dos periodos de juego y al final del tiempo reglamentario, entre el minuto 75 y el 90.



*Figura 1. Gráfico representativo del porcentaje de jugadores que participan en las transiciones ofensivas*

El número de jugadores que presenta una mayor frecuencia es 3, con un 23% del total de las acciones ofensivas.

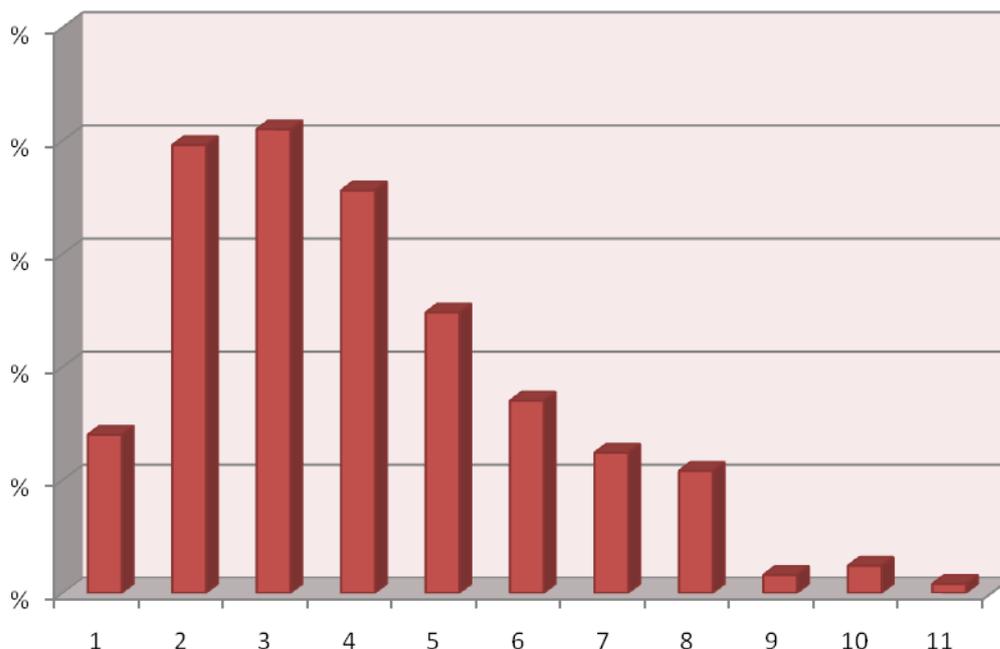


Figura 2. Porcentaje del número de jugadores que participan en las transiciones ofensivas con éxito

Las transiciones ofensivas más numerosas y eficaces son aquellas en las que participan entre 2 o 3 jugadores. Éstas suponen un 40,3% del total.

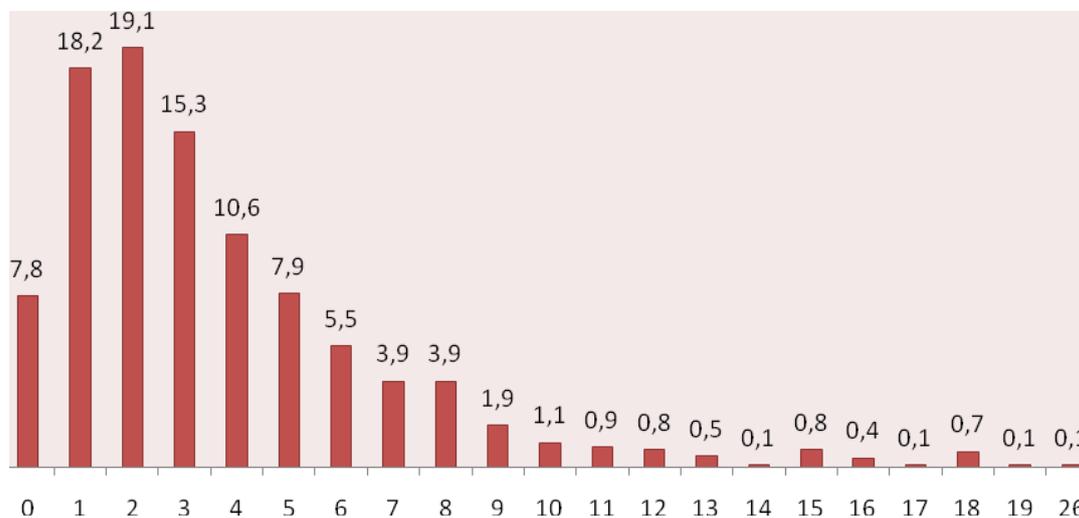


Figura 3. Porcentaje del número de pases realizados en las transiciones ofensivas

En la mayor parte de las secuencias ofensivas se realizan entre 1 y 3 pases (52,6%), siendo 2 el número de pases que se produce con mayor frecuencia (19,1%). Además, las acciones ofensivas en las que se realizan 2 pases presentan una mayor frecuencia de éxito (20,2%). El 36% de este tipo de transiciones ofensivas finalizan con éxito.

Tabla 5. Frecuencias absolutas (FA) y frecuencias relativas (FR) de cada una de las configuraciones espaciales de interacción de finalización (CEIF)

Macrocriterio final de la posesión			
Criterio	Categoría	FA	FR
Configuración espacial de interacción de finalización (CEIF)	PAF	21	2,8
	RAF	56	7,5
	RMF	10	1,3
	MRF	24	3,2
	MMF	280	37,7
	MAF	4	0,5
	ARF	312	42,0
	AMF	12	1,6
	AØF	24	3,2
TOTAL		743	100%

Las configuraciones espaciales de interacción de finalización que presentan unas mayores frecuencias son la ARF (42%) y la MMF (37,7%).

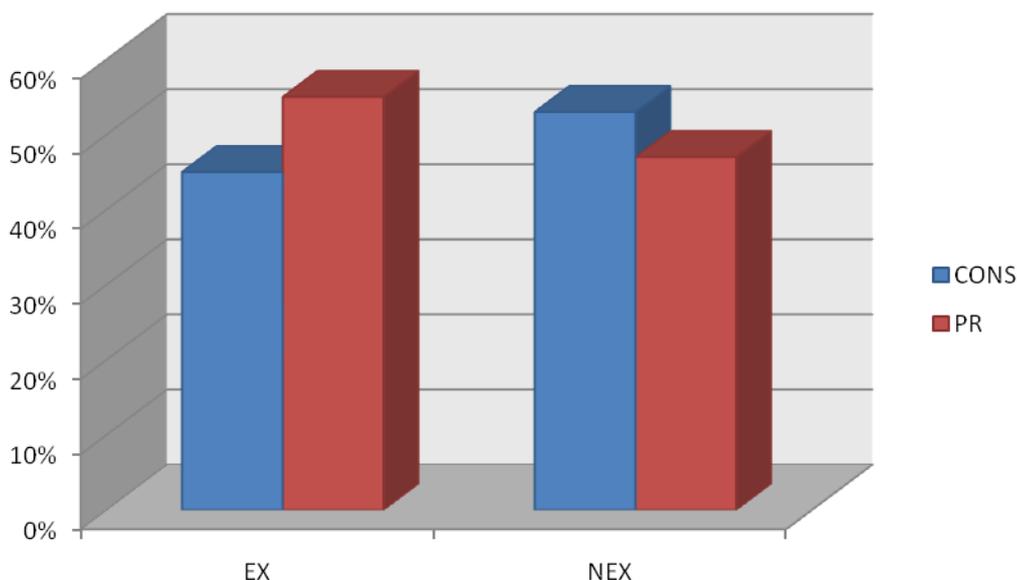


Figura 4. Porcentajes de la intención táctica del equipo y resultado de la transición ofensiva

Es importante, una vez recuperada la posesión del balón, tratar de realizar una progresión inmediata hacia la portería rival, y así tener mayores probabilidades de finalizar con éxito las transiciones ofensivas.

### Resultados del análisis predictivo

Las variables que nos van a permitir poder realizar una predicción sobre el resultado final de la transición ofensiva van a ser la configuración espacial de ambos equipos con respecto a la posición del balón, en el instante en el que el equipo observado recupera la posesión del balón (CEI). La variable INT, es decir, la intención táctica del equipo una vez recuperada la posesión del balón, y la variable D que es aquella que nos indica la duración de la transición ofensiva.

Tabla 7. Determinación del grado de asociación entre las variables explicativas y la explicada

Tipo 3 Análisis de efectos			
Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
CEI	7	17.7315	0.0132
INT	1	3.6210	0.0471
D	1	4.2146	0.0401

CEI con un valor 0,0132 es la que va a tener una mayor influencia sobre el éxito de la transición ofensiva.

INT (0,0471), va a tener una influencia menor a la anterior variable, sobre el resultado de la jugada.

D (0,0401) va a tener una influencia en el resultado final de la transición ofensiva similar al de la anterior variable.

Tabla 8. Valor del estadísticos  $\chi^2$  de Hosmer y Lemeshow

Test de bondad de ajuste de Hosmer y Lemeshow		
Chi-cuadrado	DF	Pr > ChiSq
6.3701	8	0.6059

El valor de Pr>ChiSq es alto (0,6059), permitiéndonos afirmar que el modelo se ajusta y su poder predictivo es bastante elevado.

En la tabla de clasificación podemos apreciar que el modelo seleccionado predice con exactitud el 65,3% de los casos analizados.

Tabla 9. Tabla de clasificación

Tabla de clasificación									
Nivel de prob.	Correcto		Incorrecto		Porcentajes				
	Evento (éxito)	Sin evento (no éxito)	Evento (éxito)	Sin evento (no éxito)	Correcto	Sensibilidad	Especificidad	Falso POS	Falso NEG
0.500	10	475	10	248	65.3	3,9	97,9	50	34,3

Las variables del modelo tienen una baja sensibilidad para determinar adecuadamente el éxito (3,9%). Sin embargo tienen una alta especificidad, es decir, su capacidad para detectar casos que no son éxito (97,9%). Con estos datos podemos extraer las siguientes conclusiones:

- Las tres variables seleccionadas nos van a permitir conocer el resultado final de las transiciones ofensivas en un 65,3% de los casos.
- El modelo no es adecuado para predecir un resultado positivo (éxito), pero sí es apropiado para determinar un resultado de la transición ofensiva negativo (no éxito). Esto nos servirá para determinar qué conductas debemos de evitar en el equipo, ante esta situación del juego, para tener la mayor posibilidad de finalizar con éxito la secuencia ofensiva.

A continuación, pasamos a explicar las probabilidades de predicción del resultado de la transición ofensiva, que nos ofrecen cada una de las variables con cada uno de sus diferentes valores, en términos de porcentajes.

Para la predictora, CEI sería para el valor AM,

$$100 * (0,476-1) = -52,4\%$$

A la hora de interpretar los resultados debemos tener en consideración que éstos están calculados para la probabilidad modelada de no éxito, por lo tanto:

- Un mayor porcentaje indicará mayor probabilidad de obtener no éxito.
- Un menor porcentaje será indicativo de menor probabilidad de obtener no éxito.

Aumentar en una unidad el valor AM, provocaría la disminuir en un 52,4% la probabilidad de no obtener éxito. Por lo tanto es conveniente provocar este tipo de situaciones para procurar que las transiciones ofensivas tengan éxito.

Para el valor AR,

$$100 * (0,809-1) = -19,1\%$$

Aumentar una unidad el valor AR, aumentaría la probabilidad de obtener éxito en la transición ofensiva.

Para el valor MA,

$$100 * (3,057-1) = 20,57\%$$

Aumentar una unidad el valor MA, provocaría el incremento en un 20,57% la probabilidad de obtener no éxito y por lo tanto disminuiría la probabilidad de obtener éxito en la transición ofensiva.

Para el valor MM,

$$100 * (0,874-1) = -12,6\%$$

Aumentar una unidad el valor MM, provocaría la disminución en un 12,6% la probabilidad de obtener no éxito y por lo tanto aumentaría la probabilidad de obtener éxito en la transición ofensiva.

Para el valor MR,

$$100 * (0,273-1) = -72,7\%$$

Aumentar una unidad el valor MR, provocaría la disminución en un 72,7% la probabilidad de obtener no éxito y por lo tanto aumentaría la probabilidad de obtener éxito en la transición ofensiva.

Para el valor PA,

$$100 * (1,749-1) = 74,9\%$$

Aumentar una unidad el valor PA, provocaría aumentar en un 74,9% la probabilidad de obtener no éxito.

Para el valor RA,

$$100 * (1,461-1) = 46,1\%$$

Aumentar una unidad el valor RA, provocaría incrementar en un 46,1 la probabilidad de obtener no éxito.

Para el valor CONS de la variable predictora INT,

$$100 (1,178-1) = 17,8\%$$

Aumentar en una unidad el valor CONS, supondría aumentar la probabilidad de obtener no éxito en la secuencia ofensiva de un 17,8%.

Para el valor PR de la variable predictora INT,

$$100 (0,84857-1) = -0,15\%$$

Si aumentamos en una unidad el valor PR, provocaríamos una disminución en un 0,15% la probabilidad de obtener un no éxito y por lo tanto estaríamos incrementando la posibilidad de obtener éxito en la transición ofensiva.

Para la variable predictora D,

$$100 * (0,986 -1) = -1,4 \%$$

En este caso si disminuimos en un segundo la duración de la secuencia ofensiva, la probabilidad de obtener éxito aumentaría en un 1,4%.

Una vez hallados los valores de los estimadores ya estamos en condiciones de expresar la ecuación predictiva y poder estimar la probabilidad de obtener el resultado de la jugada:

$$P_{[\text{no éxito}/\text{OEI}(\text{PA}), \text{INT}(\text{CONS}); \text{D}]} = \frac{e^{0,6802+0,5591+0,1652-0,0144}}{1 + e^{0,6802+0,05591+0,1642-0,0144}} = 0,80$$

Esto significa que las transiciones ofensivas en las que el equipo recupera la posesión del balón por medio del portero, encontrándose el equipo rival con todas las líneas de jugadores por detrás del balón, y por lo tanto con posibilidades de participar defensivamente para dificultar el ataque contrario.

El equipo, después de recuperar el balón, muestra una intención táctica de conservar la posesión del mismo sin tratar de llevar a éste rápidamente hacia la portería rival y emplea un periodo de tiempo largo para construir y finalizar el ataque. Las transiciones ofensivas que reúnen estas características tienen un 80% de probabilidades de fracasar.

$$P_{[\text{no éxito/OEI(MR),INT(PR);D(12'')}] = \frac{e^{0,6802-1,2989-0,1642-0,1728}}{1 + e^{0,6802-1,2989-0,1642-0,1728}} = 0,27$$

Como el modelo obtenido nos predice la probabilidad de no conseguir éxito en las transiciones ofensivas. Si incluimos en el modelo los valores de las variables que tienen una mayor influencia negativa en las probabilidades de obtener un no éxito podremos tener un valor de predicción con esas variables. Así, como expusimos anteriormente, las variables MR, PR y D influyen inversamente sobre el no éxito. El incremento de estas conductas en la transición ofensiva provoca que las probabilidades de obtener éxito aumenten. Podemos apreciar que, si en las transiciones ofensivas el equipo recupera el balón con la línea media, teniendo por delante sólo la línea defensiva del equipo adversario. La intención táctica del equipo es progresar y la transición ofensiva dura 12 segundos, la probabilidad de obtener no éxito es del 27%. Observamos que el descenso de la probabilidad de obtener no éxito es bastante apreciable (del 80% al 27%). Por lo tanto, éstas son las conductas que debemos proponer en el equipo para tener mayores probabilidades de conseguir éxito en las transiciones ofensivas.

#### Referencias

- Anguera, M.T., Behar, B., Blanco, A., Carreras, V., Losada, J.L., Quera, V., Riba, C. (1993). Glosario. En M.T. (Ed.), *Metodología observacional en la investigación psicológica* (pp. 587-617). Barcelona: P.P.U, vol. II.
- Ardá, A. y Casal C. (2003). *Metodología de la enseñanza del fútbol*. Barcelona: Paidotribo.

# APROXIMACIÓN INTEGRADA AL ANÁLISIS DE MARCOS INTERPRETATIVOS EN ARTÍCULOS PERIODÍSTICOS

Sergi Fàbregues<sup>1</sup> y Marie-Hélène Paré<sup>2</sup>

<sup>1</sup>Universitat Oberta de Catalunya

<sup>2</sup>University of Oxford

A pesar de que la metodología mixta ha sido definida en los últimos años como el tercer movimiento metodológico, es todavía reducido el número de estudios publicados que hayan mostrado un uso integrado y complementario de las metodologías cualitativa y cuantitativa en la fase de análisis. Hay dos argumentos que justifican el análisis integrado, siempre que éste se adecue a las preguntas de la investigación: aporta mecanismos de legitimación de los resultados obtenidos y ofrece procedimientos de análisis que permiten una mejor gestión e interpretación de los datos. Tomando en consideración esta última particularidad, esta comunicación describe un proyecto de investigación de carácter mixto sobre los *marcos* interpretativos en torno al riesgo asociado a los campos electromagnéticos en España. Empleando como material de análisis una muestra propositiva de 808 artículos periodísticos descargados de las bases de datos *Lexis/Nexis* y *Factiva/Reuters*, se presenta un ejemplo de diseño secuencial exploratorio basado en la aplicación de la estadística textual multivariante -análisis de clasificación jerárquica de palabras- en una fase previa al análisis cualitativo de contenido. El ejemplo defiende la utilidad de la metodología cuantitativa en la reducción de grandes volúmenes de información cualitativa y en la identificación sistemática y rigurosa de los *marcos* presentes en el texto. A la vez, también justifica que ambas metodologías, cualitativa y cuantitativa, lejos de ser incompatibles, pueden ser integradas con el fin de la complementariedad.

A lo largo de las tres últimas décadas ha acontecido, sobretodo en el mundo anglosajón, un auge por la combinación de las metodologías cuantitativa y cualitativa en las ciencias sociales y de la salud. El crecimiento experimentado por esta estrategia desde los años 70 ha animado a numerosos investigadores a publicar, recientemente, compilaciones (Tashakkori y Teddlie, 2003), libros (Todd, Nerlich y Clarke, 2004) y artículos que han trascendido las “*paradigm wars*” del pasado y han contribuido a reducir la brecha existente entre ambas metodologías. En palabras de Tashakkori y Teddlie (2003, p. x), la metodología mixta deviene un tercer movimiento metodológico, cuya finalidad es superar las desventajas derivadas del uso de una única aproximación y encontrar nuevos espacios de integración desde donde estudiar el mundo social.

En efecto, la importancia del uso integrado y complementario de las metodologías cualitativa y cuantitativa en el análisis de ambos tipos de datos ha sido ampliamente reconocida por distintos autores (ver Bazeley, 2009, para una extensa discusión). Las estrategias sugeridas, los acuerdos en torno a su legitimidad y el desarrollo de un amplio volumen de programas informáticos han incrementado el número de investigadores que han enfatizado la importancia de combinar diferentes aproximaciones analíticas en una misma investigación. En contraste con este reconocimiento, la presencia de análisis integrados en la literatura ha sido más bien reducida (Sandelowski, 2000). Asimismo, cuando han existido, han sido descritos de manera general y sin una reflexión detallada de los pasos empleados para su realización (Happ, Dabbs, Tate, Hricik, y Erlen, 2006). Esta ausencia ha sido todavía más acentuada en la investigación mediática basada en el “*frame analysis*” (análisis de *marcos*), a pesar de las afirmaciones de algunos autores en torno a la utilidad de este tipo de estrategias para un análisis textual sistemático y eficiente (Hertog y McLeod, 2001).

El ejemplo que se presenta a continuación, basado en el estudio del tratamiento mediático de los riesgos asociados a los campos electromagnéticos en España, se orienta a superar dichas limitaciones a partir de una aproximación integrada al análisis de *marcos* en una muestra de artículos periodísticos.

## Ejemplo empírico

### Contexto de la investigación

Los conflictos vinculados al riesgo asociado a los campos electromagnéticos generados por la telefonía móvil se fundamentan en dos factores: la ausencia de evidencia sobre sus efectos para la salud y la falta de consenso científico en torno a los niveles de emisión perjudiciales (Burgess, 2004). Ello fomenta que este riesgo no constituya un elemento material y tangible, sino que sea el resultado de una controversia pública entre diferentes *stakeholders*. En este contexto de disputa, los medios de comunicación emergen como la principal fuente de información sobre el riesgo y, al mismo tiempo, constituyen un espacio donde dichos *stakeholders* manifiestan sus posturas a través de la elaboración de una serie de mecanismos ideológicos e interpretativos en forma de *marcos* (Miller, 1997).

### Preguntas de la investigación

En esta investigación se han planteado las siguientes preguntas, de carácter cuantitativo, cualitativo y mixto, respectivamente:

1. ¿Cuáles son los principales *marcos* existentes en la cobertura mediática de los riesgos asociados a los CEM en España?
2. ¿De qué modo el *enmarcamiento* establece diferencias entre las diferentes categorías de actores y en el modo como el riesgo es evaluado, gestionado y percibido?
3. ¿Cuál es la contribución de las metodologías cuantitativa y cualitativa en el contexto de un diseño secuencial exploratorio, de carácter mixto?

### Diseño de la investigación

El diseño de esta investigación se sostiene en los fundamentos metodológicos del análisis de *marcos*. De raíz socio constructorista, los *marcos* son definidos por Gitlin (1980) como “patrones persistentes de cognición, interpretación, presentación, selección, énfasis y exclusión, a partir de los cuales, quienes definen los símbolos pueden organizar los discursos de manera rutinaria” (p. 7). Desde esta noción, los *marcos*, mediante la presencia de un conjunto de palabras clave, delimitan las fronteras del discurso en relación a un asunto, permitiendo con ello tipificar el modo cómo diferentes actores sociales estructuran la realidad social.

El uso del análisis de *marcos* se ha expandido recientemente en la literatura, hecho que ha fomentado la publicación de numerosos ejemplos empíricos. Köenig (2004) distingue entre dos aproximaciones para su realización, en el ámbito específico de los estudios de comunicación:

- Cualitativa: Incluye procedimientos hermenéuticos, lingüísticos y manuales que permiten elaborar descripciones en profundidad de los *marcos*. Esta aproximación plantea, sin embargo, dos problemas: los estudios (a) se basan a menudo en muestras pequeñas; y (b) generalmente no exponen de manera transparente el proceso de identificación de los *marcos* presentes en el texto.
- Cuantitativa: Incluye procedimientos deductivos y basados en el uso de ordenadores. Esta aproximación presenta restricciones importantes en el nivel conceptual pero, a cambio, posibilita el trabajo con muestras amplias y, asimismo, se caracteriza por presentar modelos de identificación muy explícitos y sistemáticos.

A pesar de sus fortalezas para la comprensión y la medición de los *marcos*, ambas aproximaciones presentan, pues, limitaciones notables cuando se usan de manera independiente. Por una parte, los procedimientos cualitativos permiten explorar de manera detallada el significado *latente* del texto pero, como contrapartida, tienden a carecer de transparencia, generándose con ello problemas importantes de fiabilidad. Por la otra, los procedimientos cuantitativos ofrecen medidas robustas y fiables, pero en muchas ocasiones sus resultados presentan poca validez (Matthes y Kohring, 2008).

Con el fin de superar estas limitaciones, el presente estudio parte del principio fundamental de la investigación mixta, basado en la combinación de métodos cuantitativos y cualitativos con fortalezas complementarias y con debilidades no solapadas (Johnson, Onwuegbuzie y Turner, 2007). Para ello, tal y como se expone en la figura 1, se plantea un diseño secuencial exploratorio basado en la realización, en una primera fase, de un análisis cuantitativo de contenido seguido de un análisis cualitativo temático de la misma base de datos textuales, en una segunda fase. Se justifica, en la presentación que sigue, por qué la metodología mixta contribuye a resolver una parte de los problemas inherentes a cada uno de los procedimientos de análisis de *marcos* por separado.

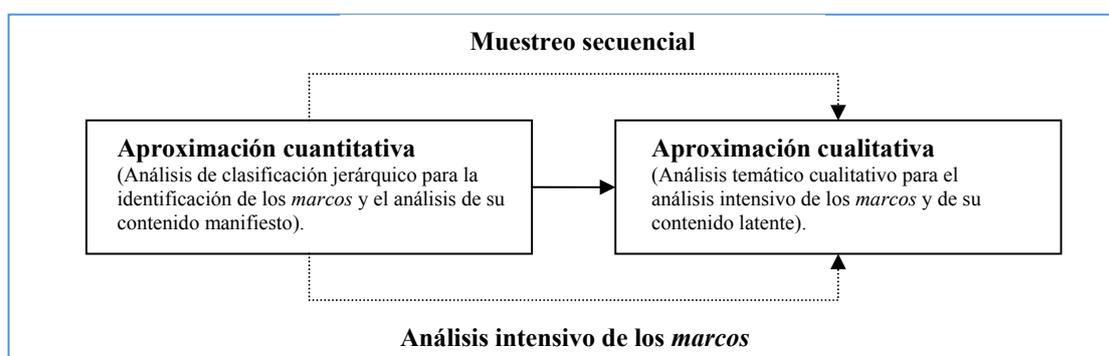


Figura 1. Diseño secuencial exploratorio

Fuente: Elaboración propia

### Obtención de los datos analizados

Los datos analizados en esta investigación han sido artículos de prensa publicados entre 1996 y 2007 en España y en lengua castellana, descargados de las bases de datos *Lexis/Nexis* y *Factiva/Reuters*. La búsqueda ha incluido un listado de palabras clave relacionadas con la temática de la investigación, que se han introducido en el buscador de manera individual o combinada. Se han obtenido inicialmente 1064 artículos, los cuales han quedado reducidos a 808 con la aplicación de un conjunto de criterios de inclusión<sup>[1]</sup>.

### Primera fase: Análisis cuantitativo de contenido

En la primera etapa del análisis de contenido, se ha procedido a la preparación de los artículos recogidos. En un inicio, se han eliminado los signos de puntuación y se han corregido los errores tipográficos presentes en el texto. Seguidamente, se ha introducido el total de artículos en una base de datos que se ha importado al programa *Wordstat* (Provalis Research, 2007). Mediante este programa, se ha hecho una *lematización* o reducción lingüística de las formas *flexionadas* a su forma canónica y se ha elaborado un listado de las palabras que finalmente se han incluido en el análisis. Para la configuración de dicho

[1] El listado de palabras clave empleadas para la búsqueda, la descripción de los criterios de inclusión y las características de los artículos analizados puede obtenerse contactando personalmente con el primer autor.

listado, se han aplicado una serie de criterios de inclusión sobre el total de palabras identificadas en el texto de los artículos y, a continuación, se ha hecho un ejercicio de validación semántica denominado “*Keyword-in-context*” (KWIC). El fin de este último procedimiento ha sido asegurar que, en todas sus apariciones en el texto, las palabras polisémicas incluidas en el diccionario hayan conservado el significado pertinente de acuerdo con el tema investigado. Finalmente, una vez organizada toda la información, se ha ejecutado un análisis de clasificación jerárquica basado en la coocurrencia de palabras, tomando el artículo como unidad de coocurrencia y usando el coeficiente de *Jaccard* como medida de proximidad. El resultado ha sido la obtención de un sistema compuesto por cinco entidades conceptuales o *marcos* que se presentan en la figura 2. Por orden de aglomeración: el marco de la regulación local, el marco de la regulación estatal y europea, el marco del desarrollo tecnológico y el marco de la acción jurídica.

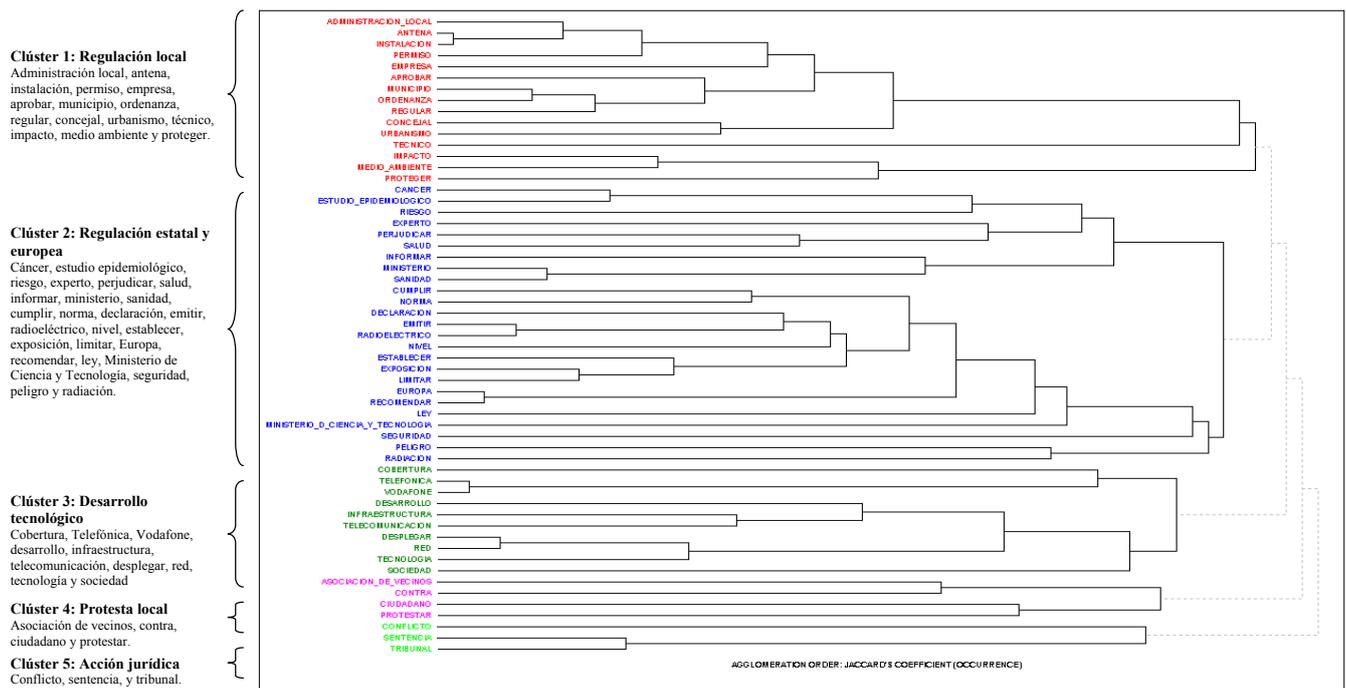


Figura 2. Dendrograma jerárquico

El análisis de clasificación jerárquica ha posibilitado identificar, de manera sistemática, transparente y fiable, los *marcos* presentes en el texto. En primer lugar, el hecho de que la configuración de los *marcos* se haya basado en criterios estadísticos ha permitido evitar las distorsiones asociadas con el proceso de identificación subjetiva de los *marcos*, a menudo presente en las aproximaciones cualitativas. Asimismo, a diferencia de lo que sucede en los enfoques manuales de detección de *marcos*, dónde en la mayoría de casos es imposible saber cómo -o bajo qué criterios- los *marcos* fueron extraídos del material analizado (Matthes y Kohring, 2008), en este caso se han podido describir minuciosamente cada uno de los pasos seguidos hasta su obtención. Finalmente, la técnica ha permitido, también, analizar un volumen muy extenso de información, la cual hubiera sido muy costosa de abordar desde un enfoque cualitativo. En contraste con estos beneficios, el análisis ha presentado, sin embargo, una limitación central: no ha permitido conocer la dimensión *latente* de los *marcos*, generándose así importantes restricciones en el nivel teórico-conceptual y explicativo.

*Segunda fase: Análisis temático cualitativo*

Tomando en consideración esta última limitación, se ha realizado, en una segunda fase, un análisis temático cualitativo del mismo corpus de datos textuales a partir del programa informático *NVivo* (QSR International, 2008). Dicho análisis ha tenido un doble fin, tal y como se indica en la figura 1.

En primer lugar, las agregaciones de palabras resultantes del análisis de clasificación jerárquica han servido para elaborar una segunda muestra cualitativa de artículos periodísticos. Mediante la opción *Text search query* se han discriminado aquéllos artículos dónde han coincidido palabras provenientes de un mismo clúster, conformándose así una muestra de artículos directamente vinculada con cada uno de los *marcos*. En segundo lugar, definida la muestra, dichos *marcos* han sido usados, en un segundo momento, como categorías conceptuales para el estudio la dimensión *latente* del problema. De este modo, se ha partido de una estructura de codificación que ha estado inicialmente configurada por tres secciones de códigos: los *marcos*, los agentes sociales protagonistas del problema y las tres dimensiones del riesgo, esto es, su evaluación, percepción y gestión. A estas secciones se han ido añadiendo, de manera inductiva, nuevos segmentos configurados por otros aspectos –procesos, actitudes, estrategias, relaciones, etc.- que han permitido redirigir el análisis hacia la respuesta a la segunda pregunta de investigación (figura 3).

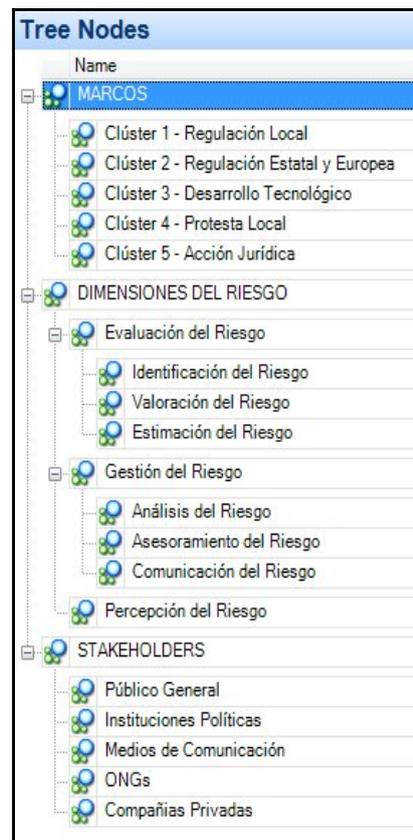


Figura 3. Libro de códigos

Fuente: Elaboración propia

## Conclusión

El ejemplo presentado evidencia las fortalezas de la combinación de las metodologías cualitativa y cuantitativa en el análisis de un mismo corpus de datos textuales. Por una parte, la detección cuantitativa de los *marcos* ha sido complementada con las fortalezas de un análisis cualitativo que ha ayudado a profundizar en el estudio de los actores y los procesos implicados. Por la otra, las dificultades de carácter técnico –la disposición de un alto volumen de información- en el contexto de un fenómeno poco estudiado –y, por tanto, de difícil aproximación deductiva- han sido parcialmente superadas mediante el uso de técnicas estadísticas que han permitido una primera aproximación a los constructos

temáticos presentes en el texto. En conclusión, el ejemplo constituye una contribución original a la literatura empírica en análisis de *marcos* y, al mismo tiempo, demuestra que, tanto en este campo como en otros de las ciencias sociales y de la salud, la combinación de ambas metodologías no es únicamente posible, sino también potencialmente necesaria.

### Referencias

- Bazeley, P. (2009). Analysing mixed methods data. En S. Andrew y E. J. Halcomb (Eds.), *Mixed methods research for nursing and the health sciences* (pp. 84-118). Oxford: Willey-Blackwell.
- Burgess, A. (2004). *Cellular phones, public fears and a culture of precaution*. London: Cambridge.
- Gitlin, T. (1980). *The whole world is watching: Mass media and the unmaking of the new left*. Berkeley: University of California Press.
- Happ, M. B., Dabbs, A. D., Tate, J., Hricik, A. y Erlen, J. (2006). Exemplars of mixed methods data combination and analysis. *Nursing Research*, 55(2 Suppl), S43-49.
- Hertog, J.K., y McLeod, D. M. (2001). A multiperspectival approach to framing analysis: A field guide. En S.D. Reese, O.H. Gandy y A. E. Grant (Eds.) *Framing public life: Perspectives on media and our understanding of the social world*. Mahwah: Lawrence Erlbaum Associates.
- Johnson, R.B., y Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Johnson, R.B., Onwuegbuzie, A.J., y Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Köenig, T. (2004). *Identification and measurement of frames*. Consultado 28 Octubre 2007 de <http://www.ccsr.ac.uk/methods/publications/frameanalysis/measurement.html>
- Matthes, J. y Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, 58(2), 258-279.
- Miller, M.M. (1997). Frame mapping and analysis of news coverage of contentious issues. *Social Science Computer Review*, 15(4), 367-378.
- Provalis Research. (2007). WordStat (Version 5.1.7). Montréal, Canada.
- QSR International. (2008). NVivo (Version 8.0.264.0). Doncaster, Australia.
- Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection and analysis techniques in mixed methods studies. *Research in Nursing and Health*, 23(3), 246-255.
- Tashakkori, A. y Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks: Sage.
- Todd, Z., Nerlich, B. y Clarke, D.D (Eds.) (2004). *Mixing methods in psychology: The integration of qualitative and quantitative methods in theory and practice*. Sussex: Taylor and Francis.

# EVALUACIÓN DE UN PROGRAMA DE HABILIDADES DE PENSAMIENTO EN NIÑOS CON SÍNDROME DE ASPERGER

Marisa Herrero<sup>1</sup>, María Frontera<sup>1</sup>, Pedro Allueva<sup>1</sup> y M. Teresa Anguera<sup>2</sup>

<sup>1</sup>Universidad de Zaragoza

<sup>2</sup>Universidad de Barcelona

Las personas con síndrome de Asperger, en su mayoría, tienen capacidad intelectual dentro del rango normal, aunque con perfil distinto de habilidades aparentes, por lo que puede resultar beneficioso para ellos ayudarles para potenciar y mejorar el desarrollo de la capacidad para orientar y mantener la atención, favorecer la adquisición de habilidades de comprensión y desarrollo de los conceptos abstractos y en el aprendizaje de habilidades en la resolución de problemas. El trabajo que presentamos tiene como objetivo evaluar un programa de intervención en habilidades metacognitivas que les ayude a conocerse mejor, planificar, predecir, controlar, verificar y conocer distintas estrategias que les permitan ser más hábiles metacognitivamente, es decir, que sean capaces de utilizar adecuada o hábilmente su conocimiento, utilizándolo para la resolución de tareas y la mejora de su propio conocimiento. El programa se aplica a un grupo de estudiantes de edades comprendidas entre 8-10 años. La evaluación del programa se lleva a cabo desde un planteamiento de complementariedad metodológica, toda vez que interesan aspectos cuantitativos y cualitativos, como son la posible existencia de patrones de conducta en la resolución de las tareas propuestas, y el análisis cualitativo de la estrategia utilizada en las distintas situaciones.

El trastorno o síndrome de Asperger es un término que designa un Trastorno Generalizado del Desarrollo, categoría diagnóstica en la que se incluyen el autismo y otros trastornos relacionados con él. Este trastorno fue oficialmente reconocido por primera vez en el CIE-10 (Organización Mundial de la Salud) en 1993 y en el Manual Diagnóstico de los Trastornos Mentales en su cuarta edición en 1994 (DSM-IV).

El Síndrome de Asperger se sitúa en la zona superior de lo que se conoce normalmente como espectro autista: representando la parte menos severa y con mayor grado de funcionamiento del mismo. Al igual que las demás condiciones registradas en dicho espectro, se cree que representa un trastorno del desarrollo con base neurológica, de causa desconocida, caracterizado por alteraciones sociales, comunicativas y patrones restrictivos, restringidos y estereotipados de comportamientos e intereses.

Las personas con síndrome de Asperger tienen, de forma característica, un nivel de inteligencia normal o superior. A pesar de ello presentan *déficits cognitivos* y un *estilo de aprendizaje peculiar* que afectan a su acceso al currículo académico y a su funcionamiento cotidiano. Entre las dificultades cognitivas que pueden obstaculizar su adaptación podemos destacar:

- *Déficit de cognición social*, manifestado en sus dificultades para comprender y atribuir estados mentales a los demás y a uno mismo, como los deseos, creencias e intenciones. El sujeto con este tipo de trastornos tiene escasa comprensión de sí mismo como persona y poca capacidad para comprender y predecir el comportamiento de los demás. Esta dificultad explica la falta de competencia social, las dificultades en la comunicación y, secundariamente, los problemas con la mediación social del aprendizaje, así como su vulnerabilidad desde un punto de vista social.
- *Déficit en habilidades de organización y planificación*. Las personas con síndrome de Asperger suelen tener serios problemas con la organización personal. Muestran dificultad para formar una representación interna del objetivo final de la tarea a realizar así como para procesar la secuencia

de pasos necesarios para resolverla. Estas dificultades derivan del déficit de función ejecutiva, característico del perfil neuropsicológico de estos sujetos. La disminuida capacidad para la organización y planificación resulta frecuentemente en dificultades que entorpecen la capacidad de aprendizaje y obstaculizan su funcionamiento cotidiano.

- *Rigidez mental*: Se interpreta como consecuencia también de un déficit en función ejecutiva. En las personas con síndrome de Asperger suele manifestarse primariamente a través de preocupaciones absorbentes referidas a temas o intereses circunscritos, sobre los que el individuo puede acumular gran cantidad de información y que se mantienen con gran intensidad y con la exclusión de otras actividades. Por otra parte, se expresa también en su tendencia a perseverar en sus respuestas de resolución de problemas, en lugar de generar nuevas hipótesis y en su dificultad para aprender de los errores. Manifiestan dificultades para contemplar distintas alternativas de solución de un problema y analizar la información desde diferentes puntos de vista.
- *Déficit en habilidades de comprensión, interpretación y evaluación crítica de la información adquirida*. Parecen presentar una dificultad específica para integrar la información y construir de ese modo representaciones significativas y contextualizadas de alto nivel (lo que se ha denominado débil coherencia central). Manifiestan, así, notables deficiencias con respecto a su capacidad para alcanzar el significado y sentido de la información adquirida, para comprenderla y evaluarla críticamente. Esto limita su capacidad de tomar decisiones, participar en las discusiones de grupo y formar sus propias ideas.
- *Dificultades con el pensamiento abstracto*. Tienen dificultad en la comprensión adecuada de las ideas definidas a través de propiedades no observables directamente. Tienen a desarrollar con mayor frecuencia destrezas en el campo técnico que en el abstracto. Su pensamiento es muy literal, adherido a lo concreto.
- *Dificultad en la formación de conceptos no verbales*. En una proporción alta de sujetos con síndrome de Asperger se observan las deficiencias neuropsicológicas descritas en el trastorno de aprendizaje no verbal o de hemisferio derecho, caracterizado por la discrepancia entre el CI verbal y el de ejecución (CI de ejecución 10 o 15 puntos inferior al CI verbal), fluidez en el lenguaje, gran memoria mecánica para la información verbal, hiperlexia, dificultades en la formación de conceptos y solución de problemas no verbales, en la pragmática y contenido del lenguaje y en el razonamiento deductivo.
- *Déficit en la capacidad de generalización*. Memoriza unos pasos concretos para resolver un problema pero le cuesta trabajo abstraer los principios que guían el pensamiento de solución de problemas.
- *Problemas de atención*. Manifiestan a menudo problemas específicos de atención selectiva mostrando una capacidad adecuada para concentrarse en las actividades de su interés. Sus dificultades atencionales parecen deberse en gran medida a una dificultad de discernir lo que es relevante, así como para desplazar de forma flexible el foco de su atención.
- *Problemas con la memoria*: Suelen tener una buena memoria mecánica. Sin embargo, cuando tienen que repetir una historia, pueden ser incapaces de explicar lo esencial de la (débil coherencia central). Tienen dificultad para estructurar y agrupar adecuadamente la información y así formar unidades significativas. Suelen tener dificultades importantes en la memoria autobiográfica y, en general, en acceder a sus recuerdos.

Estas dificultades obstaculizan el funcionamiento académico y la adaptación personal y social de estas personas, precisando una intervención educativa específica que contemple la adaptación de las enseñanzas y del entorno social, por una parte, y el desarrollo de sus competencias, por otra.

En relación con esto último, parece especialmente relevante el desarrollo de sus *habilidades metacognitivas* con el objeto de *mejorar su conocimiento de sí mismo* (su forma de comprender, aprender y saber) así como su *capacidad de control, organización y regulación* de su aprendizaje y, en general, de su propio funcionamiento.

Ello fue lo que nos llevo a plantearnos el estudio que presentamos cuyo objetivo es: Implementar un programa de intervención cognitiva (de entrenamiento en habilidades metacognitivas) y evaluar su eficacia. El programa pretende fomentar el desarrollo de habilidades básicas de pensamiento, que las teorías explicativas actuales indican que están dañadas en los SA y pueden dar cuenta de las alteraciones nucleares.

### Método

#### Participantes

El estudio se ha llevado a cabo con tres estudiantes de educación primaria (dos niños y una niña) de edades comprendidas entre 9-10 años, con diagnóstico de Síndrome de Asperger.

#### Instrumentos

##### Instrumento de observación

Se ha elaborado un instrumento (figura 1) *ad hoc*, a partir del marco teórico y de las imágenes de video producidas, caracterizado por la complementariedad entre formato de campo y sistema de categorías (Anguera, 1991; Anguera, Magnusson & Jonsson, 2007) que nos permite el estudio de la multidimensionalidad o despliegamiento de diversos niveles de respuesta que forman parte de las estrategias de habilidades metacognitivas.

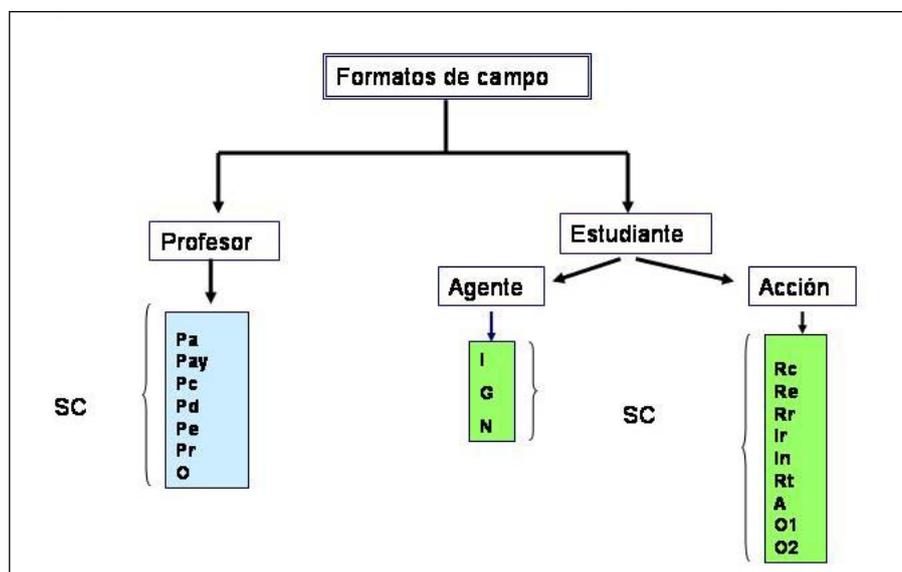


Figura 1. Esquema del Instrumento de observación

### *Instrumentos técnicos:*

- Para llevar a cabo el registro de las sesiones de observación hemos utilizado una cámara de vídeo convencional que nos ha permitido el registro y almacenamiento del total de sesiones realizadas durante los seis en que se ha llevado a cabo la implementación del programa. Para materializar la grabación se contaba con el consentimiento informado por parte de los padres de cada uno de los niños participantes.
- Programas informáticos: Para los respectivos análisis se han utilizado los siguientes programas
  - Atlas.ti: Programa informático que permite facilitar el análisis cualitativo de archivos de vídeo, ayudando al interprete humano en la segmentación y codificación de los mismos, lo que da lugar a estructura relacionales interpretables.
  - SDIS-GSEQ: Programa no comercial que permite detectar la existencia de regularidades de conducta de forma directa una vez efectuada la codificación, operando a partir de datos de naturaleza categórica, y mediante un algoritmo que se basa en el contraste entre probabilidades condicionadas e incondicionadas.

### *Procedimiento*

La implementación del programa ha tenido lugar en un aula de la Facultad de Educación a la que acudían los niños a la salida del colegio, con una periodicidad de una sesión semanal durante los meses de enero a junio, lo que hace un total de 23 sesiones de 1 hora de duración cada una, aunque en esta comunicación presentamos, únicamente los resultados de dos sesiones. El programa ha sido implementado siempre por la misma persona.

Las sesiones han sido todas grabadas en vídeo y posteriormente codificadas utilizando para ello el Atlas.ti que nos ha permitido, además, la creación de citas y memos que facilitasen el posterior análisis.

El estudio se ha planteado desde la complementariedad metodológica, integrando la vertiente cualitativa clásica con la vertiente actual, más proclive a un tratamiento cuantitativo de la información obtenida de forma cualitativa.

### Resultados

Para llevar a cabo el control de calidad del dato se ha utilizado el coeficiente Kappa de Cohen (1960, 1968), mediante el programa informático SDIS-GSEQ, que ha resultado satisfactorio. En la Tabla 1 se muestra el resultado obtenido

Tabla 1. Control de calidad del dato, realizado mediante el programa SDIS-GSEQ (Bakeman y Quera, 1996)

Kappa de Cohen = 0,9789, concordancia = 98,00%						
Filas: 01, columnas: 02						
	Pe	Pc	IRc	O1	PcG	Gre
Pe	7	0	0	0	0	0
Pc	0	17	0	0	0	0
IRc	0	0	10	0	0	0
O1	0	0	0	40	0	0
PcG	0	0	0	0	5	0
Gre	0	0	0	0	0	8
PaG	0	0	0	0	0	0
PdS	0	0	0	0	0	0
SRr	0	0	0	0	0	0
PayS	0	0	0	0	0	0
IRta	0	0	0	0	0	0

Una vez superado el control de calidad del dato el análisis se ha abordado, como ya se ha dicho, desde dos perspectivas complementarias.

### 1. Análisis de contenido, llevado a cabo mediante el programa ATLAS.ti

En la Figura 2 podemos ver un ejemplo de la codificación de los datos de video, en el que se observa la segmentación de los datos y la asignación de los mismos a los códigos correspondientes al instrumento elaborado *ad hoc*. La codificación utilizada es literal, asignándose los códigos una vez finalizada la elaboración del instrumento de observación.

<p><b>P 6: 4_planificación1_28_abril.mpg - 6:1 [4_planificación1_28_abril.mpg] (578 875) (Super)</b>            Codes: [pe]            No memos            Comment:            al comenzzar la sesion la profesora les explica de que va hoy la tarea a realiza</p>
<p><b>P 6: 4_planificación1_28_abril.mpg - 6:2 [4_planificación1_28_abril.mpg] (876 915) (Super)</b>            Codes: [Pd]            No memos            Comment:            hace una pregunta directa ¿habeis ido alguna vez al cine? (para hacerles pensar en lo que van a hacer)</p>
<p><b>P 6: 4_planificación1_28_abril.mpg - 6:3 [4_planificación1_28_abril.mpg] (916 969) (Super)</b>            Codes: [O1]            No memos            Comment:            Silencio parece que piensan que deben responder</p>
<p><b>P 6: 4_planificación1_28_abril.mpg - 6:4 [4_planificación1_28_abril.mpg] (970 1042) (Super)</b>            Codes: [IRc]            No memos            Comment:            Ines sabe lo que le preguntan es decir sabe lo que tiene que responder</p>

Figura 2. Codificación de datos de video mediante Atlas.ti

Una vez efectuada la codificación de las grabaciones de vídeo correspondientes a las diferentes sesiones de observación, se ha llevado a cabo la obtención de redes, o familias de códigos (Figura 3), que nos muestran las conexiones que se establecen entre las categorías, tanto a nivel inter-dimensional como intra-dimensional, dando lugar a una estructura muy interesante a efectos interpretativos.

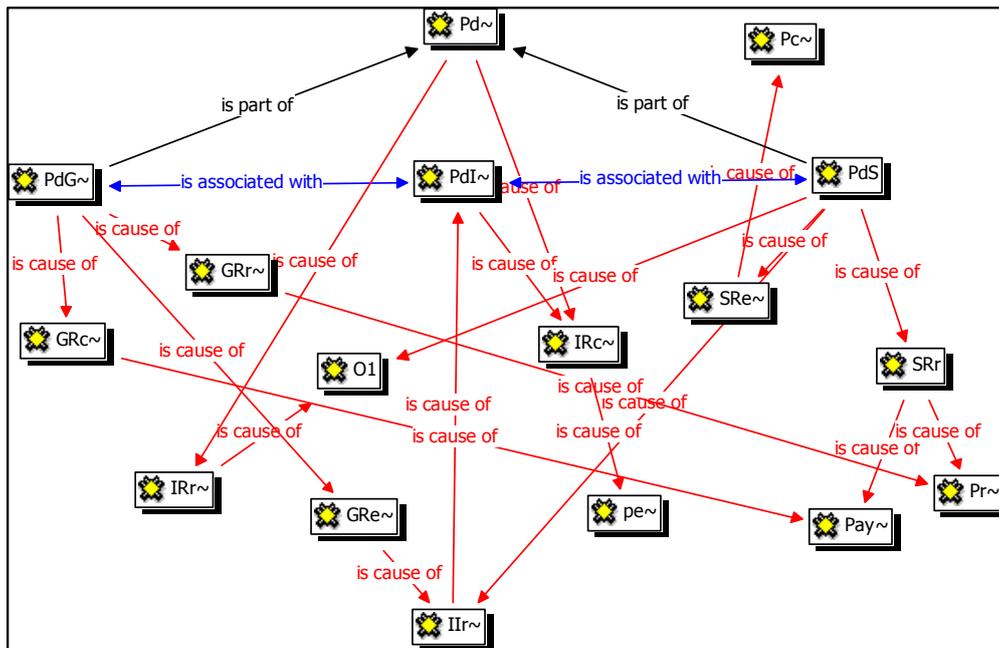


Figura 3. Familias de códigos correspondiente a los datos obtenidos a partir del instrumento de la tabla 1

## 2. Detección de patrones de conducta mediante el programa SDIS-GSEQ

El análisis secuencial de retardos es una técnica analítica desarrollada por Bakeman (Bakeman, 1978; Bakeman & Gottman, 1989), y por Sackett (1979, 1980), para detectar la posible existencia de patrones de conducta que no son otra cosa que regularidades en la sucesión de las conductas registradas. El objetivo de este tipo de análisis es la detección de patrones secuenciales de conducta, lo que se lleva a cabo mediante la búsqueda de contingencias secuenciales (Quera, 1993) entre categorías de conducta o configuraciones (teniendo en cuenta que el instrumento de observación elaborado combina formatos de campo con sistemas de categorías). En definitiva, El análisis secuencial es una de las modalidades de análisis que se puede aplicar a datos de observación tanto directa como indirecta y nos permite obtener patrones de conductas, a través de la detección de contingencias secuenciales de diferentes categorías (Arias y Anguera, 2004). Esta forma de análisis no es la única posible en estudios.

En la Figura 4 mostramos un fragmento del análisis efectuado, en el que hemos resaltado los valores significativos de los residuos ajustados, que muestran parte de las asociaciones significativas en el primer retardo, siempre teniendo en cuenta que se trata de datos, únicamente de dos sesiones.

Retardo 1. RS&J. Residuos ajustados								
Datos	Condicionados							Pa
	GRc	GRe	GRr	IRc	IRr	IRt	O1	
Pa	-0.383:	-0.585:	-0.615:	-0.615:	-0.615:	-0.341:	2.626:	0.000:
Pay	-1.241:	2.654:	0.199:	-0.531:	0.199:	-1.105:	-0.850:	0.000:
Pc	-0.621:	0.271	0.699	-0.604	2.655	-1.421:	-1.375	0.000:
Pd	2.866:	-1.313:	0.664:	2.068:	-1.442:	1.148:	-1.731:	0.000:
Pe	-0.499:	-0.763:	-0.801:	-0.801:	-0.801:	2.001:	2.431:	0.000:
Pr	-0.838:	-1.280:	-1.344:	-0.396:	-1.344:	0.832:	2.545:	0.000:

Figura 4. Fragmento del análisis secuencial de retardos efectuado

En la Figura 5 mostramos los patrones detectados mediante el análisis secuencial de retardos, a lo largo de tres retardos prospectivos, en las dos sesiones analizadas. La lectura de la tabla hemos de realizarla teniendo en cuenta que a cada uno de los códigos de la columna Cc (conducta criterio) le sigue, de forma estadísticamente significativa, los respectivos códigos de la fila, ordenados en función de los retardos calculados, en este caso 5 (Ret.1, Ret.2, Ret.3, Ret.4 y Ret.5)

Cc	Ret.1	Ret. 2	Ret. 3	Ret. 4	Ret.5
Pa	O1			IRt	
Pay	Gre		IRr		
Pc	IRr			Ser	Irt
Pd	GRc IRc		GRr		
Pe	IRt IO1				
Pr	O1	Gre Pa		O1	

Figura 5. Patrones secuenciales detectados

### Discusión

Teniendo en cuenta lo pocos datos analizados hasta el momento, no podemos determinar la eficacia o no del programa de intervención, pero si nos permiten darnos cuenta que los resultados obtenidos mediante el análisis cualitativo de los datos se ven no solo complementados sino incluso corroborados, al menos en parte, por el análisis cuantitativo de los mismos.

### Referencias

- Anguera, M.T. (1991). Proceso de categorización. En M.T. Anguera (Ed.) *Metodología observacional en la investigación psicológica*. (pp. 115-167). Vol.1. Barcelona: PPU.
- Anguera, M.T.; Magnusson, M. y Jonsson, G. (2007). Instrumentos no estándar. *Avances en medición*, 5(1), 63-82.
- Arias, E. y Anguera, M.T. (2004). Detección de patrones de conducta comunicativa en un grupo terapéutico de adolescentes. *Acción Psicológica*, 3(3), 199-206.
- Bakeman, R. (1978). Untangling streams of behavior: Sequential analysis of observation data. In Gene Paul Sackett (Ed.), *Observing Behavior, vol. 2. Data collection and analysis methods* (S. 63-78). Baltimore: University of Park Press
- Bakeman, R. y Gottman, J. (1989). *Observación de la Interacción. Introducción al análisis secuencial*. Madrid: Morata.
- Bakeman, R. y Quera, V. (1996). *Análisis de la interacción. Análisis secuencial con SDIS y GSEG*. Madrid: Ra-Ma.

- Frontera, M. (2008). Funcionamiento cognitivo en síndrome de Asperger. En *Síndrome de Asperger: Aspectos discapacitantes y valoración*. Federación Asperger España. Difusión recomendada por la Secretaría de Estado de Servicios Sociales, Familias y Discapacidad. Ministerio de Trabajo y Asuntos Sociales.
- Quera, V. (1993). Análisis secuencial. En M.T. Anguera (Ed.). *Metodología Observacional en la investigación psicológica*(pp. 341-586). Vol. II. Barcelona: P.P.U.,
- Sackett, G.P. (1979) The lag sequential analysis of Contingency and Cyclicity in Behavioral interaction research. In J. D. Osofsky (eds.), *Handbook of Infant Development* (S. 623-649). New York: Wiley.
- Sackett, Gene Paul. (1980). Lag sequential analysis as a data reduction technique in social interaction research. In D. B. Sawin, R. C. Hawkins, L. O. Walker & J- H. Penticuff (Eds.). *Exceptional infant. Psychosocial risks in infant-environment transactions* (S. 300-340). New York: Brunner/Mazel.

# ESTUDIO COMPARATIVO DE LA PERCEPCIÓN DEL RIESGO ENTRE EL PERSONAL INTERNO Y EXTERNO DE UNA OBRA DE CONSTRUCCIÓN

Elisenda López Fernández<sup>1</sup>, Lluís Maria<sup>1</sup> y Olatz López Fernández<sup>2</sup>

<sup>1</sup>Universidad Politécnica de Cataluña

<sup>2</sup>Universidad de Barcelona

La construcción es uno de los sectores de mayor siniestralidad laboral en España, por los riesgos específicos a los que se ven sometidos los trabajadores en la obra, por lo que es necesario estudiar este sector de forma diferenciada. Actualmente se dispone de un marco jurídico que establece el deber del empresario de garantizar una protección eficaz en seguridad y salud laboral, donde se incluye la obligación de evaluar los riesgos en los puestos de trabajo mediante métodos objetivos que permitan identificar y valorar dichos riesgos. Sin embargo, no se obtiene una completa valoración real de la evaluación de riesgos. El propósito de este estudio es comparar la percepción de los riesgos laborales del personal (interno) de producción de una obra y del personal (externo) experto en prevención de la misma obra, mediante entrevistas analizadas con Atlas-ti 5, en que se han observado discrepancias y similitudes en la valoración de ambos tipos de personal (no experto-experto) respecto a los riesgos de la obra. Los resultados indican concordancia general en la identificación de riesgos y sus medidas preventivas, aunque hay divergencia en cuanto a los procedimientos implicados en la reducción de los riesgos, debido a múltiples causas derivadas de la organización.

Actualmente en el panorama empresarial del país, es una necesidad social y un deber legal el hecho de trabajar en condiciones seguras en el puesto de trabajo, según lo establecido en la Ley de Prevención de Riesgos Laborales (de ahora en adelante LPRL). Esta normativa es vinculante para empresas de cualquier actividad económica, tamaño, naturaleza o ámbito geográfico, excepto para aquéllas en las que dicha actividad se regula de forma específica.

Para poder proteger de forma eficaz la integridad física de los trabajadores además de garantizar unas condiciones óptimas de seguridad en sus puestos de trabajo, se considera necesario que la organización disponga de una serie de herramientas para determinar la existencia de riesgos laborales, la valoración de los mismos y las medidas a adoptar para eliminarlos, si es posible, y si no reducirlos. Uno de los instrumentos básicos en Prevención es la evaluación de los riesgos presentes en los puestos de trabajo, entendida como el proceso dirigido a estimar la magnitud de aquellos riesgos que no hayan podido evitarse. Pero esta medida no es suficiente para incidir en la protección de la salud y la disminución de los accidentes laborales, por lo que es necesario la utilización además de otros métodos complementarios, que permitan incidir más profundamente sobre los trabajadores y la organización del trabajo, esto es, que vayan más allá del cumplimiento formal y legal de la seguridad y salud en el trabajo.

Es por este motivo que surge la necesidad del estudio de otros aspectos en el campo de la Prevención, que se centren, por un lado, en la mejora de la organización del trabajo (definición de los puestos de trabajo, del sistema productivo, del personal capacitado, de los equipos e instalaciones, entre otros) y por otro, en el llamado *factor humano*, que hace referencia a las personas y las relaciones que hay entre ellas dentro del entorno laboral para que la gestión de la seguridad y salud sea más eficaz y, a la vez, eficiente.

La presente comunicación, que forma parte de una tesis doctoral, se centra en el sector de la construcción desde la perspectiva de la psicología, debido a varios motivos:

1. La construcción es uno de los sectores con mayor índice de incidencia de accidentes de trabajo
2. El estudio del factor humano relacionado con la prevención de accidentes en la construcción *in situ* en España es limitado
3. Existe un marco legal consolidado en prevención, sin embargo la investigación en este campo es más reciente y se ha profundizado poco en este concepto
4. Es tan importante la valoración subjetiva del riesgo laboral como la evaluación objetiva del riesgo

La acción humana es básica para que el sistema técnico de la organización funcione correctamente, aunque en una empresa se establezcan unas medidas técnicas adecuadas para prevenir los accidentes de trabajo, la acción final de las personas en sus lugares de trabajo puede malograr los resultados esperados por la organización de la prevención. Y aunque desde el punto de vista *objetivo* se hayan evaluado correctamente los riesgos laborales puede ocurrir que la consecución de esas medidas preventivas implantadas por la empresa sea ineficaz y se continúen produciendo accidentes. La razón de ese fallo en el proceso, puede ser consecuencia de varias razones, por ejemplo: las actitudes erróneas sobre la seguridad y salud que tengan los trabajadores, la realización de conductas de riesgo por parte de los trabajadores o de una disonancia, entre lo que el técnico establece como riesgo y la percepción sesgada del trabajador, sobre los riesgos en su puesto de trabajo, etc.

En esta parte del estudio el contexto lo constituyen principalmente las personas, que forman parte de la obra, además de las condiciones de trabajo, la cultura preventiva, las relaciones creadas en el seno de la empresa (que comprenden a los trabajadores, mandos intermedios, empresario) y todos aquellos factores que pueden condicionar los comportamientos seguros que adopten los trabajadores.

Una forma de conocer y recoger el valor subjetivo del riesgo de las personas que trabajan en una obra de construcción (entendiéndose como tales: los trabajadores, los encargados, los técnicos de prevención, las empresas subcontratadas, los responsables de prevención de las constructoras) y completar la evaluación *objetiva* de los riesgos laborales es conocer sus opiniones éstos y las medidas preventivas y de protección que ha adoptado la empresa para evitarlos, con el fin de determinar cómo perciben los riesgos evaluados *objetivamente*. Para ello se ha utilizado la técnica de la entrevista semi-estructurada. Para extraer conclusiones se ha utilizado un software específico para analizar los datos recogidos en estas entrevistas, el ATLAS.ti Versión 5. Este tipo de datos es muy habitual en investigaciones de ámbito psicosocial, dentro del que se encuentra el ámbito objeto de estudio, el laboral. Este programa nos permite manejar datos cualitativos, básicamente los textuales. Estos datos han de ser interpretados y, para ello, primero han de ser identificados, para poder ser codificados en temas, conceptos, procesos, contextos, etc.

### *Descripción del estudio*

La obra objeto de estudio reúne las siguientes características:

- Es un tramo de la obra Línea 9 de Metro de Barcelona (6 estaciones)
- Túnel de 10 km. longitud
- La tecnología de excavación utilizada es la tuneladora
- Revestimiento del túnel con dovelas prefabricadas de hormigón
- Planta de elementos prefabricados de dovelas

El tramo de la obra objeto de estudio es la planta de prefabricado de dovelas de la UTE Línea 9, en esta planta se fabrican las dovelas, que son piezas de hormigón armado que proporcionan sostenimiento a la estructura de las paredes del túnel, lo hacen adoptando forma de anillos y son colocados por la tuneladora.

Los participantes del estudio se han clasificado en dos grupos, uno denominado grupo externo (número 1), que está formado por personas externas a la obra, que son expertos con una dilatada experiencia en el sector de la construcción y, en concreto, en el ámbito de la seguridad y salud en el trabajo y con una formación de nivel superior en prevención de riesgos laborales. El grupo denominado interno (número 2), está constituido por personal interno de la obra, que trabaja en la UTE L9, son personas que trabajan en la producción de la planta, los integrantes del servicio de prevención o del servicio médico.

Tabla 1. Distribución de expertos por Grupo de pertenencia (1 y 2)

GRUPO de participantes	%	N
1- Externo	54	7
2- Interno	46	6
TOTAL	100	13

Tabla 2. Distribución de expertos por experiencia en el sector (años)

Experiencia (años)	%	N
0-15	46	6
16-30	26	3
Más de 31	31	4
TOTAL	100	13

Los instrumentos utilizados para recoger las valoraciones y opiniones de los participantes de ambos grupos han sido entrevistas semi-estructuradas, se ha diseñado para cada grupo un guión con una parte preguntas que son iguales en ambos grupos y otras específicas, con el objeto de hacer comparaciones sobre las valoraciones de cada una de ellas y ver las diferencias y similitudes entre las respuestas de los participantes. Además de los guiones se ha utilizado una grabadora de voz, lápiz y papel.

El procedimiento para la recogida de datos ha sido:

- Para el personal de interno de la obra, se les ha entrevistado en las instalaciones de la UTE L9, en las oficinas de la obra, siguiendo el guión de la entrevista semi-estructurada *ad hoc* para este grupo.
- Para el personal externo a la obra, se les ha entrevistado en sus oficinas, siguiendo el guión de la entrevista semi-estructurada diseñada para este tipo de personal.

### Conclusión

Los expertos internos y externos a la obra establecen las causas humanas como las más influyentes y frecuentes en la producción de los accidentes de trabajo. Este dato justifica la importancia de la realización de este tipo de investigación.

La mayoría de expertos coinciden en considerar que las características del sector son, a la vez, el principal problema de las obras para reducir los accidentes de trabajo (ver Figura 1).

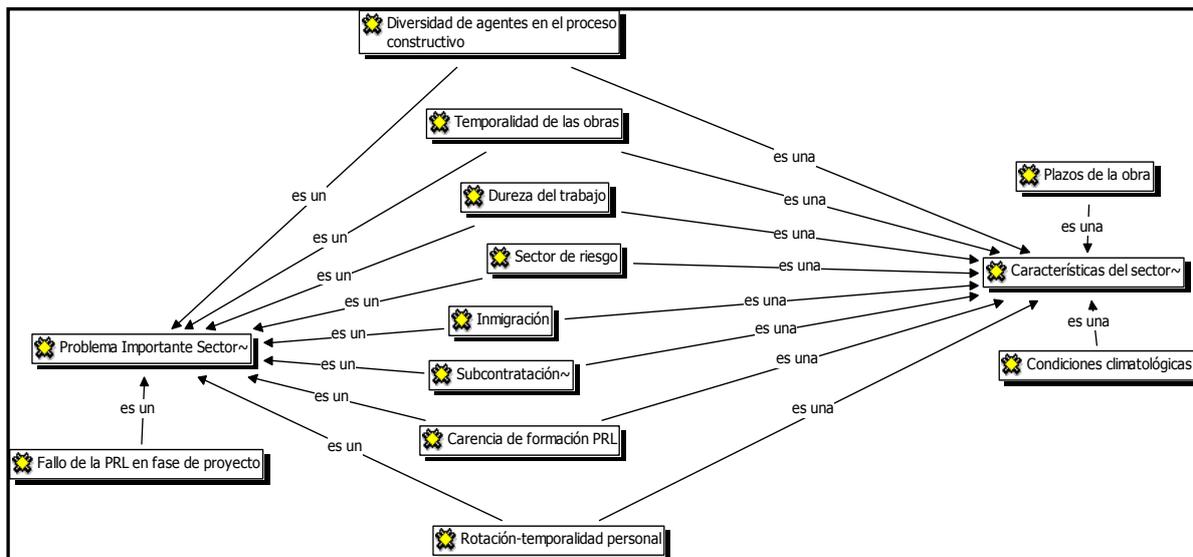


Figura 1. Network de codes “problemas del sector” y “características del sector”

Ambos grupos de participantes coinciden en que la descripción de los principales riesgos que son: riesgo de caída en altura, riesgo de sepultamiento dentro de zanjas, riesgo de atropello por vehículos, sobre esfuerzos.

El software Atlas.ti versión 5.0 es un paquete informático, en el ámbito de la prevención de riesgos laborales, que permite analizar y relacionar los contenidos de los relatos de los expertos y no expertos en prevención, relacionados con la causalidad de los accidentes en el sector de la construcción. Por lo que ha resultado ser una herramienta de gran utilidad en la investigación en seguridad y salud en el trabajo.

## Referencias

- Muñoz, J. (2005). *Análisis cualitativo de datos textuales con ATLAS.ti 5*. Barcelona: Universidad Autónoma de Barcelona.
- Sandín-Esteban, M.P. (2005). *Introducción al programa de análisis de datos cualitativos Atlas-ti 5.0*. Barcelona: Universidad de Barcelona.
- Anguera, M.T. (2001). Diseños observacionales, cuestión clave en el proceso de la Metodología Observacional. *Metodología de las Ciencias del Comportamiento*, 3 (2), 135-160.
- Anguera, M.T. (2006). ¿Cómo se lleva a cabo un registro observacional? *Butlletí LaRecerca*, Ficha 4. Barcelona: Instituto de Ciencias de la Educación de la UB.

# APLICACIÓN DE HERRAMIENTAS OBSERVACIONALES PARA EL ANÁLISIS DE LA VIOLENCIA EN MEDIOS IMPRESOS VENEZOLANOS

Zaida Márquez<sup>1</sup> y Antonio Hernández Mendo<sup>2</sup>

<sup>1</sup>Universidad Centroccidental “Lisandro Alvarado”, Venezuela

<sup>2</sup>Universidad de Málaga

El objetivo de esta investigación fue aplicar dos herramientas observacionales: un sistema de formatos de campo y un sistema de categorías emergentes para el análisis de las noticias que hacen referencia a la violencia, en la prensa escrita regional venezolana. El sistema de formatos de campo se estructuró con base en tres criterios para el análisis cuantitativo del contenido: *Forma, Texto y Contenido del Texto*; y en el sistema de categorías emergentes se definieron nueve *codes* para el análisis cualitativo con el programa computacional *ATLAS.ti*. La muestra estuvo conformada por 313 noticias contenidas en los tres periódicos regionales de mayor tirada en la ciudad de Barquisimeto, Estado Lara (*El Impulso, El Informador, El Diario Hoy*), identificándose 214 eventos de violencia en dichas noticias. Los resultados del análisis de contenido indican que la prensa escrita venezolana utiliza recursos formales denominados “derrapes” para enfatizar y espectacularizar los hechos de violencia.

El estudio de la violencia en medios de comunicación social como la televisión, ha sido abordado desde diferentes enfoques metodológicos, sin embargo, otros medios de comunicación social como los impresos, han sido menos emprendidos, de allí que en esta investigación se haya pretendido realizar una descripción del tratamiento que la prensa escrita le otorga al fenómeno de la violencia mediante la técnica de análisis de contenido. En este sentido, este tipo de análisis representa una forma sistemática y rigurosa de conocer la utilización de ciertos procesos de influencia social por parte de los *mass media*. Para ello se utilizan herramientas observacionales que permiten la estructuración y análisis del corpus.

## Método

Se utilizó la Metodología Observacional, correspondiendo esta investigación a un diseño observacional mixto *nomotético/seguimiento/multidimensional* (Anguera y Blanco Villaseñor, 2003), por cuanto se trata de un estudio de un grupo de unidades observacionales (periódicos) recogidos en un lapso de tres meses, y para lo cual se construyeron dos herramientas observacionales. En primer lugar, se estructuró una herramienta *ad hoc*, un sistema de formatos de campo, para el análisis de contenido de las noticias. Se ha optado por este sistema de registro, ya que por su flexibilidad se ajusta más a los objetivos de la investigación. No obstante, y de forma complementaria para realizar el análisis cualitativo de las noticias, se construyó un sistema de categorías emergentes. Uno de los parámetros estimados para el análisis, ha sido la frecuencia del número de noticias sobre violencia en los diarios recopilados, lo que permite conocer la importancia que la prensa escrita concede al tema de la violencia.

## Participantes

Las noticias se recogieron durante el periodo de tres meses, dos días por semana, comenzando los lunes y alternando cada semana con un día diferente, incluyendo de esta manera todos los días de la semana; y en tres periódicos distintos de tirada regional (*El Impulso, El Informador y El Diario*

*Hoy*) correspondientes a la región de Barquisimeto, Estado Lara (Venezuela). Es de resaltar, que estos periódicos son de tirada diaria, razón por la cual se utilizaron dos días a la semana para realizar el muestreo. Se analizaron 78 ejemplares, 313 noticias y 214 eventos relacionados con violencia.

### *Procedimiento*

- Recopilación de los diarios durante tres meses.
- Selección de las noticias sobre violencia.
- Lectura de las noticias e identificación de los eventos.
- Construcción de las herramientas observacionales: sistema de formatos de campo y sistema de categorías emergentes.
- Clasificación de las noticias (utilizando una base de datos, y escaneando las noticias en dos formatos: como texto (TXT) para usarlas con el ATLAS/ti, y como imagen (JPG).
- Tratamiento de los datos (interpretación, análisis cuantitativo y cualitativo).
- Utilización del paquete estadístico SPSS v 12.0, para la elaboración de una matriz de datos (N° de aparición de las noticias; eventos; periódicos- El Impulso =1, El Informador =2, El Diario Hoy =3-; días de la semana; N° de palabras; alto de la noticia; ancho de la noticia; superficie de la noticia; intensidad sugerida por la noticia, e intensidad percibida por el observador; y del programa computacional ATLAS.ti (Mühr, 1991) para el análisis cualitativo de los datos y elaboración de los *networks*.

### *Herramientas observacionales*

- El sistema de formatos de campo se estructuró tomando en cuenta tres criterios con sus respectivos niveles y subniveles. El criterio número 1, FORMA incluye aquellos aspectos relacionados con la apariencia y la superficie de la noticia, constituyendo el nivel sintáctico del texto (Van Dijk, 1980), el criterio número 2, TEXTO incluye los aspectos relacionados con la clasificación y ubicación de la noticia como texto y con la fuente que la emite y el número 3, CONTENIDO DEL TEXTO contiene los aspectos relacionados con la significación e interpretación del texto, constituyendo el nivel semiótico (intensidad de la noticia sugerida por el periódico, intensidad de la noticia percibida por el observador).
- El sistema de categorías emergentes se elaboró en función de nueve categorías diferentes, en las que se agruparon los términos o unidades lexicales identificados en las noticias. Las categorías son las siguientes:
  - *Agresión verbal*: se refiere a todos los términos plasmados en las noticias y que designan violencia verbal (palabras descalificantes, amenazantes, insultos).
  - *Agresión física*: se refiere a todos aquellos términos que designan algún daño realizado a instalaciones, ambientes o lugares.
  - *Agresión hacia personas*: se refiere a todos los términos que designan algún acto de agresión realizado hacia una persona.
  - *Daños*: se refiere a los términos que hacen alusión a algún daño producido como causa de una acción violenta.
  - *Factores de Violencia*: se identifican los términos que hacen referencia a las causas de los eventos violentos (sociales, por estimulación, psicológicos, familiares, por impulso, otros).
  - *Violencia de Acción*: se refiere a los términos que designan actos de secuestros, robos, asaltos y apropiación indebida de espacios.
  - *Intensidad*: se refiere a todos los términos que designan la intensidad del evento violento.

- *Violencia juvenil*: se refiere a los términos que aluden como violentos a los jóvenes.
- *Violencia*: se refiere a todos los términos que se utilizan en las noticias para describir de forma genérica la violencia o los eventos violentos producidos.

## Resultados

En relación con los resultados obtenidos del sistema de formatos de campo se han establecido diferencias entre los periódicos analizados, obteniéndose que *El Impulso* publica mayor número de noticias (154 noticias), *El Informador* (99 noticias) y que el *Diario Hoy* (60 noticias). Además, *El Impulso* utiliza mayor número de palabras relacionadas con violencia (838 términos) que *El Informador* (392 términos) y el *Diario Hoy* (273 términos). Otro de los criterios que se utiliza para definir la importancia es el uso de imágenes para describir la noticia. En este sentido, *El Impulso* presenta un 35% de noticias con imágenes, *El Informador* un 50.5% de sus noticias con imágenes y el *Diario Hoy* un 66.6%. De acuerdo con este dato, el *Diario Hoy* utiliza este recurso con más frecuencia que los otros periódicos para así dar más relevancia a este tipo de noticias, además las imágenes que publica podrían considerarse de impacto agresivo. Igualmente, *El Impulso* publica 10% de sus noticias violentas a color, *El Informador* un 15% de noticias, y el *Diario Hoy*, a diferencia de los anteriores, un 50% de sus noticias sobre violencia las presenta a color. En referencia a los días en que se publican este tipo de noticias, se tiene que los lunes (44.4%) son los días en que los periódicos publican más noticias sobre violencia. Asimismo, *El Impulso* publica el 95.8% de las noticias violentas en la portada o sección inicial, *El Informador* el 40.8% de sus noticias y el *Diario Hoy* el 91.6% de las noticias. En cuanto a la intensidad de la noticia sugerida por el periódico y a la intensidad de la noticia percibida por el observador, ambas resultaron graves. Por tal razón, se puede concluir que los periódicos utilizados para el primer estudio empírico otorgan gran relevancia a las noticias sobre violencia. De los tres periódicos, *El Impulso* es el que da mayor importancia a este tipo de noticias, seguido de *El Diario Hoy*, y por último *El Informador*.

Con la aplicación del sistema de categorías emergentes, se determinó que existe también una diferencia entre los periódicos: *Impulso*, *Informador* y el *Diario Hoy*, en cuanto al uso de los términos que describen la violencia, tal como se refleja en la tabla 1.

Tabla 1. Resumen de los términos utilizados en los periódicos

Periódico	Agresión Verbal		Agresión Física		Agresión hacia Personas		Daños	
	T	%	T	%	T	%	T	%
El Impulso	200	60,2	31	51	137	46,4	150	60,9
El Informador	60	18	21	35	62	21,8	59	23,9
Diario Hoy	72	21,6	8	13,3	85	29,9	37	15
Total	332	100,0	60	100,0	284	100,0	246	100,0

Periódico	Factores de Violencia		Violencia de Acción		Violencia Juvenil		Intensidad		Violencia	
	T	%	T	%	T	%	T	%	T	%
El Impulso	47	42,3	41	50	79	56,4	28	66	125	57,6
El Informador	34	30,6	26	32	36	25,7	14	33,3	80	36,8
Diario Hoy	30	27,2	14	17	25	17,8	02	4,7	12	6,4
Total	111	100,0	81	100,0	140	100	44	100	217	100,0

Tomando en cuenta estos resultados se puede considerar que el Impulso es el periódico que más términos utiliza cuando describe los acontecimientos relacionados con violencia, con un total de 838 términos lo que refleja el grado de importancia que este medio impreso le otorga a este tipo de noticias. En segundo lugar, se encuentra el Informador con un total de 392 términos, y en tercer lugar, el Diario Hoy con un total de 285 términos empleados. Además, se pudo constatar que la categoría *agresión verbal* es la que posee más términos, identificándose 332 términos referidos a insultos, amenazas, frases descalificantes y discriminativas, de los cuales el 60,2% se encuentran en el Impulso, el 18% en el Informador y el 21,6% en el Diario Hoy. En fin, la prensa escrita venezolana utiliza recursos formales denominados “derrapes” para enfatizar y espectacularizar los hechos de violencia

En el análisis cualitativo con el ATLAS/ti, y siguiendo el sistema de categorías emergentes, se relacionaron algunos *codes* (categorías) con algunas de las teorías en las que se apoya esta investigación. Además, se elaboraron mapas conceptuales o *networks* con dichas relaciones. Así por ejemplo, para la categoría *factores de violencia*, referida a los términos que designan las causas que producen la violencia, y la cual se relaciona con la hipótesis del estudio sustentada en la teoría psicosocial de John Kerr (1994), apoyada a su vez, en la teoría de Inversión de Apter (1982,1989), se registraron 111 términos como se muestra en la figura 1.

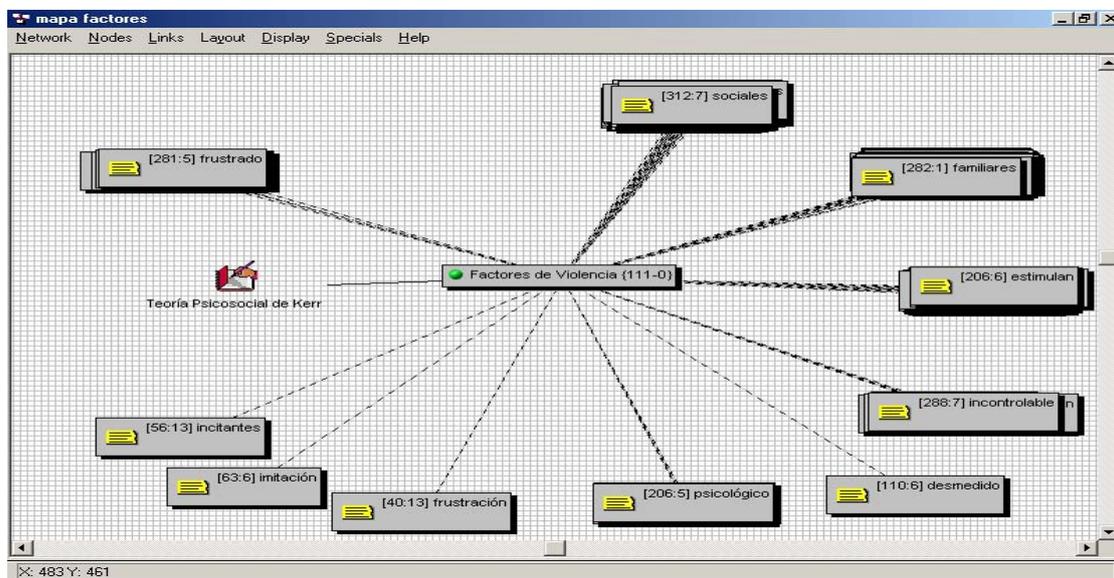


Figura 1. Mapa conceptual de la categoría factores de violencia

Como se puede apreciar los nexos más fuertes se dirigen hacia las palabras: *sociales* con 28 líneas, seguida de *familiares* con 22 nexos y *estimulan* con 21 líneas. Los términos asociados a esta categoría dan una visión importante sobre cómo los periódicos analizados describen las causas que producen la violencia, otorgando un peso relevante al factor social y reforzando la teoría psicosocial de Kerr, la cual determina que la violencia se produce por estimulación.

## Conclusión

Mediante la aplicación de las dos herramientas observacionales para el análisis de contenido, se constató que la prensa escrita venezolana trata las noticias relacionadas con violencia usando recursos formales que enfatizan dicha violencia y que se refleja en: *la frecuencia con que emplean los términos o unidades lexicales que destacan los distintos tipos de agresión, la incorporación de palabras y adjetivos con cierta connotación* (violencia política, xenofobia), *la sobrerrepresentación en la narración y descripción de los eventos violentos, el tratamiento “sensacionalista” que le dan a ciertos*

*eventos, como por ejemplo, en el asesinato de indigentes, el uso de imágenes de impacto agresivo, la exaltación del discurso agresivo, la repetición de los eventos de violencia.*

### Referencias

- Anguera, M. (Ed) (2001). *Metodología Observacional en Investigación Psicológica*. Barcelona: P.P.U.
- Ayala, R. (2001). *Una Experiencia de Análisis del Contenido sobre el mensaje publicitario de medios impresos destinados a diferente población*. Málaga. Tesis Doctoral en curso.
- Bandura, A. y Ribes, E. (1980). *Modificación de conducta. Análisis de la agresión y la delincuencia*. México: Trillas.
- Dijk, T. (1990). *La Noticia como discurso: Comprensión, estructura y producción de la información*. Barcelona: Paidós.
- Edmund, G. y Kendrick, D. (1980). *The measurement of human aggressiveness*. Chichester: Ellis Horwood. New Jersey: Prentice Hall.
- Echeburúa, E. (1998). *Personalidades Violentas*. Madrid: Pirámide. (p.p. 33-56)
- Rojas, M. (1998). *Las Semillas de la Violencia*. España: Espasa.
- Sánchez, Z. (2000). *Violencia en la Familia y el Asesoramiento Transpersonal*. UPEL. Tesis de Grado. Barquisimeto.
- Sanmartín, J. (2002). *La mente de los violentos*. Barcelona: Ariel.
- UNESCO (1990). *Las noticias extranjeras en los medios de comunicación: La información internacional en 29 países*. Francia: UNESCO.

# REVISIÓN Y COMPARACIÓN DE LA APLICACIÓN DE LA METODOLOGÍA HÍBRIDA EN TRES REVISTAS DE REFERENCIA EN CIENCIAS DEL COMPORTAMIENTO

José Francisco Molina<sup>1</sup> y Olatz López Fernández<sup>2</sup>

<sup>1</sup>Universidad de Alicante

<sup>2</sup>Universidad de Barcelona

La metodología híbrida (*mixed methods research*) consiste en el uso combinado de métodos cuantitativos y cualitativos en una misma investigación. En la actualidad está adquiriendo creciente importancia en diversas áreas de estudio, considerándose la tercera aproximación metodológica. El propósito de este trabajo es realizar una revisión y comparación de la aplicación de esta metodología en tres revistas de comportamiento: *Psicothema*, *Addictive Behaviors* y *Journal of Organizational Behavior*. Con este trabajo se perseguirán dos objetivos. En primer lugar, se revisarán todos los artículos publicados en estas tres revistas en los últimos seis años (2003-2008). El año 2003 es un año de referencia en el desarrollo de esta tercera aproximación metodológica, pues se publica el *Handbook of mixed methods in social and behavioral research* de Tashakkori y Teddlie, que ha proporcionado visibilidad y credibilidad a este enfoque. Con esta revisión, se clasificarán los trabajos publicados en dos bloques (teóricos y empíricos) y estos últimos en tres tipos (cuantitativos, cualitativos e híbridos). En segundo lugar, los trabajos híbridos serán analizados para determinar sus principales características vinculadas al propósito perseguido y al diseño híbrido utilizado, realizando comparaciones entre las revistas analizadas. Además, algunas sugerencias sobre la utilización adecuada de la metodología híbrida serán señaladas.

La integración de métodos de investigación cuantitativos y cualitativos en un mismo estudio, es decir, la utilización de métodos híbridos, está adquiriendo una creciente importancia (Creswell, 2003; Tashakkori y Teddlie, 2003). La utilización de métodos híbridos puede jugar un papel importante en la investigación, ya que los resultados obtenidos desde diferentes métodos pueden enriquecer y mejorar nuestra comprensión de las cuestiones estudiadas y generar nuevas ideas con relación a las mismas.

Creswell y Plano Clark (2007) indican que la aplicación y uso de la metodología híbrida ha sido revisada en un número pequeño de disciplinas, por lo que un tema de interés es cómo está siendo aplicada en otras áreas. El propósito de este trabajo es analizar la utilización de estos métodos híbridos en las ciencias del comportamiento, examinando los diseños utilizados y los propósitos perseguidos por los trabajos que han aplicado esta aproximación metodológica en tres revistas de referencia en diversos ámbitos de las ciencias del comportamiento: *Psicothema*, *Addictive Behaviors* y *Journal of Organizational Behavior*.

## *Métodos de investigación híbridos*

En diversos campos de investigación social está aumentando el cuerpo de literatura relativo a la metodología híbrida, tales como la evaluación, educación, sociología, ciencias de la salud o psicología. (Tashakkori y Teddlie, 2003). De hecho, sobre esta aproximación metodológica se están publicando libros, capítulos de libros, números especiales de revistas e incluso revistas específicas dedicadas a esta aproximación, por ejemplo *Journal of Mixed Methods Research* o *International Journal of Multiple Research Approaches*. Los trabajos centrados en esta metodología incluyen palabras clave, tipologías y propósitos de los diseños híbridos, cuyo conocimiento puede facilitar el análisis de estos diseños.

Existen dos factores principales que ayudan a determinar varios tipos de diseños híbridos (Creswell, 2003; Morgan, 1998; Morse, 1991):

- **Prioridad/importancia.** En un estudio híbrido, el investigador puede dar la misma prioridad o importancia a las partes cuantitativa y cualitativa, o bien otorgar mayor prioridad o relevancia a una de ellas.
- **Implantación de la recogida de datos.** Esta implantación se refiere a la secuencia con que el investigador recoge los datos cuantitativos y cualitativos. Las dos opciones consisten en recoger la información al mismo tiempo (diseño simultáneo, concurrente o paralelo) u obtener los datos en diferentes fases (diseño secuencial o en dos fases).

Para la representación de estos diseños se puede utilizar la notación propuesta por Morse (1991). En su sistema, se utiliza la abreviatura “quan” para representar la parte cuantitativa, y “qual” para la cualitativa. Además, cuando hay un método dominante o más importante, éste se representa con letras mayúsculas (QUAN, QUAL) mientras que el método menos importante aparece con letras minúsculas (quan, qual). Por otra parte, el símbolo “+” es utilizado para indicar un diseño simultáneo, mientras que la flecha “→” se refiere a un diseño secuencial. Así, podemos establecer cuatro bloques que dan lugar a nueve diseños híbridos (Johnson y Onwuegbuzie, 2004) (figura 1).

<b>DISEÑOS:</b>		<b>IMPLANTACIÓN</b>	
		<b>Simultánea</b>	<b>Secuencial</b>
<b>PRIORIDAD</b>	<b>Igual</b>	QUAL+QUAN	QUAL→QUAN QUAN→QUAL
	<b>Diferente</b>	QUAL+quan QUAN+qual	qual→QUAN QUAL→quan quan→QUAL QUAN→qual

Figura 1. Tipos de diseños híbridos

Con relación a los propósitos de llevar a cabo diseños híbridos, podemos indicar diversas razones señaladas por varios autores (Creswell, 2003; Greene, Caracelli y Graham, 1989; Morgan, 1998). Dos de las principales razones que se suelen señalar son las de triangulación y complementariedad. Con la triangulación (Jick, 1979) se persigue fundamentalmente una convergencia de los resultados a partir de ambas aproximaciones cuantitativa y cualitativa, para conseguir que esos resultados sean más fiables. Por tanto, lo que se busca es una corroboración o correspondencia de resultados desde diferentes métodos. Por su parte, con la complementariedad se busca que una de las aproximaciones (la cuantitativa o la cualitativa) complemente a la otra. En este caso, el objetivo principal que se persigue es el de clarificar, mejorar, ampliar o ilustrar a través de uno de los métodos los resultados obtenidos en el otro método. Otro propósito que se puede perseguir y alcanzar con los métodos híbridos es el de desarrollo, en los que uno de los métodos (normalmente el menos importante) ayuda en algún aspecto a mejorar la ejecución posterior del otro método (normalmente el método principal o dominante). Además, un propósito adicional que también puede permitir alcanzar los estudios que utilizan una aproximación híbrida es el de expansión, tratando de buscar un análisis y comprensión de diferentes facetas de un fenómeno, obteniendo así una comprensión más rica y profunda de ese fenómeno.

## Método

Con el objetivo de identificar y analizar los diseños híbridos utilizados y los propósitos perseguidos en el campo de las ciencias del comportamiento, hemos llevado a cabo una revisión de los artículos publicados en tres revistas de referencia en diversas temáticas de este campo, en concreto *Psicothema*, *Addictive Behaviors* y *Journal of Organizational Behavior*.

Para identificar los estudios híbridos publicados en estas revistas, todos los artículos publicados desde 2003 a 2008 fueron revisados. El año 2003 es un año de referencia en el desarrollo de esta aproximación metodológica, pues se publica el *Handbook of Mixed Methods in Social and Behavioral Research* de Tashakkori y Teddlie, que ha proporcionado visibilidad y credibilidad a este enfoque. Esta estrategia de búsqueda nos permitió, además de identificar los estudios híbridos, clasificar todos los trabajos en dos grupos, no empíricos y empíricos, y desdoblarse este grupo de artículos empíricos en tres tipos: cuantitativos, cualitativos e híbridos. Una vez realizada esta clasificación, se realizó un análisis de contenidos de los artículos híbridos identificados, determinando para cada uno de ellos el tipo de diseño utilizado en función de las características de prioridad e implantación previamente señaladas así como el propósito híbrido principal del trabajo.

Antes de pasar al siguiente apartado, hemos de indicar que determinar si un estudio es híbrido suele ser complicado, ya que los autores de estos trabajos no suelen indicar explícitamente ni en el título ni en el resumen que ese trabajo utiliza una metodología híbrida. Normalmente los trabajos híbridos fueron identificados a través de la información proporcionada en el apartado de diseño de la investigación, donde se hace referencia a las técnicas de recolección y de análisis de los datos.

## Resultados

La tabla 1 muestra el tipo de artículos identificados en cada revista. En esta tabla se observa, en primer lugar, la predominancia de artículos empíricos (89,1%) frente a los teóricos (10,9%). La revista *Addictive Behaviors* es la que más trabajos empíricos publica, en parte porque es claramente la política de la editorial, seguida de *Psicothema* y *Journal of Organizational Behavior*. En segundo lugar, la mayoría de los artículos son de carácter cuantitativo (85,7%), frente a una minoría de artículos cualitativos e híbridos (1,7% en cada caso). En resumen, esta comparación nos indica que en estos últimos años en las principales revistas con factor de impacto en ciencias del comportamiento predomina la realización y publicación de artículos empíricos, cuya metodología más habitual es de carácter cuantitativo.

Tabla 1. Comparación de los artículos publicados en las revistas analizadas (2003-2008)

Revistas	Número total de artículos	Número de artículos no empíricos	Artículos empíricos			
			Número total de artículos empíricos	Número de artículos cuantitativos	Número de artículos cualitativos	Número de artículos híbridos
Psicothema	708	81 (11.4)	627 (88.6)	608 (85.9)	10 (1.4)	9 (1.3)
AB	932	45 (4.8)	887 (95.2)	876 (94)	5 (0.5)	6 (0.7)
JOB	318	87 (27.4)	231 (72.6)	197 (62)	17 (5.3)	17 (5.3)
TOTAL	1958	213 (10.9)	1745 (89.1)	1681 (85.7)	32 (1.7)	32 (1.7)

Los valores numéricos de las casillas corresponden a frecuencias absolutas y los valores entre paréntesis son el porcentaje resultante en relación al total en cada caso.

Con relación a los estudios híbridos identificados, en la tabla 2 se sitientizan tanto las características de los diseños híbridos como los propósitos metodológicos de este tipo de artículos que integran métodos cuantitativos y cualitativos.

Tabla 2. Características de los artículos híbridos analizados (2003-2008)

	Revistas		
	Psicothema (9)	AB (6)	JOB (17)
<b>Diseños</b>			
Igual prioridad	6(66.7)	5(83.3)	7(41.2)
Diferente prioridad	3(33.3)	1(17.6)	10(58.8)
Simultáneo	2(22.2)	4(66.7)	4(23.5)
Secuencial	7(77.8)	2(33.3)	13(76.5)
<b>Propósitos</b>			
Triangulación	0(0)	3(50)	3(17.6)
Complementariedad	0(0)	3(50)	5(29.4)
Desarrollo	7(77.8)	0(0)	8(47.1)
Expansión	2(22.2)	0(0)	1(5.9)

Los valores numéricos de las casillas corresponden a frecuencias absolutas y los valores entre paréntesis son el porcentaje resultante en relación al total en cada caso.

En cuanto a los diseños híbridos, el único punto de coincidencia entre las tres revistas era la ausencia de diseños de prioridad o importancia diferente e implantación simultánea. En *Psicothema* se encontraron 9 diseños híbridos, en los que destacaban la igualdad de importancia entre los métodos cuantitativos y cualitativos además de una implantación secuencial. Los propósitos eran principalmente el desarrollo (que hace referencia a el uso de resultados de un método como ayuda para poder generar y aplicar el otro método posteriormente) y la expansión (que consiste en extender la amplitud de la pregunta de investigación mediante el uso de diferentes métodos para sus diferentes componentes). En *Addictive Behaviors* se detectaron 6 diseños híbridos, entre los que predominaban aquellos de igual importancia e implementación simultánea, siendo los propósitos metodológicos tanto la triangulación (que busca la corroboración y convergencia de los dos métodos) como la complementariedad (que trata de elaborar, ilustrar y clarificar los resultados de un método con los hallazgos del otro método). En *Journal of Organizational Behavior* se detectaron 17 diseños híbridos, de los cuales los de diferente importancia eran algo más frecuentes que los de igual prioridad. Además, destacaban los de implantación secuencial. Además, en esta última revista han aparecido un mayor número de propósitos al integrar los métodos cuantitativos y cualitativos.

### Conclusión

La realización de trabajos híbridos en el campo de las ciencias del comportamiento puede aprender determinados aspectos de los trabajos ya realizados en este y otros campos. En este sentido, Creswell, Plano Clark, Gutmann y Hanson (2003) señalan que un reto importante para los trabajos híbridos es la clarificación explícita de varios aspectos relevantes. En primer lugar, los investigadores deberían identificar claramente cuáles son las razones o propósitos principales de utilizar en su trabajo un diseño híbrido utilizando datos cuantitativos y cualitativos. Además, también debería clarificarse los factores que hemos analizado hasta ahora para determinar los tipos de diseños. Así, con relación a la prioridad, los investigadores deberían indicar claramente las decisiones tomadas relativas a la importancia o atención prestada a la parte cuantitativa y cualitativa (igual o distinta importancia), lo cual podría reflejarse en la longitud y profundidad de los comentarios y discusiones realizados para cada uno de las aproximaciones. Por otra parte, con relación a la implantación de la recogida de datos, debería determinarse claramente si el diseño es secuencial o simultáneo. Por ejemplo, si el diseño es secuencial, las dos fases de recogida y análisis de datos podrían aparecer en el documento escrito de forma separada, llevando a cabo la integración de información en las secciones de discusión y/o conclusiones. Dada la complejidad de estos aspectos, el investigador puede utilizar figuras o modelos

visuales para presentar su trabajo.

Además, los investigadores deben tener en cuenta que para un determinado estudio en principio puede establecer un diseño concreto, pero que nuevos componentes o ideas pueden surgir conforme el trabajo se va realizando, lo que puede conllevar una modificación de ese diseño previo. De esta forma, el investigador debe ser creativo y no debe limitarse a los diseños preexistentes, sino que debe incluso crear aquellos diseños que permitan contestar adecuadamente sus cuestiones de investigación. En esta línea, normalmente se señala que los diseños híbridos secuenciales tienen dos partes o fases, pero estos diseños pueden ser más complejos implicando tres o más fases (Johnson y Onwuegbuzie, 2004; Teddlie y Tashakkori, 2006).

Pensamos que como futuras líneas de investigación puede ser interesante ampliar el período analizado y realizar un análisis de otras revistas para tener una visión más amplia de la aplicación de la metodología híbrida. En nuestra opinión, este trabajo puede ayudar a difundir la utilización de la metodología híbrida en las ciencias del comportamiento. Los investigadores deben conocer la aceptación de esta metodología en sus respectivas áreas y la utilización de la misma por otros colegas. En este trabajo, se han identificado diversos trabajos híbridos que podrían servir de guía para futuras investigaciones.

### Referencias

- Creswell, J. (2003). *Research design. Qualitative, quantitative and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. y Plano-Clark, V. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Creswell, J., Plano Clark, V., Gutmann, M. y Hanson, W. (2003). Advanced mixed methods research designs. En A. Tashakkori y C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 209-240). Thousand Oaks, CA: Sage.
- Greene, J., Caracelli, V. y Graham, W. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11, 255-274.
- Jick, T. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24, 602-611.
- Johnson, B. y Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Morgan, D. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative Health Research*, 8, 362-376.
- Morse, J. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40, 120-123.
- Tashakkori, A. y Teddlie, C. (Eds.) (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage.
- Teddlie, C. y Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, 13, 12-28.

# FUENTES PSICOSOCIALES DE ESTRÉS EN MÉDICOS RESIDENTES

Micaela Ortiz, Blanca Valenzuela y Manuela Guillén  
Universidad de Sonora, México

El objetivo de este estudio fue identificar las posibles fuentes de estrés, los estilos de afrontamiento, y sus consecuencias en la salud en el ejercicio de las funciones profesionales de los Médicos Residentes del Hospital General del Estado en Hermosillo, Sonora. Para analizar lo propuesto se utilizó la metodología cualitativa, y como estrategia principal de recopilación de información, la entrevista semi-estructurada. Los resultados indican que las fuentes de estrés están en función de la jornada de trabajo, el tiempo designado para la alimentación, la cadena de mando, la división del trabajo, las demandas de la tarea, la orientación hacia la formación, las relaciones interpersonales, el paciente grave, el trabajo como clausura de la vida personal, el espacio geográfico, insuficiencia de materiales y equipo de trabajo, insuficiencia de personal de salud y la beca. Los resultados expresan que los efectos en la salud son de índole psicológicos y biológicos, prevaleciendo el agotamiento físico y emocional. Los estilos de afrontamiento están dirigidos hacia la solución de problemas, más que dirigidos a la emoción. Por lo tanto el personal médico identifica una gran variedad de factores psicosociales referidos a la organización del trabajo y a las características propias de la profesión.

El estrés es uno de los fenómenos que cada vez está teniendo más importancia en el ámbito laboral y que puede crear problemas de salud y de bienestar importantes en las personas. Se ha considerado que los profesionales de la salud toman un significado importante ya que son considerados una de las poblaciones más vulnerable al desarrollo del estrés (Arenas, 2004).

En función de lo antes citado el propósito de esta investigación fue identificar las posibles fuentes de estrés en el ejercicio de las funciones profesionales de los Médicos Residentes del Hospital General del Estado de Sonora, en Hermosillo.

La experiencia relacionada con el estrés no es nueva, ha sido considerado durante siglos, pero sólo recientemente se ha conceptualizado de forma sistemática y ha sido objeto de investigación, no ha dejado de suscitar interés entre los investigadores desde los primeros estudios realizados tras la segunda guerra mundial, representando un avance en la investigación sobre el tema dada su significación en el rendimiento en el combate (Kahn y col., 1991).

El estrés, es uno de los aspectos que más investigación ha suscitado en los últimos tiempos, probablemente como consecuencia del creciente número de personas afectadas por este complejo fenómeno, se ha convertido en uno de los conceptos más referidos, tanto por los profesionales de las diversas ciencias de la salud como en el lenguaje coloquial, para hacer referencia a una amplia gama de experiencias; por ejemplo, ansiedad, tensión, cansancio y otras sensaciones. Ello se refleja en el gran número de definiciones y la variedad de modelos teóricos, aunque no se ha conseguido llegar a un acuerdo sobre una conceptualización aceptada por toda la comunidad científica.

Uno de los principales exponentes en el tema es Hans Selye (1974) sus estudios tienen origen en la década de 1930, observó que todos los enfermos a quienes examinaba, sin importar que enfermedad padecían, presentaban síntomas comunes y generales, además de los síntomas propios de cada enfermedad: cansancio, desgano, pérdida de apetito, entre otros.

Selye llegó a concluir que diversas enfermedades (cardíacas, hipertensión arterial, trastornos mentales y emocionales) eran resultado de los cambios fisiológicos producidos por un continuo estrés y que podían estar constitucionalmente o genéticamente predeterminados. Al seguir con sus investigaciones, menciona que también las demandas sociales y las amenazas del entorno que requieren ajuste por parte del individuo provocan estrés (Pereyra, 2004).

De acuerdo a estos antecedentes, la génesis del estrés radica en el campo de las ciencias biomédicas, el cual ha sido de importancia vital para tener una visión más clara del papel que juega en el surgimiento de diferentes trastornos. Aunque en la actualidad ha quedado definido que se trata de un fenómeno que puede estar determinado por múltiples factores. La dimensión psicológica y la dimensión social toman un papel relevante, lo que ha generado información muy heterogénea. Resultado de ello es que el estrés se ha definido como respuesta fisiológica o psicológica que manifiesta un individuo ante un estresor ambiental o como estímulo, ya que actúa sobre el individuo y que da lugar a una respuesta de tensión (Ivancevich y col., 1989).

Lazarus (1990) propone que no se debe parcializar la complejidad del estrés, que deben incorporarse otras variables en las cuales se pudieran analizar las diferencias individuales, el estrés es producto de la interacción entre el organismo y su medio, considerándose los factores psicológicos que están mediando la interacción.

El estudio de las profesiones asistenciales, y de modo muy particular las profesiones de la salud, han revelado que las altas demandas del trabajo generan no sólo un cuadro general de tensión emocional que tiende a generar malestar subjetivo y a expresarse en síntomas somáticos; sino un síndrome que transcurre y se instala por etapas, cuyos efectos parciales constituyen sus componentes: el síndrome del *burnout*, que desde una perspectiva psicosocial el síndrome de quemarse por el trabajo se conceptualiza como una respuesta al estrés laboral crónico que se caracteriza porque el individuo desarrolla una idea de fracaso profesional, la vivencia de encontrarse emocionalmente agotado, y actitudes negativas hacia las personas con las que trabaja.

En un estudio realizado sobre el Síndrome de *burnout* en médicos familiares del Instituto Mexicano del Seguro Social en Guadalajara, Aranda y Col., (2004), reportaron que la prevalencia encontrada sobre estrés fue de 42.3%. El 64.8% (81 sujetos) se manifestó enfermo. Las enfermedades más frecuentes fueron músculo-esqueléticas (20%), respiratorias (19.2%), gastrointestinales (18.3%) y psicológicas (15.2%).

La OIT (2001), menciona que algunos investigadores se han interesado sobre el contenido y la duración del trabajo en la vida privada de los médicos jóvenes. Valko y col., (1975), hallaron que el 30% de los médicos residentes sufrían una crisis depresiva de una duración media de cinco meses durante su primer año de residencia. Cuatro de los 53 médicos residentes estudiados habían mostrado inclinaciones suicidas y tres de ellos habían concretado planes para privarse de la vida.

Landau y Col., (1986) hallaron que 40% de residentes reconocían tener graves problemas conyugales, de los que 72% de ellos culpaba a su trabajo.

Es importante enfatizar la enorme trascendencia histórica que asume el estrés en las diferentes sociedades pasadas y contemporáneas, puesto que representa un claro indicador directo del grado de salud propio de cada organización social, de su estructura, relaciones, medios y modos de servicio.

## Método

### *Participantes*

Las personas que formaron parte del estudio han sido médicos residentes del sexo masculino y femenino que estén realizando su especialidad en el Hospital General en Hermosillo, Sonora. Las especialidades consideradas para nuestro estudio son: Cirugía general, Anestesiología, Medicina Interna y Urgencias Médicas, ello, en virtud de que son consideradas como las especialidades que experimentan mayores eventos estresantes por la misma dinámica del servicio (Arenas, 2006).

La población quedó conformada por 21 becarios.

Para el caso de Cirugía General fueron 7 médicos entrevistados, 2 mujeres y 5 hombres.

Para el servicio de Anestesiología se establecieron los mismos criterios, quedando un total de 6 residentes, 4 mujeres y 2 hombres.

En el servicio de Medicina Interna tuvimos un total de 5 residentes, 1 mujer y 4 hombres.

Para el caso de Urgencias Médicas se seleccionó a un total de 3 médicos, 1 mujer y 2 hombres.

Sus edades comprendían entre 25 a 33 años de edad.

### *Instrumento*

La información ha sido recopilada por medio de la entrevista semi-estructurada, siguiendo los lineamientos de una guía de temas, una vez realizada la revisión de la literatura y esbozando desde la experiencia personal, se prepararon las preguntas con antelación a las entrevistas y se realizó un estudio piloto a una pequeña muestra de médicos residentes. Dicha guía se modificó y enriqueció con posterioridad a este proceso de aplicación previo. Además se utilizó la técnica de entrevista conversacional y la observación participante, tomándolas como apoyo para el análisis de los datos.

### *Procedimiento*

Antes de iniciar las entrevistas se explicó el objetivo de estudio, siempre buscando las condiciones adecuadas y propicias para la obtención de la información por cada uno de los médicos.

Las entrevistas fueron guiadas por un conjunto de preguntas pero ni la redacción exacta, ni el orden de las preguntas estaba predeterminado (Mayan, 2001).

La información proporcionada fue grabada en audio y posteriormente transcrita haciendo un análisis preliminar de los datos orientándonos en el trabajo de campo.

El proceso de análisis consistió en codificar y categorizar patrones primarios en los datos, se identificaron palabras, frases, temas o conceptos dentro de los datos, posterior, se categorizaron los datos, se recortaron las secciones del texto y se realizó la agrupación en categorías en archivos separados.

El análisis se realizó por categoría, incluyendo todos los servicios médicos explorados en el escenario empírico. Existió una diversidad en los referentes desde los cuales los sujetos entrevistados construyeron su discurso (percepción).

## Resultados

De acuerdo a los objetivos planteados los resultados indican que los estresores están en función de la jornada de trabajo, el tiempo designado para la alimentación, la cadena de mando, la división del trabajo, las demandas de la tarea, la orientación hacia la formación, las relaciones interpersonales, el paciente grave, el trabajo como clausura de la vida personal, el espacio geográfico, insuficiencia de materiales y equipo de trabajo, insuficiencia de personal de salud y la beca. Los resultados expresan que los efectos en la salud son de índole psicológicos y biológicos, prevaleciendo el agotamiento físico y emocional. Los estilos de afrontamiento están dirigidos hacia la solución de problemas, más que dirigidos a la emoción.

Otro dato importante a resaltar en esta investigación es que ciertas prácticas han sido institucionalizadas, indicando que determinadas formas de ser o actuar son parte natural de la práctica, formando parte de la cultura médica. De ahí que sean concebidas como algo “natural” y que se asuman como pertenecientes a la naturaleza misma del “ser” y el “hacer” de tales prácticas, especialmente las referidas a las jornadas intensas de trabajo, a las jerarquías y al proceso de enseñanza aprendizaje.

Además, los residentes hacen representaciones que tienen que ver sobre la realidad que viven en la institución hospitalaria, a dichas representaciones las hemos denominado imaginario, el imaginario se torna significativo ya que tiene que ver con el “deber ser” de la dinámica laboral. Es decir, a la forma en que, a su juicio, debería funcionar la institución. Lo interesante es advertir que las referencias del imaginario guardan una estrecha relación con los datos encontrados en las fuentes de estrés. Con esto queremos decir que aquello que se refiere como fuente de estrés se visualiza, asimismo, como elemento (o condición) de cambio.

Por lo tanto el personal Médico Residente identifica una gran variedad de factores psicosociales referidos específicamente a la organización del trabajo y a las características propias de la profesión.

Lo relevante para nuestro estudio es abordar al médico durante su proceso de formación, valorando el ambiente donde se desarrolla, ya que existen características propias de esta profesión, que los hace diferentes a otros trabajadores. Por ejemplo, el médico nunca es dueño de su tiempo, y sus pacientes se apoderan de este aspecto tan importante de su vida.

## Discusión

Las experiencias de estrés en el contexto laboral son producidas por una serie de situaciones ambientales o personales que podemos caracterizar como fuentes de estrés o estresores. En el escenario explorado, estas situaciones son percibidas por los médicos residentes y entonces se inicia la experiencia de estrés. Esa experiencia desencadena toda una serie de vivencias emocionales, al tiempo que se ponen en marcha una serie de procesos para afrontar esas situaciones y/o experiencias de estrés.

Podemos observar que existe una apreciación compartida de la realidad por parte de los residentes. Es decir, existen estresores compartidos por cada uno de los residentes pertenecientes a una misma especialidad, tomando un matiz propio en cada uno de los servicios analizados. Las demandas de cada servicio toman un significado particular, por lo que puede afirmarse que cada uno de ellos

tiene una identidad propia. Además, existen estresores de carácter general que son reconocidos por los médicos residentes, independientemente del servicio al que éstos se encuentran adscritos (tales como: relaciones interpersonales establecidas, cadena de mando, entre otros).

En los profesionales sanitarios (medicina, enfermería) el estrés está compuesto por una combinación de variables físicas, psicológicas y sociales. Son profesionales en los que inciden, especialmente, estresores como la escasez de personal que supone una sobrecarga laboral (Gil-Monte, 1997).

El servicio de Cirugía General cobra relevancia en los datos, ya que se ha identificado como aquél en el que tienen lugar un mayor número de fuentes de estrés, seguido de Anestesiología, Urgencias Médicas y por último Medicina Interna.

De acuerdo a las evidencias encontradas, la institución de salud no está diseñada ni estructurada para promover e impulsar el autocuidado de los médicos residentes.

### Referencias

- Aranda, C., Pando, M., Salazar, J., Torres, L., Aldrete, M. y Pérez, B. (2004). Síndrome Burnout en Médicos Familiares del Instituto Mexicano del Seguro Social. *Salud Pública de México*. 31 (2), 25-32.
- Arenas, J.O. (2006). Estrés en médicos residentes en una unidad de atención médica de tercer nivel. *Cirujano General*. 28 (2), 103-109.
- Arenas, L., Hernández, I., Valdez, R. y Bonilla, P. (2004). Las instituciones de salud y el autocuidado de los médicos. *Salud Pública de México*. 46 (4), 326-332.
- Gil-Monte, P. y Peiró, J. M. (1997). *Desgaste psíquico en el trabajo: El Síndrome de Quemarse*. Síntesis Psicología, España.
- Guillen, M. (2005). *Reconfiguración productiva, crecimiento urbano e identidades: representaciones sociales de los pobres en la periferia urbana de una ciudad del norte de México*. Tesis de doctorado. Universidad Nacional de Educación a Distancia, Facultad de Ciencias Políticas y Sociología.
- Ivancevich, J. y Matteson, M. (1989). *Estrés y trabajo. Una perspectiva gerencial*. Trillas. México.
- Kahn, R. L. y Byosiére, Ph. (1991). Stress in Organizations. En M. D. Dunnette y L. M. Hough (eds.) *Handbook of Industrial and Organizational Psychology*, 2ª ed. (vol. 1, pp. 571-650). Palo Alto, Ca: Consulting Psychologist Press.
- Laudau, G., Hall, S., Wartman, S. y Macko, M. (1986). Stress in social and family relationships during medical residency. *Journal of Medical Education*, 61, 654-660.
- Lazarus, R. (1990). Theory-based stress measurement. *Psychological inquiry*. 1(1), 3-13.
- Mayan, M. (2001). *Una introducción a los métodos cualitativos*. Edmonton: IIQM.

- Organización Internacional del Trabajo (2001). *Enciclopedia de la salud y seguridad en el trabajo, Centros y servicios de asistencia sanitaria*. Revisado el 25 de noviembre del 2005 [www.ilo.org/public/spanish/protection/index.htm](http://www.ilo.org/public/spanish/protection/index.htm).
- Pereira, M. (2004). Estrés y salud. En Oblitas, L., *Psicología de la salud y calidad de vida en España*: Thomson. P.p 223-224.
- Selye, H. (1974). *Stress without distress*. Lippincott, Filadelfia.
- Valles, M. (2000). *Técnicas cualitativas de investigación social: una reflexión metodológica y práctica profesional*. Síntesis. Madrid.
- Valko, R., Clayton, P. (1975). Depression in the internships. *Dis Nerv Syst* 36, 26-29.

# LA CALIDAD DEL REGISTRO EN SESIONES DE LARGA DURACIÓN: APLICACIÓN DE LA TEORÍA DE LA GENERALIZABILIDAD

Elena Rodríguez-Naveiras  
Universidad de La Laguna

El control de la calidad de la información del registro de una observación en ambiente natural es uno de los aspectos básicos en la investigación. La Teoría de la Generalizabilidad aporta un procedimiento de análisis que permite aislar las diversas fuentes de error de la medida, a la vez que ayuda a determinar en qué medida se pueden modificar los diversos factores, analizando el efecto en el coeficiente de generalizabilidad. En el trabajo que aquí se expone se presenta el estudio de la fiabilidad entre observadores, en un estudio dirigido a determinar las funciones docentes que realizan monitores de un programa extraescolar socioafectivo dirigido a niños y niñas con altas capacidades intelectuales, tomando como facetas los observadores (o), agentes (a) utilizando un diseño multifaceta cruzado ( $A \times O$ ).

En la intervención psicoeducativa, la evaluación de programas ocupa un lugar importante, debido a la necesidad de valorar la eficacia sistematizada de la acción. La evaluación de programas debe comenzar con una explicitación o decisión sobre las funciones que cumplirá el programa, ya que la necesidad de evaluarlo parte de la convicción de que facilita o promueve mejoras en muy diversos sentidos. A pesar de la diversidad de planteamientos y posiciones, existen elementos relevantes a considerar en cualquier modalidad de evaluación de programas: el carácter dinámico de la evaluación, la aplicación de procedimientos científicos, la recogida de información válida y fiable, la toma de decisiones, el ir referido a un programa, el tener una función ideológica y la valoración de la acción (Colás y Rebollo, 1993). La evaluación de programas sigue las reglas del método científico y, principalmente, existen dos referentes en evaluación de programas desde el punto de vista metodológico: a quiénes va dirigido el programa y el tipo de información que se obtiene, bien de carácter cambiante a lo largo del proceso de implementación del programa, o bien de carácter sincrónico o diacrónico del proceso de evaluación, dependiendo de si nos situamos en la evaluación formativa o sumativa (Veney y Kluzny, 1984). De esta forma, los resultados obtenidos a través de la evaluación pueden ser útiles, tanto desde una perspectiva externa como interna.

En el ámbito de las altas capacidades intelectuales se desarrollan programas de intervención tanto intra como extraescolares, centrados en dar una formación acorde a las necesidades de estos niños y niñas, desarrollar los talentos que poseen o revalorizar las actividades académicas. El Programa Integral para Altas Capacidades (PIPAC) es extraescolar y se crea en el año 2004 con el objetivo de apoyar a niños y niñas con altas capacidades, contribuyendo a su desarrollo completo, basándose en modelos sociales (Monks, 1994; Ziegle y Heller, 2000). Tiene un enfoque de intervención psicoeducativa, ya que se sitúa en una perspectiva más preventiva que centrada en los problemas, desarrollando las capacidades de los usuarios, sus contenidos son aspectos de corte socioafectivos y se realiza con una programación previa, evaluando el programa al inicio, en su desarrollo y al final.

Dicha evaluación es interna y anual, siendo tanto formativa, donde se analiza la implementación del programa, la valoración del monitor y el proceso seguido (Borges, Hernández-Jorge, Rodríguez-Naveiras, 2005; Borges, Rodríguez-Naveiras, Hernández-Jorge, Fernández, 2007), como sumativa, determinándose su eficacia y eficiencia (Rodríguez-Naveiras, Borges, Hernández-Jorge, 2007).

Una evaluación formativa más completa supone analizar los patrones comportamentales que se producen en las sesiones, siendo para ello la metodología de elección la observacional, que permite acceder no sólo a la descripción de conductas, sino también al tipo de interrelaciones conductuales en el flujo de comportamientos en una situación natural. Por otra parte, puesto que toda evaluación es, por concepto, un proceso de toma de decisiones, que implica, necesariamente, una recogida de datos, en este caso para analizar los procesos de enseñanza-aprendizaje, la observación se pone al servicio de la evaluación.

El proceso de enseñanza-aprendizaje descansa en la evaluación que el docente desarrolla dentro del aula en cualquier proceso educativo, del cual la intervención psicoeducativa no es ajena. Hay diversos modelos que explican estas formas de acción de los docentes (Brophy y Good, 1986; Hernández, 1991; Hernández y García, 1991; Cooper, 1993; Justicia, 1996 y Alonso, 1997), en el presente trabajo se asume el de (Hernández, 1991), que señala ocho funciones que quedan expresadas en la tabla número 1.

*Tabla 1. Funciones Docentes*

Funciones Docentes	
Organización	Planificación previa de la enseñanza y la organización sobre el contexto. Recoge dos aspectos: organización interna y externa
Comunicabilidad docente	Capacidad para comunicar los contenidos de forma comprensible
Motivadora	Capacidad del docente para estimular al alumnado al aprendizaje
Control comportamental	Se relaciona con la gestión, el orden y la disciplina
Orientación y asesoría	Ayudas que otorga el docente al aprendizaje del alumnado
Interacción y cooperación	Corrige errores y ampliar información
Personalización	Adaptación de la enseñanza a diferentes cualidades del alumnado
Evaluación	Evaluación de los resultados del programa de formación

Con este objetivo de evaluación de proceso, nuestro equipo de trabajo ha elaborado un instrumento basado en las funciones docentes, apoyándose en la estructura teórica del modelo de funciones docentes previamente descrito. La metodología observacional, que fija su atención en la conducta espontánea del individuo, exigiendo la máxima expresión de naturalidad en el contexto donde se desarrolla la investigación, supone también, en muchas ocasiones, una serie de limitaciones que afectan a la calidad de los registros en la observación. Entre los sesgos más importantes se pueden señalar, al menos, tres fuentes de error: debidos a la categorización y codificación de conductas (adecuación de la definición de cada código, complejidad de la codificación, frecuencia de aparición de la conducta a observar), debidas al observador (influencia de personas ajenas, fatiga del observador, conocimiento de lo observado) o bien problemas debidos a los medios técnicos de registro (aspectos referidos a la grabación, utilización de software de codificación).

Estos sesgos exigen, en primer lugar, utilizar instrumentos de observación debidamente depurados, para eliminar lo relativo a defectos de categorización; en segundo lugar, prestar una

adecuada atención tanto al sistema de registro como de codificación y, obligatoriamente, el cálculo de la fiabilidad de los observadores.

El estudio de la fiabilidad ha estado ligado al estudio de las diferencias individuales medidas a través de pruebas psicométricas. Las observaciones de los individuos en situaciones naturales han ido sustituyendo estas pruebas, pasando a registrar el comportamiento de los individuos a través de observadores (Blanco-Villaseñor, 1989).

Uno de los procedimientos más ricos en el estudio de la calidad del dato es la Teoría de la Generalizabilidad (Cronbach, Gleser, Nanda y Rajaratnam, 1972), que asume que hay múltiples fuentes de variación integradas en una estructura global, permitiendo aplicaciones particulares de la teoría estadística del muestreo. La TG ha sido concebida para unificar la calidad del dato en fiabilidad, validez y precisión. Una medición observacional puede estar influenciada por diversos aspectos diferentes de la situación de observación (diferentes observadores, diferentes ocasiones, diferentes formas de registro, diferentes instrumentos de registro), incluyendo las diferencias individuales entre sujetos. Se parte de un modelo de “*simetría*”, donde cada faceta de un diseño es seleccionada como objeto de estudio, pudiendo considerarse en cada análisis de generalizabilidad como un instrumento de medida o una condición de evaluación en el estudio de las restantes facetas (Blanco-Villaseñor y Anguera, 2000).

En cualquier situación de medida, las múltiples fuentes de variación conllevan estrategias para reducir su influjo sobre la medida, siendo por tanto, imprescindible identificar y medir los componentes de varianza que aportan error a la estimación. Las fases de análisis, según la TG, son cuatro (Blanco-Villaseñor, 1993):

- Plan de Observación: Identificación, organización de los datos, elección de variables y facetas.
- Plan de Estimación: Elección del modelo de estimación.
- Plan de Medida: Precisión de la intención de medida y que faceta o facetas constituyen el objeto de estudio.
- Plan de Optimización: Identificación de la mejor adecuación posible en los procedimientos de medida.

El objetivo de esta investigación es el estudio de la fiabilidad entre observadores, como parte inicial de control del dato en la evaluación de proceso.

## Método

### *Participantes*

En este estudio se han registrado los comportamientos de dos monitoras y los niños y niñas integrantes del Programa Integral para Altas Capacidades, cuyas características se muestran en la tabla número 2.

*Tabla 2. Características monitoras y participantes*

Agentes	Nivel de estudios	Edad	Sesión	Actividades
Monitora principal	Licenciada Psicología	25	2	2 y 3 por sesión
Monitora básica	Estudiante Psicología	19		
Participantes	Primaria-Secundaria	9-12		

Los comportamientos de las monitoras y niños y niñas del programa han sido registrados por cuatro observadores, pertenecientes al Grupo de Trabajo e Investigación en Superdotación cuyas características se muestran en la tabla número 3.

Tabla 3. Características de los observadores

Obs	Sexo	Edad	Nivel estudios	Parejas observadores	Sesiones observadas
1	M	21	Estudiante Psicología	Pareja 1	1 sesión con 2 actividades
2	V	19	Estudiante Psicología		
3	M	29	Licenciada Psicología	Pareja 2	1 sesión con 3 actividades
4	M	21	Estudiante Psicología		

### Procedimiento

A partir del encuadre teórico basado en las funciones docentes, se ha elaborado un instrumento de observación, El Protocolo de Observación de Funciones Docentes (PROFUNDO) que está organizado de forma jerárquica, formado en un plano superior por seis macrocategorías, que se despliegan en criterios, esto es, especificaciones concretas de estas funciones y que, finalmente, se operativizan en códigos directamente observables. Además de las seis macrocategorías, para completar el flujo continuo de los comportamientos de las monitoras, se hace necesario incluir una categoría instrumental que recoge comportamientos inobservables y otro tipo de comportamientos (véase tabla número 4).

Tabla 4. Protocolo de Observación de Funciones Docentes

MACROCATEGORÍAS	CRITERIOS	CÓDIGOS	
Organización	Organización externa	Organización del contexto didáctico	OD
		Organización del alumnado	OA
Comunicabilidad docente	Conductas de dirección de actividades	Exposición individual	EI
		Exposición grupal	EG
Motivadora	Integración en la actividad propuesta	Participación del monitor	PM
	Refuerzo	Refuerzo individual	RI
		Refuerzo grupal	RG
Control comportamental	Contingencias de control	Control individual	CI
		Control grupal	CG
Orientación y asesoramiento	Guía, asesoría y retroalimentación	Guía individual	GI
		Guía grupal	GG
	Revisión no verbal de la tarea	Revisión no verbal	RN
Interacción	Entre monitor y participantes	Interacciones generales	IG
		No responde	NR
	Entre monitores	Interacción entre monitores	IM
	Intervenciones de los participantes	Responde al monitor	RP
		Se dirige al monitor	DM
		Disrupciones de los participantes	DP
CATEGORÍAS INSTRUMENTALES (CI)		Otros comportamientos	X
		Inobservable	Y

La formación de observadores se llevó a cabo a través de un entrenamiento desarrollado a lo largo de 20 horas teóricas y prácticas, que recoge aspectos generales y específicos del instrumento. Por otro lado, este entrenamiento persigue que los observadores se ajusten a las normas del protocolo, lo apliquen correctamente y, realicen registros fiables.

Para la realización de las pruebas de fiabilidad se han formado dos parejas de dos observadores cada una, codificando una sesión de fiabilidad cada una de ellas. Las sesiones de observación a codificar han sido de una hora de duración, y se han fraccionado para su codificación por actividades realizadas en la hora de sesión del programa. La decisión de analizarlo sobre actividades y no sobre sesiones se plantea por de 2 tipos de problemas: de una parte problemas técnicos, pues el software utilizado para el registro, *ThemeCoder* (Borrie, Jonsson y Magnusson, 2001, 2002; Anguera y Jonsson, 2003) que ralentizaba y dificultaba la codificación de los vídeos, pues estos tenían una duración de entre una hora y hora y media y, en segundo lugar, el registro de una sesión de una hora de duración, que supone, al menos, siete horas de codificación, lleva al cansancio del observador, lo que, produce una merma en la fiabilidad.

La primera pareja ha observado un vídeo compuesto por dos actividades y la segunda pareja otro vídeo formado por tres actividades y que, en todos los casos, corresponden a una sesión completa de intervención. Por tanto, se presentan tanto los resultados correspondientes a cada una de las actividades que forman la sesión de intervención, como, los resultados relativos a la sesión completa de observación, formada por las actividades registradas de manera independiente.

El *Plan de Observación* ha sido diseñado para el cálculo de la fiabilidad tanto a un nivel de códigos como de funciones, para cada una de las actividades y para la sesión completa de intervención. Comprende tres facetas, dos de diferenciación: observadores, con dos niveles, agentes, con tres niveles, que corresponden a la monitoras principal y básica y los participantes, siendo el objeto de medida a generalizar los 20 códigos de conducta a observar y las seis funciones docentes y la categoría instrumental, dando lugar a el plan de observación multifaceta cruzado  $A \times O$ . También se realizó un segundo *Plan de Observación* para las sesiones de intervención completas, abarcando 4 facetas, tres de diferenciación: observadores, con dos niveles, agentes, con tres niveles que corresponden a la monitoras principal y básica y los participantes, y actividades realizadas en la sesión de intervención, utilizando como objeto de medida a generalizar, los 20 códigos de conducta a observar y las seis funciones docentes y la categoría instrumental, dando lugar a el plan de observación multifaceta cruzado  $A \times O \times A$ .

El modelo de estimación se ha realizado a través del análisis de varianza de efectos aleatorios, considerando infinito el muestreo en cada nivel de las facetas. Los resultados de los Planes de Observación, Estimación y Optimización se han obtenido a través del programa GT (Ysewijn, 1996).

## Resultados

Se presentan a continuación los datos relativos a los Planes de Estimación: A) (Agentes x Observadores) para cada una de las parejas de observadores, analizando primero por actividades y luego por sesión completa y, B) (Agentes x Observadores x Actividades) para cada una de las parejas de observadores, analizando la sesión completa. Estos planes permitirán posteriormente llevar a cabo las optimizaciones: 1) A/O ¿será necesario aumentar el número de observadores para mejorar la fiabilidad? 2) A/OA ¿Será necesario aumentar el número de actividades dentro de la sesión de intervención para aumentar la fiabilidad de los observadores? Las tablas número 5 y 6 corresponden a la primera pareja, por actividades y la sesión completa.

Tabla 5. Planes de estimación de la pareja 1 por actividades y sesión completa. Por códigos

Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 1					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	1461.734	722.55000	16.83429	95
Observadores	1	78.408	20.59150	21.69711	3
PA	2	16.634	16.63350	11.76166	2
Coeficiente G relativo = 0,989 Coeficiente G absoluto = 0,975					
Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 2					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	1388.508	677.92500	91.04766	93
Observadores	1	85.008	17.44983	24.38320	2
PA	2	32.658	32.65850	23.09305	4
Coeficiente G relativo = 0,976 Coeficiente G absoluto = 0,964					
Plan de estimación aleatorio (Agentes x Observadores). SESIÓN COMPLETA					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	5016.160	2447.40000	73.99933	95
Observadores	1	100.833	-6.84183	39.64024	0
PA	2	121.359	121.35850	85.81342	5
Coeficiente G relativo = 0,976 Coeficiente G absoluto = 0,976					

Tabla 6. Planes de estimación de la pareja 1 por actividades y sesión completa. Por funciones

Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 1					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	4003.166	1972.40500	15.48347	95
Observadores	1	168.000	36.54767	47.74791	2
PA	2	58.357	58.35700	41.26463	3
Coeficiente G relativo = 0,985 Coeficiente G absoluto = 0,977					
Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 2					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	3258.500	1534.38100	54.00513	89
Observadores	1	4.567	-61.69034	44.73971	0
PA	2	189.738	189.73800	34.16503	11
Coeficiente G relativo = 0,942 Coeficiente G absoluto = 0,942					
Plan de estimación aleatorio (Agentes x Observadores). SESIÓN COMPLETA					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	2091.070	5981.64300	75.07753	95
Observadores	1	771.429	214.54780	12.10577	3
PA	2	127.785	127.78550	90.35799	2
Coeficiente G relativo = 0,989 Coeficiente G absoluto = 0,972					

Para la pareja número 1 los planes de estimación, tanto por códigos como por funciones para cada una de las actividades y para la sesión de intervención completa, apenas difieren en los resultados obtenidos correspondientes a los coeficientes absoluto y relativo. Dichos coeficientes muestran alta fiabilidad entre los dos observadores iniciales. En las tablas 7 y 8 se presentan los planes de estimación por códigos y por funciones relativos a la segunda pareja de observadores.

Para la segunda pareja de observadores, al igual que para la primera, los resultados obtenidos apenas difieren en los coeficientes absoluto y relativo en los análisis realizados tanto para los códigos como para las funciones, mostrando alta fiabilidad entre los dos observadores.

*Tabla 7. Planes de estimación de la pareja 2 por actividades y sesión completa. Por códigos*

Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 1					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	1031.059	509.64170	64.55800	92
Observadores	1	102.675	30.30000	28.08208	5
PA	2	11.775	11.77500	8.32618	2
Coeficiente G relativo = 0,989 Coeficiente G absoluto = 0,960					
Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 2					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	423.034	210.56680	49.56644	96
Observadores	1	19.200	5.76667	5.24473	3
PA	2	1.900	1.90000	1.34350	1
Coeficiente G relativo = 0,996 Coeficiente G absoluto = 0,982					
Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 3					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	550.009	274.84170	94.45740	100
Observadores	1	0.300	-0.00833	0.11196	0
PA	2	0.325	0.32500	0.22981	0
Coeficiente G relativo = 0,999 Coeficiente G absoluto = 0,999					
Plan de estimación aleatorio (Agentes x Observadores). SESIÓN COMPLETA					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	5820.160	2899.71700	57.75035	97
Observadores	1	226.875	68.7167	61.94048	2
PA	2	20.725	20.72500	14.65479	1
Coeficiente G relativo = 0,996 Coeficiente G absoluto = 0,985					

*Tabla 8. Planes de estimación de la pareja 2 por actividades y sesión completa. Por funciones*

Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 1					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	2960.643	1462.59500	46.82042	92
Observadores	1	288.095	84.21416	78.85345	5
PA	2	35.453	35.45250	25.06870	2
Coeficiente G relativo = 0,988 Coeficiente G absoluto = 0,961					
Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 2					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	1152.667	567.33330	107.57884	93
Observadores	1	85.714	22.57133	23.71105	4
PA	2	18.000	18.00000	12.72792	3
Coeficiente G relativo = 0,984 Coeficiente G absoluto = 0,965					

Plan de estimación aleatorio (Agentes x Observadores). ACTIVIDAD 3					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	1571.453	785.26200	55.59246	100
Observadores	1	0.857	-0.02383	0.31984	0
PA	2	0.928	0.92850	0.65655	0
Coeficiente G relativo = 0,999 Coeficiente G absoluto = 0,999					
Plan de estimación aleatorio (Agentes x Observadores). SESION COMPLETA					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	6560.165	8248.54700	54.94495	97
Observadores	1	672.000	202.97620	83.49841	2
PA	2	63.072	63.07150	44.59829	1
Coeficiente G relativo = 0,996 Coeficiente G absoluto = 0,984					

Para la pareja 2 la precisión de generalización es satisfactoria con los dos observadores iniciales, al existir alta fiabilidad entre ellos, tanto a nivel de códigos como de funciones. Cualquier modificación supone incrementar innecesariamente el costo.

A continuación se presentan los datos relativos a los planes de optimización a nivel de códigos y a un nivel de funciones (A/O) por sesión completa de observación para las parejas 1 y 2.

Tabla 9. Planes de optimización de la pareja 1

Plan de Optimización aleatorio para observadores (A/O). Códigos							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	3	4	5	6	10
N total de observaciones = 6			9	12	15	18	30
Coeficiente G relativo = 0,976			0,984	0,988	0,990	0,992	0,995
Coeficiente G absoluto = 0,976			0,984	0,988	0,990	0,992	0,995
Plan de Optimización aleatorio para observadores (A/O). Funciones							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	3	4	5	6	10
N total de observaciones = 6			9	12	15	18	30
Coeficiente G relativo = 0,989			0,993	0,995	0,996	0,996	0,996
Coeficiente G absoluto = 0,972			0,981	0,986	0,989	0,991	0,994

Tabla 10. Planes de optimización de la pareja 2

Plan de Optimización aleatorio para observadores (A/O). Códigos							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	3	4	5	6	10
N total de observaciones = 6			9	12	15	18	30
Coeficiente G relativo = 0,996			0,998	0,998	0,999	0,999	0,999
Coeficiente G absoluto = 0,985			0,990	0,992	0,994	0,995	0,997
Plan de Optimización aleatorio para observadores (A/O). Funciones							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	3	4	5	6	10
N total de observaciones = 6			9	12	15	18	30
Coeficiente G relativo = 0,996			0,997	0,998	0,998	0,999	0,999
Coeficiente G absoluto = 0,984			0,989	0,992	0,994	0,995	0,997

Los coeficientes de generalizabilidad de este diseño están por encima de 0.90, lo que indica un menor costo en la investigación, al no ser necesario aumentar el número de observadores en esta sesión de observación, para alcanzar valores iguales o superiores a 0.90 en los coeficientes relativo y absoluto.

Se presentan a continuación los datos relativos al Plan de Estimación (Agentes x Observadores x Actividades), para cada una de las parejas de observadores, analizando la sesión completa, con el objetivo de comprobar posteriormente, en el plan de optimización, si el aumento del número de actividades dentro de la sesión suponen un aumento o una disminución en la fiabilidad de los observadores.

*Tabla 11. Planes de estimación de la pareja 1 por sesión completa. Por códigos y funciones*

Plan de estimación aleatorio (Agentes x Observadores x Actividades). CÓDIGOS					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	2846.738	700.13750	03.30323	94
Observadores	1	163.350	20.09992	22.88358	3
Actividades	1	1.067	0.13333	0.58593	0
AO	2	45.987	21.34175	16.30095	3
AA	2	3.504	0.10000	1.77897	0
OA	1	0.067	-1.07900	0.77897	0
AOA	2	3.304	3.30400	2.33628	0
Coeficiente G relativo = 0,984 Coeficiente G absoluto = 0,970					
Plan de estimación aleatorio (Agentes x Observadores x Actividades). FUNCIONES					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	7883.010	1908.71400	94.26785	90
Observadores	1	657.440	65.82125	94.44703	3
Actividades	1	33.440	4.15458	5.19549	0
AO	2	255.940	123.10730	90.55394	6
AA	2	1.940	-3.89275	3.50643	0
OA	1	16.298	2.19067	4.99311	0
AOA	2	9.726	9.72600	6.87732	0
Coeficiente G relativo = 0,968 Coeficiente G absoluto = 0,950					

*Tabla 12. Planes de estimación de la pareja 2 por sesión completa. Por códigos y funciones*

Plan de estimación aleatorio (Agentes x Observadores x Actividades). CÓDIGOS					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	1928.011	314.55270	27.24932	82
Observadores	1	75.625	5.52219	7.12246	1
Actividades	2	263.886	34.39162	31.43600	9
AO	2	6.434	0.88342	1.68213	0
AA	4	38.044	17.13062	11.03667	4
OA	2	23.275	6.49725	5.53407	2
AOA	4	3.783	3.78325	2.18426	1
Coeficiente G relativo = 0,979 Coeficiente G absoluto = 0,934					

Plan de estimación aleatorio (Agentes x Observadores x Actividades). FUNCIONES					
Fuentes de variación	Grados de libertad	CM	Componente de Varianza Aleatorio	Error Estándar	% Variabilidad Total
Agentes	2	5315.720	869.08720	26.55647	81
Observadores	1	203.175	13.14692	19.88969	1
Actividades	2	784.222	100.46030	93.67757	9
AO	2	12.531	-2.64692	4.92413	0
AA	4	109.139	44.33337	32.05521	4
OA	2	92.793	24.10708	22.22366	2
AOA	4	20.472	20.47225	11.81966	2
Coeficiente G relativo = 0,979					
Coeficiente G absoluto = 0,933					

Los resultados obtenidos para ambas parejas tanto para los análisis realizados para códigos como para las funciones, muestran alta fiabilidad entre los dos observadores.

A continuación se presentan los datos relativos a los planes de optimización de la sesión completa a nivel de códigos y a un nivel más global (A/OA) por sesión de intervención.

*Tabla 13. Planes de optimización, pareja 1 (A/OA). Por códigos y funciones*

Plan de Optimización aleatorio para observadores (A/OA). CÓDIGOS							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	2	2	2	2	2
Actividades	n = 2	N = ∞	3	4	5	6	7
N total de observaciones = 12			18	24	30	36	42
Coeficiente G relativo = 0,984			0,984	0,984	0,984	0,985	0,985
Coeficiente G absoluto = 0,970			0,970	0,971	0,971	0,971	0,971
Plan de Optimización aleatorio para observadores (A/OA). FUNCIONES							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	2	2	2	2	2
Actividades	n = 2	N = ∞	3	4	5	6	7
N total de observaciones = 12			18	24	30	36	42
Coeficiente G relativo = 0,968			0,968	0,968	0,968	0,968	0,968
Coeficiente G absoluto = 0,950			0,951	0,951	0,952	0,952	0,952

*Tabla 14. Planes de optimización, pareja 2 (A/OA). Por códigos y funciones*

Plan de Optimización aleatorio para observadores (A/OA). CÓDIGOS							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	2	2	2	2	2
Actividades	n = 3	N = ∞	4	5	6	7	8
N total de observaciones = 18			24	30	36	42	48
Coeficiente G relativo = 0,979			0,984	0,987	0,989	0,990	0,991
Coeficiente G absoluto = 0,934			0,948	0,956	0,961	0,965	0,968
Plan de Optimización aleatorio para observadores (A/OA). FUNCIONES							
Facetas	Niveles	Tamaño	Opt. 1	Opt. 2	Opt. 3	Opt. 4	Opt. 5
Agentes	n = 3	N = ∞	3	3	3	3	3
Observadores	n = 2	N = ∞	2	2	2	2	2
Actividades	n = 3	N = ∞	4	5	6	7	8
N total de observaciones = 18			24	30	36	42	48
Coeficiente G relativo = 0,979			0,985	0,988	0,990	0,991	0,992
Coeficiente G absoluto = 0,933			0,947	0,956	0,962	0,966	0,969

La precisión de generalización es satisfactoria con las actividades iniciales propuestas en las dos sesiones de observación, al existir alta fiabilidad tanto a nivel de códigos como de funciones. Cualquier aumento en el número de actividades que forman la sesión de intervención supone incrementar el costo en el programa.

### Discusión

La calidad del dato es una medida incuestionable en un estudio con metodología observacional. Sin embargo, en investigación aplicada es preciso enfrentar problemas prácticos que van desde los de tipo técnico, como pueden ser limitaciones en los software de registro o bien la fatiga del observador. Los problemas ya planteados (limitaciones del software y fatiga del observador) obligaron a utilizar una unidad de medida diferente a la de la sesión de observación completa, utilizando entidades completas (actividades), con una duración menor, sin necesidad de sacrificar la fiabilidad. De esta forma, la utilización de esta unidad de medida nos ha permitido el cálculo de la fiabilidad, tanto para actividades como para la sesión completa de intervención, lo cual nos permite ver si las observaciones hechas por los dos observadores en períodos relativamente cortos de tiempo conllevan a una buena medida de fiabilidad evitando así las dificultades ya mencionadas a la hora de realizar una observación.

El objetivo planteado en este estudio alcanza los estándares al alcanzar en todos los casos, coeficientes de generalizabilidad superiores a 0,90. Por otra parte este procedimiento de análisis, que ayuda al cálculo de la eficiencia de los recursos, afirma que no es necesaria la utilización de más de dos observadores para el cálculo de la calidad del dato en una sesión de observación, tanto desde una perspectiva de códigos como de funciones y, por otro lado, el número inicial de actividades planteadas en la sesión de intervención permiten su generalización. Un aumento del número de actividades, que supondría un aumento en la temporalización de la sesión de intervención, sin conllevar un aumento de generalización en los registros de los observadores.

Por tanto, la evaluación de proceso partiendo de los resultados obtenidos en el estudio de generalizabilidad, nos sirve para comprobar que partimos de un buen diseño de investigación tanto, en cuanto al número de observadores para el cálculo de la fiabilidad cuanto al número de actividades que forman la sesión de intervención. Este estudio sirve para el futuro diseño de investigaciones más amplias, en las que analizar si el número de sesiones desarrolladas a lo largo del programa son las adecuadas, o si el número de monitoras que intervienen en cada sesión de intervención es suficiente, o bien si el número de actividades por área trabajada y por trimestre son las recomendables para el programa de intervención socioafectivo, entre otros posibles enfoques. Es innegable la aportación que la Teoría de la Generalizabilidad supone para el análisis de la calidad del dato, así como también para el estudio de la eficiencia del programa.

### Referencias

- Alonso Tapia, J. (1997). *Motivar para el aprendizaje: teorías y estrategias*. Barcelona. Edebé
- Anguera, T. y Jonsson, G. (2003). *Detection on real-time patterns in Sports: Interactions in football*. 1st Meeting of Complex Systems and Sport and 4th International Conference of Computer Science in Sport. Barcelona: INEFC.
- Blanco Villaseñor, A. (1993). Fiabilidad, precisión, validez y generalización de los diseños observacionales. En M.T. Anguera (Ed.), *Metodología Observacional en la investigación psicológica* (Vol 2 Fundamentación, pp 151-261). Barcelona: PPU.
- Blanco, A. y Anguera, M.T. (2000). Evaluación de la calidad en el registro del comportamiento: Aplicación a deportes de equipo. En E. Oñate, F. García-Sicilia y L. Ramallo (Eds.), *Métodos*

*numéricos en Ciencias Sociales* (pp. 30-48). Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería (CIMNE).

- Borges, A., Hernández-Jorge, C. y Rodríguez-Naveiras, E. (2005, Marzo). *Evaluación formativa en una escuela de padres y madres de niños y niñas superdotados*. Comunicación presentada en el IX Congreso Internacional de Educación Familiar celebrado en Las Palmas de Gran Canaria.
- Borges, A., Rodríguez-Naveiras, E., Hernández-Jorge, C. y Fernández, R. (2007, Junio). *Programa Integral para Altas Capacidades: Evaluación Formativa 2004-2006*. Comunicación presentada en la IX Conferencia Española y I Encuentro Iberoamericano de Biometría, celebrado en Salamanca.
- Borrie, A., Jonsson, G.K. y Magnusson, M.S. (2001). Application of T-pattern detection and analysis in sports research. *Metodología de las Ciencias del Comportamiento*, 3 (2), 215-226.
- Borrie, A., Jonsson, G.K. y Magnusson, M.S. (2002). Temporal pattern analysis and its applicability in sport: An explanation and exemplar data. *Journal of Sports Sciences*, 20, 845-852.
- Brophy, J. y Good, T.L. (1986). Teacher behavior and student achievement. En Wittrock, M. (Ed) *Handbook of Research on Teaching*. MacMillan. Nueva York.
- Colás, M.P y Rebollo, M.A. (1993). *Evaluación de programas: Una guía práctica*. Editorial Kronos.
- Cooper, J. M. (1993): El maestro y la toma de decisiones. En J. M. Cooper (Ed) *Estrategias de enseñanza: Guía para una mejor instrucción*. Limusa. Méjico.
- Cronbach, L.J., Gleser, G.C., Nanda, H., y Rajaratnam, N. (1972). *The dependability of behavioral measurerents: theory of generalizability for scores and profiles*. New York: John Wiley and
- Hernández, P. (1991). *Psicología de la Educación: corrientes actuales y teorías aplicadas*. Trillas. Méjico. Hernández, P. y García, L. A. (1991): *Psicología y enseñanza del estudio*. Pirámide. Madrid.
- Justicia, F. (1996). El profesor: los procesos de pensamiento. En A.Barca, J.A. González-Pienda, R. González y J. Escoriza (Eds.). *Psicología de la Instrucción*. Vol. 3: Componentes contextuales y relacionales del aprendizaje escolar. EUB. Barcelona.
- Monks, F.J. (1994). Desarrollo socio-emocional de los niños superdotados. En Y. Benito (dir) *Intervención e investigación socioeducativa en alumnos superdotados*, 139-152. Salamanca: Amarú Ediciones.
- Rodríguez-Naveiras, E., Borges del Rosal, A. y Hernández-Jorge, C. (2007, Febrero). *Problemas metodológicos en la evaluación de programas repetidos con los mismos sujetos: resultados de tres años de implementación del programa integral para altas capacidades*. Comunicación presentada en el X Congreso de Metodología de las Ciencias Sociales y de la Salud, celebrado en Barcelona.
- Veney, J.E. y Kaluzny, A.D. (1984). *Evaluation and decision making for health services program*. Englewood Cliff, N.J: Prentice Hall.
- Ysewjin, P. (1996). *Software for generalizability studies*. Mimeografía.
- Ziegler, A. y Heller, K.A. (2000). Conceptions of giftedness from a met-theoretical perspective. En K.A. Heller, F.J. Mönks, R.J. Stenberg y R.F. Subotnik (Eds) *International handbook of giftedness and talent*, 3-21. Oxford: Elsevier.

# ANÁLISIS DE LAS EXPERIENCIAS SUBJETIVAS DE PERSONAS CON TRASTORNOS MENTALES GRAVES EN PROCESOS DE “RECUPERACIÓN”

Francisco Javier Saavedra  
Universidad de Sevilla

En la última década ha aumentado el interés por la investigación de las experiencias subjetivas y los procesos de construcción de significados en las personas que padecen enfermedades mentales graves a la luz del concepto de “recuperación”. Este concepto, aunque con menos recorrido en España, es la piedra angular del diseño de los programas de salud mental comunitaria en Europa. De hecho, en el reciente Plan Integral de Salud Mental de Andalucía 2008-2012 se aboga por la inclusión de este concepto en el diseño de los recursos de salud mental comunitaria. Los investigadores que estudian el proceso de recuperación apuestan por la metodología cualitativa mediante el análisis de relatos en primera persona, narrativas, informes diarios, etc. El uso de metodología cualitativa en este campo no sólo permite descubrir los recursos de las personas para afrontar la enfermedad sino que facilita el análisis de los procesos de negociación entre los discursos sociales sobre la enfermedad mental y las definiciones de ésta por parte de los afectados. El objetivo de la comunicación es presentar el concepto de “Recuperación”, mostrar las metodologías cualitativas utilizadas en este campo, defender la pertinencia de su uso y exponer muy sintéticamente algunos resultados de las investigaciones citadas.

Si consultamos los manuales de buenas prácticas en el ámbito de la salud mental resulta evidente que son los meta-análisis de ensayos con grupos control aleatorizados con muestras significativas los diseños que se describen como los que ofrecen mayor evidencia científica. En el extremo opuesto se encontrarían los relatos personales, “anécdotas” o narrativas. Así, las decisiones y las prácticas sanitarias deberían guiarse a partir de los primeros.

Sin embargo, en la última década numerosas revistas de carácter cualitativo en el campo de salud mental han aumentado su impacto y su influencia y cientos de artículos han sido publicado sobre el análisis cualitativo de las experiencias de las personas que sufren trastornos mentales graves. Tan sólo es necesario escribir en cualquier buscador “Narrativas” y “Enfermedad Mental” para percatarse de este impacto. ¿Cuál es la causa de este auge, a pesar del reconocimiento oficial de su falta de evidencia científica?

## *“Narrative Turn” y Post-modernismo*

En la última década ha sido denominado como “Narrative Turn” (Brown, Nolan, Crawford & Lewis, 1996) el giro epistemológico y metodológico que ha situado las narrativas personales como el núcleo de la experiencia humana y, por lo tanto, como objeto de estudio e investigación en las ciencias sanitarias y sociales. Las narrativas podrían entenderse como núcleos de significados atribuidos a eventos personales mediante los cuales ordenamos nuestra realidad: el conocimiento de la “realidad” se entiende como una historia sobre nosotros mismos (Bruner, 1991). Los significados son negociados en un espacio social y las narrativas son co-construidas en interacción. Sin lugar a dudas, esta aproximación teórica y metodológica se apoya en una perspectiva postmodernista para la cual no existe una verdad ahí fuera que pueda ser descubierta y medida independiente de los observadores, sino que es construida por las personas interactuando con su contexto. La verdad o realidad es siempre dependiente del contexto y de naturaleza provisional.

## *El concepto de “Recuperación” y “Análisis de narrativas” en salud mental*

Aparte de la importancia del paradigma post-modernista en el contexto académico, la aparición y uso del concepto de “Recuperación” en los sistemas de gestión de salud mental como complemento a los tradicionales de “rehabilitación” e “integración” ha sido, sin lugar a dudas, un impulso para el auge de las investigaciones sobre las experiencias subjetivas de las personas con enfermedad mental. La importancia de este concepto es tal que es imposible entender los sistemas de gestión de la salud mental, en especial en los países anglosajones, sin entender esta idea. En los últimos cinco años el paradigma de la “Recuperación” se está empezando a incluir en los planes integrales de salud mental en España.

Veamos como un documento oficial proveniente del Reino Unido nos define “La Recuperación”:

“Una serie de valores sobre los derechos de la persona a construir un vida significativa para si misma, con o sin la presencia continua de síntomas de salud mental. La recuperación está basada en las ideas de la auto-determinación y el control de si mismo. Hace énfasis en la importancia de la esperanza en el mantenimiento de la motivación y la ayuda con el objetivo de la consecución una vida llena de sentido” (*The future vision coalition*, 2009; p.22).

Como se puede observar algunos conceptos de esta definición son de naturaleza subjetiva y muy difícil de objetivar. Recuérdese el carácter oficial del documento y que es la “recuperación” uno de los objetivos esenciales del sistema de salud. Han sido las propia asociaciones de afectados, clientes, familiares o enfermos las que presionado para que sus experiencias y opiniones las que sean tenidas en cuenta en las decisiones de los técnicos y políticos.

Veamos a continuación de forma muy resumida la utilidad que la investigación sobre las narrativas o experiencias subjetivas de pacientes con enfermedad mental:

La investigación cualitativa nos permiten descubrir qué recursos utilizan las personas con TMG para enfrentarse a la enfermedad mental desde sus propias voces (Mezzina, y cols, 2006). También nos posibilitan estudiar cómo las personas se apropian y elaboran los significados y las herramientas culturales que los contextos socio-sanitarios ofrecen (Saavedra, 2009). El análisis de estas experiencias se constituye como un nuevo campo en el cual analizar las relaciones entre mente y cultura. Además, la exploración de los caminos recorridos por las personas que sufren TMG para lograr dar sentido a sus vidas, al fin y al cabo, nos remiten a la búsqueda de sentido a la vida de todos los seres humanos. Y por último, desde una perspectiva histórica el concepto de “Recuperación” puede ser comprendido como un nuevo marco ideológico a partir del cual construir y diseñar los servicios de salud mental.

La metodología empleada puede ser muy variada e incluso integrar aspectos cualitativos y cuantitativos, aunque siempre haciendo énfasis en los primeros: Análisis del discurso, análisis conversacional, análisis del contenido, análisis de posicionamientos sociales, etnografía, análisis mixto cuantitativo (corpus)-cualitativo, grupos focales.

Algunos ejemplos en los últimos años de investigación cualitativa pueden ser los siguientes.

- Jenkins y Carpentenr-Song (2006) estudiaron con métodos etnográficos en clínicas, espacios públicos, restaurantes u hogares a 90 pacientes ambulatorios de esquizofrenia, que tomaban fármacos de segunda generación.
- Ridgway (2001) examinó los informes autobiográficos de 25 narrativas de mujeres pacientes de TMG publicados, apoyándose metodológicamente en la “*grounded Theory*”.
- Lührman (2007) realiza una descripción etnográfica de los clientes de albergues en el norte de Chicago.

- Dorvil, Morin, Beaulieu y Robert (2007) analizan, mediante entrevistas, el control percibido de 21 residentes en dos modelos distintos de alojamientos protegidos.
- Pejler, Asplund y Norberg (1995) entrevistaron a 10 pacientes, 17 enfermeras y 8 familiares en tres momentos. El objetivo de la investigación fue analizar los cambios en la calidad de los cuidados en el paso de un contexto hospitalario a otro de residencia protegida.
- Roe and Davidson. (2005) reflexionan *sobre la importancia del paradigma narrative en el campo de la esquizofrenia y la recuperación*.
- Mezzina y cols. (2006) realizan un exhaustivo repaso de las implicaciones del concepto de “Recuperación” en la práctica clínica e investigadora.
- Saavedra, (2009) y Saavedra, Cubero, y Crawford, (2009) estudiaron exhaustivamente 10 pacientes graves de esquizofrenia paranoide residentes en casas hogar para explorar como el contexto socio-cultural de los recursos donde vivían afectaba a su proceso de recuperación.
- Saavedra (2010) analizó con una metodología cualitativa y cuantitativa los criterios de coherencia y cohesión de las narrativas de vida de 18 pacientes de esquizofrenia paranoide.

Sin la intención de ser exhaustivo me gustaría ofrecer algunos resultados de las investigaciones cualitativas sobre las experiencias de personas con enfermedad mental en proceso de recuperación.

La sensación de control sobre la propia vida y la capacidad de tomar decisiones son dos de los factores fundamentales para el éxito de la recuperación. El contexto de la vida cotidiana ha sido identificado por diversos investigadores como esencial para que los pacientes adquieran algún grado de control sobre sus vidas. Por ejemplo, decidir quedarse unos minutos más en la cama. La negociación del significado y el simbolismo de la enfermedad con el objetivo de construir una identidad social adaptativa resultan esenciales en el proceso de integración social. La conciencia de enfermedad no es simplemente una cognición aislada, sino también un elemento de una historia vital y una comprensión narrativa de la vida del paciente (Lysaker y cols., 2002). Desde una perspectiva narrativa aquello que nos parecen simples negaciones de la enfermedad, complejas delusiones o conductas extrañas, en algunas ocasiones son intentos activos de dar sentido a los síntomas o la búsqueda de experiencias significativas y seguras. Mezzina y cols., (2006) observan distintas estrategias para negociar con el estigma: enfatizar algunos componentes específicos del contenido semántico de este concepto como el de “problemas para aprender” o “malo de los nervios”, hasta la lucha por los propios derechos como enfermo y persona. “Tácticas lingüísticas”. Los recursos residenciales con apoyo profesional las 24 horas, como las “Casas Hogares”, facilitan algún grado de sensación de pertenencia y posibilitan la creación de relaciones de amistad y reciprocidad entre sus residentes y profesionales. Unido a la sensación de pertenencia aparece otro importante sentimiento: la responsabilidad respecto a los otros, respecto a nuestros semejantes. Éste es un aspecto al que se le ha prestado poca atención en relación a las personas que padecen TMG (Saavedra, 2009a). Para lograr la recuperación es necesario considerar a la persona con enfermedad mental no como un paciente sino como “ciudadano”, “empleado” y “consumidor” (Mezzina y cols., 2006).

### *Fiabilidad, validez y metodología cualitativa*

No es necesario señalar las importantes diferencias metodológicas en lo que respecta al concepto de validez y fiabilidad que conllevan una investigación cualitativa enfocada a analizar experiencias subjetivas de personas y un diseño experimental. El diseño experimental implica una concepción

“realista”. Es decir que existe una realidad que puede ser observada y esta realidad es cognoscible a través de la investigación. Así, la fiabilidad se logra dentro de la investigación en el proceso de repetición y confirmación de los resultados. Sin embargo, la investigación cualitativa supone, en general, una visión interpretativa. O lo que es lo mismo, que existen varios caminos para entender la vida social y la realidad, incluso múltiples realidades. La fiabilidad se logra mediante un proceso de negociación entre los investigadores y los destinatarios de los resultados de la investigación.

Esto no quiere decir que la validez y fiabilidad en la investigación cualitativa no sea esencial. Cohen y Crabtree, (2008) y Mays y Pope, (2000) apuntan los siguientes métodos para obtener una fiabilidad y una validez aceptable en las investigaciones cualitativas. Tras una clara exposición de la selección de la muestra, recolección de datos y método de análisis son posibles las siguientes técnicas. Me limito simplemente a citarlas: Triangulación, *peer Review/debriefing*, auditores externos, “*member Checking*”, reflexividad, “*fair dealing*”, “atención a los casos negativos”.

*A modo de conclusión: ¿Es posible integrar lo cualitativo y cuantitativo?*

La “medicina basada en la evidencia” utiliza indicadores como “recaídas” o “días de hospitalización”, “ratios de síntomas psiquiátricos”, que difícilmente podrán ser compatibles con los métodos y objetivos de una perspectiva desde la recuperación (Anthony, 2001). Otros investigadores (Frese, Stnaley, Kress y Vogerl-Scibita, 2001), sin embargo, intentan integrar las prácticas basadas en la evidencia con el modelo de recuperación. El debate está abierto, mientras algunos no muestran ningún interés en integrarse en las prácticas basadas en la evidencia, por considerarlas iatrogénicas, otros opinan que no sólo son compatibles, sino que es necesaria la integración.

En mi opinión, y sin la intención de cerrar el debate, las dos perspectivas serían compatibles si las situamos en dos planos distintos. En una dimensión de decisión política, de gestión sanitaria pública, son imprescindibles investigaciones que se acerquen al grado máximo de excelencia, según “la medicina basada en la evidencia”. Sin embargo, al aterrizar en una dimensión más micro, la de las prácticas sociales y sanitarias del día a día, la dimensión de los complejos encuentros sociales que acontecen en escenarios socioculturales concretos, es necesario aprender de las investigaciones que se señalaron en esta comunicación, cómo las personas comprenden y dan significados a las decisiones, planes y prácticas. Sólo a partir de estas investigaciones cualitativas es posible que los profesionales sanitarios y sociales adquieran los recursos, habilidades y competencias necesarias para sus prácticas diarias. Sólo a partir del diálogo entre una y otra dimensión de análisis y de decisión, es posible lograr un sistema sanitario y social eficaz y humano.

## Referencias

- Anthony, W.A. (2001). Need for recovery compatible evidence-based practices. *Mental Health Weekly*.
- Brown, B., Nolan, P., Crawford, P. & Lewis, A. (1996). Interaction, Language and “Narrative Turn” in Psychotherapy and Psychiatry. *Social Science and Medicine*, 43 (11), 1569-1578.
- Bruner, J. (1991). *Actos de significado*. Madrid: Alianza Editorial.
- Cohen D.J. y Crabtree B.F. (2008) Evaluative criteria for qualitative research in health care: controversies and recommendations. *Ann Fam Med*, 6(4), 331-339
- Dorvil, H., Morin, P., Beaulieu, A. y Robert. D. (2005). Housing as a social integration factor for people classified as mentally ill. *Housing Studies*, 20(3), 497-519.

- Frese, F., Stanley, J., Kress, K. y Vogel-Scibita, S. (2001). Integrating evidence-based practices and the recovery model. *Psychiatric Services* 52(11), 1462-1468.
- Jenkins, J.H. y Carpenter-Song, E. (2006). The new paradigm of recovery from schizophrenia: cultural conundrums of improvement without cure. *Culture, Medicine and Psychiatry*, 29, 379-413.
- Lurhman, T.M. (2007). Social defeat and the culture of chronicity: or., why schizophrenia does so well over there and so badly here. *Culture, Medicine and Psychiatry*, 31, 135-172.
- Lysaker, P.H., Clements, C.A., Plascak-Hallberg, C.D., Knipscheer, S.J. y Wright, D.E. (2002). Insight and personal narrative of illness in schizophrenia. *Psychiatry*. Tomo 65, N° 3, Pág. 197.
- Mays, N. y Pope C. (2000). Qualitative research in healthcare: assessing quality in qualitative research. *British Medical Journal*, 320, 50-52.
- Mezzina, R., Davidson, L., Borg, M., Marin, I., Topor, A. y Sells, D. (2006). Social Nature of Recovery: Discussion and Implications for Practice. *American Journal of Psychiatric Rehabilitation*, 9, 63-80.
- Pejlert, A., Asplund, K. & Norberg, A. (1995) Stories about living in a hospital ward as narrated by schizophrenic patients. *Journal of Psychiatric and Mental Health Nursing*, 2, 269-277.
- Ridgway, P. (2001). Restorying Psychiatric Disability: Learning from First Person Recovery Narratives. *Psychiatric Rehabilitation Journal*, 24(4), 335-343.
- Roe, D. y Davidson, L. (2005). Self and Narrative in Schizophrenia: Time to Author a New Story. *Medical Humanities* 31, 89-94.
- Saavedra, J. (2009). Schizophrenia, Narrative and Change: Andalusian Care Homes as Novel Socio-cultural Context. *Culture, Medicine and Psychiatry*, 33(2), 163-184.
- Saavedra, J., Cubero, M. y Crawford, P. (2009). Incomprehensibility of narratives of persons with schizophrenia. *Qualitative Health Research* (in press).
- Saavedra, J. (2010). Quantitative criteria of narrative coherence and complexity in persons with paranoid schizophrenia. *Journal of Nervous and Mental Disease* (in press).
- The future Vision Coalition. *A future vision of mental health*. Reino Unido. 2009. [18 de Agosto 2009]; accesible en: [http://www.newvisionformentalhealth.org.uk/A\\_future\\_vision\\_for\\_mental\\_health.pdf](http://www.newvisionformentalhealth.org.uk/A_future_vision_for_mental_health.pdf).

# PRÁCTICAS DE PLANIFICACIÓN, GESTIÓN, EJECUCIÓN Y CONTROL DE LA PRODUCCIÓN DE ALIMENTOS EN SITUACIONES SIMULADAS Y REALES

Marta Tenutto<sup>1</sup>, Cristina Brutti<sup>2</sup> y Sonia Algora<sup>2</sup>

<sup>1</sup>Facultad Latinoamericana de Ciencias Sociales

<sup>2</sup>Escuela de Nivel Medio y Superior de Educación Técnica N° 1 de Nogoyá, Argentina

Se presenta una investigación cualitativa, concretada en un estudio de caso. El objetivo: indagar los efectos generados en el aprendizaje de Procesos Productivos y Tecnología de Gestión en alimentos, propiciados por una enseñanza basada en estrategias indirectas (estudio de casos y en Aprendizaje en base a problemas) en el Profesorado de Educación Tecnológica de la Escuela de Nivel Medio y Superior de Educación Tecnológica N°1 de Nogoyá, Entre Ríos, Argentina. La investigación partió de la identificación de dificultades originadas por las bajas posibilidades de transferencia de los contenidos aprendidos al ámbito laboral. Se realizaron entrevistas a los egresados y a profesores. En base de los resultados se diseñó una propuesta de enseñanza. Se realizaron encuestas y entrevistas a los alumnos de la experiencia y a graduados así como entrevistas a actores claves y a profesores involucrados en el proyecto. Se relevaron las diferencias percibidas por todos los actores involucrados durante todo el proceso y al final. Dato adicional: estudiantes que habían aprobado las materias pidieron integrarse a la propuesta. Los resultados indican que esta propuesta de enseñanza propicia un saber hacer, favorece el compromiso y el trabajo colaborativo y le proporciona al egresado mayores herramientas para su práctica profesional.

La Escuela de Nivel Medio y Superior de Educación Tecnológica N° 1 de Nogoyá, cuenta en su oferta educativa con el Profesorado de Educación Tecnológica desde el año 2001. El análisis de las prácticas de residencia identificó falencias de los alumnos al enfrentar clases (en talleres o laboratorios) en las que debían realizar procesos productivos. Las posibles causas de la inseguridad en su desempeño fue atribuida a la falta de práctica en la elaboración y concreción de proyectos productivos así como falencias en el manejo del equipamiento e instrumental pertinente. Se vincularon estos problemas con el uso de propuestas de enseñanza basadas exclusivamente en la exposición oral y a la falta de oportunidad de aprendizaje de conocimientos integrados y de prácticas en situaciones reales.

Los supuestos básicos<sup>[1]</sup> que sostienen esta indagación son:

- En relación con el saber y el conocimiento, el primero es producido por el sujeto confrontado por otros sujetos en “marcos metodológicos en tanto “el conocimiento constituye como resultado de una experiencia personal ligada a la actividad de un sujeto dotado de cualidades afectivo-cognitivas” (Charlot, 2007:100) La práctica es una forma de saber y hay un saber en la práctica.
- El aprendizaje 1) se dice que es auténtico a aquel que se produce cuando los alumnos tienen oportunidad de construir activamente sus concepciones sobre la realidad, y se producen en un proceso situado y cultural; 2) la implicación de los alumnos en prácticas con los otros favorece la apropiación de los bienes culturales y simbólicos (Vigotsky, 1981; Rogoff, 1993; Bruner, 1997) y 3) la curiosidad, el interés y la comprensión facilitan el aprendizaje.

[1] Para elaborar estos supuestos se tomó como marco el aportado por: Joaquín García en “¿Qué significa investigar en el aula?”. Cooperación educativa, páginas 22 a 27; y por Charlot, B. (1997). *La relación con el saber. Elementos para una teoría*, Buenos aires, Libros del Zorzal. 2007.

- En el marco de la enseñanza el profesor tiene que trabajar para: 1) ofrecer un ambiente flexible, facilitador y culturalmente intenso y provocador de nuevos desafíos; 2) fomentar el desarrollo autónomo y crítico de los estudiantes; 3) generar un proceso reflexivo orientado a que el alumno asuma el control de su conocimiento; 4) propiciar la comunicación escrita y oral de las actividades y de los procesos, la presentación de justificaciones a las decisiones asumidas (o por asumir) así como de las dudas e incertidumbre en espacios de intercambio; 5) habilitar espacios donde se formulen hipótesis, se pongan a prueba y se enuncien conclusiones y 6) habilitar espacios donde la evaluación se constituya en un proceso y se integre al trabajo cotidiano. Tomando el marco teórico antes enunciado se formuló como hipótesis de la investigación que la planificación y ejecución de proyectos de producción propician un saber hacer que favorece la vinculación con el saber y le proporciona al egresado mayores herramientas para su práctica profesional incentivando el interés por la investigación en tecnología de gestión y productivas, elevando su capacitación técnica y minimizando los temores en las Prácticas de Residencia producto de las carencias de conocimientos procedimentales sobre producción.

Dado que la escuela cuenta con una planta para la elaboración de alimentos la propuesta se vinculó con la producción de los mismos involucrando a los espacios Procesos Productivos y Tecnología de Gestión.

### Método

Se utilizó un diseño de investigación cualitativo concretado en estudio de caso. La población bajo estudio estuvo constituida por el total del alumnado sujeto a la práctica propuesta de los espacios Procesos Productivos II y Tecnología de Gestión pertenecientes al cuarto año de la carrera.

Se usaron como estrategias de enseñanza privilegiadas el estudio de casos y el aprendizaje basado en problemas (Torp y Sage, 1998; Wassermann, 2006). Qué alimento producir, cómo, en qué condiciones son algunos de los problemas que los alumnos debieron enfrentar a fin de formular un proyecto de producción, concretar su ejecución y evaluar los resultados obtenidos.

Durante el primer cuatrimestre, en el espacio Procesos Productivos, se realizaron trabajos prácticos que requirieron del análisis de problemas en toda su complejidad y de reconocimiento de la necesidad de adquirir nuevos conocimientos para efectuar este análisis en profundidad y así llegar a una resolución posible y eficaz. Estos trabajos fueron ejecutados bajo modalidad grupal. El producto final fue la elaboración del proyecto de producción. El rol del docente fue ejercer un rol de tutoría. Las instancias de enseñanza directa se dieron en las prácticas de laboratorio y planta.

Durante el segundo cuatrimestre se realizaron prácticas en planta laboratorio sobre la base de las necesidades definidas por los alumnos para cada una de las producciones seleccionadas. Hacia el final del cuatrimestre se realizaron las producciones bajo el plan construido por los alumnos.

Como actividad no programada se realizaron microclases destinadas a los alumnos residentes en las cuales los alumnos afectados a la investigación actuaron como docentes de sus pares.

En el espacio tecnología de Gestión se realizó un diagnóstico de conocimientos de la Administración de empresas, Economía y Política Económica del siglo XX en Argentina. Se realizaron trabajos prácticos referidos al análisis de unidades de producción y al estudio de sistemas de gestión de las organizaciones productivas teniendo como estructura de recorrido la producción de alimentos. Los trabajos de clase en la primera etapa tuvieron carácter individual con presentación escrita y/u oral, luego estos se desarrollaron en forma grupal en número de dos estudiantes por grupo.

La recolección de datos de las diferentes actividades ejecutadas en ambos espacios se realizó mediante: entrevistas a egresados, evaluaciones de desempeño realizadas por los docentes y auto evaluaciones de los alumnos. Se trabajó con una propuesta de evaluación de proceso, tomando como estrategia privilegiada el portafolio (Tenutto, 2005). Al finalizar el año se propuso un portafolio reflexivo, los alumnos realizaron una autoevaluación mediante pruebas de aprovechamiento y escala de Likert (Bravin y Pievi, 2008).

La evaluación del desempeño de los alumnos en el diseño y ejecución del proyecto de producción se realizó mediante la observación de las habilidades intelectuales tales como capacidad de planificar, de aplicar y transferir conocimientos, búsqueda y procesamiento de la información, etc. antes y después de la ejecución de los trabajos prácticos producidos.

## Resultados

A mediados de año surgió una situación no prevista en las actividades programadas: la experiencia llevada a cabo por los alumnos afectados a esta nueva modalidad de enseñanza fue conocida por los residentes quienes cursaron las materias afectadas al proyecto de investigación el año anterior. Esto motivó el interés de los mismos quienes solicitaron participar de las actividades. Los alumnos manifestaban falta de conocimiento en el manejo del instrumental de laboratorio y equipamiento de planta que conducía a problemas e inseguridades en sus prácticas de residencia. Como respuesta a esta solicitud se propusieron tres microclases en las cuales los alumnos participes del proyecto actuaron como docentes de sus compañeros residentes.

Las microclases permitieron observar el desempeño de los alumnos en la enseñanza del uso del instrumental de laboratorio donde debieron abordar los conceptos científico-tecnológicos que lo sustentan y demostrar el uso de la información obtenida. Los contenidos de las clases fueron los involucrados en los diversos análisis físico químicos realizados en los tres procesos de producción. Como actividad de desarrollo cada uno de los residentes realizó el análisis correspondiente y como actividad de cierre socializaron los resultados.

El dominio de los conceptos científico-tecnológico fue el objetivo donde mayor grado de dificultad se observó (Fig.1). Sólo el 28 % de los alumnos mostraron un sólido dominio de los mismos evidenciándose en los restantes la necesidad de consolidarlos. Estos resultados evidenciarían la ausencia de aprendizaje significativo en las materias física y química a cuyos conceptos los alumnos debieron recurrir.

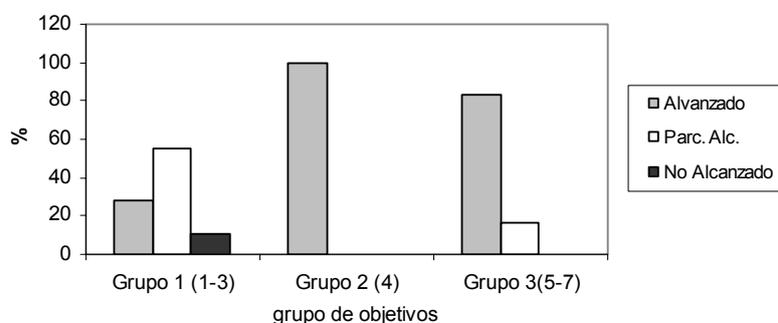


Figura 1. Desempeño de los alumnos por grupo de objetivos (fr%)

### Objetivos:

1. Empleo de vocabulario científico-técnico.
2. Integración de conocimientos de física y química.
3. Precisión y claridad en la explicación de conceptos científico-tecnológicos.
4. Uso solvente de materiales e instrumentos de medición.
5. Claridad en las consignas de trabajo propuestas.
6. Socialización de la interpretación de los resultados obtenidos en los análisis efectuados.
7. Obtención de conclusiones útiles para el proceso de producción seleccionado.

No se observaron falencias en el uso de los equipos de laboratorio en ninguno de los alumnos indicando la facilidad en recibir una capacitación técnica la cual no indica el conocimiento de los principios de su funcionamiento sino sólo su operatividad.

Finalmente el uso de la información obtenida en las determinaciones fisico-químicas, fue satisfactorio por cuanto el 80% de los alumnos aplicaron correctamente los datos recogidos al contexto de la producción seleccionada (Fig. 1).

Los resultados de los trabajos prácticos de ambos espacios orientados hacia la formulación y ejecución del proyecto de producción permitieron obtener los siguientes resultados:

#### a- Espacio Procesos Productivos

Los alumnos lograron, mediante esta metodología de enseñanza, elaborar y ejecutar proyectos de producción de alimentos con el logro de los objetivos propuestos. La evaluación de su desempeño evidencia la existencia de aprendizaje significativo. Todos los alumnos mostraron una mejora en el desarrollo de las capacidades intelectuales y técnicas evaluadas (Fig. 2, Fig. 3). La capacidad de aplicar y transferir conocimientos al proyecto fue la habilidad intelectual de mayor mejora en todos los alumnos. Los conocimientos previos fueron insuficientes. Se evidenció ausencia de aprendizaje significativo en los conocimientos científicos de física y química. Se propone aplicar la metodología ABP (aprendizaje basado en problemas) como organizador curricular. Las capacidades técnicas previas de los alumnos eran superficiales lo cual producía inseguridad. Consolidar dichas capacidades condujo a un cambio de actitud en el espacio. El temor fue sustituido por una indagación responsable.

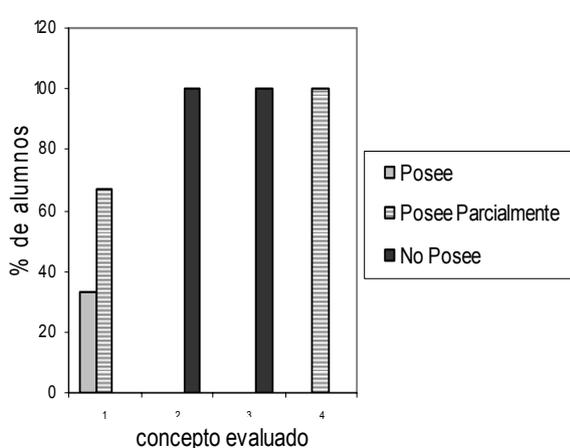


Figura 2. Valoración de conocimientos y habilidades técnicas de los alumnos antes de la aplicación del ABP

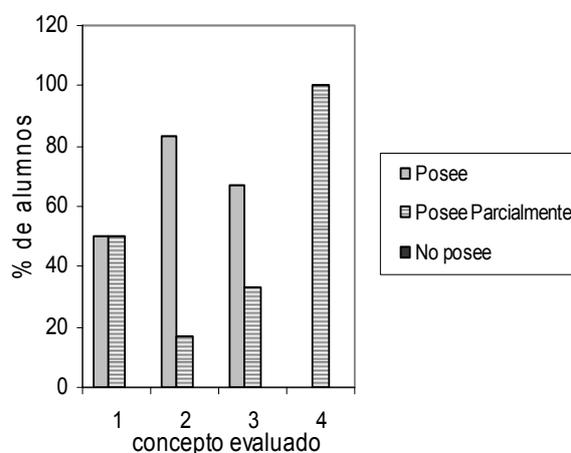


Figura 3. Valoración de conocimientos y habilidades técnicas de los alumnos después de la aplicación del ABP

### *Conceptos evaluados:*

1. *Dominio de conocimientos científico-tecnológico para resolver el diseño del proyecto*
2. *Capacidad de operación de instrumentos de laboratorio*
3. *Conocimiento de equipos de planta de elaboración de alimentos*
4. *Empleo de Autocad*
5. *ABP: Aprendizaje Basado en Problemas*

### *b-. Tecnología de Gestión*

Los estudiantes pudieron, mediante esta metodología de enseñanza elaborar y ejecutar proyectos reales que los pusieron frente a sus dificultades en el manejo de procedimientos técnico administrativo y financieros que deberían dominar al egresar con título con Habilitación Docente para esta área de aprendizaje.

Los estudiantes se mostraron comprometidos y hacedores de todas las instancias del proyecto, pero advirtieron que les demandó más tiempo de actividad extra áulica que con otras metodologías de enseñanza y aprendizaje, por lo que se propone el trabajo institucional en equipo de docentes para la capitalización del tiempo empleado tanto por los estudiantes como por el propio cuerpo de profesores.

En su mayoría de los estudiantes la actitud hacia el aprendizaje basado en la intuición, con poco recorrido previo de las fuentes bibliográficas, cambió volcándose a una indagación reflexiva en el contacto con las fuentes.

### *Autoevaluaciones*

#### *a. Pruebas de Aprovechamiento*

El total de alumnos consideró que la experiencia les permitió alcanzar los siguientes objetivos:

- conocer las operaciones que deben realizar para obtener un producto
- conocer el manejo del instrumental y del grupo
- definir parámetros
- lograr el manejo del equipo respetando el instructivo
- comprender los análisis microbiológicos

#### *b. Escala de Likert*

Sobre las afirmaciones que integraron la escala el total de los alumnos acordó que:

- se produjo un uso correcto del vocabulario técnico
- que solo se logró parcialmente la integración en los diferentes espacios
- que tuvieron problemas en el manejo de equipos solo ocasionalmente
- que realizaron un buen manejo de la información disponible

En el resto de las respuestas se observó cierta dispersión. Así, el desempeño referido a la obtención de la información fue percibido de manera diferente en los distintos grupos siendo para algunos grupos muy bueno y para otros una meta parcialmente lograda. Del mismo modo la afirmación referida al correcto desempeño de los alumnos en las prácticas fue acordada por el 75 % de los mismos en tanto que el 25% restante acordó parcialmente con esta afirmación.

### *c. Dificultades encontradas. Modificaciones propuestas*

A fin de año se propuso recuperar la experiencia realizada mediante un relato breve de lo realizado en cada una de las prácticas de Procesos Productivos y Tecnología de Gestión.

Indicando: cuáles fueron las dificultades encontradas, qué rescata de dichas experiencia, qué modificaciones propondría.

- Dificultades encontradas: de organización a pesar de que en cada uno de los proyectos de producción las acciones estaban planificadas cuando estuvieron frente a los equipos en las producciones se desorganizaron, existía cierta falta de profundidad en la lectura de los textos (aunque llegaron a abordarlos a todos), información específica insuficiente sobre la producción seleccionada por uno de los grupos que hizo retrasar en la tarea.
- Qué rescata de dichas experiencia: la forma de aprender así como lo que se aprendió, enfatizan que rescatan los conocimientos, experiencia muy gratificante y repleta de emociones, les resulta muy significativo descubrir que lo pueden llevar a la práctica, la experiencia en la planta la consideran un aporte sustancial para su futuro profesional.
- Qué modificaciones propondría: trabajar las consignas con más anticipación, que las todas materias tengan teoría y práctica.

## Discusión

Los resultados obtenidos en la adquisición de competencias vinculadas a la elaboración y ejecución de proyectos productivos mediante la estrategia de enseñanza propuesta coinciden con lo enunciado por Wassermann (2006) y Torp y Sage (1998). En la presente propuesta se evitó el monismo metodológico siendo importantes las instancias de aplicación de estrategias indirectas de enseñanza tal como lo señala David Ausubel (2002). Las carencias evidenciadas en la construcción de estructuras cognitivas de las ciencias exactas (física-química) indicarían la necesidad de emplear una combinación de estrategias para su enseñanza debiendo identificar los procedimientos más efectivos para la obtención de los fines perseguidos tal como lo señala Zabalza (2002). La planificación por competencias propuesta entre otros por Yáñez y Villardón (2008) definiría dichos fines acorde al perfil profesional. Las mejoras observadas en las habilidades intelectuales indican un dominio cognoscitivo relacional según la escala de Biggs (2005) donde el término comprensión tiene un sentido académicamente relevante.

Los resultados indican el valor del aprendizaje en situación y la posibilidad trabajar la teoría y la práctica de un modo no disociado en tanto todo concepto teórico está cargado de experiencia y toda experiencia es leída desde alguna teoría, por lo que resulta necesario trabajar teoría y práctica y propiciar un enfoque que tenga presente tanto el producto como el proceso.

## Conclusión

Sobre la base de los resultados expuestos se concluye que los alumnos lograron mediante esta metodología de enseñanza, elaborar y ejecutar proyectos de producción de alimentos con el logro de los objetivos propuestos. La evaluación de su desempeño evidencia la existencia de aprendizaje significativo. Todos los alumnos mostraron una mejora en el desarrollo de las capacidades intelectuales y técnicas evaluadas.

Es importante señalar el compromiso y la responsabilidad que estos estudiantes asumieron. Los proyectos promovieron en los alumnos, el descubrimiento, dominio y comprensión de la realidad por cuanto fueron no sólo diseñados sino ejecutados. Para lograrlo, el docente tiene que asumir un riesgo y un desafío: no se trata de facilitar solo información, es necesario propiciar el aprendizaje donde saber hacer se encuentra atravesado por el saber, en lugar de abordar teoría y práctica como pares en disputa.

### Referencias

- Biggs, J. (2005). *Calidad del Aprendizaje Universitario*. Colección Universitaria. Madrid: Ed. Narcea de S.A. Ediciones.
- Bravin, C. y Pievi, N. (2008). *Documento Metodológico orientador para la investigación educativa*, Buenos Aires, Ed. INFD. Ministerio de Educación. OEI. UNICEF
- Bruner, J. (1997). *La educación, puerta de la cultura*, Volumen CXXV de la colección Aprendizaje, Madrid: Editorial Visor.
- Charlot, B. (2007). *La relación con el saber. Elementos para una teoría*, Buenos Aires: Libros del Zorzal.
- Tenutto, M. (2005). *Herramientas de evaluación en el aula*, Buenos Aires: Editorial Magisterio del Río de la Plata.
- Torp, L. y Sage, S. (1998). *El Aprendizaje basado en problemas*, Buenos Aires: Amorrortu Editores.
- Vigotsky, L. (1979). *Pensamiento y Lenguaje*, Buenos Aires: Ed. La Pléyade
- Wassermann, S. (2006). *El estudio de casos como método de enseñanza*, Buenos Aires: Amorrortu editores.
- Yáñez y Villardón. (2008). *Planificar desde las competencias para promover el aprendizaje*. Cuadernos Monográficos del Instituto de Ciencias de la Educación N°12. Bilbao: Publicaciones Universidad de Deusto.
- Zabalza. (2002). *La enseñanza universitaria. El escenario y sus protagonistas*, Ed. Narcea de S.A. Ediciones.

# EL CONSENSO ENTRE EXPERTOS COMO MÉTODO PARA EVALUAR LA VALIDEZ DE CONTENIDO

Ana Andrés, Juana Gómez-Benito y Carmina Saldaña  
Universidad de Barcelona

Recientes investigaciones apuntan a la utilidad del método Delphi como técnica para analizar la validez de contenido de cuestionarios mediante el consenso de expertos. En el presente estudio, el método Delphi ha sido aplicado para validar el contenido de una escala que evalúa procesos de cambio en personas con sobrepeso y obesidad. Se elaboró una primera versión del cuestionario *Procesos de Cambio en personas con Sobrepeso y Obesidad*, formado por 107 ítems. Un total de 341 expertos en el ámbito de la obesidad fueron invitados a participar. De ellos, 66 expertos de 29 países formaron el panel de expertos. El estudio se llevó a cabo mediante una página web (en castellano y en inglés), mediante la cual se pidió a los expertos que dieran su opinión acerca de la representatividad y claridad de los ítems a lo largo de 3 rondas. Después de cada ronda se incorporaron los cambios sugeridos (ítems que había que eliminar, reformular o añadir). En cada ronda recibieron una respuesta estadística grupal acerca de la opinión del panel de expertos en cada uno de los ítems. Como resultado de las opiniones recibidas fue posible refinar la escala, obteniendo una versión final del cuestionario de 63 ítems.

## *Validez de contenido*

El proceso de validación de un cuestionario representa la acumulación de evidencia científica que sustente la correcta interpretación de sus puntuaciones así como su relevancia para el objetivo que es utilizado (AERA, APA y NCME 1999). Puesto que el objetivo de la aplicación de un determinado cuestionario será la evaluación de un constructo, es necesario analizar su validez con el objetivo de examinar en qué medida el cuestionario mide los diferentes aspectos del constructo. La American Psychological Association (AERA, APA y NCME 1999) propone un concepto unitario de validez, la cual puede ser evaluada desde diversos puntos de vista o evidencias de la misma.

La validez de contenido representa una de las evidencias de validez. Ésta se refiere al análisis de la relación entre el contenido del cuestionario y el constructo que se pretende medir. La importancia de llevar a cabo un correcto análisis de la validez de contenido radica en el hecho de que el constructo será evaluado a partir de la interpretación de los resultados del test.

La evidencia basada en el contenido del cuestionario puede incluir análisis lógicos o empíricos acerca del grado en que el contenido del test representa adecuadamente el dominio del contenido a evaluar. Asimismo, tendrá en cuenta el grado en que el contenido del cuestionario es relevante para la interpretación de las puntuaciones obtenidas en el cuestionario. Paralelamente, la evidencia basada en el contenido puede basarse en las opiniones de jueces expertos acerca de la relación entre el contenido del test y el constructo. El juicio de expertos puede ser requerido para evaluar el cuestionario de distintas formas: desde la asignación de los diversos ítems a distintas categorías hasta la evaluación de la representatividad de un conjunto de ítems (AERA, APA y NCME 1999).

### *El método Delphi*

El método Delphi tiene por objetivo estructurar la comunicación grupal permitiendo que un grupo de personas, como un todo, solucionen un problema complejo (Linstone y Turnoff, 1975).

Las principales características de esta técnica han sido descritas por Dalkey y Helmer (1969). En primer lugar, implica un proceso iterativo en donde los participantes del estudio tienen que dar su opinión en más de una ocasión. De esta manera, los participantes tienen la oportunidad de reconsiderar su opinión. Además, es un proceso anónimo, de forma que los participantes no conocen quiénes son los demás jueces que forman el panel de expertos, lo que evitará la posible inhibición de algunos participantes. Otra característica propia de esta técnica es el *feedback* controlado que recibe el panel de jueces por parte del equipo coordinador. De esta manera, los participantes reciben, entre una ronda y la siguiente, una medida estadística grupal. Disponiendo de esta información, los participantes pueden modificar su opinión de una ronda a la siguiente.

El objetivo de la aplicación del método Delphi es la consecución del consenso entre los jueces expertos en la resolución del problema presentado. En consecuencia, esta técnica ha sido ampliamente utilizada en las ciencias médicas con el objetivo de alcanzar el consenso acerca de los criterios diagnósticos de enfermedades. Sin embargo, recientemente ha sido aplicado al análisis de la validez de contenido de cuestionarios (Hagen, Stiles, Nekolaichuk y cols., 2008; Morgan, Lam-McCulloch, Herold-McIlroy y Tarshis, 2007; Petry, Maesm y Vlaskamp, 2007), dadas sus ventajas como técnica grupal.

### *El sobrepeso y la obesidad*

El rápido incremento de la prevalencia del sobrepeso y la obesidad en los países desarrollados los sitúa entre unos de los principales problemas de salud alcanzando proporciones de epidemia, como apunta la Organización Mundial de la Salud (WHO, 2000).

Actualmente, el modelo transteórico del cambio de conducta (TTM, Prochaska y DiClemente, 1983) es uno de los modelos más prometedores acerca de la adquisición de conductas saludables. El modelo sugiere que las personas que adquieren una nueva conducta se mueven a través de los estadios de cambio (DiClemente, Prochaska, Fairhurst, Velicer, Velasquez y Rossi 1991), los cuáles representan la primera dimensión del modelo. El primero de los estadios, *precontemplación*, comprende aquellas personas que no tienen intención alguna de cambiar en un futuro próximo. Muchas personas en este estadio no consideran que tengan un problema que haya que cambiar. El siguiente estadio, *contemplación*, englobaría aquellas personas que consideran que tienen un problema y que están pensando seriamente en abordarlo. Sin embargo, estas personas no tienen intención de llevar a cabo un cambio en los próximos 6 meses. En el estadio de *preparación* se combinan intención y comportamiento. Las personas de este estadio tienen intención de pasar a la acción durante el próximo mes, y puede que ya hayan intentado llevar a cabo pequeños intentos. En el siguiente estadio, de *acción*, estarían aquellas personas que ya han modificado su conducta para abordar su problema por un período inferior a 6 meses. Finalmente, en el último estadio de *mantenimiento* se sitúan las personas que han adquirido una conducta saludable durante más de 6 meses y que están intentando prevenir las recaídas. Por lo tanto, los estadios representan una dimensión temporal hacia la adquisición de conductas saludables.

La segunda dimensión del modelo son los procesos de cambio. Éstos son mediadores de los estadios de cambio y representan principios básicos del cambio de conducta (DiClemente y Prochaska, 1985). Específicamente, los procesos de cambio son actividades y experiencias abiertas o encubiertas que las personas llevan a cabo cuando intentan modificar su conducta no saludable (Prochaska,

DiClemente y Norcross, 1992). Se han definido hasta doce procesos de cambio, algunos de ellos siendo experienciales (toma de conciencia, implicaciones emocionales, reevaluación del entorno, autreevaluación y presión social) y otros comportamentales (control del reforzamiento, apoyo social, autoconfianza, contracondicionamiento, control de estímulos, abuso de sustancias y sistemas de control interpersonal). La importancia de los procesos de cambio radica en el hecho de que son potentes predictores del cambio de conducta (Prochaska, DiClemente y Norcross, 1992). Sin embargo, no disponemos de instrumentos específicos para evaluar los procesos implicados en la pérdida de peso en personas con sobrepeso u obesidad. Horwath (1999) destaca la necesidad de crear instrumentos de medida específicos para la evaluación de estos constructos, haciendo especial énfasis en la validez de su contenido.

El presente estudio tiene por objetivo validar el contenido de un cuestionario para evaluar procesos de cambio para la pérdida de peso en personas con sobrepeso y obesidad mediante un estudio Delphi.

## Método

### *Participantes*

Los participantes de este estudio son expertos en el ámbito de la obesidad (Figura 1). Se invitó a participar a los miembros del comité organizador de las diversas sedes de asociaciones internacionales relacionadas con la obesidad y la endocrinología (IASO, IOTF y ESE). Asimismo, con el objetivo de ser lo más exhaustivos posible, se invitó a participar a autores de literatura especializada tanto en el ámbito de la obesidad como en el del modelo transteórico del cambio de conducta. El panel de expertos a los que se invitó a participar en este estudio fue de 341 personas de 65 países.



Figura 1. Panel de jueces expertos en el ámbito de la obesidad

### *Instrumentos*

Se elaboró un cuestionario para evaluar procesos de cambio para la pérdida de peso, así como una breve escala para evaluar estadios de cambio en sobrepeso y obesidad.

El cuestionario de procesos de cambio se diseñó en función de los 12 procesos de cambio identificados por Prochaska y DiClemente (1985). El procedimiento para su elaboración fue el siguiente. En primer lugar, se reunieron las definiciones que diversos autores habían dado de los

procesos de cambio (Bock, Marcus, Pinto y Forsyth, 2001; Callaghan, Eves, Norman, Chang y Lung, 2002; Marcus, Rossi, Selby, Niaura y Abrams, 1992; Nigg y Courneya, 1998; Peipert y Ruggiero, 1998; Pinto, Lynn, Marcus, DePue y Goldstein, 2001; Prochaska y DiClemente, 1985; Prochaska, DiClemente y Norcross, 1992).

Puesto que el modelo transteórico ha sido aplicado a diversos problemas de salud, se adaptaron las definiciones de los procesos de cambio al caso concreto de la pérdida de peso. A partir de estas definiciones se elaboraron los ítems del cuestionario. Finalmente se incorporaron ítems de otros cuestionarios, ya validados, que miden procesos de cambio en otras problemáticas. Concretamente, se tomaron como referencia un cuestionario de procesos de cambio aplicado a la realización de ejercicio físico (Marcus, Rossi, Selby, Niaura y Abrams, 1992) y otro aplicado a tabaquismo (Prochaska, Velicer, DiClemente y Fava, 1988). De éstos, se seleccionaron los ítems que podían ser relevantes para la pérdida de peso y se adaptaron a esta problemática. Finalmente, se obtuvo un total de 107 ítems aplicados a la pérdida de peso para evaluar los 12 procesos de cambio propuestos por la literatura.

El cuestionario de estadios de cambio en sobrepeso y obesidad se elaboró de acuerdo con las definiciones de los estadios encontradas en la literatura (Prochaska, 1994; Prochaska, DiClemente y Norcross, 1992; Prochaska, Velicer, Rossi y cols., 1994). Un total de 5 ítems, correspondientes a los 5 estadios de cambio, fueron creados con el objetivo de medir la motivación para la pérdida de peso en personas con sobrepeso y obesidad.

Ambos cuestionarios fueron elaborados en castellano y en inglés, simultáneamente.

### Procedimiento

El presente estudio se llevó a cabo mediante una página web. Puesto que los cuestionarios se elaboraron en inglés y castellano, la página web fue presentada también en ambas lenguas. De esta manera, los participantes podían escoger en qué idioma preferían evaluar el contenido de los cuestionarios. En primer lugar se presentó el cuestionario de procesos de cambio, mientras que fue en la última fase del estudio cuando se presentó el cuestionario de estadios de cambio.



Figura 2. Evaluación del cuestionario mediante página web

Con el objetivo de invitar a las personas seleccionadas a participar en el estudio Delphi, se les envió un correo electrónico a todos ellos, invitándoles a visitar la página web. La tarea que tenían que realizar consistía en evaluar el contenido del cuestionario de procesos de cambio en sobrepeso y obesidad (Figura 2), a lo largo de sus 12 subescalas. Seleccionando cada una de las subescalas, era posible ver su definición y los ítems que la formaban (Figura 3). Su tarea consistía en evaluar, para cada uno de los ítems, los siguientes aspectos:

- **Representatividad.** Entendida como el grado en que el ítem representa el contenido de la subescala. Cada ítem debía ser evaluado entre 1 y 5. Por tanto, en el caso de que los jueces consideraran que el ítem era muy representativo de la definición de la subescala a la que pertenecía (en su totalidad o referente a una parte de la definición de la subescala) debían valorarlo con un 5. Por otra parte, en el caso de que el ítem no representara de ninguna manera el contenido de la subescala, debía ser valorado con un 1.
- **Claridad.** Los expertos debían valorar si el ítem, tal y como se presentaba, estaba expresado con claridad. Nuevamente, los ítems tenían que ser valorados en una escala de 1 a 5, obteniendo la menor puntuación en el caso de que el ítem fuera confuso, poco claro. Por el contrario, si consideraban que el ítem podía ser fácilmente comprendido por las personas a las que va dirigido el cuestionario, éste debía ser valorado con un 5. En el caso de que el ítem fuera valorado como poco claro, se animaba a los participantes a especificar el por qué de su valoración en el apartado de observaciones.
- **Observaciones.** Finalmente, los participantes tenían la opinión de añadir comentarios específicos acerca de alguno de los ítems o de la subescala.

1. Consciousness raising		
The items of this subscale assess whether the person realizes that he or she has a weight problem. In other words, whether he or she seeking information about the consequences of obesity and if he or she knows the benefits of healthy living habits.		
Item	Representativeness	Clarity
1. I know realize I have a weight problem	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
2. I know the benefits of a well-balanced diet	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
3. I seek information about the benefits of healthy living habits	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
4. I am aware of the benefits of making regular physical exercise a part of my life	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
5. I think I should eat food with less fat and fewer calories	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
6. A sedentary lifestyle damages health	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
7. I know that I should exercise	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
8. The benefits of a healthy lifestyle are very important	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
9. I am aware of the benefits that losing weight would have on my daily life	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
10. I seek information related to excess weight and obesity	● 1 ● 2 ● 3 ● 4 ● 5	● 1 ● 2 ● 3 ● 4 ● 5
<b>Observations:</b>		
<input type="text"/>		
<input type="button" value="Next subscale"/>		

Figura 3. Primera ronda

Una vez recogidas las opiniones de los participantes en la primera ronda del estudio, se llevó a cabo un análisis grupal de los datos. En la segunda ronda, sólo se invitó a participar a aquellos jueces expertos que ya hubieran participado en la primera ronda. En esta segunda ocasión, los participantes disponían tanto de su respuesta individual dada en la ronda anterior como de la respuesta grupal. La respuesta del grupo era proporcionada en base a una medida de centralidad (mediana) y de dispersión (rango intercuartílico). Disponiendo de esta información, se pedía a los participantes que volvieran a valorar los ítems del test (Figura 4).

Nuevamente, se integraron las opiniones de esta segunda fase, con el objetivo de que los participantes evaluaran los ítems en una tercera fase del estudio. En esta última evaluación se incluyeron también los ítems del cuestionario de estadios de cambio, acerca de los cuáles debían también evaluar su representatividad y claridad.

### *Análisis de datos*

Los análisis llevados a cabo entre una ronda y la siguiente fueron tanto cuantitativos como cualitativos.

Los análisis cuantitativos se basaron en las medianas y los rangos intercuartílicos obtenidos tanto en relación a la representatividad como claridad de los ítems, con el objetivo de eliminar

aquéllos que fueran peor valorados. Asimismo, se analizó el consenso entre los jueces de una ronda a la siguiente, siguiendo el criterio utilizado por Hagen, Stiles, Nekolaichuk y cols. (2008), definiendo un consenso aceptable siempre que por lo menos el 50% de los participantes estuvieran de acuerdo. En este caso, el consenso se calculó en base a la proporción de participantes que consideraron que el ítem era representativo o claro, y que por tanto, fue evaluado con 4 ó 5 puntos. Por tanto, el análisis del consenso se realizó tanto para representatividad y claridad, así como para cada una de las rondas.

Además, se llevó a cabo un análisis cualitativo en base a los comentarios y sugerencias de los participantes.

Mediante el análisis cuantitativo y cualitativo llevado a cabo a lo largo de las tres rondas fue posible depurar el cuestionario de forma iterativa.

		Representativeness					Clarity				
		1	2	3	4	5	1	2	3	4	5
The items of this subscale assess whether the person realizes that he or she has a weight problem. In other words, whether he or she is seeking information about the consequences of obesity and if he or she knows the benefits of healthy living habits.											
1. I now realize I have a weight problem		5	4	3	2	1	5	4	3	2	1
	Overall response of the group	3.6 ± 1.0					3.6 ± 1.0				
	Once response given to the previous stage	3					3				
2. I know the benefits of a well-balanced diet		5	4	3	2	1	5	4	3	2	1
	Overall response of the group	3.8 ± 1.0					3.8 ± 1.0				
	Once response given to the previous stage	3					3				
3. I am aware of the benefits of making regular physical exercise a part of my life		5	4	3	2	1	5	4	3	2	1
	Overall response of the group	3.6 ± 1.0					3.6 ± 1.0				
	Once response given to the previous stage	3					3				
4. I think I should eat food with less fat		5	4	3	2	1	5	4	3	2	1
	Overall response of the group	3.8 ± 1.0					3.8 ± 1.0				
	Once response given to the previous stage	3					3				
5. A sedentary lifestyle damages health		5	4	3	2	1	5	4	3	2	1
	Overall response of the group	3.6 ± 1.0					3.6 ± 1.0				
	Once response given to the previous stage	3					3				

Figura 4. Segunda ronda

## Resultados

### Primera ronda

Los participantes de la primera ronda del estudio fueron 66 expertos en el ámbito de la obesidad. Por tanto, el porcentaje de participación fue del 19.35%, considerando las 341 personas que fueron invitadas a participar en el estudio. La formación de los participantes era básicamente en medicina (60.6%), dietética y nutrición (33.3%) y psicología (15.2%). La experiencia de los participantes, tanto en la práctica clínica como en investigación, era mayoritariamente de más de 15 años, por tanto se trataba de personas con una amplia trayectoria en el ámbito de la obesidad.

Una vez reunidas las opiniones dadas en la primera ronda, se llevó a cabo una primera depuración de los ítems, eliminando algunos ítems según los siguientes criterios:

- Mediana en la medida de representatividad  $\leq 3$ , puesto que significa que el panel de jueces consideran que el ítem no corresponde al contenido de la subescala.
- Rango intercuartílico en representatividad  $\geq 2$ , lo que indica que no hay consenso acerca de la adecuación del ítem a la subescala.
- Por sugerencia de varios participantes.

Se añadieron nuevos ítems ante las sugerencias de algunos participantes de cubrir otros contenidos que debían de tenerse en consideración. Además, se hicieron los cambios de vocabulario necesarios, en el caso de que el ítem fuera poco claro (mediana en claridad  $\leq 3$  o por sugerencia de los participantes).

Se obtuvo una nueva versión del cuestionario formada por 87 ítems. El análisis del consenso se llevó a cabo para cada uno de los ítems tanto para representatividad como claridad. Asimismo, se calculó un porcentaje global de consenso, alcanzando el 72.2% en representatividad y el 78.4% en claridad.

### *Segunda ronda*

Sólo los participantes de la primera fase fueron invitados a participar. De éstos, 22 personas participaron en esta segunda ronda, lo que supuso un 33.3% de participación. La formación de los participantes fue similar a la ronda anterior, mayoritariamente en medicina, dietética y nutrición y psicología (54.4, 36.8 y 36.4%, respectivamente). Nuevamente, se trataba de participantes con una amplia trayectoria profesional, de más de 15 años.

En esta segunda ronda, los participantes debían volver a valorar los ítems en relación a su representatividad y claridad. Sin embargo, a diferencia de la primera ronda, en este caso disponían de información adicional: su respuesta dada en la ronda anterior para cada uno de los ítems, y la respuesta grupal. De esta manera, los participantes podían cambiar su opinión.

Nuevamente, se eliminaron ítems siguiendo los criterios aplicados en la primera fase, y se añadieron contenidos que los jueces expertos consideraron necesarios, obteniendo una nueva versión del cuestionario formada por 67 ítems. El consenso global en esta ronda fue del 83.5% para representatividad y de 85.1% para claridad.

### *Tercera ronda*

En esta tercera ronda, el panel de jueces evaluó únicamente los ítems que fueron añadidos durante las rondas anteriores, puesto que los demás ítems ya habían sido evaluados en dos ocasiones. Además, se añadió el breve cuestionario de estadios de cambio en sobrepeso y obesidad.

Sólo los 22 participantes de la ronda anterior fueron invitados a participar, obteniendo una participación del 100% en esta tercera ronda. Nuevamente, se eliminaron los ítems necesarios, siguiendo los criterios establecidos. Se obtuvo una versión final del test de 63 ítems, junto con un nivel de consenso global del 87.3% para representatividad y del 88.2% para claridad.

Respecto al cuestionario de estadios de cambio, se hicieron los cambios de vocabulario necesarios. El consenso global de este cuestionario se situó en el 80% para representatividad y el 80.9% para claridad.

## Discusión

El presente estudio pone de manifiesto la utilidad del método Delphi como método de análisis de la validez de contenido.

Las ventajas que el método Delphi proporciona en comparación al tradicional análisis de la validez de contenido han sido demostradas a lo largo de este estudio. En primer lugar, ha permitido la obtención de opiniones de un numeroso panel de jueces expertos, procedentes de una gran diversidad de países, y especialistas en diversas disciplinas. Además, mediante esta técnica se ha dado la posibilidad a los participantes a cambiar su opinión a lo largo de las sucesivas rondas, dando lugar a un proceso dinámico e iterativo.

Otro de los puntos fuertes de este estudio ha sido la validación de los cuestionarios en inglés y castellano al mismo tiempo. Puesto que el contenido presentado al panel de jueces era el mismo en ambos idiomas, y que las modificaciones del cuestionario se realizaron simultáneamente en ambas formas, se han obtenido los cuestionarios de procesos y estadios de cambio tanto en castellano como en inglés.

A pesar de que no es posible establecer un número óptimo de participantes para un estudio de este tipo (Galanc y Mikus, 1986), se ha demostrado que contra mayor es el número de personas que participan en el panel de expertos, más precisa es la estimación grupal (Dalkey, Brown y Cochran, 1970). En consecuencia, podemos afirmar que el número de participantes de este estudio es adecuado (n=66), en comparación con otros estudios de similares características (Callaghan y cols., 2002; Hagen, Stiles, Nekolaichuk y cols., 2008; Petry, Maes y Vlaskamp, 2007).

Puesto que el objetivo del método Delphi es conseguir solucionar un problema mediante el consenso entre jueces expertos, el análisis del consenso obtenido a lo largo de las rondas es una medida de gran importancia. En el presente estudio, el consenso ha ido en aumento de una ronda a la siguiente, alcanzando más del 87% en relación al cuestionario de procesos de cambio en la fase final del estudio. Un nivel de consenso similar es el que encontramos en relación al cuestionario de estadios de cambio, situándose en torno al 80%.

El cuestionario que ha sido validado mediante el presente estudio representa el primer cuestionario basado en el consenso de expertos en el ámbito de la obesidad. Éste cubre una necesidad en este ámbito, puesto que hasta el momento, no se disponía de ningún cuestionario específico para medir procesos y estadios de cambio en el ámbito del sobrepeso y la obesidad. De esta investigación se deriva la necesidad de aplicar estos cuestionarios a personas con sobrepeso y obesidad, con el objetivo de analizar cuáles son sus propiedades psicométricas.

## Referencias

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. American Educational Research Association, Washington, DC.; American Psychological Association, Washington, DC.; National Council on Measurement in Education, Washington, DC.
- Bock, B.C., Marcus, B.H., Pinto, B.M. y Forsyth, L.H. (2001). Maintenance of physical activity following an individualized motivationally tailored intervention. *Annals of Behavioral Medicine*, 23, 79-87.
- Callaghan, P., Eves, F.F., Norman, P., Chang, A.M. y Lung, C.Y. (2002). Applying the transtheoretical model of change to exercise in young Chinese people. *British Journal of Health Psychology*, 7, 267-282.
- Dalkey, N.C., Brown, B. y Cochran, S. (1970). The Delphi method, III: Use of self-ratings to improve group estimates. *Technology Forecast Society Change*, 1, 283-291.
- Dalkey, N.C. y Helmer, O. (1969). An experimental application of the Delphi method to the use of experts. *Management Science*, 9, 458-467.
- DiClemente, C.C. y Prochaska, J.O. (1985). Processes and stages of self-change: coping and competence in smoking behavior change. In: Shiffman, S., Wills, T., (eds.) *Coping and substance abuse*. Academic Press: San Diego, CA.

- DiClemente, C.C., Prochaska, J.O., Fairhurst, S.K., Velicer, W.F., Velasquez, M.M. y Rossi, J.S. (1991). The process of smoking cessation: An analysis of precontemplation, contemplation, and preparation stages of change, *Journal of Consulting and Clinical Psychology* 59, 295–304.
- Galanc, T. y Mikus, J. (1986). The choice of an optimum group of experts. *Technological Forecast Society*, 20, 245-250.
- Hagen, N.A., Stiles, C., Nekolaichuk, C., Biondo, P., Carlson, L.E., Fisher, K. y Fainsinger, R. (2008). The Alberta Breakthrough Pain Assessment Tool for cancer patients: a validation study using a Delphi process and patient think-aloud interviews. *Journal of Pain Symptom and management*, 35, 136-152.
- Horwath, C.C. (1999). Applying the transtheoretical model to eating behaviour change: challenges and opportunities. *Nutrition Research Reviews*, 12, 281-317.
- Linstone, H. y Turnoff, M. (1975). *The Delphi method: Techniques and applications*. Addison-Wesley Publishing: Boston, MA.
- Marcus, B., Rossi, J., Selby, V., Niaura, R. y Abrams, D. (1992). The stages and processes of exercise adoption and maintenance in a worksite sample. *Health Psychology*, 11, 386-395.
- Morgan, P.J., Lam-McCulloch, J., Herold-McIlroy, J. y Tarshis, J. (2007). Simulation performance checklist generation using the Delphi technique. *Canadian journal of Anaesthesia*, 54, 992-997.
- Nigg, C.R. y Courneya, K.S. (1998). Transtheoretical model: examining adolescent exercise behavior. *Journal of Adolescent Health*, 22, 214-224.
- Peipert, J.F. y Ruggiero, L. (1998). Use of the transtheoretical model for behavioral change in women's health. *Womens Health Issues*, 8, 304-309.
- Petry, K., Maes, B. y Vlaskamp, C. (2007). Operationalizing quality of life for people with profound multiple disabilities: a Delphi study. *Journal of Intellectual Disabilities Research*, 51, 334-339.
- Pinto, M.P., Lynn, H., Marcus, B.H., DePue, J. y Goldstein, M.G. (2001). Physician-based activity counseling: intervention effects on mediators of motivational readiness for physical activity. *Annals of Behavioral Medicine*, 23, 2-10.
- Prochaska, J.O. (1994). Strong and weak principles for progressing from precontemplation to action on the basis of twelve problem behaviors. *Health Psychology*, 13, 47-51.
- Prochaska, J.O. y DiClemente, C.C. (1983). Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51, 390-395.
- Prochaska, J.O. y DiClemente, C.C. (1985). Common processes of self-change in smoking, weight control, and psychological distress. En: Shiffman, S., Wills, T. (eds). *Coping and Substance Abuse*. Academic Press: San Diego, CA.
- Prochaska, J.O., DiClemente, C.C. y Norcross, J.C. (1992). In search how people change: applications to addictive behaviors. *American Psychologist*, 47, 1102-1114.
- Prochaska, J.O., Velicer, W.F., DiClemente, C.C. y Fava, J. (1988). Measuring processes of change: applications to the cessation of smoking. *Journal of Consulting and Clinical Psychology*, 56, 520-528.

Prochaska, J.O., Velicer, W.F., Rossi, J.S., Goldstein, M.G., Marcus, B.H., Rakowski, W., Fiore, C., Harlow, L.L., Redding, C.A., Rosenbloom, D. y Rossi, S.R. (1994). Stages of change and decisional balance for 12 problem behaviors. *Health Psychology*, *13*, 39-46.

World Health Organization, WHO (2000). Obesity: preventing and managing the global epidemic. Report of a WHO consultation, *World Health Organization Technical Report Series*, *894*, 1-253.

# CALIBRACIÓN *ONLINE* EN TEST ADAPTATIVOS INFORMATIZADOS

Andrés Antivilo, Julio Olea y Francisco J. Abad  
Universidad Autónoma de Madrid

La actualización de un banco de ítems es una actividad indispensable cuando se administra frecuentemente un Test Adaptativo Informatizado (TAI). Sin embargo, la creación y calibración de nuevos ítems suele ser costosa. Una opción que facilita dicho proceso es la utilización del diseño de Calibración Online. Con el objetivo de evaluar qué método de calibración basado en la estimación máximo verosímil marginal vía algoritmo *Expectation-Maximization (EM)* alcanza estimaciones más precisas, se realizó un estudio de simulación que compara cuatro métodos de calibración con parámetros fijos (*NWU-OEM*, *NWU-MEM*, *OWU-OEM* y *OWU-MEM*). Se manipula el formato del test (adaptativo o convencional), el tamaño muestral (500, 1000 o 2000 casos) y la ratio ítems pretest/ítems operativos (.10, .17 y .23). Los resultados indican que, para ambos tipos de test, las estimaciones desarrolladas por los métodos con múltiples ciclos EM resultan más precisas, especialmente cuando hay pocos ítems pretest y una muestra amplia.

Wainer (2000) señala como el avance de la tecnología ha facilitado un crecimiento exponencial en el uso de TAIs desde finales de los 90's. Sin embargo, esto puede resultar problemático para el banco de ítems si alguno de sus elementos se utiliza sistemáticamente. Por ello, se implementan procedimientos para garantizar que los ítems con mejores características no sean sobre-utilizados. Métodos como Sympon-Hetter, Randomesque o el Progresivo-Restringido, entre otros, controlan la tasa de exposición pero resultan insuficientes para solventar otras amenazas que debe enfrentar un TAI. Así, el uso frecuente de ítems compromete la seguridad de la evaluación (Van der Linden, 2006). Además sus propiedades pueden cambiar con el tiempo, generándose una pérdida de precisión en posteriores aplicaciones (Guo y Wang, 2003; Wainer y Mislevy, 2000). Por estas razones la actualización del banco de ítems es un proceso indispensable que se debe abordar desde la planificación del test. Periódicamente -incluso cada nueve meses (Rojas, 2001)- se deben desarrollar, calibrar y añadir nuevos ítems al banco, evitándose la presentación de reactivos obsoletos o sobreexposados.

Pese a ser necesario, no es sencillo realizar la actualización del banco. La primera dificultad se encuentra en la creación de ítems y su posterior calibración. Se debe añadir que la adquisición de un número suficiente de examinados puede ser problemática (Stocking, 1988). Adicionalmente, la estimación que se realiza a partir de la versión "papel y lápiz" del test puede no ser comparable con aquellas estimaciones calculadas a partir del TAI (Ban, Hanson, Wang, Yi y Harris, 2001). Todas estas particularidades hacen inviable una administración de los ítems inéditos mediante el formato tradicional. Un método que permite superar estos inconvenientes es la calibración asumiendo parámetros fijos (Kim, 2006), también conocida como Calibración *Online*. En ella, con cada presentación del TAI se incorporan nuevos ítems, llamados *ítems pretest*, junto a los *ítems operativos*, administrados habitualmente (Ban, Hanson, Yi y Harris, 2002). El resultado es un vector de respuestas a ítems cuyos parámetros son conocidos y algunas respuestas a nuevos ítems, cuyos parámetros deben ser estimados (Ban et al., 2002; Stocking, 1988; Wainer y Mislevy, 2000). Precisamente, el conocimiento de los parámetros de los ítems operativos permite establecer los parámetros pretest, incluso si los evaluados responden conjuntos de ítems distintos, algo propio a los TAIs (Abad y Olea, 2008). Pero, ¿Cómo se calibran los ítems pretest junto a los operativos? En la práctica, existen diversas soluciones, la mayoría basadas en la Estimación Máximo Verosímil Marginal (MMLE) via Algoritmo *Expectation-Maximization*" (EM). Como exponen Wainer y Mislevy (2000, pag. 87), los parámetros del set de ítems operativos  $\beta_{\text{antiguo}}$  son conocidos y solo queda por determinar los parámetros de los ítems pretest  $\beta_{\text{nuevo}}$ .

Si las respuestas del examinado  $j$  a ambos sets de ítems no superpuestos las denotamos como  $(x_{j,antigua}$  y  $x_{j,nueva}$ , de modo que  $x_i = (x_{antigua}, x_{nueva})$ , la expresión para la verosimilitud correcta toma la misma forma que la MMLE usual, excepto porque se necesita maximizar con respecto a  $\beta_{nuevo}$ . Formalmente:

$$L(\beta_{nuevo} | X, \beta_{antiguo}) = \prod_j \int p(x_j | \theta, \beta) p(\theta) d\theta \tag{1}$$

La ecuación (1) puede resolverse usando la técnica de estimación algoritmo EM. En el procedimiento pueden desarrollarse variaciones que originan hasta cinco métodos para ejecutarlo, lo que depende del el número de veces que se actualiza la distribución de habilidad a priori (ninguna, una o múltiples veces) y el número de ciclos EM incluidos (uno o varios).

Tabla 1. Diseños de calibración online

Número de actualizaciones de la distribución inicial ( <i>prior weight updating</i> )	Número de Ciclos EM	
	Uno (OEM)	Varios (MEM)
Ninguna	NWU – OEM	NWU – MEM
Una	OWU – OEM	OWU – MEM
Varias	No es posible	MWU – MEM

Para realizar las estimaciones del nivel de habilidad  $\theta$ , el método *No prior weight updating – One EM cycle* (NWU-OEM) utiliza solo un ciclo EM. El método *No prior weight updating – Multiple EM cycles* (NWU-MEM) es una extensión del anterior, ya que introduce múltiples ciclos EM para desarrollar la estimación. En *One prior weight updating – One EM cycle* (OWU-OEM) se actualiza la distribución inicial para la distribución de habilidad a priori en un ciclo EM. En *One prior weight updating – Multiple EM cycles* (OWU-MEM) se estiman los parámetros pretest a través de múltiples ciclos EM, pero solo con una actualización de la distribución inicial. Finalmente, en *Multiple prior weight updating – Multiple EM cycles* (MWU-MEM) se realizan múltiples actualizaciones de la distribución inicial y múltiples ciclos EM. La Tabla 1 (tomada de Kim, 2006) ofrece una clasificación de los procedimientos reseñados. Una descripción más técnica y detallada puede consultarse en Kim (2006) y Wainer y Mislevy (2000).

*Objetivo*

Determinar qué diseño de calibración *online* basado en la estimación MMLE vía algoritmo EM (*NWU-OEM, NWU-MEM, OWU-OEM y OWU-MEM*) realiza las estimaciones más precisas de ítems pretest ajustados al modelo 3PL, tanto para un TAI como para un Test Convencional (TC).

Método

*Participantes*

Se simularon tres grupos de distinto tamaño: 500, 1000 y 2000 casos. El valor de  $\theta$  que se asume como verdadero para cada sujeto se asignó a partir de una distribución normal que fluctúa entre -4 y 4, a intervalos de 0.5.

### *Instrumentos*

Se simuló la administración de un TAI y un TC. Los ítems operativos del TAI fueron extraídos del banco de 197 ítems del test *eCAT*, todos calibrados según el modelo 3PL (Abad y Olea, 2008). Como criterio de parada se utilizó una extensión fija de 30 ítems, seleccionados mediante máxima información. Por otra parte, los ítems operativos que conforman el TC (donde todos los evaluados responden a las mismas preguntas) fueron simulados a partir de distribuciones estadísticas conocidas:  $a_j$  presenta una distribución uniforme (0.4, 1.4);  $b_j$  una distribución normal (0,1) y  $c_j$  una distribución uniforme (0.05, 0.25). En ambos contextos los examinados respondieron una determinada cantidad de ítems pretest. En la práctica, los tests podían incluir tres, cinco o siete ítems pretest, lo que generó una *ratio ítems pretest / ítems operativos* de .10, .17 y .23. En estos ítems se establecieron valores constantes para los parámetros  $a_j$  (1.20) y  $c_j$  (0.20), mientras que para  $b_j$  se asignaron valores equidistantes en 0.5 puntos, con media cero, de modo que los valores para la condición de 3 ítems fueron -0.5, 0.0 y 0.5; con 5 ítems, fueron -1.0, -0.5, 0.0, 0.5 y 1.0; y para la última condición variaron entre -1.5 y 1.5.

### *Diseño de Calibración Online*

Se simuló la administración de un TAI y de un TC a una muestra (500, 1000 o 2000 examinados) cuyo  $\theta$  se distribuye normalmente. Ambos instrumentos contenían 30 ítems más una cantidad específica de ítems pretest (tres, cinco o siete). El resultado de cruzar los factores mencionados (tipo de test, muestra y cantidad de ítems pretest) fueron 18 condiciones, en las que se aplicaron los métodos de calibración *NWU-OEM*, *NWU-MEM*, *OWU-OEM* y *OWU-MEM*. Cada condición fue replicada diez veces, lo que permitió evaluar la capacidad de los métodos para recuperar adecuadamente los parámetros pretest originales. Los métodos fueron programados usando las subrutinas que provee el programa ICL (Hanson, 2002).

### *Medida de Precisión*

Para cada parámetro se calculó la raíz cuadrada del error cuadrático medio (RMSE).

## Resultados

Las Tablas 2 y 3 resumen la recuperación de los parámetros considerando el tipo de test simulado, la *ratio ítems pretest/ítems operativos*, el tamaño muestral y el método de calibración. En ambas tablas los valores medio de RMSE para los parámetros  $a_j$ ,  $b_j$  y  $c_j$  se encuentran bajo las columnas RMSE-A, RMSE-B y RMSE-C, respectivamente.

La Tabla 2 revela un resultado esperable para el TAI en función de investigaciones previas (Ban et al., 2001; Ban et al., 2002; Kim, 2006): los métodos que incluyen múltiples ciclos EM logran una estimación más precisa de los parámetros  $a_j$  y  $b_j$  que los métodos que tienen solo un ciclo, lo que se refleja en un RMSE más pequeño. Este contraste ya no resulta tan evidente para RMSE-C. Pese a ello, desde una perspectiva general, es posible afirmar que los métodos OEM aparecen con un rendimiento inferior a los métodos MEM. Se debe resaltar que en varios casos (en especial, para  $a_j$ ) los métodos OEM con *ratio ítem pretest/ítem operativo* de .10 resultan más imprecisos que los métodos MEM, incluso cuando en ellos la *ratio* es de .23. Se observa también una tendencia en la condición de tamaño muestral 2000 a alcanzar valores más pequeños para RMSE-A y RMSE-B que con 500 casos, sin importar con qué método se desarrolle la estimación. Eso sí, este decremento es escaso para los métodos OEM, siendo pronunciado solo para los métodos MEM. Si se usa como criterio de

precisión el RMSE-A, las estimaciones más adecuadas aparecen ligadas a NWU-MEM, cuyos valores son ligeramente más pequeños que OWU-MEM. Algo más compleja es la descripción para  $b_j$ . Si bien es indiscutible que ambos métodos MEM recuperan de mejor forma los parámetros, ninguno de ellos presenta sistemáticamente los mejores resultados. Finalmente, es en  $c_j$  donde resulta más difícil sugerir un método concreto, ya que en todas las condiciones la mayoría de las estimaciones podrían considerarse satisfactorias.

Para el TC (ver Tabla 3) se verifica que -en casi todos los casos- NWU-MEM y OWU-MEM entregan mejores estimaciones. Destaca la similitud que alcanzan sus resultados en las distintas condiciones, siendo frecuente una discrepancia entre NWU-MEM y OWU-MEM inferior a .002. El nivel de precisión que presentan los métodos MEM en detrimento de los OEM se aprecia de manera contundente para los parámetros  $a_j$  y  $c_j$ , no importando la ratio ítems pretest/ítems operativos. En este sentido, el utilizar como medida de precisión el RMSE-A o el RMSE-C, permite señalar que el mejor método para desarrollar la calibración es NWU-MEM, ya que logra estimaciones levemente superiores a OWU-MEM (variaciones que podrían ser calificadas de triviales).

La excepción se encuentra en la estimación de  $b_j$ , ya que -inesperadamente- hay casos en que los métodos OEM ostentan niveles de precisión equivalentes a los entregados por los métodos MEM. Finalmente, en  $a_j$  se observan errores de estimación severos en los métodos OEM, con RMSE medios superiores a .30, situación inédita en el TAI. Tampoco se observa con nitidez una disminución del RMSE a medida que aumenta el tamaño muestral.

Tabla 2. Valores del RMSE medio de cada parámetro en el Test Adaptativo Informatizado

MÉTODO	RMSE-A			RMSE-B			RMSE-C		
	Número de ítems pretest (ratio)			Número de ítems pretest (ratio)			Número de ítems pretest (ratio)		
	3 (.10)	5 (.17)	7 (.23)	3 (.10)	5 (.17)	7 (.23)	3 (.10)	5 (.17)	7 (.23)
<i>N = 500</i>									
NWU-OEM	.222	.219	.282	.098	.097	.144	.017	.022	.031
NWU-MEM	.051	.070	.151	.096	.137	.090	.024	.021	.026
OWU-OEM	.214	.251	.353	.109	.103	.139	.031	.027	.021
OWU-MEM	.060	.093	.156	.099	.073	.093	.032	.026	.021
<i>N = 1000</i>									
NWU-OEM	.226	.220	.243	.069	.099	.144	.011	.036	.043
NWU-MEM	.088	.087	.117	.074	.096	.098	.018	.046	.033
OWU-OEM	.233	.214	.275	.078	.107	.173	.021	.049	.030
OWU-MEM	.047	.100	.123	.072	.079	.132	.021	.037	.034
<i>N = 2000</i>									
NWU-OEM	.223	.267	.259	.032	.073	.134	.028	.024	.025
NWU-MEM	.035	.049	.074	.026	.044	.071	.015	.015	.018
OWU-OEM	.230	.261	.263	.028	.083	.134	.025	.020	.026
OWU-MEM	.028	.051	.086	.031	.053	.072	.013	.016	.019

Tabla 3. Valores del RMSE medio de cada parámetro en el Test Convencional

MÉTODO	RMSE-A			RMSE-B			RMSE-C		
	Número de ítems pretest (ratio)			Número de ítems pretest (ratio)			Número de ítems pretest (ratio)		
	3 (.10)	5 (.17)	7 (.23)	3 (.10)	5 (.17)	7 (.23)	3 (.10)	5 (.17)	7 (.23)
<i>N = 500</i>									
NWU-OEM	.258	.270	.328	.093	.109	.165	.038	.050	.042
NWU-MEM	.120	.172	.173	.115	.124	.130	.023	.040	.035
OWU-OEM	.274	.281	.338	.112	.114	.165	.035	.050	.042
OWU-MEM	.118	.174	.176	.118	.114	.136	.021	.040	.035
<i>N = 1000</i>									
NWU-OEM	.229	.235	.242	.104	.127	.132	.047	.037	.028
NWU-MEM	.107	.178	.180	.091	.077	.149	.042	.027	.016
OWU-OEM	.227	.233	.242	.103	.090	.132	.047	.026	.028
OWU-MEM	.107	.179	.183	.095	.079	.153	.043	.026	.017
<i>N = 2000</i>									
NWU-OEM	.225	.246	.255	.130	.124	.180	.042	.021	.027
NWU-MEM	.066	.139	.130	.082	.088	.156	.029	.023	.016
OWU-OEM	.237	.260	.255	.153	.142	.180	.040	.020	.027
OWU-MEM	.076	.140	.133	.096	.073	.156	.029	.026	.017

### Conclusión

El objetivo era evaluar qué método de calibración *online* basado en MMLE vía algoritmo *EM* recupera de forma más precisa los parámetros originales de los ítems pretest. Se concluye que –independiente del tipo de test- es necesario incluir métodos con múltiples ciclos EM en el proceso de estimación para lograr resultados satisfactorios. Asimismo, los valores RMSE en el TAI disminuyen a medida que aumenta el tamaño de la muestra, pero aumentan cuando se adicionan ítems pretest con parámetros distantes de la media (por ejemplo, cuando  $b_i = 1.5$  o  $-1.5$ ). En el TC el patrón de descenso en el valor de los parámetros relacionado al tamaño del grupo no se reproduce de forma tan nítida como en el TAI. Aun más, en varios casos no se detecta descenso alguno; incluso, se llega a observar un ligero aumento, independiente del tamaño seleccionado. Esto puede ocurrir debido al número limitado de réplicas realizadas, que no permitirían alcanzar la estabilidad en la estimación de parámetros lograda para el TAI. A su vez la ratio ítems pretest/ítems operativos es un factor crítico en la recuperación de los parámetros cuando se administra el TC, dado que la estimación se vuelve más imprecisa cuando aumenta la ratio. En este sentido, pese a que un resultado similar ya fue reportado por Kim (2006), no deja de llamar la atención el deterioro que se produce en la estimación cuando se incluyen ítems pretest con valores distantes a la media, pues es en estos ítems donde las estimaciones alcanzan su máximo nivel de error. Los resultados informados deben ser interpretados con cautela, ya que el número de réplicas utilizado es limitado; queda por evaluar qué tendencia mostrarían las estimaciones de los ítems pretest del test convencional si se repite el proceso reiteradamente. Finalmente, futuros estudios sobre diseños de calibración *online* deberían analizar el posible efecto de incluir ítems pretest con parámetros extremos o ítems desajustados al modelo 3PL.

## Referencias

- Abad, F. y Olea, J. (2008). *Actualización de los parámetros de los ítems de eCAT y propuesta de ampliación del banco*. Universidad Autónoma de Madrid.
- Ban, J.C., Hanson, B.H., Wang, T., Yi, Q. y Harris, D.J. (2001). A comparative study of on-line pretest item calibration-scaling methods in computerized adaptive testing. *Journal of educational measurement*, 38, 191-212.
- Ban, J., Hanson, B., Yi, Q. y Harris, D. (2002). Data Sparseness and on-line pretest item calibration-scaling methods in CAT. *Journal of educational measurement*, 39, 207 - 218.
- Guo, F. y Wang, L. (2003). *Online calibration and scale stability of a CAT program*. NCME. Chicago: IL.
- Hanson, B.A. (2002). *IRT command language*. Monterey, CA: Author.
- Kim, S. (2006). A comparative study of IRT fixed parameter calibration methods. *Journal of educational measurement*, 43, 355-381.
- Rojas, A. (2001). Pasado, presente y futuro de los test adaptativos informatizados: entrevista con Isaac I. Bejar. *Psicothema*, 13, 685-690.
- Stocking, M.L. (1988). *Scale drift in on-line calibration* (Educational Testing Service research Rep. No 88-28). Princeton, NJ: ETS.
- Van der Linden, W. (2006). A formal characterization of and some alternatives to Sympon-Hetter item-exposure control in computerized adaptive testing. *Law school admission council computerized testing report 02-05*.
- Wainer, H. (2000). *CATs: Whither and whence*. (Educational Testing Service research Rep. No 00-12). Princeton, NJ: ETS.
- Wainer, H. y Mislevy, R.J. (2000). Item response theory, item calibration, and proficiency estimation. In Wainer, H. (Ed.). *Computer adaptive testing: A primer* (Capítulo 4, pg. 61-99). Hillsdale, NJ: Lawrence Erlbaum.

# ELABORACIÓN DE UNA ESCALA DE LIDERAZGO EN DEPORTES DE EQUIPO

Constantino Arce, Julio Torrado, Elena Andrade, Javier Garrido, Cristina de Francisco e Iria Arce  
Universidad de Santiago de Compostela

Mayoritariamente, en los contextos deportivos se ha tomado como referencia de liderazgo la figura formal del entrenador. El presente estudio, sin embargo, se ha focalizado en un liderazgo más informal que algunos jugadores ejercen en los equipos deportivos. Las investigaciones precedentes presentan ciertas limitaciones en cuanto a la metodología utilizada. Mientras unos autores utilizaron escalas elaboradas originalmente para otro contexto o tipo de liderazgo, otros trataron de desarrollar instrumentos específicos, aunque sin ofrecer datos sobre análisis realizados para su construcción. Además, las investigaciones se realizaron con muestras muy específicas en el tipo de deporte, género y rango de edades. Con el fin de superar estas limitaciones se elaboró un listado de 54 ítems que definen conductas que se atribuyen a líderes de equipos deportivos, relativos a 5 hipotéticas dimensiones conceptuales: *Empatía, Apoyo social, Influencia en la toma de decisiones, Valores y Orientación a la tarea*. Este listado se administró a una amplia muestra de jugadores de fútbol federado. Se realizaron análisis factoriales exploratorios, se evaluó la consistencia interna de los factores y se analizó la fiabilidad de los ítems, con resultados que avalan las hipótesis planteadas previamente. El modelo se sometió a análisis factorial confirmatorio, obteniéndose índices de ajuste satisfactorios.

## *Objetivo*

Desarrollar una escala para la evaluación del liderazgo ejercido por deportistas entre sus compañeros de equipo.

## *Antecedentes*

El modelo teórico tradicional sugiere que el liderazgo que ejercen algunos deportistas entre sus compañeros de equipo se define en base a dos grandes dimensiones: a) orientación social, referida a la atención a las relaciones entre los compañeros, la cohesión y el clima grupal y b) orientación hacia la tarea, definida por la actitud que el líder toma ante el deporte, los entrenamientos y la competición (Kim, 1992; Rees, 1983; Rees y Segal, 1984). Varios estudios han tratado de profundizar en cada una de las dos dimensiones, si bien en el aspecto más social es el que ha recogido mayor atención por parte de los investigadores (Moran y Weiss, 2006; Arce, Torrado, Andrade, Garrido y de Francisco, 2008; Weiss y Stuntz, 2004; Yukelson, Weinberg, Richardson y Jackson, 1983).

Investigaciones previas que tuvieron entre sus objetivos evaluar este liderazgo presentan ciertas limitaciones metodológicas. Loughead y Hardy (2005) realizaron un estudio comparativo del liderazgo de los entrenadores y de los deportistas utilizando el LSS (Leadership Scale for Sport; Chelladurai y Saleh, 1980), diseñado originalmente para la evaluación de entrenadores. Glenn y Horn (1993) elaboraron el SLBI (Sport Leadership Behavior Inventory) en un estudio con jugadoras de fútbol en edad escolar. Las propias autoras señalan en la publicación del trabajo la especificidad del inventario y la necesidad de profundizar con otras muestras y en otros contextos. Moran y Weiss (2006) replicaron este estudio añadiendo hombres a la muestra, aunque manteniendo el mismo deporte y el mismo rango de edad (14-18 años). Kozub y Pease (2001) se asentaron sobre el modelo teórico predominante y

elaboraron el PLS (Player Leadership Scale) para un estudio sobre la influencia del liderazgo del entrenador sobre la aparición de un jugador líder en equipos de baloncesto de institutos pero no ofrecen datos sobre las garantías psicométricas de la escala.

## Método

### *Participantes*

Formaron parte de la muestra 407 sujetos, pertenecientes a equipos de fútbol federados de la Comunidad Autónoma de Galicia. La media de edad de los participantes fue de 23,9, con una desviación típica de 4,4. El rango de edad fue de 16-38 años.

### *Instrumento de medida. Propuesta inicial.*

A partir de un modelo teórico de dos grandes dimensiones y cinco subfactores, elaborado en base a la investigación existente y a aportaciones y estudios exploratorios previos (Arce et al., 2008; Torrado, 2007), se diseñó una escala inicial de 54 ítems agrupados en cinco factores: Apoyo social, Influencia en la toma de decisiones, Valores deportivos, Empatía (referidos a la dimensión Orientación Social) y Orientación a la tarea (referido a la dimensión del mismo nombre).

La escala de respuesta para la escala inicial constaba de cinco alternativas: “casi nunca”, “pocas veces”, “algunas veces”, “a menudo” y “casi siempre”. Para la ordenación de los ítems se utilizó un método de rotación consistente en alternar un ítem de cada dimensión.

### *Procedimiento de recogida de datos*

Se aplicó a los jugadores la propuesta inicial de escala siguiendo un procedimiento estandarizado. Se contactó con los entrenadores de los equipos para acudir a un entrenamiento y, siempre previamente al inicio de la sesión, se explicaba a los jugadores el objetivo de la investigación y se administraba la escala de forma colectiva. A modo de instrucciones se indicaba a los jugadores que pensarán en los compañeros que habían tenido hasta el momento en equipos de categoría senior y escogieran, de entre todos ellos, al que creyeran que mejor había ejercido como líder. Una vez elegido debían evaluarlo utilizando los ítems de la escala.

## Resultados

### *Análisis Factorial Exploratorio*

Se realizó Análisis Factorial Exploratorio, mediante el programa estadístico SPSS 13.0, con método de extracción de ejes principales y método de rotación oblicuo. Combinando criterios sustantivos (interpretación teórica de las dimensiones), estadísticos (basados en análisis de ítems), de consistencia interna de cada factor y de aplicabilidad, se optó por una solución de 30 ítems agrupados en cinco factores (seis ítems en cada uno) que se corresponden con las cinco dimensiones de primer orden hipotetizadas inicialmente. La varianza explicada, siempre previa a la rotación, fue de 47,187%. En la Tabla 1 se ofrece la matriz estructural y el valor de Alpha de Cronbach para cada factor.

Tabla 1. Matriz estructural y Alpha de Cronbach para cada factor

		Factor				
		1	2	3	4	5
EMPATÍA	Se identifica con los sentimientos de los compañeros	<b>,748</b>	,083	-,227	-,230	,346
	Siente lo que le pasa a los compañeros	<b>,731</b>	,005	-,289	-,304	,397
	Comprende los problemas de los compañeros	<b>,723</b>	,074	-,274	-,335	,358
	Comprende el sentir de los compañeros	<b>,709</b>	,116	-,187	-,352	,331
	Percibe los problemas de los compañeros	<b>,684</b>	,014	-,279	-,146	,344
	Comparte los sentimientos de los compañeros	<b>,631</b>	,005	-,162	-,244	,236
	<i>Autovalor: 6,264 Varianza explicada del factor: 20,881%</i>		<i>Alpha de Cronbach: 0,849</i>			
ORIENTACIÓN A LA TAREA	Lo da todo en los partidos	,032	<b>,767</b>	-,167	-,198	,130
	Se esfuerza en los partidos	-,047	<b>,724</b>	-,167	-,212	,103
	Se emplea a fondo en los partidos	-,039	<b>,707</b>	-,051	-,222	,092
	Entrena duro	,059	<b>,684</b>	-,205	-,136	-,003
	Se prepara a conciencia para competir	,111	<b>,669</b>	-,235	-,173	,047
	Se toma los entrenamientos en serio	,084	<b>,630</b>	-,182	-,205	,041
	<i>Autovalor: 3,029 Varianza explicada del factor: 10,097%</i>		<i>Alpha de Cronbach: 0,842</i>			
INFLUENCIA EN LA TDD	Su punto de vista es importante a la hora de tomar decisiones entre los compañeros	,271	,160	<b>-,766</b>	-,224	,105
	Los compañeros tienen en cuenta su opinión	,318	,145	<b>-,732</b>	-,248	,251
	Su opinión ejerce influencia sobre los compañeros	,106	,202	<b>-,714</b>	-,123	,099
	Tiene mucho peso en las decisiones entre compañeros	,153	,150	<b>-,705</b>	-,121	,026
	Su opinión es respetada por los compañeros	,192	,223	<b>-,680</b>	-,214	,114
	Su punto de vista convence a sus compañeros	,282	,132	<b>-,630</b>	-,180	,219
	<i>Autovalor: 2,124 Varianza explicada del factor: 7,080%</i>		<i>Alpha de Cronbach: 0,847</i>			
VALORES DEPORTIVOS	Se comporta de forma honesta	,226	,140	-,118	<b>-,709</b>	,183
	Actúa de manera razonable	,208	,152	-,215	<b>-,709</b>	,144
	Muestra un comportamiento maduro	,231	,232	-,233	<b>-,664</b>	,101
	Hace las cosas con honradez	,415	,169	-,213	<b>-,649</b>	,287
	Actúa de manera reflexiva	,411	,116	-,264	<b>-,558</b>	,089
	Se comporta de manera responsable	,117	,254	-,101	<b>-,509</b>	,171
	<i>Autovalor: 1,587 Varianza explicada del factor: 5,290%</i>		<i>Alpha de Cronbach: 0,792</i>			
APOYO SOCIAL	Da señales de satisfacción cuando los compañeros lo hacen bien	,337	,065	-,196	-,215	<b>,706</b>
	Aprovecha los buenos momentos para dar confianza a los compañeros	,385	,062	-,191	-,172	<b>,669</b>
	Felicita a los compañeros cuando lo hacen bien	,207	,024	-,029	-,118	<b>,636</b>
	Muestra entusiasmo cuando se están haciendo las cosas bien	,227	,125	-,161	-,172	<b>,630</b>
	Defiende a los compañeros en situaciones difíciles	,408	,077	-,176	-,221	<b>,530</b>
	Apoya a los compañeros en los momentos malos	,374	,020	-,110	-,091	<b>,472</b>
	<i>Autovalor: 1,155 Varianza explicada del factor: 3,849%</i>		<i>Alpha de Cronbach: 0,776</i>			

### Análisis Factorial Confirmatorio

La solución obtenida se sometió a análisis factorial confirmatorio, mediante el programa LISREL 8.72 (versión 2005, basado en Jöreskog y Sörbom, 1993). Se especificaron las relaciones entre cada ítem y el factor en el que mayor carga presentaba y se permitieron las relaciones entre factores. El método de estimación empleado fue el de máxima verosimilitud. Para la evaluación del ajuste global del modelo se siguieron las indicaciones de Schermelleh-Engel, Moosbrugger y Müller (2003). Los índices de ajuste obtenidos se pueden observar en la Tabla 2.

Tabla 2. Índices de ajuste del AFC

Índice	Valor	Evaluación del ajuste
Cociente: $\chi^2 / \text{gl}$	1,96	Bueno
Root Mean Square Error of Aproximation (RMSEA)	0,052	Aceptable
Root Mean Residual (SRMR)	0,056	Aceptable
Non normed fit index	0,95	Aceptable
Comparative fit index	0,96	Aceptable

En la Tabla 3 se recogen las correlaciones entre factores ofrecidas por el análisis factorial confirmatorio.

Tabla 3. Correlaciones entre factores

Factor	EMP	OTA	ITD	VD	AS
EMP	1				
OTA	0,08	1			
ITD	0,36	0,25	1		
VD	0,48	0,28	0,33	1	
AS	0,58	0,12	0,27	0,33	1

Si bien el patrón de las correlaciones estimadas probablemente no nos conduciría a un modelo mejor ajustado, hemos querido probar si estas tendencias sugieren que a estas cinco dimensiones puedan subyacer dos de segundo orden. Como se puede comprobar en la Tabla 4, un análisis factorial exploratorio de segundo orden nos indica que los factores Empatía, Influencia en la toma de decisiones, Valores deportivos y Apoyo social se agrupan en una dimensión superior, que podemos denominar Orientación Social, mientras que el factor Orientación a la tarea compondría un segundo factor de orden superior con la misma denominación de Orientación a la Tarea, avalando así la hipótesis previamente establecida.

Tabla 4. Matriz estructural del AFE de 2º orden

	Factor	
	1	2
Empatía	,693	,431
Valores Deportivos	-,640	,207
Influencia en la TDD	-,605	,207
Apoyo Social	,590	,455
Orientación a la Tarea	,409	-,757

### Conclusión

Se ha elaborado una escala de evaluación del liderazgo de equipos deportivos con el objetivo de superar las limitaciones de las investigaciones previas a nivel metodológico. La construcción de la escala ha constado de un análisis factorial exploratorio, que ofreció una solución de 5 factores de primer orden (*Apoyo Social, Orientación a la tarea, Influencia en la toma de decisiones, Responsabilidad y Empatía*) y dos grandes dimensiones de segundo orden (Orientación a la Tarea, que recoge el subfactor del mismo nombre, y Orientación Social, que agrupa a todos los demás subfactores). Los valores obtenidos del análisis de consistencia interna de los factores, así como de los índices de ajuste del modelo ofrecidos

por el análisis factorial confirmatorio resultaron satisfactorios. La construcción de esta escala supone un avance con respecto a los trabajos precedentes con las siguientes aportaciones:

- a) Elaboración de una herramienta específica, a partir de un trabajo estadístico, para la evaluación del liderazgo que ejercen algunos deportistas entre sus compañeros de equipo.
- b) Contribución al desarrollo del marco teórico de esta línea de investigación, avalando el modelo tradicional de liderazgo.

La investigación futura se debiera orientar a mejorar las propiedades psicométricas de la escala, así como investigar sobre su validez comprobando la capacidad de contribuir a la detección y evaluación de líderes de equipos. Es necesario, además, profundizar en la dimensión de orientación a la tarea para estudiar una posible estructura subfactorial. También se abre la posibilidad de estudiar posibles diferencias de género en el liderazgo deportivo y la necesidad de generalizar la estructura factorial encontrada en el presente estudio a otras muchas modalidades deportivas.

Agradecimientos. La presente investigación ha sido realizada con la ayuda de un proyecto de investigación financiado por la Consellería de Economía e Industria de la Xunta de Galicia (PGIDIT06PXIB211187PR).

#### Referencias

- Arce, C., Torrado, J., Andrade, E., Garrido, J. y de Francisco, C. (2008). Elaboración de una escala para la medida de la capacidad de liderazgo de los deportistas de equipo. *Psicothema*, 20(4), 913-917.
- Chelladurai, P. y Saleh, S.D. (1980). Dimensions of leader behavior in sports: development of a leadership scale. *Journal of Sport Psychology* 2, 34-45.
- Glenn, S.D. y Horn, T.S. (1993). Psychological and personal predictors of leadership behavior in female soccer athletes. *Journal of Applied Sport Psychology*, 5, 17-34.
- Jöreskog, K.G. y Sörbom, D. (1993). *Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software.
- Kozub, S.A. y Pease, D.G. (2001). Coach and placer leadership in high school basketball. *Journal of Sport Pedagogy: Teaching and Coaching in Sport*, 7,1-15.
- Loughead, T.M. y Hardy, J. (2005). An examination of coach and peer leader behaviors in sport. *Psychology of Sport and Exercise*, 6(3), 303-312.
- Moran, M.M. y Weiss, M. R. (2006). Peer leadership in sport: Links with friendship, peer acceptance, psychological characteristics and athletic ability. *Journal of Applied Sport Psychology*, 18, 97-113.
- Rees, C. R. (1983). Instrumental and expressive leadership in team sports: A test of leadership role differentiation theory. *Journal of Sport Behavior* 6, 17-27.
- Rees, C. R. y Segal, M. W. (1984). Role differentiation in groups: The relationship between instrumental and expressive leadership. *Small Group Behavior*, 15, 109-123.

- Schermelleh-Engel, K., Moosbrugger, H. y Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online*, 8(2), 23-74.
- Torrado, J. (2007). Desarrollo de una escala de liderazgo deportivo. *Trabajo de Investigación Tutelado no publicado*. Facultad de Psicología. Universidad de Santiago de Compostela.
- Yukelson, D., Weinberg, R., Richardson, P. y Jackson, A. (1983). Interpersonal attraction and leadership within collegiate sport teams. *Journal of Sport Behavior*, 6, 28-36.

# ELABORACIÓN DE UN CUESTIONARIO DE ACCESIBILIDAD DE LOS INMIGRANTES AL SISTEMA SANITARIO

Isabel Benítez<sup>1</sup>, Juana Gómez-Benito<sup>2</sup>, José Luis Padilla<sup>1</sup>, Palmira Tejero<sup>3</sup>  
y Montserrat Martín-Baranera<sup>3</sup>

<sup>1</sup>Universidad de Granada

<sup>2</sup>Universidad de Barcelona

<sup>3</sup>Hospital General de l'Hospitalet

El crecimiento progresivo de los flujos migratorios plantea importantes retos al sistema sanitario que debe adaptarse a las nuevas demandas. En este contexto, las dificultades más relevantes en la prestación de servicios proceden de las diferencias entre la población inmigrante y la población autóctona en aspectos relacionados principalmente con la accesibilidad al sistema sanitario, lo que manifiesta la necesidad de conocer las características de esta población. En este trabajo se presenta el proceso de elaboración de un cuestionario para evaluar la accesibilidad de los inmigrantes al sistema sanitario. En este proceso se implementaron, además de los procedimientos habituales, otros métodos para optimizar la calidad de las preguntas, como fueron las entrevistas a profesionales y los grupos focales durante la fase de elaboración del instrumento, y métodos cognitivos como las entrevistas cognitivas y la codificación del comportamiento durante la fase de pretest del cuestionario. El objetivo de este estudio es mostrar la utilidad de las entrevistas cognitivas y la codificación del comportamiento como “asesores” del proceso de elaboración de cuestionarios durante la fase de pretest. Los datos obtenidos mediante la codificación del comportamiento identificaron como problemáticas las preguntas 4, 11, 12 y 14, encontrando además un peor ajuste de los participantes procedentes de China y Cuba. Las entrevistas cognitivas ofrecieron información sobre las dificultades de los participantes para responder a algunas de las preguntas del cuestionario. Los resultados muestran la complementariedad de ambos métodos así como su utilidad para proporcionar información cualitativa sobre la causa de las dificultades que los participantes pueden tener para proporcionar una respuesta adecuada.

Las migraciones constituyen un fenómeno de gran relevancia por las implicaciones demográficas, económicas y socioculturales que supone tanto para el país emisor como para el receptor. España ha sido principalmente país emisor durante los primeros tres cuartos del S.XX y país receptor durante el último cuarto, estando su población compuesta según el censo elaborado por el Instituto Nacional de Estadística (INE) en 2008 por un 11.3% de residentes de nacionalidad extranjera, lo que supone 5.220.660 habitantes censados.

Este fenómeno implica la organización de la sociedad de forma que pueda adaptarse a una población diversa. Por ello, una de las preocupaciones más importantes que supone la inmigración es la capacidad del sistema sanitario para hacer frente a la llegada de los inmigrantes y a sus necesidades. Dar respuesta a las demandas de la población supone conocer sus características y sus experiencias y el uso que los inmigrantes hacen de los servicios sanitarios, es decir, supone realizar una evaluación.

Tradicionalmente, las evaluaciones realizadas en este contexto se han centrado en la calidad y la eficacia del sistema sanitario medidas a través de la opinión (Gasquet, Villemint, Estaquio, Durieux, Ravaud y Falissard, 2004), o la satisfacción de los usuarios, sin tener en cuenta experiencias concretas que reflejen aspectos como la información, la coordinación o el apoyo recibido, fundamentales, como indican Coulter y Cleary (2001), para evaluar la calidad. También se han realizado estudios sobre la accesibilidad al sistema sanitario pero centrados en servicios específicos como puede ser el servicio de urgencias y dirigidos a la población general (Nasra, Makhdoom y Jaafar, 1996).

La inexistencia de instrumentos centrados en la población inmigrante y que abarquen la evaluación de las experiencias de los usuarios además de aspectos como la calidad y eficacia del sistema, manifestó la necesidad de elaborar un instrumento destinado a tal fin.

Tradicionalmente el proceso de elaboración de cuestionarios ha estado guiado por los siguientes pasos (Padilla, González y Pérez, 1998):

1. Determinación de los objetivos.
2. Planificación.
3. Elaboración de las preguntas.
4. Análisis de la calidad de las preguntas.
5. Disposición provisional de las preguntas en el cuadernillo.
6. Análisis de la fiabilidad y validez.
7. Edición final.

Sin embargo, este modelo ha recibido críticas con respecto a: las dificultades en la identificación de dimensiones e indicadores y al uso exclusivo de métodos cuantitativos para evaluar la versión piloto de los cuestionarios (Willis, 2005). En el caso de la primera crítica, el modelo anterior no recoge un paso específico dedicado a la identificación de dimensiones e indicadores, aunque se asume incluido en la fase de planificación. En otros modelos la atención se centra en las características de las preguntas y en el diseño del cuestionario (Casas, Repullo y Donado, 2003) prestando menos atención a las dimensiones e indicadores. En relación a la segunda crítica, la mayoría de los estudios que implican la elaboración de un instrumento aplican estudios piloto cuantitativos, basándose exclusivamente en las características psicométricas de la escala (Traver y García-López, 2007) o en las respuestas numéricas dadas por una muestra (Fogliatto y Pérez, 1990).

Para solucionar estas dificultades, en el estudio se incluyeron algunos aspectos novedosos, dividiéndose la elaboración del cuestionario en dos partes: la primera que comienza con la determinación del objetivo y termina tras la redacción del borrador del cuestionario, y la segunda centrada en la evaluación de dicho borrador que finaliza con la versión final del cuestionario. Durante la primera parte se incluyeron, para realizar la identificación de dimensiones e indicadores, entrevistas a profesionales del ámbito sanitario que actuaron como expertos y grupos focales con población inmigrante y autóctona. Mediante ambos métodos se obtuvo una definición detallada tanto de las dimensiones como de los indicadores que debían incluirse en el cuestionario. Durante la segunda parte, dedicada a la evaluación del cuestionario, se implementaron métodos cognitivos como son las entrevistas cognitivas y la codificación del comportamiento con el fin de obtener información cualitativa de los participantes. A continuación se describen algunos aspectos relevantes relacionados con los métodos cognitivos por el papel central que éstos desempeñan en el estudio.

### *Métodos cognitivos*

Los métodos cognitivos para la evaluación de cuestionarios (Woolley, Bowen, y Bowen, 2006) son utilizados para obtener información sobre cómo los entrevistados realizan el proceso de respuesta a las preguntas y sobre potenciales errores ocurridos en alguna de las fases del proceso. Estos métodos permiten, como indica Snijkers (2002) obtener evidencias sobre la validez de la información aportada por el cuestionario.

Como señalan Presser et al. (2004), los métodos de pretests están basados en la consideración de que los problemas del cuestionario se detectarán mediante las respuestas que provocan las preguntas (por ejemplo, respuestas “no sé” o rechazos), o mediante otras consecuencias visibles (por ejemplo,

dudas o malestar al responder). Según estos autores, los métodos de pretest cognitivo, dirigidos a la identificación de las causas de error, son la única forma de evaluar si el cuestionario causa problemas para el entrevistador o para el entrevistado antes de elaborar una versión final de éste. Entre los métodos comúnmente agrupados bajo la etiqueta de métodos cognitivos para el pretest de cuestionarios se encuentran las entrevistas cognitivas y la codificación del comportamiento, ambos aplicados en este estudio.

La entrevista cognitiva es uno de los procedimientos más utilizados en los pretests cognitivos de los cuestionarios, donde se busca identificar errores sistemáticos en el proceso pregunta-y-respuesta (Willis, 2005). La entrevista cognitiva es una entrevista semi-estructurada en la que se pretende conseguir información sobre las dificultades que ha encontrado la persona que responde al cuestionario mediante la aplicación de técnicas de indagación. Uno de los objetivos de las entrevistas cognitivas es entender que piensan las personas mientras responden a las preguntas de la encuesta, es decir, conocer las estrategias usadas por los participantes para responder. Además, proporciona evidencias sobre problemas en la comprensión de términos clave, fallos en la recuperación de la información, errores en la elaboración de la respuesta y desajuste en la elección de la opción de respuesta.

Por otro lado, la codificación del comportamiento es un procedimiento basado en la observación sistemática del comportamiento del entrevistador y del entrevistado durante la administración de un cuestionario, cuyo objetivo es encontrar los puntos débiles del instrumento que estamos evaluando (Ongena, 2005). La finalidad de la codificación del comportamiento es detectar la presencia de fuentes potenciales de error de medida mediante el análisis de aspectos específicos como pueden ser la forma en que los entrevistadores presentan las preguntas o la reacción de los entrevistados ante éstas. Es decir, se utiliza la codificación del comportamiento para identificar preguntas problemáticas a partir del análisis de los comportamientos de los entrevistadores y de los entrevistados durante el proceso de entrevista.

En este estudio, los métodos cognitivos han sido utilizados como “asesores” durante la elaboración del cuestionario, ya que se han implementado durante la evaluación de un borrador del cuestionario y la información aportada se utilizó para elaborar la versión final, mientras que tradicionalmente se han usado para “probar” el cuestionario. Para ello, se aplicó la codificación del comportamiento sobre los datos obtenidos durante las entrevistas, lo que proporcionó información sobre preguntas con posibles fuentes de error. A continuación se analizó la información aportada por las entrevistas cognitivas con respecto a estas preguntas obteniendo posibles interpretaciones y guías para solucionar los problemas ocurridos y obtener así la versión final del cuestionario. Por tanto, el objetivo de este estudio es presentar el uso de los métodos cognitivos para optimizar la elaboración del cuestionario durante la fase de pretest.

## SECCIÓN 1: ELABORACIÓN DEL BORRADOR DEL CUESTIONARIO

### Método

#### *Participantes*

22 personas participaron en el proceso de elaboración del cuestionario, 15 de ellas en los grupos focales y 7 profesionales en entrevistas individuales. La tabla 1 muestra las características sociodemográficas de estos participantes.

*Tabla 1. Características sociodemográficas de los participantes*

Fase	Género		Edad		
	Muj.	Hom.	18-30	30-45	45-60
Grupo Focal Inmigrantes*	2	4	2	2	2
Grupo Focal Autóctonos	7	2	3	4	2
Entrevistas a profesionales	6	1	0	4	3

\* El grupo focal de inmigrantes estuvo compuesto por 2 participantes procedentes de África y 4 procedentes de Sudamérica.

### *Instrumentos*

Para la realización tanto de las entrevistas como de los grupos focales se utilizó un guión semi-estructurado elaborado a partir de los aspectos que según la bibliografía debían incluirse para evaluar la accesibilidad. Las entrevistas y los grupos fueron grabados en vídeo y audio.

### *Procedimiento*

Una vez especificado el objetivo del cuestionario “evaluar la accesibilidad de los inmigrantes al sistema sanitario” se llevó a cabo la identificación de dimensiones e indicadores por medio de tres métodos. En primer lugar se realizó la habitual revisión bibliográfica en la que se encontraron estudios focalizados en la evaluación de la accesibilidad a los sistemas sanitarios. Se utilizaron como punto de partida las dimensiones e indicadores, o comportamientos observables, incluidos en estos estudios. A partir de ellos se elaboraron guiones para coordinar los grupos focales y las entrevistas a los profesionales. Durante las entrevistas y los grupos focales se discutieron estos aspectos obteniendo un esquema de la estructura del cuestionario.

## Resultados

Los grupos de discusión y las entrevistas a profesionales proporcionaron información que permitió elaborar un esquema de dimensiones e indicadores a partir del cuál se elaboraron las preguntas del cuestionario. La tabla 2 contiene a modo de ejemplo una de las seis dimensiones identificadas con algunos de sus indicadores.

*Tabla 2. Esquema de dimensiones e indicadores*

Dimensiones	Indicadores
Necesidades de los profesionales y técnicos sanitarios	- Coordinación entre profesionales - Organización - Sensibilidad en el trato a los pacientes

Toda esta información dio lugar a la elaboración de la primera versión del cuestionario cuya evaluación mediante los métodos cognitivos se detalla en la siguiente sección.

## SECCIÓN 2: EVALUACIÓN DEL BORRADOR DEL CUESTIONARIO

## Método

*Participantes*

Para la realización de las entrevistas cognitivas se reclutaron participantes de diferentes países siendo mayor la proporción de participantes procedentes de América por su mayor presencia en la población inmigrante censada en España. La tabla 3, muestra detalladamente las características de estos participantes en función del sexo, la edad, el continente de procedencia y el nivel de estudios.

*Tabla 3. Variables sociodemográficas participantes entrevistas cognitivas*

Género		Continente de procedencia		
Mujeres	Hombres	África	América	Asia
16	14	4	25	1
Edad		Nivel de estudios		
18-30	30-45	Básico	Medio	Sup.
11	11	5	14	11

*Instrumentos*

Los técnicos responsables de la realización de las entrevistas dispusieron de cuestionarios adaptados y protocolos de actuación. Los cuestionarios adaptados contenían las 21 preguntas del cuestionario que fueron seleccionadas por los expertos para ser pretestadas, denominadas preguntas target, además de las preguntas sociodemográficas habituales. Por otro lado, los protocolos de actuación contenían indicaciones para el entrevistador sobre cual debía ser su comportamiento además de las pruebas de indagación de la entrevista cognitiva. Las entrevistas cognitivas fueron grabadas en audio y video para su posterior análisis.

*Procedimiento*

El diseño general de las entrevistas cognitivas fue “concurrente”. Primero el entrevistador formulaba la pregunta tal y como se realizará en el trabajo de campo; y después procedía a la aplicación de la técnica cognitiva de indagación. La codificación del comportamiento se aplicó a partir de las transcripciones de las entrevistas.

En primer lugar se realizaron los análisis de la codificación del comportamiento que se llevaron a cabo mediante el programa Sequence Viewer versión 4.4.a (Dijkstra, 2008). Para ello, fue necesaria la introducción de cada una de las transcripciones en el programa, transcripciones que posteriormente fueron divididas en secuencias. Una secuencia comienza con la lectura de una pregunta y termina cuando comienza la lectura de la pregunta siguiente. Estas secuencias fueron analizadas mediante la asignación de códigos en función de los comportamientos ocurridos durante la interacción entrevistador-entrevistado. Los códigos utilizados en cada estudio dependen de los aspectos que el investigador desee registrar. En este estudio se han utilizado códigos para registrar por un lado comportamientos del entrevistador como la realización de indagaciones; por otro lado comportamientos del entrevistado no relacionados con la respuesta como son comentarios, o solicitudes de clarificación o repetición; y por último el tipo de respuesta dado por el entrevistado. La tabla 4 muestra las categorías utilizadas para clasificar las respuestas de los participantes.

Tabla 4. Categorías para la clasificación de las respuestas de los participantes

Códigos	Significado
R. no ajustada	La respuesta no coincide con ninguna de las alternativas ofrecidas
R. inválida	La respuesta no está relacionada con la pregunta
R. no sé	El entrevistado no sabe la respuesta
R. limitada	La respuesta indica incertidumbre en la precisión
R. adecuada	La respuesta se ajusta al objetivo de la pregunta

Dependiendo de los códigos asignados, las secuencias pregunta-y-respuesta son clasificadas como: “secuencias paradigmáticas”, “secuencias no paradigmáticas-no problemáticas” y “secuencias no paradigmáticas-problemáticas”. Una secuencia paradigmática es la secuencia ideal, desde el punto de vista metodológico, en el proceso pregunta-y-respuesta. Una secuencia ideal es aquella en que la formulación de la pregunta se realiza como indica el cuestionario y la respuesta del entrevistado se ajusta a una de las alternativas de respuesta (Ongena y Dijkstra, 2006). Por otro lado, las secuencias no paradigmáticas se dividen en problemáticas y no problemáticas dependiendo del efecto anticipado sobre la calidad de los datos. En este caso, se consideraron no problemáticas aquellas secuencias que contenían las desviaciones de la secuencia ideal identificadas mediante los códigos “comentarios” e “indagación”, mientras que las secuencias fueron clasificadas como problemáticas cuando incluían alguno de los siguientes códigos: “respuesta no ajustada”, “respuesta inválida”, “respuesta limitada”, “respuesta no sé”, y “solicitud de clarificación”.

Una vez asignados los códigos el análisis consistió en el cómputo, para cada una de las preguntas analizadas, de las frecuencias de aparición de cada tipo de secuencia y de cada uno de los códigos. Una vez obtenidas las frecuencias, la codificación del comportamiento permite obtener información sobre el funcionamiento de las preguntas mediante la clasificación de aquellas que deben ser revisadas por la posible presencia de fuentes de error. Un criterio ampliamente aceptado, es considerar el porcentaje de secuencias no paradigmáticas, revisando aquellas preguntas en que el porcentaje sea superior al 60% (Van der Zouwen y Dijkstra, 2002).

A continuación, se llevó a cabo el análisis de las entrevistas cognitivas que consistió en la realización de análisis cualitativos destinados a localizar similitudes y diferencias entre los distintos participantes en relación a la interpretación de los conceptos incluidos en el cuestionario. Estas interpretaciones realizadas por los participantes se compararon con las interpretaciones previstas por los investigadores. El análisis se focalizó en las preguntas clasificadas por la codificación del comportamiento como problemáticas, aunque también se extrajeron otros datos relevantes.

## Resultados

### *Codificación del comportamiento*

En primer lugar, se realizó un análisis de frecuencias de cada uno de los tipos de secuencias incluidas en el estudio en función de las características sociodemográficas de los participantes. No se encontraron diferencias significativas en función del sexo ni del estado civil en el tipo de secuencias producidas por los participantes. Sin embargo, sí se encontraron diferencias significativas entre los distintos niveles de estudios, obteniendo los participantes con estudios superiores el mayor porcentaje de secuencias paradigmáticas y los participantes con estudios medios el mayor porcentaje de secuencias no paradigmáticas (tanto problemáticas como no problemáticas). Se encontraron diferencias significativas en el tipo de secuencias producidas por participantes procedentes de distintos países. La tabla 5 muestra estos porcentajes.

Tabla 5. Porcentajes de secuencias en función del país de nacimiento

País de Nacimiento	Tipo de secuencia		
	SP*	NP/NP*	NP/P*
Ecuador	42.86	29.52	27.62
Perú	48.57	31.43	20
Colombia	61.90	14.88	23.21
China	19.05	14.29	66.67
Cuba	38.10	28.57	33.33
Marruecos	47.62	30.95	21.43
México	47.62	14.29	38.10
Bolivia	58.33	19.05	22.62
Chile	42.86	14.29	42.86

$\chi^2 = 47.200$ ;  $df = 16$ ;  $p = 0.0001$

\*SP: Secuencias paradigmáticas

\*NP/NP: No paradigmáticas no problemáticas

\*NP/P: No paradigmáticas problemáticas

Como muestra la tabla 5, existen diferencias significativas entre las secuencias producidas por participantes de diferentes países. El porcentaje de secuencias paradigmáticas más bajo fue el alcanzado por el participante de China que además registró el mayor porcentaje de secuencias problemáticas. También los participantes de Cuba obtuvieron un porcentaje de secuencias paradigmáticas inferior al 40%. Por otra parte, los participantes de Colombia son los que mejor se ajustaron alcanzando un alto porcentaje de secuencias paradigmáticas. A continuación se realizaron estos análisis para cada una de las preguntas del cuestionario. La tabla 6 muestra los porcentajes obtenidos.

Tabla 6. Porcentajes de secuencias de preguntas

Preguntas target	Tipo de secuencia		
	SP	NP/NP	NP/P
P.1.	93.33	3.33	3.33
P.2.	50	20	30
P.3.	66.67	13.33	20
P.4.	13.33	33.33	53.33
P.5.	53.33	13.33	33.33
P.6.	53.33	16.67	30
P.7.	30	43.33	26.67
P.8.	53.33	26.67	20
P.9.	50	36.67	13.33
P.10.	33.33	33.33	33.33
P.11.	43.33	40	16.67
P.12.	30	50	20
P.13.	66.67	10	23.33
P.14.	26.67	43.33	30
P.15.	56.67	10	33.33
P.16.	43.33	20	36.67
P.17.	80	10	10
P.18.	63.33	16.67	20
P.19.	43.33	3.33	53.33
P.20.	63.33	16.67	20
P.21.	53.33	26.67	20

Como muestra la tabla 6, las preguntas 4, 7, 10, 12 y 14 alcanzan porcentajes de secuencias no paradigmáticas superiores al 60%, y por tanto deben ser revisadas según el criterio establecido. Además, las preguntas 4 y 19 alcanzan porcentajes muy altos de secuencias no paradigmáticas problemáticas.

A continuación se analizaron los porcentajes de cada código registrado para cada una de las preguntas del cuestionario. La tabla 7 resume esta información mostrando las preguntas que obtuvieron los mayores y los menores porcentajes de aparición para cada uno de los códigos.

*Tabla 7. Preguntas con porcentajes extremos en cada código*

Códigos	Preguntas con el porcentaje más bajo	Preguntas con el porcentaje más alto
Códigos de respuesta		
- Adecuada	4 (46.67%)	1 (100%)
- No ajustada	1,19 y 21 (0%)	12 (53.33%), 11 y 14 (46.67%)
- Inválida	18 (0%)	4 (40%)
- No sé	1,9,15,17 (0%)	12 (30%) 14 (23.33%)
- Limitada	1, 5, 6, 7, 8, 9, 13 y 17 (0%)	19 (20%)
Otros códigos		
- Solicitudes	1 (0%)	11 (60%) y 7 (53.33%)
- Indagaciones	1 (0%)	12 y 14 (56.67%)
- Comentarios	1 y 17 (6.67%)	12 (90%) y 11 (70%)

Como muestra la tabla 7 las preguntas 4, 11, 12 y 14 son las que muestran un mayor porcentaje de respuestas problemáticas por un lado, y un mayor porcentaje de aparición de otros comportamientos como solicitudes e indagaciones. En cuanto a las características de las preguntas, la pregunta 4 es una pregunta en la que se pide al sujeto que realice una ordenación de los servicios más usados por él, mientras que las otras preguntas tienen en común que la tarea que se pide al sujeto es que realice una evaluación en una escala de 1 a 10.

Más específicamente es importante señalar que el alto porcentaje de respuestas no ajustadas encontrado para las preguntas 11, 12 y 14 podría estar indicando la incapacidad de las alternativas de respuesta ofrecidas para reflejar la respuesta de los participantes, ya que los contenidos de las respuestas de los participantes fueron adecuados pero no se ajustaron a las alternativas proporcionadas. Es posible que la escala utilizada no sea la más adecuada para el contenido de estas preguntas. Por otro lado, la pregunta 4 alcanza un alto porcentaje (40%) de respuestas inválidas, es decir, que no responden al objetivo de la pregunta.

En cuanto a los códigos no referidos a la respuesta se observa un alto porcentaje de solicitudes en las preguntas 11 y 7, lo que indica la presencia de conceptos de difícil comprensión para los participantes. También se observó un alto porcentaje de indagaciones en las preguntas 12 y 14 lo que muestra los intentos de los entrevistadores por obtener una respuesta adecuada por parte del participante. En relación a las preguntas con los porcentajes más bajos, la pregunta 1.

### *Entrevistas Cognitivas*

Los análisis realizados con las entrevistas cognitivas se centraron en las preguntas detectadas por la codificación del comportamiento con el fin de localizar posibles causas de las dificultades ocurridas. Los datos de las entrevistas cognitivas mostraron lo siguiente.

En la pregunta 4 los participantes tuvieron dificultades para comprender la tarea que debían realizar, comentando en muchos casos que sólo iban al médico general pero sin realizar la ordenación de los servicios demandada en la pregunta. También tuvieron problemas en la comprensión de los conceptos “médico de familia” y “especialista”, considerando el primero como el médico que tradicionalmente había atendido a su familia y sin conocer en algunos casos los médicos incluidos en el término especialistas.

En cuanto a la pregunta 11 las dificultades de los participantes se basaron en la incapacidad para diferenciar varios conceptos incluidos en la pregunta como fueron “listas de espera” y “cumplimiento de horarios”, lo que explica la alta aparición de solicitudes de clarificación. Lo mismo ocurrió en la pregunta 12, en la que los participantes confundieron “campañas de información” y “servicios de información”, dando respuestas no ajustadas e inválidas en un alto porcentaje. El alto porcentaje de comentarios pudo deberse a la aparición de explicaciones y reflexiones por parte del entrevistado destinadas a asegurarse la adecuada comprensión de la pregunta.

Por último, en la pregunta 14 los participantes mostraron dificultades para comprender la expresión “señalización para acceder a los distintos servicios”, lo que de nuevo converge con los datos obtenidos mediante la codificación del comportamiento que muestran un alto porcentaje de respuestas no ajustadas e inválidas por parte del entrevistado y de indagaciones por parte del entrevistador.

Por otro lado, las entrevistas cognitivas mostraron que las dificultades de los participantes en las preguntas en que la tarea consistía en realizar una evaluación en una escala de 1 a 10 no se debía exclusivamente a la escala, sino que además los participantes mostraron dificultades para comprender los diferentes aspectos que debían evaluar.

## Discusión

El objetivo del estudio era presentar el uso de los métodos cognitivos para optimizar la elaboración del cuestionario durante la fase de pretest. Para ello, se implementaron entrevistas cognitivas y la codificación del comportamiento durante la fase de evaluación del borrador de un cuestionario para evaluar la accesibilidad de los inmigrantes al sistema sanitario.

Por un lado, la codificación del comportamiento señaló aquellas preguntas que debían ser revisadas por resultar problemáticas, tanto por el porcentaje de secuencias no paradigmáticas ocurrido como por la alta aparición de códigos inadecuados referidos a las respuestas y a otros comportamientos analizados. Estas preguntas identificadas fueron las siguientes: 4, 11, 12 y 14.

A partir de los comportamientos ocurridos durante estas preguntas se formularon hipótesis sobre las posibles causas de la aparición de dificultades. En el caso de la pregunta 4, las hipótesis se centraron en la complejidad de la tarea mientras que en el caso de las preguntas 11, 12 y 14 las explicaciones se centraron en la falta de ajuste de la escala propuesta para realizar la evaluación.

Estas hipótesis se contrastaron con la información obtenida mediante las entrevistas cognitivas. En el caso de la pregunta 4 se confirmó la complejidad de la tarea y la falta de comprensión de ésta, y además se obtuvieron datos sobre las dificultades de los participantes para comprender conceptos incluidos en la pregunta. Con esta información se propuso modificar la tarea de forma que se preguntara la frecuencia de uso de los servicios de forma separada; también se propusieron alternativas a los términos “médico de familia” y se evaluó la posibilidad de añadir algunos ejemplos de especialistas para clarificar el tipo de profesional al que se hacía referencia la pregunta.

En el caso de las preguntas 11, 12 y 14 se observó que los participantes habían comprendido la escala en la que debían hacer la evaluación, sin embargo las dificultades procedían de la falta de comprensión de algunas de las expresiones que debían evaluar. Se propuso incluir aclaraciones que especificaran a qué hacían referencia dichas expresiones de forma que se asegurara la convergencia entre las interpretaciones realizadas por los participantes y las interpretaciones previstas por los investigadores.

Por otro lado, se encontraron diferencias significativas en el tipo de secuencias producidas por los participantes en función del país de nacimiento. Se observó que el participante de China obtuvo un porcentaje muy bajo de secuencias paradigmáticas (19.05%) y un porcentaje muy alto de secuencias no paradigmáticas- problemáticas (66.67%). Las entrevistas cognitivas mostraron que las dificultades se debían al escaso dominio del español que tenía el participante. También los participantes de Cuba obtuvieron un bajo porcentaje de secuencias paradigmáticas, debido a la alta frecuencia de comentarios y solicitudes que causaron el incremento de las secuencias no paradigmáticas.

Todos estos datos muestran no sólo la convergencia entre la información obtenida mediante las entrevistas cognitivas y la codificación del comportamiento, sino también la complementariedad de ambos métodos. Esta complementariedad se basa en aspectos como la capacidad de las entrevistas cognitivas para ofrecer posibles interpretaciones a los datos proporcionados por la codificación del comportamiento. La codificación del comportamiento es una técnica muy potente con ventajas como la sistematicidad y la objetividad de los datos que ofrece, pero una de sus limitaciones es que no puede explicar el por qué de las dificultades que detecta, por lo que es fundamental la contribución de las entrevistas cognitivas. En este mismo sentido la codificación del comportamiento complementa a las entrevistas cognitivas identificando los momentos de la interacción en que han aparecido dificultades, y focalizando el análisis en dichas situaciones. Además, los datos proporcionados por las entrevistas cognitivas permiten formular propuestas de modificación encaminadas a solucionar los problemas detectados.

Por tanto es importante señalar la posibilidad de obtener una amplia y rica información que ofrece la implementación de ambos métodos durante los procesos de pretest de cuestionarios. Además, esta implementación puede realizarse no sólo para probar el cuestionario, sino también para asesorar el proceso de elaboración de forma que la información obtenida sea utilizada para elaborar la versión final.

Por último, destacar el hecho de que la utilización de los métodos cognitivos no impide la utilización de estudios piloto cuantitativos, cuyas aportaciones también son necesarias. En este estudio se realizó un estudio piloto sobre la versión final elaborada a partir de los datos presentados en apartados anteriores. Futuras investigaciones se centrarán en la comparación de los datos obtenidos y en la complementariedad de los métodos cuantitativos y cualitativos.

## Referencias

- Casas, J., Repullo, J.R. y Donado, J. (2003). La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I). *Atención Primaria*, 31 (8), 527-538.
- Coulter, A. y Cleary, P.D. (2001). Patients' experiences with hospital care in five countries. *Health Affairs* 20, 24-52.
- Dijkstra, W. (2008). Sequence Viewer (version 4.4a). Free University of Amsterdam, Netherlands.

- Flogliatto, H.M. y Pérez, E. (1990). Construcción del cuestionario de intereses profesionales. *Revista de psicología General y Aplicada*, 43 (4), 533-541.
- Gasquet, I.; Villemintot, S., Estaquio, C., Durieux, P., Ravaud, P. y Falissard, B. (2004). Research Construction of a questionnaire measuring outpatients' opinion of quality of hospital consultation departments. *Health and Quality of Life Outcomes*, 1, 2-43.
- Instituto Nacional de Estadística. Revisión del Padrón municipal 2008. [www.ine.es](http://www.ine.es)
- Ongena, Y.P. (2005). *Interviewer and Respondent Interaction in Survey Interviews*. Unpublished doctoral dissertation. Amsterdam Vrije Universiteit.
- Ongena, Y.P. y W. Dijkstra (2006). Methods of Behavior Coding of Survey Interviews. *Journal of Official Statistics* 22, 419-451.
- Padilla, J. L., González, A. y Pérez, C. (1998). Elaboración del cuestionario. En A. J. Rojas, J. S. Fernández y C. Pérez. (Eds.), *Investigar mediante encuestas. Fundamentos teóricos y aspectos prácticos*. Madrid: Síntesis (Psicología).
- Presser, S., Rothgeb, J.M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J. y Singer, E. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley-Interscience.
- Snijkers, G. (2002). *Cognitive Laboratory Experience: On Pre-testing Computerised Questionnaires and Data Quality*. Ph.D thesis. Utrecht University, Utrecht, and Statistics. Netherlands, Heerlen.
- Traver, J. A. y García-López, R. (2007). Construcción de un cuestionario-escala sobre actitud del profesorado frente a la innovación educativa mediante técnicas de trabajo cooperativo (CAPIC). *Revista Electrónica de Investigación Educativa*, 9 (1). Consultado el día 21 de Julio de 2009 en: <http://redie.uabc.mx/vol.9no1/contenido-traver.html>
- Van der Zouwen, J. y Dijkstra, W. (2002). Testing questionnaires using interaction coding. En D. Maynard, H. Houtkoop-Steenstra, N. Schaeffer y J. Van der Zouwen J (Eds.). *Standardization and Tacit knowledge: Interaction and Practice in the Survey interview*. New York: Wiley.
- Willis, G. B. (2005). *Cognitive interviewing*. Thousand Oaks: Sage Publications.
- Wolley, M. E., Bowen, G. L. y Bowen, N. K. (2006). The development and evaluation of procedures to assess child self-report item validity. *Educational and psychological measurement*, 66 (4), 687-700.

# DETECTING HALO EFFECTS IN PERFORMANCE-BASED EXAMINATIONS<sup>[1]</sup>

Timo M. Bechger y Gunter Maris  
Cito, Arnhem, University of Amsterdam

The purpose of this paper is to demonstrate how halo-effects may be detected and quantified using two independent ratings of the same person.

So-called *productive abilities* (e.g., speaking, writing) that require active behaviour of examinees are usually measured via human judgement. That is, examinees demonstrate their ability on a number of assignments/exercises and experts are used to assess the quality of each response. This simple fact gives rise to a myriad of complications. The most conspicuous one being that judges will usually disagree.

Here, we focus on the *halo-effect* which occurs when judgements of one rated characteristic influences judgements of other characteristics in a positive or negative direction. Thus, ratings are influenced by former ratings causing dependencies that cannot be explained by the tendency of examinee’s to produce responses of similar quality.

Our main purpose is to demonstrate how halo-effects may be detected and quantified using an *incomplete block design* with two independent ratings of each examinee’s performance. This design is often used in large-scale examinations when many trained raters are available, but it is not practical to have all raters rate all examinees. Unlike a fully crossed design, where each rater rates every examinee, generalizability theory cannot be used to identify or control halo-effects in this case.

		Raters					
		1	2	3	4	5	6
Examinees	1	X			X		
	2	X				X	
	3	X					X
	4		X		X		
	5		X			X	
	6		X				X
	7			X	X		
	8			X		X	

*Figure 1. The Assignment of Raters to Examinees in an Incomplete Block Design. When a Rater is Assigned to an Examinee There is a Cross*

[1] A much extended version of this paper is accepted for publication in *Applied Psychological Measurement*. Correspondence should be addressed to: Timo M. Bechger, Cito, Nieuwe Overstraat 50, NL-6811JB, Arnhem, The Netherlands; e-mail: timo.bechger[at]cito.nl

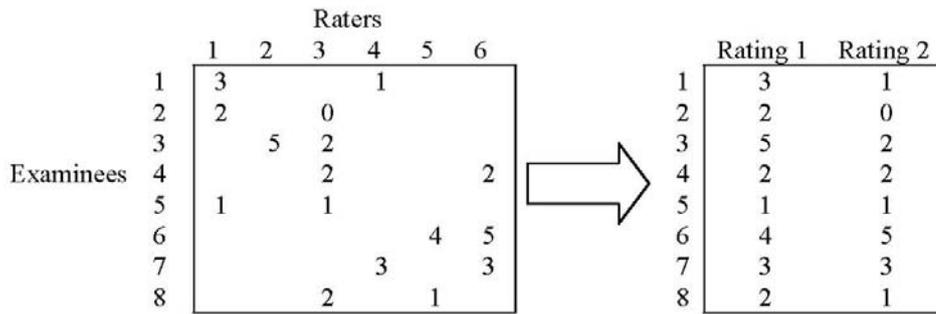


Figure 2. From Incomplete Data with Raters to Complete Data with Ratings Preliminaries

As illustrated in Figure 1, the incomplete block design gives rise to a sparse data set with many missing observations. Before we can analyze the data we take four steps (see Figures 1-3):

1. Assign raters *at random* to examinees in such a way that each of the examinees is assigned two different raters while at the same time the workload of each rater is controlled.
2. Construct a complete data set with two independent *ratings*,  $R_1$  and  $R_2$ , for each examinee. To this aim, one simply places the available ratings in two columns. Due to random assignment, each rating represents the average rater. Formally, the ratings are *exchangeable*.
3. Divide the examination into two half-tests:  $T_1$  and  $T_2$ . Based upon a single rating, the two half-test are scored separately for each examinee.
4. Calculate correlations between the four ratings. The resulting correlation matrix (in Figure 4) will be the main unit of analysis in the sequel.

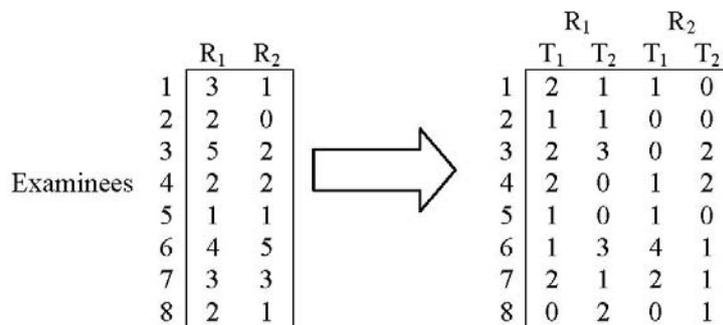


Figure 3. Constructing the Data to Calculate the Correlations

		R <sub>1</sub>		R <sub>2</sub>	
		T <sub>1</sub>	T <sub>2</sub>	T <sub>1</sub>	T <sub>2</sub>
R <sub>1</sub>	T <sub>1</sub>	1			
	T <sub>2</sub>	$\rho_{21}$	1		
R <sub>2</sub>	T <sub>1</sub>	$\rho_{31}$	$\rho_{32}$	1	
	T <sub>2</sub>	$\rho_{41}$	$\rho_{42}$	$\rho_{43}$	1

Figure 4. Correlation Matrix with Two Ratings and Two Test-Halves

*Two Reliabilities*

Suppose the half-tests are *parallel*. Based the first rating, the two half-test are scored separately for each examinee and the correlation coefficient  $\rho_{21}$  is computed between these two sets of scores. Since the half-tests are parallel,  $\rho_{21}$  equals the reliability of each half-test. Using the *Spearman-Brown (SB) prophecy formula*, the reliability of the full-length examination is calculated as:

$$\rho_{xx'} = \frac{2\rho_{21}}{1 + \rho_{21}} \quad (1)$$

When each examinee is rated by two independent raters there are two scores for each half-test (see Figure 3). Schematically, the correlation matrix between these scores is given in Figure 5. In Figure 5,  $\rho_{31} = \rho_{42} = \rho_3$  which is a consequence of having parallel test-halves. There are now *four* different split-half correlations:  $\rho_{21}$ ,  $\rho_{41}$ ,  $\rho_{32}$ , and  $\rho_{43}$ . The correlation  $\rho_3$  between two ratings of the same test-half is a measure of rater reliability. Unfortunately, all these correlations would be affected by halo-effects which complicates their interpretation.

Differences between  $\rho_{21}$  and  $\rho_{43}$ , or between  $\rho_{41}$  and  $\rho_{32}$ , are due to differences between raters. When the two raters that rate each student's product have been chosen *at random*, there will be no such differences and the correlations show the pattern in Figure 6.

		$R_1$		$R_2$	
		$T_1$	$T_2$	$T_1$	$T_2$
$R_1$	$T_1$	1			
	$T_2$	$\rho_{21}$	1		
$R_2$	$T_1$	$\rho_3$	$\rho_{32}$	1	
	$T_2$	$\rho_{41}$	$\rho_3$	$\rho_{43}$	1

Figure 5. Correlation Matrix with two Raters and Parallel Test-Halves

		$R_1$		$R_2$	
		$T_1$	$T_2$	$T_1$	$T_2$
$R_1$	$T_1$	1			
	$T_2$	$\rho_1$	1		
$R_2$	$T_1$	$\rho_3$	$\rho_2$	1	
	$T_2$	$\rho_2$	$\rho_3$	$\rho_1$	1

Figure 6. Correlation Matrix with Parallel Test-Halves and Two Exchangeable Raters

Now, we are left with *two*, possibly different, split-half correlations:  $\rho_1$  and  $\rho_2$ . However, differences between  $\rho_1$  and  $\rho_2$  have a simple interpretation. Specifically, a halo-effect will increase or decrease the dependencies among ratings by the same rater so that  $\rho_1$  will become different from  $\rho_2$ . Hence,

$$\rho^*_{xx'} = \frac{2\rho_2}{1 + \rho_2} \neq \rho_{xx'} = \frac{2\rho_1}{1 + \rho_1}, \quad (2)$$

with equality when there are no halo-effects. Thus, compared to the  $\rho_{XX'}^*$ ,  $\rho_{XX'}$  may be inflated or deflated due to halo-effects.

### *An Effect-Size Measure*

To quantify the size of a halo-effect we employ the general form of the Spearman-Brown formula:

$$\rho_{XX'} = \frac{k\rho_{XX'}^*}{1 + (k-1)\rho_{XX'}^*}, \quad (3)$$

where  $k$  is the number of times the test would have to be lengthened to raise  $\rho_{XX'}$  to the value of  $\rho_{XX'}^*$ . Solving for  $k$  gives:

$$k = \frac{\rho_{XX'}(1 - \rho_{XX'}^*)}{\rho_{XX'}^*(1 - \rho_{XX'})}. \quad (4)$$

Thus, the reliability of the test is over- (or under) estimated due to halo-effects and  $k$  expresses this effect in terms of examination length, where examination length is defined in terms of the number of ratings of each examinee.

### *Estimation and Testing*

The correlations  $\rho_i$  are population quantities. In practice, we estimate  $\rho_{XX'}$  and  $\rho_{XX'}^*$  using sample correlations:  $r_{ij}$ . To keep things simple, we simply average the sample correlations. Hence,  $\rho_1$  is estimated as the average of the two *within-rating* correlations, and  $\rho_2$  is estimated as the average of the two *between-rating, between-halves* correlations. Formally, this gives ordinary least-squares estimates.

The simplest way to test for the presence of halo-effects is to compare  $r_{21}$  with  $r_{32}$ : A halo effect is present when  $r_{21} \neq r_{32}$ . These correlations are based on independent data so that standard tests can be used.

### Conclusion

We have discussed how a halo-effect can be detected and how its size can be expressed in terms of examination length. It is important to note that our approach requires no assumptions on the underlying behaviour of the judges. Uni-dimensionality, for example is not assumed. This is useful because very little is known about rater behaviour. This lack of knowledge hampers the use of more sophisticated approaches.

## ESTRUCTURA FACTORIAL DEL CUESTIONARIO DE VIOLENCIA ESCOLAR COTIDIANA (CUVECO)

María J. Blanca, F. Javier Fernández-Baena, Milagros Escobar y María V. Trianes  
Universidad de Málaga

En estudios previos del Cuestionario de Violencia Escolar Cotidiana (CUVECO), construido a partir del *California School Climate and Safety Survey*, con una muestra de 954 estudiantes de segundo y tercer curso de Enseñanza Secundaria Obligatoria, se obtuvo una estructura factorial de dos factores, denominados experiencia personal de sufrir violencia (8 ítems) y violencia observada en el centro (6 ítems). El primer factor evalúa los comportamientos violentos padecidos por los estudiantes. El segundo factor evalúa la violencia entre iguales observada en el contexto escolar. El objetivo del presente trabajo es averiguar si esta estructura factorial se mantiene con estudiantes de primer curso de educación secundaria ( $N = 569$ ). Para ello, se realizó un análisis factorial de componentes principales con rotación Oblimin. Éste arrojó una medida de adecuación muestral de 0,89, con la prueba de Bartlett estadísticamente significativa, y una solución factorial de dos factores, que explican el 37,09% y el 15,11% de la varianza, coincidente con la anterior, incluyendo los mismos ítems y con saturaciones aproximadas. Estos resultados muestran evidencia de la estabilidad de la estructura factorial del instrumento.

El estudio de la conducta agresiva de los estudiantes en relación con sus compañeros recibe una mayor atención a partir de los trabajos de Olweus (1978), centrado particularmente en el comportamiento de acoso o *bullying*. Sobre este fenómeno se han aportado datos de su incidencia en diversos países, tipologías de alumnos implicados y correlatos comportamentales asociados (Griffin y Gross, 2004; Smith et al., 1999). En España cabe destacar los estudios de Ortega (1994, 1998), Cerezo (1997), los Informes del Defensor del Pueblo (2000, 2007), Mora-Merchán, Ortega, Justicia y Benítez (2001), del Barrio, Martín, Montero, Gutiérrez y Fernández (2003), y Avilés y Monjas (2005). Estos estudios han puesto de manifiesto que experiencias de sufrir violencia, o incluso, observarla, tienen consecuencias muy negativas para el desarrollo infantil y adolescente.

Los efectos perniciosos de la violencia escolar de alta intensidad se muestran en el ámbito psicopatológico (Greve y Wilmers, 2003), afectivo-social (Bernstein y Watson, 1997; Pellegrini, Bartini y Brooks, 1999), en el aprendizaje y rendimiento académico (Beale, 2001), y está relacionada con el incremento de conductas agresivas en chicos y chicas (O'Keefe, 1997). Igualmente, un ambiente de violencia escolar de alta intensidad se asocia al deterioro del clima social de clase, afectando negativamente a todo el contexto educativo e impidiendo el normal desarrollo de la docencia y de las actividades de clase (Sánchez, Rivas y Trianes, 2006; Westling, 2002).

La violencia escolar de baja o moderada intensidad, que denominamos violencia cotidiana entre iguales, también es relevante en el contexto educativo. Ésta ocurre cuando una persona se ve insultada, físicamente agredida, socialmente excluida o aislada, amenazada o atemorizada por otro/s de forma puntual y no reiterada (Ortega, del Rey y Mora-Merchán, 2001). Algunos estudios muestran que sufrir violencia cotidiana, sin episodios de acoso o maltrato, puede predecir psicopatología e inadaptación personal (Cangas, Gázquez, Pérez-Fuentes, Padilla y Miras, 2007; Trianes, 2004). Por ello, la evaluación de la violencia cotidiana es un aspecto fundamental como paso previo al diseño de intervenciones psicoeducativas dirigidas a combatir las consecuencias mencionadas. En esta línea, Fernández-Baena et al. (en prensa) elaboraron el Cuestionario de Violencia Escolar Cotidiana, construido a partir del *California School Climate and Safety Survey* (Rosenblatt y Furlong, 1997),

con una muestra de 954 estudiantes de segundo y tercer curso de Enseñanza Secundaria Obligatoria (ESO). El cuestionario mostró una estructura factorial de dos factores, denominados *experiencia de sufrir violencia* (8 ítems) y *violencia observada en el centro* (6 ítems). El objetivo del presente trabajo es completar el trabajo previo, analizando la estructura factorial con estudiantes de primer curso, con el fin de disponer de un instrumento que pueda ser utilizado en todos los niveles de la ESO.

## Método

### *Participantes*

Participaron 526 estudiantes de primer curso de la ESO (286 niños y 240 niñas), con una media de edad de 12,54 años ( $DT = 0,69$ ), procedentes de siete centros públicos de Málaga capital y provincia.

### *Instrumentos*

Se administró el Cuestionario de Violencia Escolar Cotidiana (CUVECO; Fernández-Baena et al., en prensa). Este cuestionario consta de 14 ítems con un formato de respuesta tipo Likert de 5 puntos (1 = nunca, 5 = casi siempre), organizados en dos factores denominados *experiencia de sufrir violencia* (8 ítems) y *violencia observada en el centro* (6 ítems). El primer factor evalúa comportamientos violentos padecidos por los estudiantes, mientras que el segundo evalúa el nivel de violencia entre iguales observada en el contexto escolar.

### *Procedimiento*

La administración de la prueba ha sido realizada por dos investigadores que supervisaron todo el proceso y respondieron a las dudas que los estudiantes plantearon sobre los ítems del cuestionario. Los estudiantes completaron la prueba de forma colectiva, voluntaria y anónima, durante una hora lectiva en la que los tutores no estuvieron presentes. Al inicio del estudio se contactó con los padres para ofrecerles información sobre la investigación y solicitar su autorización y conformidad.

## Resultados

Se ha realizado un análisis de componentes principales con rotación Oblimin. Éste arrojó una medida de adecuación muestral de 0,89, con la prueba de Bartlett estadísticamente significativa,  $\chi^2(91) = 2785,64$ ;  $p < 0,01$ . Del análisis se extrajeron dos factores que explican el 37,09% y el 15,11% de la varianza, respectivamente. En la tabla 1 se exponen las saturaciones factoriales de cada ítem.

Tabla 1. Saturaciones de los ítems del CUVECO en sus respectivos factores

Ítems	Saturaciones
<i>Factor 1. Experiencia personal de sufrir violencia</i>	
1. Me han empujado (en el pasillo, patio, etc.)	0,71
2. Me han dado puñetazos o patadas	0,80
3. Me han golpeado con un objeto	0,71
4. Me han robado algún objeto	0,53
5. Me han dicho que me iban a lastimar o pegar	0,77
6. Me han roto cosas	0,50
7. Me han amenazado	0,71
8. Se han burlado de mí o me han despreciado	0,64
<i>Factor 2. Violencia observada en el centro</i>	
9. Los estudiantes usan drogas	-0,62
10. Los estudiantes destrozan cosas	-0,83
11. Los estudiantes se meten en peleas	-0,84
12. Los estudiantes roban cosas	-0,82
13. Los estudiantes amenazan a otros estudiantes	-0,73
14. Se dicen palabras malsonantes	-0,72

Los factores pueden ser identificados como:

Factor 1. Experiencia personal de sufrir violencia, compuesto por ocho ítems que reflejan las agresiones de tipo físico, verbal o psicológico que padece el alumnado víctima de violencia cotidiana entre iguales en el contexto escolar.

Factor 2. Violencia observada en el centro, compuesto por seis ítems referentes principalmente a comportamientos violentos entre iguales que son observados por los estudiantes en el centro educativo.

La correlación entre ambos factores del CUVECO fue de -0,41. En la tabla 2 se presentan los estadísticos descriptivos de los ítems del cuestionario, junto con el índice de homogeneidad corregido con su respectivo factor, el cual supera el 0,40 en todos los ítems. El coeficiente alfa de consistencia interna para los dos factores fue, respectivamente, 0,83 y 0,86.

Tabla 2. Media (M), desviación típica (DT) e índice de homogeneidad corregido (IH) de los ítems del CUVECO

Ítems	M	DT	IH
1. Me han empujado (en el pasillo, patio, etc.)	1,53	0,87	0,57
2. Me han dado puñetazos o patadas	1,30	0,76	0,61
3. Me han golpeado con un objeto	1,31	0,68	0,59
4. Me han robado algún objeto	1,44	0,82	0,48
5. Me han dicho que me iban a lastimar o pegar	1,44	0,87	0,66
6. Me han roto cosas	1,43	0,74	0,48
7. Me han amenazado	1,21	0,66	0,57
8. Se han burlado de mí o me han despreciado	1,59	1,06	0,54
9. Los estudiantes usan drogas	1,61	1,14	0,50
10. Los estudiantes destrozan cosas	2,10	1,20	0,70
11. Los estudiantes se meten en peleas	2,71	1,28	0,74
12. Los estudiantes roban cosas	2,25	1,23	0,71
13. Los estudiantes amenazan a otros estudiantes	2,15	1,21	0,67
14. Se dicen palabras malsonantes	3,56	1,42	0,58

## Discusión

Este estudio se enmarca dentro de las necesidades de evaluación de la violencia escolar interpersonal de intensidad baja o moderada en los centros de secundaria. El CUVECO recoge información sobre la violencia escolar cotidiana entre iguales, al margen de la evaluación del acoso entre iguales, el cual ha recibido una mayor atención en la literatura. En un estudio de Trianes (2004), se analiza la incidencia de la violencia escolar cotidiana entre iguales ocurrida en centros de enseñanza secundaria de Andalucía, enfatizándose y poniéndose de relieve la importancia de la evaluación y prevención de este tipo de violencia, ya que sus consecuencias individuales y grupales, a través del clima del aula, son perjudiciales y entorpecedoras del óptimo desarrollo socioemocional de los adolescentes, así como del logro de un aprendizaje de calidad.

El objetivo del presente trabajo ha sido analizar la estructura factorial del CUVECO (Fernández-Baena et al., en prensa) con estudiantes de primer curso de ESO. El análisis muestra dos factores estables que pueden ser identificados como *experiencia de sufrir violencia* y *violencia observada en el centro*, respectivamente. El primer factor evalúa los comportamientos violentos padecidos por los estudiantes. El segundo factor evalúa la violencia entre iguales observada en el contexto escolar. Este factor está estrechamente relacionado con el constructo de clima social de centro, referido a la percepción de los estudiantes sobre la calidad de las relaciones entre los compañeros, y sobre la seguridad y no violencia del centro (Trianes, Blanca, de la Morena, Infante y Raya, 2006; Welsh, 2000).

Ambos factores presentan una adecuada consistencia interna así como índices de homogeneidad superior a 0,40 en todos sus elementos. Tanto la estructura factorial como los ítems que saturan en los factores son coincidentes con el estudio previo (Fernández-Baena et al., en prensa), lo que proporciona evidencia de la estabilidad de la estructura factorial del instrumento en los diversos cursos de la ESO.

Estos resultados avalan un instrumento de evaluación de la violencia cotidiana entre iguales que puede ser utilizado para la detección de este tipo de violencia en contextos educativos, así como para la elaboración, diseño y evaluación de programas psicoeducativos en los que se incida sobre los procesos individuales, sin olvidar los procesos grupales y contextuales, que se encuentran presentes en las situaciones de violencia escolar.

Agradecimientos. Este trabajo ha recibido financiación del Ministerio de Ciencia e Innovación. Proyecto I+D+i. Ref. SEJ2007-61447/PSIC.

## Referencias

- Avilés, J. M. y Monjas, I. (2005). Estudio de la incidencia de la intimidación y el maltrato entre iguales en la educación secundaria obligatoria mediante el cuestionario CIMEI. *Anales de Psicología*, 21, 27-41.
- Beale, A. V. (2001). 'BullyBusters': Using drama to empower students to take a stand against bullying behavior. *Professional School Counseling*, 4, 300-305.
- Bernstein, J. Y. y Watson, M. W. (1997). Children who are targets of bullying: A victim pattern. *Journal of Interpersonal Violence*, 12, 483-498.
- Cangas, A. J., Gázquez, J. J., Pérez-Fuentes, M. C., Padilla, D. y Miras, F. (2007). La evaluación de la violencia escolar y su afectación personal en una muestra de estudiantes europeos. *Psicothema*, 19, 114-119
- Cerezo, F. (1997). *Conductas agresivas en edad escolar*. Madrid: Pirámide.

- Defensor del Pueblo (2000). *Informe del Defensor del Pueblo sobre Violencia Escolar*. Madrid: Publicaciones de la Oficina del Defensor del Pueblo.
- Defensor del Pueblo (2007). *Violencia escolar: El maltrato entre iguales en la educación secundaria obligatoria 1999-2006*. Madrid: Publicaciones de la Oficina del Defensor del Pueblo.
- Del Barrio, C., Martín, E., Montero, I., Gutiérrez, H. y Fernández, I. (2003). La realidad del maltrato entre iguales en los centros de secundaria españoles. *Infancia y Aprendizaje*, 26, 25-47.
- Fernández-Baena, F. J., Trianes, M. V., de la Morena, M. L., Escobar, M., Infante, L. y Blanca, M. J. (en prensa). Estudio sobre violencia escolar cotidiana. Desarrollo de un instrumento de evaluación. *Anales de Psicología*.
- Greve, W. y Wilmers, N. (2003). School violence and self-esteem: Moderating influences of coping resources among offenders and victims. *Psychologic in Erziehung und Unterricht*, 50, 353-368.
- Griffin, R. y Gross, A. (2004). Childhood bullying: current empirical findings and future directions for research. *Aggressive and Violent Behaviour*, 9, 379-400.
- Mora-Merchán, J., Ortega, R., Justicia, F. y Benítez, J. L. (2001). Violencia entre iguales en escuelas andaluzas. Un estudio exploratorio utilizando el cuestionario general europeo TMR. *Revista de Educación*, 325, 323-338.
- O'Keefe, M. (1997). Adolescents' exposure to community and school violence: Prevalence and behavioral correlates. *Journal of Adolescent Health*, 20, 368-376.
- Olweus, D. (1978). *Aggressions in the school: Bullying and whipping boys*. Washington DC: Hemisphere.
- Ortega, R. (1994). Violencia interpersonal en los centros educativos de enseñanza secundaria. Un estudio sobre el maltrato e intimidación entre compañeros. *Revista de Educación*, 304, 253-280.
- Ortega, R. (1998). *La convivencia escolar: qué es y cómo abordarla*. Sevilla: Consejería de Educación y Ciencia. Junta de Andalucía.
- Ortega, R., del Rey, R. y Mora-Merchán, J. (2001). Violencia entre escolares. Concepto y etiquetas verbales que definen el fenómeno del maltrato entre iguales. *Revista Interuniversitaria de Formación del Profesorado*, 41, 95-113.
- Pellegrini, A. D., Bartini, M. y Brooks, F. (1999). School bullies, victims, and aggressive victims: Factors relating to group affiliation and victimization in early adolescence. *Journal of Educational Psychology*, 91, 216-224.
- Rosenblatt, J. y Furlong, M. J. (1997). Assessing the reliability and validity of student self-reports of campus violence. *Journal of Youth and Adolescence*, 26, 187-201.
- Sánchez, A. M., Rivas, M. T. y Trianes, M. V. (2006). Eficacia de un programa de intervención para la mejora del clima escolar: algunos resultados. *Revista Electrónica de Investigación Psicoeducativa*, 4, 353-370.

- Smith, P. K., Morita, Y., Junger-Tas, J., Olweus, D., Catalano, R. y Slee, P. (1999). *The nature of school bullying. A cross-national perspective*. Londres: Routledge.
- Trianes, M. V. (2004, Octubre). *Contextos de la violencia juvenil en España*. Ponencia presentada en la VIII Reunión Internacional sobre Biología y Sociología de la violencia: Violencia y Juventud. Centro Reina Sofía para el Estudio de la Violencia, Valencia.
- Trianes, M. V., Blanca, M. J., de la Morena, L., Infante, L. y Raya, S. (2006). Un cuestionario para evaluar el clima social de la clase y centro. *Psicothema*, 18, 272-277.
- Welsh, W. N. (2000). The effects of school climate on school disorder. *Annals of the American Academy of Political and Social Science*, 567, 88-107.
- Westling, M. (2002). A two level analysis of classroom climate in relation to social context, group composition, and organization of special support. *Learning Environments Research*, 5, 253-274.

# ESTRUCTURA FACTORIAL DE LA ESCALA SSI SOBRE RESPUESTAS DE ESTRÉS

Sergio Cantos, F. Javier Fernández-Baena, María V. Trianes, Milagros Escobar y María J. Blanca  
Universidad de Málaga

La evaluación de las manifestaciones del estrés en la adolescencia es un aspecto importante tanto para la investigación como para la intervención psicoeducativa. En la investigación sobre estrés ha sido frecuente el empleo de cuestionarios referidos a la aparición o no de situaciones estresantes, siendo el análisis de las respuestas asociadas al estrés una dimensión menos estudiada. El propósito del presente trabajo es traducir al castellano y presentar algunas propiedades psicométricas del cuestionario desarrollado por Moulds (2003), a partir del *Student Stress Inventory*, para identificar las respuestas de estrés dominantes en adolescentes. La escala se ha aplicado a una muestra de 381 estudiantes de ambos sexos de educación secundaria con edades comprendidas entre los 12 y 16 años. El resultado del análisis factorial en la muestra española del cuestionario presenta una estructura de cinco factores, que han sido etiquetados como: manifestaciones de ira, reacción de confusión, manifestaciones somáticas, manifestaciones de inadaptación escolar y manifestaciones de ansiedad. El coeficiente alfa de consistencia interna de los factores oscila entre 0,63 y 0,70. Estos factores no coinciden con los obtenidos en la escala original y evidencian la necesidad de futuros estudios confirmatorios de su estructura factorial.

La investigación sobre estrés en la infancia y la adolescencia ha generado diversos instrumentos de medida para su evaluación, predominando los cuestionarios referidos a la frecuencia de situaciones estresantes y siendo el análisis de las respuestas asociadas al estrés una dimensión menos estudiada. La evaluación de esta dimensión resulta de interés, ya que permitiría complementar la información referida a los estresores sufridos contrastando el impacto emocional, conductual y fisiológico de los mismos, tanto en adolescentes de muestras clínicas como normativas. Asimismo, la información derivada de la evaluación puede proporcionar la base para programar intervenciones psicoeducativas con adolescentes que presenten alguna problemática relacionada con el estrés.

El propósito del presente trabajo es traducir al castellano y presentar la estructura factorial y consistencia interna del *Student Stress Inventory* (SSI) desarrollado por Fimian, Fastenau, Tashner y Cross (1989) y modificado por Moulds (2003). Este cuestionario se centra en la evaluación en adolescentes de las manifestaciones derivadas del estrés de tipo emocional, comportamental y fisiológico.

## Método

### *Participantes*

Han participado 381 estudiantes (190 niños y 191 niñas) procedentes de los cursos de 1º y 2º de Educación Secundaria Obligatoria de tres centros escolares de Málaga capital, siendo la edad media de 13,20 años ( $DT = 1,02$ ).

## Material

Se ha administrado el *Student Stress Inventory* (SSI; Fimian et al., 1989), atendiendo a las modificaciones propuestas por Moulds (2003). La versión elaborada por Moulds recoge los ítems del SSI referidos a las manifestaciones de estrés en tres ámbitos: manifestaciones emocionales, manifestaciones comportamentales y manifestaciones fisiológicas.

El cuestionario consta de 22 ítems con un formato de respuesta tipo Likert de cinco puntos. El estudio original de su dimensionalidad (Moulds, 2003) identifica tres factores empíricos referidos a manifestaciones de estrés: a) Ansiedad, se compone de 10 ítems, explica el 26,5% de la varianza y presenta un coeficiente  $\alpha$  de 0,84, b) Ira, se compone de 6 ítems, explica el 9,8% de la varianza y presenta un coeficiente  $\alpha$  de 0,72 y c) Arousal, se compone de 6 ítems, explica el 9% de la varianza y presenta un coeficiente  $\alpha$  de 0,69. Estos tres factores se corresponden, respectivamente, a las tres manifestaciones anteriormente citadas.

## Procedimiento

Se siguió una metodología de traducción directa atendiendo a las recomendaciones de la Comisión Internacional de Tests. En una primera fase, se llevó a cabo la traducción al castellano del cuestionario original en inglés por parte de un traductor oficial cuya lengua materna era el castellano. En una segunda fase, el equipo de investigación, junto con el traductor, revisó de forma sistemática cada uno de los ítems, obteniéndose así una versión definitiva consensuada en castellano.

Se administró la prueba con la ayuda de los orientadores de los centros y en horario escolar. En todos los casos se recogieron los permisos pertinentes de la familia y del centro escolar.

## Resultados

De acuerdo con el procedimiento de análisis seguido por Moulds (2003), se ha realizado un análisis de componentes principales con rotación Varimax. El análisis arrojó una medida de adecuación muestral de 0,88, con la prueba de Bartlett estadísticamente significativa [ $\chi^2(231) = 2126,54$ ;  $p < 0,01$ ]. Se extrajeron cinco factores que explican un total de 51,17% de la varianza (27,51%, 7,50%, 6,08%, 5,46% y 4,60%, respectivamente), que pueden ser identificados como *Manifestaciones de ira*, *Reacción de confusión*, *Manifestaciones somáticas*, *Manifestaciones de inadaptación escolar* y *Manifestaciones de ansiedad* (tabla 1).

Por otra parte, se ha calculado el coeficiente alfa de Cronbach para cada uno de los factores, hallándose unos valores comprendidos entre 0,63 y 0,70. El análisis de ítems muestra que el índice de homogeneidad corregido a su respectivo factor fue aceptable, con valores mayores a 0,30 en todos los ítems. Los valores de alfa al eliminar el ítem indican que la eliminación de cualquier ítem no produce un aumento en la consistencia interna de la escala (tabla 1).

Tabla 1. Estructura factorial de la adaptación al castellano de los ítems del SSI referidos a las manifestaciones de estrés (N = 381)

Ítem	Saturación	Correlación ítem-total corregida	Alfa al eliminar ítem
<i>Factor 1: Manifestaciones de ira. <math>\alpha = 0,70</math></i>			
1. Irritarse	0,66	0,53	0,61
7. Actuar a la defensiva con los demás	0,62	0,40	0,69
18. Enfadarse	0,68	0,59	0,57
19. Tomarla con alguien	0,65	0,44	0,66
<i>Factor 2: Reacción de confusión. <math>\alpha = 0,68</math></i>			
2. Comer más o menos de lo habitual	0,56	0,37	0,66
3. Dejar las cosas para otro día	0,49	0,40	0,65
4. Asustarse	0,46	0,32	0,67
5. Preocuparse	0,50	0,50	0,61
12. Ser indeciso	0,63	0,41	0,64
15. No saber qué hacer	0,55	0,47	0,62
<i>Factor 3: Manifestaciones somáticas. <math>\alpha = 0,63</math></i>			
9. Sentir un sudor frío	0,54	0,40	0,57
10. Sentir picor u hormigueo por todo el cuerpo	0,63	0,48	0,51
14. Quedarse afónico/a	0,53	0,35	0,60
20. Tener dolor de estómago	0,68	0,42	0,56
<i>Factor 4: M. de Inadaptación escolar. <math>\alpha = 0,67</math></i>			
11. Sentirse incapaz de ir bien en el colegio	0,48	0,39	0,65
16. Hablar mal de compañeros/as o profesores/as	0,61	0,48	0,58
21. Responder mal a los profesores/as	0,61	0,59	0,50
22. Fingir estar enfermo/a	0,62	0,36	0,65
<i>Factor 5: Manifestaciones de Ansiedad. <math>\alpha = 0,63</math></i>			
6. Sentirse ansioso/a	0,49	0,47	0,51
8. Sentirse presionado/a	0,44	0,49	0,49
13. Descuidar las amistades	0,52	0,39	0,57
17. Tener palpitaciones o taquicardia	0,68	0,30	0,63

## Discusión

Los resultados del análisis exploratorio del SSI ha determinado la presencia de cinco factores que se han denominado *Manifestaciones de ira*, *Reacción de confusión*, *Manifestaciones somáticas*, *Manifestaciones de inadaptación escolar* y *Manifestaciones de ansiedad*. Los valores de consistencia interna de los factores (entre 0,63 y 0,70) pueden considerarse aceptables, siendo algo menores que los obtenidos por Moulds (2003). Los índices de homogeneidad también son adecuados, revelando una buena cohesión entre los respectivos ítems. No obstante, la estructura factorial encontrada no se ajusta a los factores informados por Fimian et al. (1989) referidos a manifestaciones emocionales, conductuales y fisiológicas y nombradas por Moulds (2003) como ansiedad, ira y arousal, respectivamente. En la tabla 2 se exponen las diferencias en la composición de los factores entre este último estudio y el obtenido aquí con la muestra española. La denominación de los factores encontrados se ha ajustado a su contenido, siguiéndose para ello un criterio conceptual. Sin embargo, se aprecia la existencia de ítems que conceptualmente deberían estar contenidos en otros factores, como por ejemplo, “descuidar las amistades” y “comer más o menos de lo habitual”. Futuras investigaciones aportarán más evidencia acerca de la dimensionalidad del cuestionario así como de los indicios de validez.

Tabla 2. Comparación entre los factores originales y los de la adaptación al castellano

Factor 1: ANSIEDAD	Factor 1: M. DE IRA
2. Comer más o menos de lo habitual	1. Irritarme
3. Dejar las cosas para otro día	7. Actuar a la defensiva con los demás
4. Asustarme	18. Enfadarme
5. Preocuparme	19. Tomarla con alguien
6. Sentirme ansioso/a	
8. Sentirme presionado	Factor 2: REACCIÓN DE CONFUSIÓN
11. Sentirme incapaz de ir bien en el colegio	2. Comer más o menos de lo habitual
12. Ser indeciso	3. Dejar las cosas para otro día
15. No saber qué hacer	4. Asustarme
18. Enfadarme	5. Preocuparme
	12. Ser indeciso
Factor 2: IRA	15. No saber qué hacer
1. Irritarme	
7. Actuar a la defensiva con los demás	Factor 3: M. SOMÁTICAS
16. Hablar mal de compañeros o profesores	9. Sentir un sudor frío
19. Tomarla con alguien	10. Sentir picor u hormigueo por todo el cuerpo
21. Responder mal a los profesores	14. Quedarme afónico/a
22. Fingir estar enfermo	20. Tener dolor de estómago
Factor 3: ACTIVACIÓN	Factor 4: M. DE INADAPTACIÓN ESCOLAR
9. Sentir un sudor frío	11. Sentirme incapaz de ir bien en el colegio
10. Sentir picor u hormigueo por todo el cuerpo	16. Hablar mal de compañeros o profesores
13. Descuidar las amistades	21. Responder mal a los profesores
14. Quedarme afónico/a	22. Fingir estar enfermo
17. Tener palpitaciones o taquicardia	
20. Tener dolor de estómago	Factor 5: M. DE ANSIEDAD
	6. Sentirme ansioso/a
	8. Sentirme presionado
	13. Descuidar las amistades
	17. Tener palpitaciones o taquicardia

Agradecimientos. Este trabajo ha recibido financiación del Ministerio de Ciencia e Innovación. Proyecto I+D+i. Ref. SEJ2007-61447/PSIC.

### Referencias

- Fimian, M. J., Fastenau, P. A., Tashner, J. H. y Cross, A. H. (1989). The measure of classroom stress and burnout among gifted and talented students. *Psychology in the Schools*, 26, 139-153.
- Moulds, J. (2003). Stress manifestation in high school students: An australian sample. *Psychology in the Schools*, 40, 391-402.

# ¿CÓMO EVALUAMOS? ANÁLISIS DE ÍTEMS DE OPCIÓN MÚLTIPLE Y SU RELACIÓN CON ERRORES EN LA CONSTRUCCIÓN

Carmen García García, Vicente Ponsoda y Alejandro Sierra  
Universidad Autónoma de Madrid

En la evaluación del aprendizaje de nuestros estudiantes es habitual emplear ítems de opción múltiple, de ahí la importancia de conocer las consecuencias que su construcción y/o análisis defectuoso pueden acarrear para las calificaciones que otorgamos. En este trabajo recogimos una muestra de 35 exámenes, y sus respuestas, de asignaturas impartidas en la licenciatura de Psicología y de Informática de la Universidad Autónoma de Madrid (UAM) durante los cursos 2007/08 y 2008/09. Para cada examen se realiza un análisis de los ítems obteniendo sus índices de dificultad y de discriminación y se analiza si éstos se ven afectados, y cómo, por el número de formas del examen, la posición de la opción correcta y por fallos en la redacción al incumplirse alguna de las recomendaciones propuestas por Haladyna, Downing y Rodríguez (2002). A pesar de que la muestra de exámenes no es representativa, ofrecemos datos sobre la calidad psicométrica de los exámenes con los que realmente evaluamos en la universidad. No conocemos ningún trabajo que aporte tal información.

Los ítems de elección múltiple tienen un papel importante en la evaluación en los niveles educativos superiores. Como muestra de ello diremos que sólo en la convocatoria de febrero de 2008 en la UAM más de 30.000 estudiantes fueron evaluados mediante este tipo de ítems. Se podría pensar que los profesores tenemos la habilidad de escribir buenos ítems sin necesidad de recibir instrucción para ello. Sin embargo, tal y como muestran los análisis que presentaremos más tarde esto dista mucho de ser así. Ítems defectuosos tienen como consecuencia exámenes con baja fiabilidad, pero ¿cuáles son las consecuencias de que un examen tenga poca fiabilidad? para mostrar los efectos de la falta de fiabilidad se obtuvieron, por simulación estadística, las puntuaciones de 1.000 estudiantes en tres exámenes de opción múltiple de 50 preguntas. Uno de fiabilidad 0,89, otro de 0,65 y el tercero de 0,32<sup>[1]</sup>. El objetivo era ver como cambiaba la proporción de estudiantes que eran erróneamente aprobados o suspendidos. Se fijó el punto de corte para el aprobado en 20 aciertos. Aprueba, por tanto quien tiene una puntuación observada ( $X$ ) de 20 o más puntos. La Figura 1, que corresponde a un examen cuyo coeficiente era 0,89, muestra que el porcentaje de estudiantes que aprobaría el examen sería del 84%, pero sólo un 80% tendría una puntuación verdadera ( $T$ ) superior a 20, por lo que un 4% de estudiantes lo aprobaría, pero tendría una  $T$  de suspenso. Igualmente tendríamos un 4% de suspensos con una  $T$  de aprobado. En resumen, en el examen más fiable suspenderíamos o aprobaríamos por error al 8% de los estudiantes. Como se muestra en la Tabla 1 a medida que la fiabilidad del examen baja la proporción de estudiantes mal calificados aumenta.

*Tabla 1. Errores en la calificación de los estudiantes en función de la fiabilidad del examen*

Coeficiente de fiabilidad	Aprobados que deberían suspender	Suspensos que deberían aprobar	Calificaciones erróneas
0,89	0,04	0,04	0,08
0,65	0,06	0,10	0,16
0,32	0,06	0,17	0,23

[1] Estos valores coincidían con el coeficiente alfa más alto, medio y menor obtenidos en la muestra de exámenes analizados.

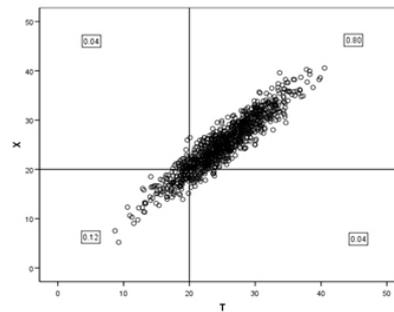


Figura 1. Diagrama de dispersión de las puntuaciones observadas ( $X$ ) y verdaderas ( $T$ ) obtenidas por simulación, en un examen de 1.000 estudiantes cuyo coeficiente de fiabilidad era 0,89

Conscientes de la importancia que este sistema de evaluación tiene, las universidades están implantando sistemas de evaluación de la calidad de los exámenes. Por ejemplo, Muntinga y Schuil (2007) estudian las consecuencias de un sistema de análisis automático de ítems que su universidad les pide que utilicen para determinar qué ítems de cada examen han de ser anulados y no contar en la evaluación. Lo aplicaron a sucesivas convocatorias de una misma materia. El sistema clasifica cada ítem en una de tres categorías: Retener, Posible eliminación y Eliminar. Los criterios utilizados son el índice de dificultad y de discriminación corregido. Los autores muestran como los ítems que son aplicados en más de un examen no reciben una clasificación consistente; es decir, para un mismo ítem el sistema recomienda que sea retenido en algunas ocasiones y eliminado en otras. Un sistema de análisis de ítems que funcione adecuadamente debiera ser consistente en sus decisiones. Una posible explicación para la falta de consistencia, es que los indicadores obtenidos, que son los habitualmente empleados para rechazar preguntas de los exámenes, varíen de forma considerable a causa de ser calculados en muestras no demasiado grandes. Con el objetivo de explorar el efecto del tamaño de la muestra en la que se aplica el examen sobre los valores de los índices de dificultad y de discriminación, hemos realizado un estudio de simulación para determinar qué tamaño ha de tener la muestra para que se pueda esperar que las características de los ítems cambien poco si los volviésemos a aplicar en otra muestra similar, de modo que podamos esperar resultados consistentes al ser aplicados en muestras similares. En este estudio aplicamos un test de 20 ítems a 10 muestras de sujetos simulados similares. Obtenemos, por ejemplo, los índices de dificultad ( $p$ ) de los 20 ítems en cada muestra. Para ver si estamos obteniendo valores similares en las 10 muestras obtuvimos las 45 correlaciones de Pearson que resultan de formar todos los pares posibles de dos columnas a partir de las 10 muestras. Si los valores de  $p$  de los 20 ítems que se obtienen en las distintas muestras son similares entre sí, la media de las 45 correlaciones será alta y tanto mayor cuanto mayor sea la similitud. El procedimiento descrito se ha aplicado a muestras de distinto tamaño (50, 100, 200 y 400 sujetos simulados) y obtuvimos el índice de dificultad ( $p$ ) y la correlación biserial puntual corregida ( $r_{jX}$ ). En la Tabla 2 se muestra el mayor, el menor y el valor medio de las 45 correlaciones ( $r$ ) obtenidas con cada tamaño muestral.

Tabla 2. Estabilidad de los índices de dificultad ( $p$ ) y de discriminación corregido ( $r_{jX}$ ) en función del tamaño muestral ( $n$ )

$n$		$r$ media	$r$ mínima	$r$ máxima
50	$p$	0,73	0,55	0,91
	$r_{jX}$	0,32	-0,02	0,67
100	$p$	0,82	0,68	0,93
	$r_{jX}$	0,42	-0,05	0,77
200	$p$	0,91	0,82	0,96
	$r_{jX}$	0,56	0,37	0,78
400	$p$	0,94	0,87	0,97
	$r_{jX}$	0,74	0,60	0,94

De la Tabla 2 pueden extraerse varias conclusiones. Las correlaciones ítem-resto del test de los 20 ítems obtenidos en una muestra de 50 sujetos pueden diferir bastante de las obtenidas en otra de iguales características (la correlación media es 0,32 y entre dos muestras se ha obtenido una correlación de solo -0,02). La correlación media entre los indicadores de dificultad es más alta (0,73). Hay que ir a una muestra de 400 sujetos simulados para que la correlación media de las correlaciones ítem-resto del test alcance ese valor. Por lo tanto, con muestras de 50 estudiantes podemos fiarnos de la ordenación que hayamos encontrado de los valores de  $p$ : los ítems que hayan resultado más fáciles es probable que resulten más fáciles en otra aplicación de similares características, pero eso no ocurre con el indicador de discriminación. Hacen falta muestras mayores para sacar conclusiones en este caso. Burton (2001) concluye que tanto el índice  $p$  como las correlaciones ítem-test son muy poco estables a no ser que se obtengan en muestras mucho mayores de las que habitualmente tenemos en los contextos educativos. Su utilidad debiera ceñirse a permitirnos comprobar las características de los ítems que resultan diagnosticados como muy buenos o muy malos. Estos últimos son los más interesantes porque pueden revelar que hay algún error en la clave de respuestas. En las siguientes secciones se presentan resultados obtenidos para éstos y otros indicadores en la muestra de exámenes recogida.

## Método

### Muestra

Se analizaron las respuestas a 35 exámenes de opción múltiple realizados en la UAM entre los cursos 2005/6 y 2008/9. Veintiséis eran exámenes de 9 asignaturas distintas de la licenciatura de Psicología. Los 9 restantes eran exámenes realizados en la Escuela Politécnica Superior y correspondían a 4 asignaturas distintas. Todos los exámenes estaban compuestos por ítems de opción múltiple con 3 alternativas. El número de ítems de los exámenes oscilaba entre 18 ítems (1 examen) y 50 ítems (5 exámenes). Como media los exámenes tenían 32 ítems. Un 34%, es decir 12 exámenes, tenían 30 ítems. El número de estudiantes que respondían a cada examen era muy variable. La muestra más pequeña era de 31 estudiantes y la mayor de 477. La mitad de los exámenes habían sido respondidos por 114 estudiantes o menos. Sólo un 25% de los exámenes habían sido respondidos por más de 202 estudiantes.

## Resultados

### Media por ítem

Si dividimos la puntuación media obtenida en el examen entre el número de ítems del examen obtenemos la media por ítem para cada examen, que vendría a ser un indicador de la facilidad media de los ítems del examen. Valores próximos a 1 nos indicarían ítems muy fáciles y próximos a 0 ítems muy difíciles. McAlpine (2002a) y Johnstone (2005) han sugerido como media de una evaluación aceptable valores entre el 0,50 y 0,60. La Tabla 3 muestra los principales resultados.

Tabla 3. Estadísticos descriptivos de la media por ítem en los 35 exámenes

	Mínimo	Máximo	Media	$S_x$
Media por ítem	0,33	0,75	0,57	0,084

La media por ítem más baja fue 0,33 y la más alta 0,75. Solo 16 de los exámenes (45,7%) obtuvieron valores comprendidos en el rango recomendable. Cinco exámenes (14,3%) obtuvieron medias inferiores a 0,5 y catorce exámenes (40%) superiores a 0,6.

### Índice de discriminación corregido

Como indicador de la discriminación de los ítems se calculó la correlación del ítem con el resto del test después de restarle la puntuación del ítem ( $r_{jX}$ ). La Tabla 4 muestra los resultados. El valor mínimo obtenido fue 0,07 y el máximo 0,39. Treinta y uno de los 33 exámenes presentan valores menores a 0,3. Catorce (40%) de los exámenes presentan una correlación media no significativa.

Tabla 4. Medias de los índices de discriminación corregidos en los 35 exámenes

	Mínimo	Máximo	Media	$S_x$
$r_{jX}$	0,07	0,39	0,21	0,07

### Coeficiente alfa de Cronbach

McAlpine (2002b) sugiere que las evaluaciones con los ítems de opción múltiple deberían tener un alfa de al menos 0,70. En nuestros datos, 17 exámenes (48,6%) presentan un alfa inferior. El alfa mínimo fue 0,32 y el máximo 0,89. Una de las variables que influye sobre el valor de  $\alpha$  es la longitud del test, si todos ellos tuviesen la misma longitud, 30 ítems, serían 19 (54,3%) los exámenes que no alcanzarían el valor propuesto. La correlación entre el coeficiente alfa y la varianza de las puntuaciones en el examen fue significativa ( $r = 0,73, p < 0,01$ ). La Tabla 5 muestra los principales resultados.

Tabla 5. Estadísticos descriptivos del coeficiente alfa de Cronbach en los 35 exámenes

	Mínimo	Máximo	Media	$S_x$
$\alpha$	0,32	0,89	0,65	0,14

### Efecto del número de formas del examen

Encontramos un efecto significativo del número de formas del examen sobre el valor de alfa<sup>[2]</sup>  $F(2, 32) = 3,744, p < 0,04$ , pero no sobre el valor de la dificultad media de los ítems  $F(2, 32) = 2,603, p > 0,05$ . La prueba de Tukey nos indica que los exámenes con 2 formas consiguen alfas significativamente superiores, pero no había diferencias significativas entre 1 ó 4, ni tampoco entre 2 ó 4 formas. La Tabla 6 muestra los resultados.

Tabla 6. Coeficiente alfa en función del número de formas del examen

	N	Media	$S_x$
1 forma	12	0,57*	0,15
2 formas	15	0,71*	0,12
4 formas	8	0,66	0,10
Total	35	0,65	0,14

\*  $p < 0,05$

### Posición de la opción correcta

Attali y Bar-Hillel (2003) encuentran que hay una tendencia a colocar la opción correcta en las posiciones intermedias y que como consecuencia los ítems en los que la opción correcta ocupa esta

[2] También se obtuvo un efecto, tal y como era de esperar, sobre  $r_{jX}$ .

posición tienden a resultar más fáciles y menos discriminativos. La Tabla 7 muestra la distribución de la opción correcta en una muestra formada por 669 ítems de 3 opciones que correspondían a 21 exámenes diferentes. Como se puede observar la distribución encontrada no es uniforme. La opción correcta se coloca en la primera posición con menor frecuencia que en la segunda y la tercera  $\chi^2$  (2,  $N = 669$ ) 6,47,  $p < 0,04$ ).

Tabla 7. Frecuencia con que la opción correcta ocupa cada opción del ítem. ( $N = 669$ )

Posición de la opción correcta	nº de ítems
Primera	192
Segunda	239
Tercera	238

No se obtuvieron efectos de la posición de la opción correcta sobre la dificultad ( $p$ ) ni sobre la discriminación ( $r_{jX}$ ). Los resultados se pueden observar en la Tabla 8.

Tabla 8. Valores medios de  $p$  y  $r_{jX}$  en función de la posición de la opción correcta

	Posición de la opción correcta	Media	$S_x$
$p$	Primera	0,55	0,23
	Segunda	0,54	0,23
	Tercera	0,55	0,23
$r_{jX}$	Primera	0,20	0,16
	Segunda	0,21	0,16
	Tercera	0,20	0,15

### Análisis de los errores de redacción

Sierra y Ponsoda (en preparación) extrajeron de forma automática 40 características de los enunciados de los ítems relacionadas con la calidad de redacción, según las recomendaciones de Haladyna (p. ej: la varianza del número de caracteres de las opciones, la presencia de determinantes en los distractores, las palabras compartidas entre el enunciado y la opción correcta, etc.). Estudiaron mediante regresión lineal y un algoritmo genético qué variables de las 40 ayudaban a predecir las características psicométricas de los ítems (sus valores  $p$  y  $r_{jX}$ ). La capacidad predictora de los indicadores resultó muy pobre, no encontrándose ninguno claramente asociado a las propiedades psicométricas, en línea con otros resultados (como los de Boldt, 1998).

### Conclusión

Es práctica común para los profesores poner un examen, puntuarlo y olvidarse de la evaluación hasta la siguiente convocatoria. Sin embargo, el análisis de las respuestas de los estudiantes a los exámenes es esencial para mejorar la calidad de la evaluación. Este análisis es especialmente importante para los ítems de respuesta seleccionada ya que estos suelen ser reutilizados en cursos siguientes. Con una reflexión formal sobre las respuestas a los exámenes además de mejorar los ítems con los que se realiza la evaluación se podría mejorar el rendimiento de los estudiantes. Las preguntas que mostrasen ser más adecuadas deberían usarse no solo para la evaluación, también en las etapas de aprendizaje. Sin embargo, muchos profesores no son especialistas en medición y no analizan rutinariamente sus evaluaciones usando los estándares asociados a la fiabilidad y la validez. Un profesor no tiene la habilidad de escribir buenos ítems sin recibir instrucción. Muchos ítems de opción múltiple están mal

escritos, intentan medir un objetivo para el que no son adecuados, contienen pistas de la respuesta correcta o están redactados de modo ambiguo. Un análisis apropiado de las respuestas de los estudiantes a las evaluaciones es un paso esencial para mejorar la calidad de la evaluación en si misma, así como de la docencia y del aprendizaje.

### Referencias

- Attali, Y. y Bar-Hillel, M. (2003). Guess Where: The Position of Correct answers in Multiple-Choice Tests Items as a Psychometric Variable. *Journal of Educational Measurement*, 40 (2), 109-128.
- Boldt, R. F. (1998). *GRE Analytical Reasoning Item Statistics Prediction Study*. GRE Board Report N° 94-02P.
- Burton, R. F. (2001). Quantifying the Effects of Chance in Multiple-Choice and True/False Tests: Item Selection and Guessing of Answers. *Assessment and Evaluation in Higher Education*, 26, 41-50.
- Haladyna, T. M., Downing, S. M. y Rodríguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Johnstone, A. (2005). *Evaluation of Teaching: Practice Guide*. The Higher Education Academy. Physical Sciences Centre.
- McAlpine, M. (2002a). *Principles of Assessment*. Computer Assisted Assessment Centre. University of Luton.
- McAlpine, M. (2002b). *A Summary of Methods of Item Analysis*. Computer Assisted Assessment Centre. University of Luton.
- Muntinga, J. H. J. y Schuil, H. A. (2007). Effects of Automatic Item Eliminations Based on Item Test Analysis. *Advances in Physiology Education*, 31, 247-252.
- Sierra, A. y Ponsoda, V. (en preparación). Calidad de la redacción de los ítems de opción múltiple y sus propiedades psicométricas.

# ESTUDIO EXPERIMENTAL SOBRE LA INFLUENCIA DE LA REDACCIÓN DE LOS ÍTEMS EN LAS RESPUESTAS DE LOS SUJETOS

M. Rocío Garrido, Manuela A. Fernández Borrero, Estrella Villalba,  
María de Vega y Pedro J. Pérez Moreno  
Universidad de Huelva

En el presente trabajo se muestra el estudio experimental desarrollado sobre la influencia de la redacción de los ítems en las respuestas de los sujetos. Se empleó para ello un diseño transversal multifactorial con dos variables independientes, el enunciado positivo o negativo de los ítems y la presencia o no de cuantificadores en estos enunciados. Han participado un total de 92 sujetos, repartidos por igual a las condiciones experimentales del estudio. La conclusión principal a la que se ha llegado es que en los ítems donde se hace uso de la negación el efecto es mayor si están presentes los cuantificadores también en el enunciado del ítem.

La investigación desarrollada sobre la influencia de la redacción de los ítems en la respuesta de los sujetos es un tema de relevancia en Psicología, donde las escalas son instrumentos utilizados para generar conocimiento científico. Según García, Martín, Martínez y Puerto (1998), “un aspecto de gran trascendencia es la redacción de los ítems en las escalas de medida, ya que su redacción puede condicionar el comportamiento del sujeto”.

Las Variables Independientes son el tipo de enunciado de los ítems (afirmativo/negativo) y la utilización o no del cuantificador muy/mucho. Se han elegido estas variables por las recomendaciones de su no utilización en psicometría: “en tests de ejecución típica hay que evitar el uso de cuantificadores, porque pueden ser subjetivos” (Sánchez, 2006), “En general es preferible no abusar de los enunciados negativos” (Barbero, Vila y Suárez, 2003).

La Variable Dependiente es el grado de acuerdo con los ítems, medida a través de cuatro tipos del mismo cuestionario sobre bienestar (condiciones experimentales).

Las hipótesis son, que el grado de desacuerdo cuando los ítems están en negativo, será mayor que el grado de acuerdo cuando están en positivo, y, que en los ítems en con negación el efecto será mayor si los cuantificadores están presentes.

García, Martín, Martínez y Puerto (1998) llevaron a cabo una investigación sobre el efecto de la redacción de los ítems. Sus variables eran el sentido de la frase, afirmativo o negativo (sin utilizar “no”, invirtiendo palabras), el uso o no de la negación y el uso o no del cuantificador poco o mucho. Los resultados fueron que la inversión del sentido de los enunciados (sin la inclusión del “no”) no genera diferencias significativas, que la aparición explícita “no” genera más desacuerdo que acuerdo genera la misma frase afirmativa, y, que “la existencia de un cuantificador genera diferencias con respecto a su ausencia”.

## Método

### *Participantes*

Los 92 sujetos (23 por condición experimental) pertenecen a los turnos mañana y tarde de primero y segundo de Psicología de la Universidad de Huelva. El criterio de selección ha sido por accesibilidad.

La edad está entre 18 y 24 años.

### *Diseño*

Es un diseño transversal factorial 2x2, con cuatro condiciones experimentales: 1) ítems en afirmativo sin cuantificador; 2) ítems afirmativos con cuantificador; 3) ítems negativos sin cuantificador; 4) ítems negativos con cuantificador.

El orden en el que se repartieron los cuestionarios fue adjudicado aleatoriamente a través de SPSS.

### *Instrumentos*

Se han empleado el *Cuestionario de Aproximaciones a la felicidad* de Peterson (2002), diferenciándose las cuatro condiciones experimentales (anexo): 1) escala con todos los ítems formulados de modo afirmativo y sin presencia de cuantificador; 2) escala con todos los ítems formulados de modo afirmativo con presencia de cuantificador; 3) escala con todos los ítems formulados de modo negativo y sin cuantificador; 4) escala con todos los ítems formulados de modo negativo y con presencia de cuantificador.

El cuestionario está compuesto por 15 ítems tipo Likert de grado de acuerdo, con valores entre 1 y 6, siendo 1 total desacuerdo y 6 acuerdo total. Se han eliminado tres ítems para adaptarlo a la investigación.

Hay dos ítems sociodemográficos (sexo y edad) y se ha usado el software SPSS 14 para la informatización y análisis estadísticos.

### *Procedimiento*

La fase de administración del instrumento se llevó a cabo en el contexto del aula, informando acerca del modo de cumplimentar el cuestionario. Para no incluir sesgos, no se informó sobre el objetivo real, indicándoles que versaba sobre bienestar.

El cuestionario fue autocumplimentado. Las investigadoras presentes (dos), fueron las mismas siempre.

## Resultados

Recodificados los ítems necesarios, se han utilizado las puntuaciones promedio de cada sujeto como variable dependiente.

El análisis ha sido factorial de varianzas (ANOVA) y comparaciones *a priori* de los efectos simples. A continuación, se exponen y analizan los datos.

Tabla 1. Estadísticos descriptivos

	Cuantificador	Media	Desv. típ.	N
Afirmativo	Sin	5,4145	,86829	23
	Con	4,9971	,81041	23
	Total	5,2058	,85685	46
Negativo	Sin	5,7014	,66935	23
	Con	5,9739	,92654	23
	Total	5,8377	,81099	46
Total	Sin	5,5580	,78017	46
	Con	5,4855	,99228	46
	Total	5,5217	,88838	92

La tabla 1 resume las variables de sujetos por condición experimental. La mayor media se corresponde con la condición negativo con cuantificador.

Según la prueba de Levene ( $F(3, 88) = 0.472, p = 0.702$ ), nada se opone a aceptar la homocedasticidad, que permite interpretar el estadístico F de ANOVA.

Tabla 2. Efectos inter-sujetos

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
Modelo corregido	12,041(a)	3	4,014	5,908	,001
Intersección	2805,043	1	2805,043	4129,331	,000
VI1	9,183	1	9,183	13,519	,000
VI2	,121	1	,121	,178	,674
VI1 * VI2	2,736	1	2,736	4,028	,048
Error	59,778	88	,679		
Total	2876,862	92			
Total corregida	71,819	91			

a R cuadrado = ,168 (R cuadrado corregida = ,139)

La tabla 2 muestra el valor de F y la significación estadística de los efectos principales y del efecto de interacción. El efecto de interacción resulta significativo ( $F$  de ANOVA (1,88)= 4.028,  $p = 0.048$ ), representado en los Gráficos 1 y 2 del Anexo.

Esta significación permite analizar los efectos simples. Se pueden hacer dos conjuntos de comparaciones ortogonales, analizando el efecto principal de la variable afirmativo-negativo y los efectos simples de la variable cuantificador, y viceversa.

Tabla 3. Coeficientes de los contrastes

Contraste	Condición experimental			
	Afirmativo sin cuantificador	Afirmativo con cuantificador	Negativo sin cuantificador	Negativo con cuantificador
1	1	1	-1	-1
2	1	-1	0	0
3	0	0	1	-1

Tabla 4. Pruebas para los contrastes

		Contraste	Valor del contraste	Error típico	t	gl	Sig. (bilateral)
VD	Igualdad varianzas	1	-1.2638	.34371	-3.677	88	.000
		2	.4174	.24304	1.717	88	.089
		3	-.2725	.24304	-1.121	88	.265

La tabla 3 muestra los coeficientes de las comparaciones que vamos a realizar para el primer grupo de comparaciones ortogonales, éstos se refieren al efecto principal de afirmativo-negativo (contraste 1), los efectos simples del cuantificador cuando los ítems están redactados en afirmativo (contraste 2), y los efectos simples del cuantificador cuando los ítems están redactados en negativo (contraste 3). La Tabla 4 muestra la significación para cada uno de estos contrastes resultando significativo solo el primero ( $p = 0.000$ ).

Tabla 5. Coeficientes contrastes

Contraste	C.experimental			
	Afirmativo sin cuantificador	Afirmativo con cuantificador	Negativo sin cuantificador	Negativo con cuantificador
1	1	-1	1	-1
2	1	0	-1	0
3	0	1	0	-1

Tabla 6. Pruebas contrastes

		Contraste	Valor del contraste	Error típico	t	gl	Sig. (bilateral)
VD	Igualdad varianzas	1	.1449	.34371	.422	88	.674
		2	-.2870	.24304	-1.181	88	.241
		3	-.9768	.24304	-4.019	88	.000
	No igualdad varianzas	1	.1449	.34371	.422	83.836	.674
		2	-.2870	.22860	-1.255	41.324	.216
		3	-.9768	.25667	-3.806	43.234	.000

Las tablas 5 y 6 corresponden al otro grupo ortogonal. Los contrastes expuestos en la Tabla 5, atañen al efecto principal de la variable cuantificador (contraste 1), los efectos simples de la variable afirmativo-negativo cuando los ítems están redactados sin cuantificador (contraste 2), y los efectos simples de la variable afirmativo-negativo cuando los ítems están redactados con cuantificador (contraste 3). La Tabla 6 nos muestra la significación para cada uno de estos contrastes, siendo estadísticamente significativa la comparación 3 ( $p = 0.000$ ).

## Discusión

Tras el análisis de los datos, debemos rechazar nuestra hipótesis de que “el grado de desacuerdo cuando los ítems están enunciados en negativo, será mayor que el grado de acuerdo cuando están en positivo”, ya que se refiere a un efecto principal y no se puede concluir sobre él porque la interacción entre ambas variables es significativa. No obstante, con los datos sí podemos mantener nuestra hipótesis acerca del efecto simple de la variable afirmación-negación: “en los ítems en los que hagamos uso de la negación el efecto será mayor si los cuantificadores están presentes”, ya que el contraste resulta estadísticamente significativo.

Esto nos da pie a proponer sugerencias a la hora de construir ítems, tales como que se intente no enunciar los ítems en negativo, sobretodo si éste presenta cuantificadores, porque esto crea una tendencia de respuesta de los sujetos independientemente del constructo que se mida (influenciada por la redacción del ítem en vez del lugar en el que se sitúe el sujeto con respecto al constructo que se mide).

En contraste con los resultados de la investigación llevada a cabo por García, Martín, Martínez y Puerto (1998), coinciden los referentes a que la aparición de la negación explícita “no” en la redacción del ítem genera más desacuerdo que acuerdo genera la misma frase afirmativa (sin tener en cuenta la interacción), pero no coinciden con que “la existencia de un cuantificador, por sí mismo, parece generar diferencias con respecto a su ausencia” (García, Martín, Martínez y Puerto, 1998), ya que no hemos encontrado diferencias significativas en las respuestas de los sujetos en los ítems con cuantificador y sin cuantificador. No obstante, esta diferencia puede deberse a que nosotros solo hemos empleado el cuantificador “muy/mucho” y ellos emplearon los cuantificadores “muy/mucho y poco”.

Las limitaciones que presenta el diseño -y que deben tenerse en cuenta a la hora de replicarse este experimento- son que la muestra ha sido seleccionada por accesibilidad, lo que puede llegar a afectar a la validez externa y, por otra parte, que el tamaño de la muestra no es demasiado amplio, por lo que sería mejorable este aspecto de la investigación.

## Referencias

- Barbero, M. I., Vila, E. y Suárez, J. C. (2003). *Psicometría*. Madrid: Universidad de Educación a Distancia (UNED).
- García, A., Martín, M. J., Martínez, R. y Puerto, A. (1998). *Efectos de la redacción de ítems en las respuestas de los encuestados: una aplicación con escalas tipo Likert*. Investigación del comportamiento. Sevilla. Innovaciones metodológicas y estratégicas de docencia. Cáp.7, 65-75.
- Peterson, C. (2003). *Cuestionario de Aproximaciones a la felicidad. Values in Action Institute (VIA)*, Extraído el 3 de Diciembre de 2006 desde [http://www.authentic happiness.sas.upenn.edu/tests/SameAnswers\\_t.aspx?id=266](http://www.authentic happiness.sas.upenn.edu/tests/SameAnswers_t.aspx?id=266)
- Sánchez, M. (2006). *Psicometría, Tema 2: Construcción de tests*. Apuntes de clase, Facultad de Ciencias de la Educación Universidad de Huelva.

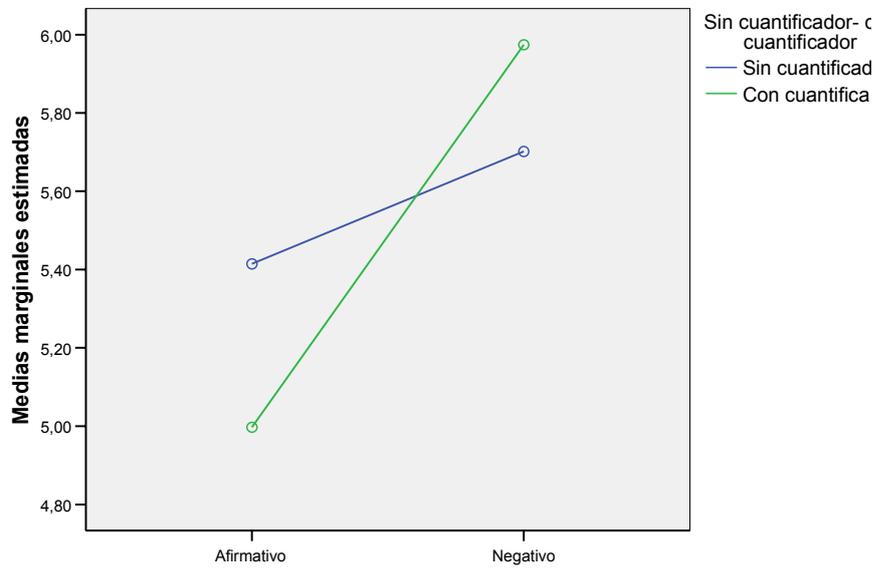
## ANEXO 1

Afirmativo sin cuantificador	Afirmativo con cuantificador	Negativo sin cuantificador	Negativo con cuantificador
Sea lo que sea lo que esté haciendo, el tiempo se me pasa deprisa	Muy deprisa	No se me pasa	No se me pasa muy deprisa
Busco situaciones que desafíen mis destrezas y capacidades	Desafíen mucho	No desafíen	No desafíen mucho
Al decidir qué hacer tengo en cuenta si beneficiará a otras personas	Tengo muy en	No tengo	No tengo muy
Ya sea en el trabajo o en el tiempo libre, normalmente estoy abstraído en lo que hago y sin ser consciente de mí mismo	Muy abstraído	No estoy	No estoy muy
Frecuentemente estoy concentrado en lo que hago	Muy concentrado	No estoy	No estoy muy
Hago todo lo posible para sentirme bien	Muy bien	No hago	No hago... muy bien
Al decidir qué hacer, tengo en cuenta si puedo entregarme a ello	Muy en cuenta	No tengo	No tengo muy
Rara vez me distraigo con las cosas que hay a mi alrededor	Distraigo mucho	No me distraigo	No me distraigo mucho
Me siento responsable de hacer del mundo un lugar mejor	Muy responsable	No me	No me siento muy
Al decidir lo que hacer, tengo en cuenta si será agradable	Muy en cuenta	No tengo	No tengo muy
Lo que hago importa a la sociedad	Importa mucho	No importa	No importa mucho
Comparto el mensaje de la frese: "no dejes para mañana lo que puedas disfrutar hoy"	Comparto mucho	No comparto	No comparto mucho
Me gusta hacer cosas que estimulen mis sentidos	Gusta mucho	No me	No me gusta mucho
He empleado tiempo pensando en lo que la vida significa y en cómo encajo yo en ella	Mucho tiempo	No he	No he empleado mucho
Para mí, la buena vida es una vida llena de placeres	Muy llena	No es	No es una vida muy

### ANEXO 2: Gráficos

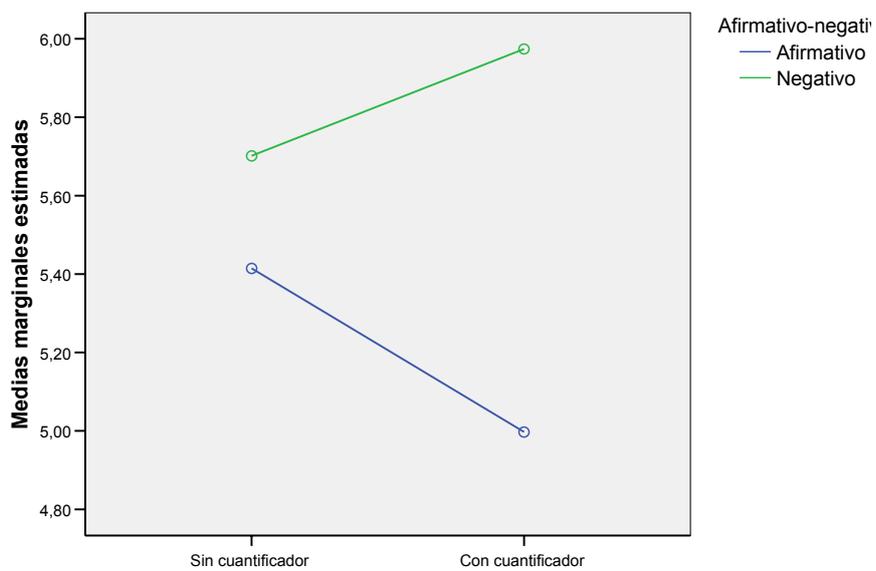
1

#### Medias marginales estimadas de mediaVD



2

#### Medias marginales estimadas de mediaVD



# SESIONES SUPERVISADAS Y NO SUPERVISADAS EN LA ADMINISTRACIÓN DE SUBTEST DE ANCLAJE A TRAVÉS DE INTERNET

Javier López-Cuadrado, Tomás A. Pérez Fernández, José A. Vadillo y Julián Gutiérrez  
Universidad del País Vasco

Se ha calibrado un banco de 250 ítems de nivel de euskara según el modelo logístico de tres parámetros, con objeto de incluirlo en un generador de test adaptativos informatizados de ingreso del sistema educativo Hezinet. Siguiendo un diseño de anclaje de grupos no equivalentes con ítems comunes, se han repartido los ítems en 6 subtest, cada uno de los cuales se ha administrado a una muestra de más de 500 sujetos. Posteriormente se han obtenido y equiparado estimaciones de los parámetros de dificultad, discriminación y pseudoacierto. Esta comunicación se centra en la administración de los subtest, que ha sido realizada a través de Internet mediante una herramienta informática diseñada para tal efecto. Durante esta fase se han distinguido dos tipos de administraciones: las sesiones supervisadas, en las que una persona encargada controlaba la realización de las pruebas, y las sesiones no supervisadas, en las que voluntarios anónimos completaron las pruebas desde sus casas o centros de trabajo sin ningún tipo de vigilancia. Se ha efectuado un análisis de funcionamiento diferencial de los ítems para identificar diferencias entre las administraciones supervisadas y no supervisadas, habiéndose detectado 62 ítems con posible funcionamiento diferencial.

Hezinet (Pérez, 2000) es un sistema hipermedia adaptativo para el aprendizaje de la lengua vasca que ha servido de punto de partida para el trabajo que se presenta. Este sistema está en uso desde el año 2000 en más de 60 centros de enseñanza del euskara repartidos por todo el mundo. Al convertirse Hezinet en un producto comercial, los autores, interesados en mejorar la arquitectura mediante la inclusión de nuevas funcionalidades, rebautizaron el sistema. A día de hoy la evolución de Hezinet se conoce con el nombre de ELSA (Armendariz, et al., 2003), acrónimo de *E-Learning Systems Architecture*. Se trata de una arquitectura orientada a la evaluación, que incorpora nuevos elementos con respecto a Hezinet (López-Cuadrado, Pérez, et al., 2008), entre los que cabe destacar la herramienta de autor ADISTI (López-Cuadrado, Armendariz & Pérez, 2003) para la edición y el almacenamiento de ítems, el asistente para la calibración de bancos de ítems CALLIE (López-Cuadrado, Armendariz, Pérez & Arruabarrena, 2008), y el generador de TAI de ingreso que obtiene el nivel de los nuevos usuarios (López-Cuadrado, 2008).

El generador de TAI de ingreso a ELSA se alimenta de un banco que inicialmente tenía 252 ítems, de los que 46 fueron retirados durante el proceso de calibración. La siguiente sección está dedicada precisamente a la calibración según el modelo 3PL del banco, para la que se utilizó una muestra de más de 500 respuestas por ítem. Después, se presentarán algunos resultados obtenidos tras el proceso, dando lugar a la sección de conclusiones, con la que terminará esta comunicación.

## *El proceso de calibración*

Con objeto de establecer la base desde la que realizar un diseño de anclaje homogéneo, se consultó a varios profesores y filólogos de la lengua vasca. A cada experto se le solicitó tres valoraciones por cada ítem: la respuesta correcta, la destreza lingüística evaluada y el nivel de dificultad. La primera sirvió como elemento de control, mientras que las otras dos constituyeron las dimensiones según las que clasificar los ítems. Se usó una muestra de al menos 5 valoraciones por ítem para establecer una

clasificación conforme a las dos dimensiones citadas. En este punto, 2 ítems fueron retirados del banco por no ajustarse a ninguna de las destrezas lingüísticas contempladas.

Los 250 ítems restantes se distribuyeron siguiendo un diseño de anclaje de grupos no equivalentes de ítems comunes (Kolen & Brennan, 1995). Se definieron 6 subtest, cada uno con 60 ítems, 22 de los cuales eran comunes. El reparto de ítems entre los conjuntos de anclaje y los propios de cada subtest se realizó lo más homogéneo posible, tratando de mantener las proporciones de dificultad y destrezas lingüísticas identificadas para el banco de ítems en su totalidad. El algoritmo utilizado para determinar los ítems del conjunto de anclaje fue una variante del procedimiento descrito por Olea, Abad y Ponsoda (2002) adaptada al caso concreto del banco de ítems para el TAI de ingreso (López-Cuadrado, 2008).

Los subtest se administraron por ordenador, no sólo para eliminar problemas de logística y organización, sino además para que la aplicación de ítems fuese lo más parecida posible al funcionamiento del TAI (López-Cuadrado, Armendariz & Pérez, 2005). Se contemplaron dos tipos de administración: las sesiones supervisadas (SV), que se realizaron bajo el control de una persona encargada en centros de enseñanza secundaria y universitaria, y las sesiones no supervisadas (NSV), en las que voluntarios anónimos completaron por su cuenta las pruebas a través de Internet. En total se completaron 2341 administraciones SV, de las que 2268 se consideraron válidas. Por su parte, para el desarrollo de las sesiones NSV se enviaron mensajes a listas de distribución relacionadas con el mundo universitario, el deporte, la cultura y la lengua vasca, invitando a sus suscriptores a que realizaran un subtest a través de Internet. Finalmente, se validaron 975 de las 1635 sesiones NSV registradas, correspondientes a voluntarios anónimos que, tanto por teléfono como por correo electrónico, aseguraron que las condiciones en que realizaron el test fueron adecuadas. Tras la fase de administración de los subtest se dispuso de una muestra de 3243 administraciones válidas ( $n_1=545$ ,  $n_2=542$ ,  $n_3=537$ ,  $n_4=539$ ,  $n_5=539$ ,  $n_6=541$ ), de las que, para cada subtest, al menos 365 correspondían a sesiones SV y 150 a sesiones NSV.

Los procesos de análisis de unidimensionalidad, estimación de parámetros, ajuste al modelo 3PL y equiparación se realizaron mediante software específico. El tándem PRELIS-LISREL se utilizó para comprobar que el banco de ítems era unidimensional; el paquete estadístico SPSS sirvió para efectuar los análisis de fiabilidad, como resultado de los que se retiraron del banco 46 ítems cuya correlación ítem-subtest era muy próxima a cero; los parámetros de los ítems se estimaron mediante el procedimiento bayesiano MAP implementado en el programa XCALIBRE; y una simple hoja de cálculo MS EXCEL permitió equiparar los parámetros mediante el método media-sigma (Marco, 1977).

## Resultados

Tras efectuar la calibración se retiraron del banco 48 ítems (el 19% del total). Los parámetros de los 204 ítems remanentes proporcionan un grado de ajuste aceptable al modelo 3PL. Los valores de discriminación están comprendidos en el intervalo (0.6, 1.7), los de dificultad en el intervalo (-2.82, 1.66), y los de pseudoacierto en el intervalo (0.11, 0.24). Aunque no hay ítems de dificultad extrema, más del 87% tienen un parámetro  $b$  negativo, lo que significa que los TAI producidos a partir de este banco no van a ser capaces de medir habilidades superiores a un cierto valor. La función de información del banco de ítems, que se muestra en la figura 1, ofrece una cota de la precisión que puede obtenerse mediante los TAI.

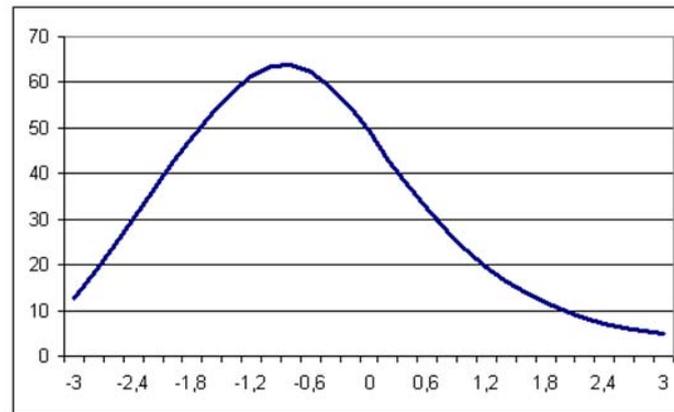


Figura 1. Función de información del banco de ítems

Como puede observarse, su máximo está cerca del valor de dificultad  $-0,8$ , de ahí que los TAI trabajarán mejor si los examinados tienen un nivel bajo o medio del idioma. Afortunadamente, éste es precisamente el perfil de usuario esperado por el TAI de ingreso a Hezinet. Sin embargo, para conseguir un TAI de ingreso altamente operativo, será necesario calibrar y añadir al banco nuevos ítems de dificultad alta. Durante las pruebas piloto realizadas al sistema, se ha comprobado que el TAI necesita cerca de 20 ítems para obtener una estimación para las habilidades comprendidas en el intervalo  $[-1,5, 0,0]$ . Sin embargo, el número de ítems necesarios aumenta cuando se quieren estimar habilidades mayores, hasta el punto de que se aplica prácticamente todo el banco cuando la habilidad objetivo es mayor que 2, tal y como puede verse en la figura siguiente.

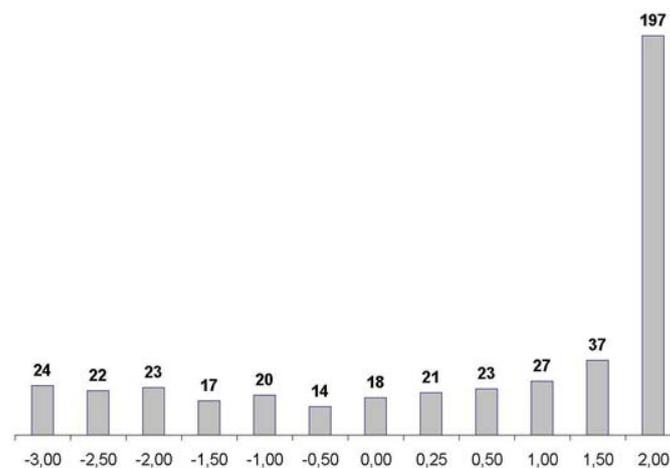


Figura 2. Número de ítems utilizados por el TAI para estimar una habilidad

Antes de incluir el banco calibrado en el generador de TAI de ingreso al sistema, se han efectuado análisis de funcionamiento diferencial de los ítems (FDI) para buscar disparidades entre las muestras SV y NSV. El primer estudio se basa en las recomendaciones de Renom y Doval (1999), quienes sugieren estimar dos veces los parámetros de los ítems, empleando en cada caso una parte de los examinados, para obtener y comparar dos curvas características por cada ítem. Estas superficies se han calculado mediante sumas de Riemann sobre una división del intervalo de habilidades en subintervalos de tamaño 0.2. Se han identificado 39 ítems con posible FDI. Adicionalmente, se ha realizado una clasificación en niveles ETS mediante el software ItemQual y se han encontrado otros 23 ítems con FDI destacable. Estos ítems no se han retirado del banco porque la detección de FDI puede haberse debido a la diferencia de tamaños muestrales en ambos conjuntos, y no a errores propios de los ítems.

## Conclusión

Se ha presentado el proceso de calibración del banco de ítems para el TAI de ingreso al sistema de aprendizaje de la lengua vasca Hezinet, que obtiene estimaciones aceptables de habilidades bajas y medias con sólo 20 ítems, si bien aplica prácticamente todo el banco cuando la habilidad objetivo es alta. En consecuencia, se hace necesario calibrar e incorporar más ítems difíciles al banco con el fin de que el TAI funcione ante cualquier patrón de respuesta. El hecho de que el 30% de los ítems que permanecen en el banco puedan ofrecer FDI también obliga a desarrollar y calibrar más ítems.

Se han dado los primeros pasos en el estudio de la precisión y validez del generador de TAI, pero aún queda mucho trabajo por delante. Se está trabajando en un módulo de simulación que estará integrado en la nueva arquitectura *e-learning* y que generará miles de sujetos virtuales para administrarles TAI simulados y, por ejemplo, evaluar la convergencia del error estándar y el sesgo para diferentes valores discretos de habilidad; contrastar las frecuencias de aplicación de cada ítem para cada método de control de la exposición; medir la precisión del TAI mediante simulación de pruebas con criterio de parada fijo de diferentes tamaños y definir así la longitud más adecuada para una prueba; determinar cuántos ítems hacen falta para alcanzar un error estándar determinado; o determinar la proporción de sujetos que se clasifican en el nivel educativo correcto. También se prevé utilizar los datos de las administraciones de subtest para estudiar la validez de los TAI. Así, por ejemplo, puede establecerse una clasificación ordenada de formación a partir de la información que se solicitó a los sujetos durante la calibración (nivel educativo, posesión o no de títulos, etcétera), y estudiar la capacidad del TAI para discriminar los niveles del idioma para los diferentes niveles educativos. También se prevé estudiar el comportamiento de los ítems sobre factores como el sexo o el domicilio del examinado, que en este caso tiene especial interés de cara a comprobar la dependencia de dialectos en los resultados del TAI.

Por último, se necesita aplicar muchos más TAI de ingreso a personas cuyo nivel del idioma es conocido *a priori* para poder ajustar mejor los puntos de corte entre los cursos de Hezinet, es decir, para refinar cuáles son los intervalos de habilidad en términos de la TRI que se corresponden con los diferentes niveles del sistema *e-learning*.

## Referencias

- Armendariz, A. J., López-Cuadrado, J., Tapias, A., Villamañe, M., Sanz-Lumbier, S., & Sanz-Santamaría, S. (2003). *Learning environments should follow standards: ELSA does*. Paper presented at the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education (E-Learn 2003), Phoenix, Arizona (USA).
- Kolen, M. J. & Brennan, R. L. (1995). *Test equating: methods and practices*. New York (USA): Springer-Verlag.
- López-Cuadrado, J. (2008). *Evaluación mediante test adaptativos informatizados en el contexto de un sistema adaptativo para el aprendizaje de la lengua vasca (Tesis Doctoral)*. San Sebastián (España): Universidad del País Vasco (UPV-EHU).
- López-Cuadrado, J., Armendariz, A. J. & Pérez, T. A. (2003). ADISTI: an authoring tool for creating and managing exercises in e-learning systems. In A. Méndez Vilas, J. A. Mesa González & J. Mesa González (Eds.), *Advances in technology-based education: towards a knowledge-based society* (Vol. 3, pp. 1555-1559). Badajoz (España): Junta de Extremadura (CECT).

- López-Cuadrado, J., Armendariz, A. J. & Pérez, T. A. (2005). A supporting tool for the adaptive assessment of an e-learning system. In A. Méndez Vilas, B. Gonzalez Pereira, J. Mesa González & J. A. Mesa González (Eds.), *Recent research developments in learning technologies* (Vol. 1, pp. 295-299). Cáceres (España): Formatex Research Center.
- López-Cuadrado, J., Armendariz, A. J., Pérez, T. A. & Arruabarrena, R. (2008). *Helping tools for item bank calibration and development of computerized adaptive tests*. Paper presented at the International Technology, Education, and Development Conference (INTED2008), Valencia (España).
- López-Cuadrado, J., Pérez, T. A., Sanz-Santamaría, S., Armendariz, A. J., Gutiérrez, J. & Vadillo, J. Á. (2008). Some Innovations for Keeping an e-Learning System Alive. In M. Iskander (Ed.), *Innovative techniques in instruction technology, e-learning, e-assessment and education* (pp. 223-228). New York (USA): Springer Science+Business Media B.V.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.
- Olea, J., Abad, F. J. & Ponsoda, V. (2002). Elaboración de un banco de ítems, predicción de la dificultad y diseño de anclaje. *Metodología de las Ciencias del Comportamiento, Volumen Especial*, 427-430.
- Pérez, T. A. (2000). *Un hiperentorno adaptativo para el aprendizaje instructivo/ constructivo (Tesis Doctoral)*. San Sebastián (España): Lenguajes y Sistemas Informáticos (Universidad del País Vasco UPV-EHU).
- Renom, J. & Doval, E. (1999). Tests adaptativos informatizados: estructura y desarrollo. In J. Olea, V. Ponsoda & G. Prieto (Eds.), *Tests informatizados: fundamentos y aplicaciones* (pp. 127-162). Madrid (España): Ediciones Pirámide.

# ANÁLISIS FACTORIAL CONFIRMATORIO DE LA *STATISTICAL ANXIETY SCALE*

Lorena Pérez-Gómez, José Carmona y José Luís Sánchez Ramos  
Universidad de Huelva

El objetivo del trabajo es presentar evidencias de la validez de la *Statistical Anxiety Scale* (SAS de Vigil-Colet, Lorenzo-Seva, y Colon, *Psicothema* 2008). Según sus autores la SAS tiene una estructura compuesta de 3 subescalas: Ansiedad hacia los Exámenes, Ansiedad a pedir Ayuda y Ansiedad a la Interpretación de datos estadísticos. La estructura factorial asociada a esta composición está avalada por los análisis factoriales exploratorios realizados por Vigil-Colet et al. En este trabajo se realiza un análisis factorial confirmatorio (AFC) de las respuestas a la SAS de una muestra compuesta por 195 estudiantes de 1º de Psicología y de 1º de Enfermería. Los resultados del AFC muestran un ajuste aceptable de los datos a la estructura factorial probada cuando se permite la correlación entre los errores asociados a tres ítems de la subescala Ansiedad a pedir Ayuda ( $\chi^2/df = 1,62$ ;  $CFI = ,94$ ;  $RMSEA = ,06$ ). Se propone una justificación para la existencia de esa asociación entre errores basada en ciertas diferencias de contenido entre esos tres ítems y los otros cinco ítems de la subescala.

Las investigaciones relativas a la predicción del rendimiento académico han ido evolucionando desde estudios centrados casi exclusivamente en el análisis de los efectos de las habilidades cognitivas hasta estudios que incorporan variables propias del dominio afectivo. En el ámbito de las asignaturas con contenido estadístico el rendimiento parece estar relacionado, entre otros factores, con la extraversión (Furnham y Chamorro-Premuzic, 2004), la motivación (Tremblay, Gardner, y Heipel, 2000), las actitudes hacia la estadística (Nasser, 2004) o la ansiedad hacia la estadística (Onwuegbuzie y Seaman, 1995).

La ansiedad hacia la estadística es una de las características de los estudiantes con mayor impacto en el proceso de enseñanza-aprendizaje de asignaturas con contenido estadístico. De hecho se han demostrado sus efectos no sólo en el rendimiento en los exámenes de las asignaturas, sino también en la percepción que tienen los estudiantes de sus competencias (Hanna y Dempster, 2009) o en la procrastinación de tareas académicas (Onwuegbuzie, 2004). La importancia de la ansiedad hacia la estadística en el desarrollo de la docencia de estas asignaturas es aún más evidente cuando se tiene en cuenta que el porcentaje de alumnos que experimentan este tipo de ansiedad suele ser muy elevado (hasta un 80%, según estimaciones de Onwuegbuzie y Wilson, 2003).

Zeidner (1990) define la ansiedad hacia la estadística como una reacción emocional, caracterizada por una gran preocupación, pensamientos intrusivos, desorganización mental, tensión y excitación psicológica, producida al enfrentarse a contenidos estadísticos o a situaciones de enseñanza o evaluación relacionadas con la estadística. Su carácter situacional parece ser una de las propiedades definitorias de la ansiedad hacia la estadística y constituye el fundamento de la dimensionalidad de este constructo, de manera que es habitual distinguir entre ansiedad hacia el examen, ansiedad en clase o ansiedad al resolver problemas, entre otros tipos de ansiedad hacia la estadística que pueden surgir en situaciones académicas específicas (Carmona, 2004).

En una revisión de la literatura, Onwuegbuzie y Wilson (2003) señalan algunos antecedentes de la ansiedad hacia la estadística, entre los que destacan la experiencia previa en matemáticas (Baloglu, 2003) y los estilos de aprendizaje (Onwuegbuzie, 1998). Por otra parte, diversas investigaciones han mostrado una relación inversa entre la ansiedad y las actitudes hacia la estadística (Mji y Onwuegbuzie, 2004).

Uno de los objetivos básicos en este campo es el desarrollo de un instrumento de medida válido y fiable de la ansiedad hacia la estadística. Entre los instrumentos usados para medir este tipo de ansiedad destaca el *Statistical Anxiety Rating Scale* (STARS; Cruise, Cash y Bolton, 1985). El STARS es un cuestionario compuesto de 51 ítems que evalúan la ansiedad que surge en los estudiantes en distintas situaciones académicas relacionadas con la estadística. Hasta fechas recientes el STARS ha sido el instrumento más utilizado para evaluar la ansiedad hacia la estadística. Sin embargo, sus propiedades psicométricas han sido puestas en duda por algunos autores (Baloglu, 2002). Con el objetivo de proponer una alternativa al mencionado STARS, Vigil, Lorenzo y Condon (2008) desarrollaron un nuevo instrumento de medida de la ansiedad hacia la estadística: *Statistical Anxiety Scale* (SAS). Según dichos autores, apoyándose en los resultados de un análisis factorial exploratorio, el SAS permite evaluar tres dimensiones relacionadas con la ansiedad hacia la estadística: Ansiedad en el examen, Ansiedad a pedir ayuda, y Ansiedad a la interpretación.

El objetivo de nuestro trabajo es aportar evidencias adicionales de la validez del SAS. En primer lugar, mediante la confirmación de la estructura factorial propuesta mediante técnicas de Análisis Factorial Confirmatorio. Además se aportan evidencias de su validez a través del estudio de las relaciones entre las puntuaciones en el SAS y otras variables teóricamente relacionadas con este constructo.

## Método

### *Participantes*

Los participantes fueron 195 estudiantes, de los cuales 150 estaban matriculados en Análisis de Datos en Psicología I de 1º de Psicología (119 en la Universidad de Huelva y 31 en la Universidad de Sevilla), y 45 estaban matriculados en Bioestadística de 1º de Enfermería en la Universidad de Huelva. La media de edad de los participantes fue de 19,46 años (DT 2,81), y el porcentaje de estudiantes mujeres era del 83,6%.

### *Instrumentos*

*Statistical Anxiety Scale (SAS)*: Instrumento compuesto por 24 ítems que describen situaciones potencialmente generadoras de ansiedad. Cada ítem tiene 5 opciones de respuesta, desde 1 ‘Ninguna ansiedad’ hasta 5 ‘Mucha ansiedad’. Los autores de dicho instrumento, como se ha señalado, proponen una estructura de 3 dimensiones: Ansiedad ante el examen, Ansiedad a pedir ayuda y Ansiedad a la interpretación. Los ítems 5, 21 y 24 fueron modificados respecto a la versión original de forma que se sustituyó “profesor de apoyo”, que en el contexto utilizado no tenía aplicación, por “experto en estadística”.

*Survey of Attitudes Toward Statistics (SATS)*: (Schau, Stevens, Dauphinee y Del Vecchio, 1995) para medir actitudes hacia la estadística. Consta de 28 ítems tipo Likert con 7 opciones de respuesta. Aunque pueden identificarse 4 dimensiones, en este trabajo sólo se utilizó la puntuación total.

*Approaches and Study Skill Inventory for Students (ASSIST, versión reducida)*: (Tait y Entwistle, 1996) para evaluar estilos de aprendizaje. La versión reducida consta de 18 ítems tipo Likert con 4 opciones de respuesta. Proporciona información acerca de 3 dimensiones: aprendizaje profundo, aprendizaje superficial y aprendizaje estratégico.

*Prerequisiti di Matematica per la Psicometria* (PMP) (Galli, Chiesi & Primi, 2008). Prueba de conocimientos matemáticos básicos para el aprendizaje de estadística. Consta de 30 preguntas de elección múltiple, con 4 opciones de respuesta.

*Procedimiento*

Los instrumentos fueron administrados como parte de una práctica voluntaria: SATS y PMP en la primera semana de curso, SAS y ASSIST a falta de tres semanas para finalizar el curso.

Resultados

*Análisis Factorial Confirmatorio*

La matriz de varianzas-covarianzas de las puntuaciones en los 24 ítems del SAS fue sometida a un AFC para poner a prueba la estructura simple de 3 factores propuesta por Vigil-Colet et al. (2008). Para ello se usó el programa EQS 6.1 y un método de estimación de máxima verosimilitud robusto (Satorra y Bentler, 1994). El valor de chi-cuadrado Satorra-Bentler para este modelo fue de 569.61 ( $gl = 249, p < .001$ ). Las otras medidas de ajuste usadas mostraban que el modelo de 3 factores presentaba un pobre ajuste a los datos:  $SB\chi^2/gl = 2.29, CFI = .86, RMSEA = .08$ . Sin embargo, el análisis de los índices de modificación y de los residuales estandarizados indicaba que el ajuste del modelo podía ser claramente mejorado si se incluían las covarianzas entre los términos de error correspondientes a los ítems 5 (“Pedir a un experto en estadística que me explique un tema del que no he entendido nada”), 21 (“Pedir ayuda a un experto en estadística para interpretar una tabla de resultados”), y 24 (“Pedir a un experto en estadística que me explique cómo se hace un ejercicio”). Dada la aparente semejanza en los contenidos de estos ítems, la inclusión de la covarianza entre los errores de estos ítems parece tener cierta justificación. Este modelo modificado mostró un ajuste aceptable ( $SB\chi^2 = 399.74, gl = 246, p < .001, SB\chi^2/gl = 1.62, CFI = .94, RMSEA = .06$ ). Además, el modelo modificado ajustaba significativamente mejor que el modelo original ( $\Delta SB\chi^2 = 86.37, df = 3, p < .01$ )<sup>[1]</sup>. Las estimaciones estandarizadas del modelo modificado se presentan en la figura 1. Todas estas estimaciones fueron estadísticamente significativos ( $p < .001$ ).

Las escalas elaboradas como suma de las puntuaciones en los ítems de los 3 factores del SAS mostraron una alta consistencia interna ( $\alpha_1 = .91, \alpha_2 = .94, \text{ y } \alpha_3 = .83$ , para SAS-Examen, SAS-Ayuda y SAS-Interpretación, respectivamente).

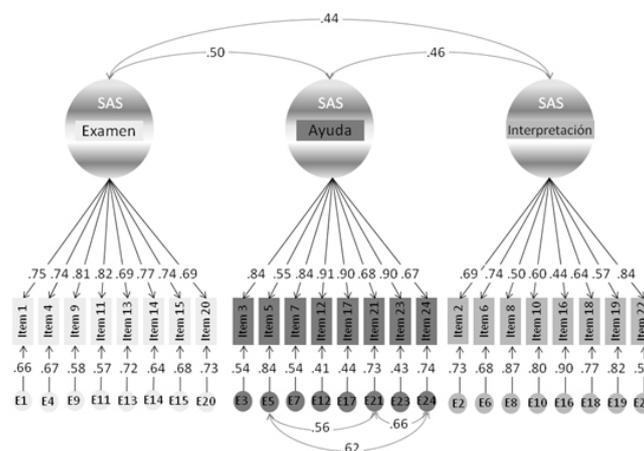


Figura 1. Solución estandarizada del Modelo de 3 factores modificado

[1] Para la comparación entre modelos anidados se usó la prueba de la diferencia entre  $SB\chi^2$  (Satorra y Bentler, 2001).

### Relaciones del SAS con otras variables

Los coeficientes de correlación entre las puntuaciones en el SAS y las otras variables analizadas se presentan en la tabla 1. Como era esperable se encontraron correlaciones inversas significativas de todas las escalas del SAS con las actitudes hacia la estadística medida a través del SATS. En cambio, los conocimientos matemáticos (PMP) sólo parecen estar relacionados débilmente con la escala SAS-Interpretación. Con respecto a las estrategias de aprendizaje, cuanto mayor sea el uso de estrategias de aprendizaje superficial más intensa es la ansiedad evaluada a través de las tres dimensiones del SAS. Por último, el SAS no parece mostrar una importante capacidad predictiva de las notas en el examen final de la asignatura, encontrándose únicamente una correlación débil entre SAS-Examen y las notas.

Tabla 1. Coeficientes de correlación de Pearson entre las escalas del SAS y otras variables

	SATS	PMP	ASSIST-S	ASSIST-P	ASSIST-E	Nota
SAS Examen	-,47**	-,03	,48**	,09	,23**	-,18*
SAS Ayuda	-,27**	,04	,36**	,06	-,04	-,03
SAS Interpretación	-,37**	-,18*	,49**	,00	,01	-,13

Nota: ASSIST-S, ASSIST-P y ASSIST-E corresponden a las escalas del ASSIST Aprendizaje Superficial, Aprendizaje Profundo y Aprendizaje Estratégico, respectivamente. \* $p < 0.01$ , \*\* $p < 0.01$

### Discusión

Los resultados obtenidos en este trabajo representan un apoyo a la validez del SAS como medida de la ansiedad hacia la estadística. En primer lugar, la estructura de 3 factores propuesta por Vigil-Colet et al. (2008) ha sido parcialmente confirmada. La modificación realizada a la estructura simple, la inclusión de las covarianzas entre los términos de error de los ítems 5, 21 y 24, puede ser explicada además por la diferencia de contenido de estos tres ítems respecto al resto de ítems de la escala SAS-Ayuda. En concreto, mientras los ítems problemáticos planteaban situaciones en las que se pedía ayuda a algún *experto en estadística*, los otros ítems de la escala preguntaban acerca de situaciones en las que se pedía ayuda al *profesor de estadística*. Hay que tener en cuenta, además que los ítems 5, 21 y 24 fueron modificados en la versión administrada, debido a que las referencias a un *profesor de apoyo* no tenían aplicación en nuestro contexto. Además, las escalas del SAS mostraron una alta consistencia interna, lo que supone un apoyo para la fiabilidad de las mismas.

Por otra parte, las correlaciones de las puntuaciones en el SAS con las otras variables analizadas se ajustan por lo general a lo encontrado en la literatura cuando se han usado otras medidas de ansiedad hacia la estadística. Así, por ejemplo, se encontraron correlaciones en el sentido esperado entre las actitudes hacia la estadística y el SAS. Sin embargo, estas correlaciones tenían una intensidad menor a las encontradas en trabajos en los que se usó el STARS como medida de la ansiedad hacia la estadística (Watson, Kromrey, Ferron, Lang y Hogarty, 2003; Baloglu, Koçak, y Zelhart, 2007). Esto puede ser debido a que del SAS se excluyeron algunos factores evaluados en el STARS (Valor de la estadística, Auto-concepto en cálculo) que son probablemente aspectos más relacionados con la actitud que con la ansiedad hacia la estadística. Desde nuestro punto de vista, la eliminación de estos factores en el SAS propicia una deseable mayor diferenciación entre ambos constructos.

La asociación encontrada entre las estrategias de aprendizaje superficial y las tres escalas del SAS también se ajusta a las expectativas teóricas. Esto coincide además con lo encontrado en un dominio conductual muy relacionado, la ansiedad hacia las matemáticas (Bessant, 1995). La capacidad predictiva del SAS, sin embargo, no es muy elevada, encontrándose sólo una débil relación entre la ansiedad hacia los exámenes de estadística y las notas. En cualquier caso, esta falta de capacidad predictiva suele ser común en otros instrumentos para evaluar la ansiedad hacia la estadística (ver Carmona 2004). Igualmente débil es la relación encontrada entre los conocimientos matemáticos al iniciar el curso y la ansiedad. Esto supone, sin duda, el hallazgo más inesperado del presente trabajo y requeriría de futuras investigaciones que permitieran confirmar su consistencia.

## Referencias

- Baloglu, M. (2002). Psychometric properties of the Statistics Anxiety Rating Scale. *Psychological Reports, 90*, 315-325.
- Baloglu, M. (2003). Individual differences in statistics anxiety among college students. *Personality and Individual Differences, 34*, 855-865.
- Baloglu, M., Koçak, R. y Zelhart, P.F. (2007). The Relationship between statistics anxiety and attitudes toward statistics. *Ankara University, Journal of Faculty of Educational Sciences, 40(2)*, 23-39.
- Bessant, K.C. (1995). Factors associated with types of mathematics anxiety in college students. *Journal for Research in Mathematics Education, 26*, 327-345.
- Carmona, J. (2004). Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística. *Statistics Education Research Journal, 3*, 5-28.
- Cruise, R. J., Cash, R. W. y Bolton, D. L. (1985). Development and validation of an instrument to measure statistical anxiety. *ASA Proceedings of the Section on Statistical Education* (pp. 92-97). Alexandria, VA: American Statistical Association.
- Furnham, A. & Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statistics examination grades. *Personality and Individual Differences, 37*, 943-955.
- Galli, S., Chiesi, F. & Primi, C. (2008). The construction of a scale to measure mathematical ability in psychology students: an application of the Rasch model. *Testing Psicometria Metodologia, TPM, 15*, 1-16.
- Hanna, D. y Dempster, M. (2009). The effect of statistics anxiety on students' predicted and actual test scores. *Irish Journal of Psychology, 30*, 201-209.
- Mji, A. y Onwuegbuzie, A. (2004). Evidence of score reliability and validity of the statistical anxiety rating scale among technikon students in South Africa. *Measurement and Evaluation in Counselling and Development, 36*, 238-251.
- Nasser, F. (2004). Structural model and effects on cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education, 12*. [<http://www.amstat.org/publications/jse/v12n1/nasser.html>].
- Onwuegbuzie, A.J. (1998). Statistics anxiety: a function of learning style? *Research in the Schools, 5*, 43-52.

- Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education*, 29, 3-18.
- Onwuegbuzie, A.J. y Seaman, M.A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *Journal of Experimental Education*, 63, 115-124.
- Onwuegbuzie, A. J. y Wilson, V. A. (2003). Statistics anxiety: nature, etiology, antecedents, effects and treatments-a comprehensive review of the literature. *Teaching in Higher Education*, 8, 195-209.
- Satorra, A. y Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. En A. Von Eye, & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Satorra, A. y Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Schau, C., Stevens, J., Dauphinee, T. L. y Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55, 868-875.
- Tait, H. y Entwistle, N. J. (1996). Identifying students at risk through ineffective study strategies. *Higher Education*, 31, 99-118.
- Tremblay, P. F., Gardner, R. C. y Heipel, G. (2000). A model of the relationship among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioural Science*, 32(1), 40-48.
- Vigil-Colet, A., Lorenzo-Seva, U. y Condon, L. (2008). Development and validation of the statistical anxiety scale. *Psicothema*, 20(1), 174-186.
- Watson, F., Kromrey, J., Ferron, J., Lang, T. y Hogarty, K. (2003). *An assessment blueprint for Encstat: A statistics anxiety intervention program*. Comunicación presentada al AERA Annual Meeting, San Diego, CA, EE. UU.
- Zeidner, M. (1990). Does test anxiety bias scholastic aptitude test performance by gender and socio-cultural group? *Journal of Genetical Psychology*, 150, 175-185.

# DISCRIMINACIÓN DE LOS ÍTEMS DE LA SUBESCALA DE COMBINACIÓN DE EMOCIONES DEL *MAYER-SALOVEY-CARUSO EMOTIONAL INTELLIGENCE TEST* (MSCEIT)

Manuel Sánchez-García<sup>1</sup>, Natalio Extremera<sup>2</sup> y Pablo Fernández-Berrocal<sup>2</sup>

<sup>1</sup>Universidad de Huelva

<sup>2</sup>Universidad de Málaga

Las medidas de ejecución se consideran el mejor método para evaluar la inteligencia emocional como una habilidad cognitiva (Mayer, Salovey y Caruso, 2002). De entre las medidas de ejecución, el MSCEIT es la prueba más utilizada y validada entre los investigadores. Este test se compone de 141 ítems agrupados en 8 subescalas o tareas, que a su vez se agrupan en 4 ramas, dos áreas y una puntuación global. Usualmente los análisis psicométricos realizados con este test, salvo los cálculos de la fiabilidad de las agrupaciones, se realizan siempre a nivel de subescala o superior, pero no se informa acerca de las propiedades de los ítems. El objetivo de este trabajo es analizar la capacidad discriminativa de los 12 ítems que forman la subescala G (Combinación de emociones) del MSCEIT en una muestra de 946 sujetos (con edades comprendidas entre 16 y 58 años). Los resultados muestran las buenas propiedades psicométricas de los 12 ítems que integran la subescala en cuanto a su capacidad para discriminar entre sujetos con bajo o elevado nivel de inteligencia emocional. No obstante, se aprecia que algunos ítems aparecen como menos adecuados en todos los análisis realizados. Se comentan y discuten las implicaciones de los resultados y las posibilidades de mejora de los ítems.

Hace ya casi 20 años que la inteligencia emocional (IE) apareció como un nuevo concepto psicológico con repercusiones importantes para el funcionamiento personal y social de las personas (Salovey y Mayer, 1990). Gran parte de la década de los noventa estuvo dedicada a la formulación de diferentes perspectivas teóricas (Fernández-Berrocal y Extremera, 2006). Sin embargo, a finales del siglo pasado y a principios de éste, el interés se ha centrado en el desarrollo de medidas de evaluación del constructo, especialmente en población adulta (Mayer, Caruso y Salovey, 2000a).

Actualmente no existe un único acercamiento teórico que sea seguido de forma consensuada y unánime por los investigadores. No obstante, muchos consideran que la perspectiva de Mayer y Salovey (1997) es actualmente la que tiene mayor aval científico (Matthews, Zeidner y Roberts, 2007). Esta formulación considera la IE como un nuevo sistema de inteligencia basado en el procesamiento emocional de la información, concretado en un conjunto de habilidades emocionales de percepción, asimilación, comprensión y regulación (Mayer y Salovey, 1997). Este acercamiento es denominado como modelo de habilidad de IE (Mayer, Salovey y Caruso, 2000). Esta perspectiva ha generado dos formas diferentes de evaluación, en los momentos iniciales de su investigación confiaron en la evaluación de las habilidades emocionales mediante medidas de auto-informe, sin embargo, posteriormente consideraron que la mejor manera para evaluar las dimensiones de la IE es mediante tareas de ejecución (Mayer et al., 2000a) y desarrollaron instrumentos que operacionalizan su acercamiento teórico tales como el Mayer-Salovey-Caruso-Emotional-Intelligence Test (MSCEIT) y sus precursores.

El MSCEIT se compone de 141 ítems que se agrupan en 8 tareas, que a su vez se agrupan en 4 ramas (Percepción, Asimilación o Facilitación, Comprensión y Regulación o Manejo): dos tareas por cada rama. A su vez, las ramas de Percepción y Facilitación se agrupan en el área Experiencial y las ramas de Comprensión y Manejo en el área Estratégica. Por último, las dos áreas citadas se integran en una puntuación global. Así pues, a partir de las puntuaciones en los ítems de cada sujeto, se generan 15 puntuaciones organizadas jerárquicamente (ver figura 1).

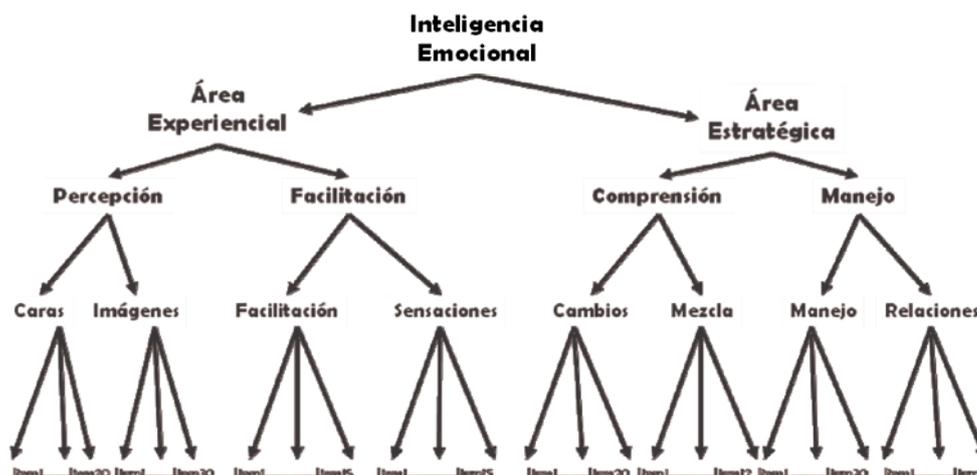


Figura 1. Estructura jerárquica de las puntuaciones MSCEIT

Los ítems originales del test (Mayer et al., 2002) son variables nominales –p.e., distintas emociones o combinación de sentimientos- u ordinales –grado de adecuación de una determinada respuesta ante una situación, en algunos casos- con 5 opciones de respuesta. Las respuestas serán consideradas correctas o incorrectas de acuerdo a ciertos criterios: las puntuaciones cuantitativas de cada uno de los sujetos en cada uno de los ítems corresponden a la proporción de sujetos (de una muestra normativa o de expertos) que han considerado que esa es la respuesta correcta. Por ejemplo, si en el ítem 4 de la tarea de combinación (o mezcla) de emociones el 74.38% de los sujetos de la muestra normativa consideraron que la respuesta correcta era la opción E; a cada uno de los sujetos que eligieron esa opción E en ese ítem se le asignará una puntuación de 0.7438. Las puntuaciones de las tareas, ramas, áreas y total serán las medias de las puntuaciones de rango inferior que la integran.

En la versión original del MSCEIT (Mayer et al., 2002) las puntuaciones cuantitativas de los sujetos se obtenían a partir de dos criterios: consenso general (muestra de 5000 sujetos) y expertos (21 miembros de la International Society Research on Emotion). Los datos que se presentan en este trabajo se obtienen después de corregir los ítems a partir de los criterios de una muestra normativa española de 2066 sujetos y el instrumento de medida corresponde a la versión española del MSCEIT (Extremera y Fernández-Berrocal, 2009).

En concreto, en este trabajo estamos interesados en la rama 3 del modelo que hace referencia a la comprensión emocional. Esta habilidad abarcaría la capacidad para comprender las emociones junto con un mayor conocimiento de las conexiones entre pensamientos y sentimientos. Además, comprendería la capacidad para designar o etiquetar las emociones y reconocer las conexiones entre la palabra y el estado emocional, así como entender las relaciones existentes entre las diversas emociones y las diferentes situaciones a las que obedecen. Igualmente, incluye la habilidad para interpretar el significado de las emociones complejas, por ejemplo, las generadas durante una situación interpersonal, así como la destreza para reconocer las transiciones de unos estados emocionales a otros y la aparición de sentimientos simultáneos. En concreto, en la medición a través del MSCEIT de la habilidad de comprensión se realiza mediante dos tareas: cambios y mezclas. La tarea de cambios está compuesta por veinte situaciones que evalúan la comprensión sobre cómo las emociones varían o se transforman a través del tiempo y de diferentes contextos (i.e., Tatiana estaba enfadada porque un compañero de trabajo se adjudicó el protagonismo en un proyecto, cuando este compañero lo hizo de nuevamente, ¿ella se sintió...?). La tarea de mezclas está compuesta por doce ítems con opciones múltiples donde los participantes identifican emociones que podrían estar combinadas o formar otro tipo de sentimiento más complejo (i.e., miedo, júbilo, sorpresa y vergüenza son parte del sentimiento de...?). En este trabajo analizamos el grado de discriminación de los ítems correspondientes a la subescala de combinaciones (subescala G) de la rama de comprensión del MSCEIT.

El objetivo de este trabajo es calcular e interpretar distintos índices de discriminación de los 12 ítems que integran la subescala G de combinación o mezcla de emociones. Intentaremos demostrar que los 12 ítems permiten clasificar con precisión a los sujetos en grupos de ‘alto’ o ‘bajo’ nivel de inteligencia emocional.

## Método

### *Participantes*

Los análisis se realizaron a partir de las respuestas de 946 personas con edades comprendidas entre los 16 y los 58 años. La media de edad se situó en los 20 años (D.T.=5.55). El 54% de los participantes eran mujeres.

### *VARIABLES*

Las variables de análisis son las puntuaciones cuantitativas de los sujetos en cada uno de los 12 ítems que forman la escala. Las puntuaciones medias en los ítems oscilan entre 0.29 para el ítem 6 y 0.53 para el ítem 12. Las varianzas extremas (0.014 y 0.106) también pertenecen a los ítems 6 y 12, respectivamente.

## Resultados

Cuando trabajamos con ítems cuyas respuestas son correctas o incorrectas y se corrigen y asignan puntuaciones en función del grado de corrección, un modo de saber si un determinado ítem permite diferenciar con precisión entre sujetos con alto y bajo nivel en IE es comparar la proporción de aciertos de los sujetos más ‘listos’ (los más inteligentes, hábiles o los que tienen más conocimientos) con la proporción de aciertos de los más ‘torpes’ (o menos inteligentes).

Para ello vamos a clasificar a los sujetos en grupos extremos según 3 criterios: primero con dos grupos muy separados entre sí [el 25% superior frente al 25% inferior] y después con dos grupos más próximos [33% superior frente a 33% inferior] según su puntuación en la subescala objeto de estudio. Además intentaremos conocer en qué medida estos ítems contribuyen a la clasificación y discriminación de los sujetos inteligentes frente a aquellos con menor nivel de inteligencia emocional a partir de la puntuación global del test [25% superior frente 25% inferior].

Calcularemos el estadístico de contraste  $t$ , correspondiente a una prueba de comparación de medias de grupos independientes. Al trabajar con tamaños muestrales tan elevados, como es el caso, es muy probable que diferencias mínimas sean, a pesar de todo, estadísticamente significativas. Así pues, para determinar no sólo la significación estadística sino también la relevancia de las diferencias, calcularemos distintos tamaños del efecto ( $d$ ,  $r$  y  $r^2$ , Cohen, 1988). Se acepta que valores de ‘ $d$ ’ superiores a 0.80 identifican diferencias de medias sustantivamente importantes. Lo mismo cuando el valor de ‘ $r$ ’ es mayor de 0.5. En casi todos los ítems se cumplen ambas condiciones.

Tabla 1. Comparación de medias. Grupos extremos (25% superior e inferior) en base a sus puntuaciones en la tarea de combinación de emociones

Ítems	Diferencia de medias	<i>p</i>	Tamaño del efecto		
			<i>d</i>	<i>r</i>	<i>r</i> <sup>2</sup>
1	0.28	,00	1.69	0.64	0.42
2	0.32	,00	1.73	0.65	0.43
3	0.25	,00	1.22	0.52	0.27
4	0.24	,00	1.50	0.60	0.36
5	0.32	,00	1.74	0.66	0.43
6	0.12	,00	0.98	0.44	0.19
7	0.42	,00	1.58	0.62	0.38
8	0.23	,00	1.03	0.46	0.21
9	0.41	,00	1.82	0.67	0.45
10	0.24	,00	1.43	0.58	0.34
11	0.30	,00	1.58	0.62	0.38
12	0.47	,00	1.91	0.69	0.48

En el primer caso, comparación del 25% superior-inferior en la tarea analizada, las diferencias de medias oscilan entre 0.12 (ítem 6) y 0.47 (ítem 12). Todas las diferencias son estadísticamente significativas y con tamaños de efecto superiores a 0.80. Si analizamos los valores de *r* y *r*<sup>2</sup> sólo los ítems 6 y 8 aparecen ligeramente por debajo de los valores de corte. Por el contrario, vemos que, para los 467 sujetos que forman los dos grupos extremos, el 48% de la variabilidad del ítem 12 es explicada por el hecho de pertenecer a uno u otro grupo.

No obstante, el 25% superior e inferior son grupos realmente extremos y no es extraño que obtengan puntuaciones muy distintas en cada uno de los ítems. Así pues, decidimos repetir los mismos análisis con grupos no tan alejados: 33% superior frente al 33% inferior. De nuevo se aprecia que casi todos los ítems discriminan bien entre sujetos con elevado y bajo nivel de inteligencia emocional: las diferencias de medias son estadísticamente significativas y relevantes. También se aprecia que los ítems 6 y 8 son los peores.

Tabla 2. Comparación de medias. Grupos extremos (33% superior e inferior) en base a sus puntuaciones en la tarea de combinación de emociones

Ítems	Diferencia de medias	<i>p</i>	Tamaño del efecto		
			<i>d</i>	<i>r</i>	<i>r</i> <sup>2</sup>
1	0.22	,00	1.20	0.51	0.26
2	0.28	,00	1.44	0.58	0.34
3	0.23	,00	1.14	0.50	0.25
4	0.19	,00	1.11	0.49	0.24
5	0.25	,00	1.27	0.54	0.29
6	0.10	,00	0.88	0.40	0.16
7	0.37	,00	1.37	0.56	0.32
8	0.21	,00	0.95	0.43	0.18
9	0.36	,00	1.45	0.59	0.34
10	0.20	,00	1.17	0.51	0.26
11	0.27	,00	1.36	0.56	0.32
12	0.40	,00	1.48	0.59	0.35

Por supuesto, la capacidad de estos ítems para discriminar entre sujetos con alta y baja puntuación global en IE se reduce. Aunque todas las diferencias son significativas y todos los valores del tamaño del efecto son superiores a 0.80, excepto para los ítems 6 y 8. Globalmente podemos considerar que estos ítems discriminan razonablemente bien.

Tabla 3. Comparación de medias. Grupos extremos (25% superior e inferior) en base a sus puntuaciones globales en inteligencia emocional

Ítems	Diferencia de medias	<i>p</i>	Tamaño del efecto		
			<i>d</i>	<i>r</i>	<i>r</i> <sup>2</sup>
1	0.19	,00	1.00	0.45	0.20
2	0.22	,00	1.04	0.46	0.21
3	0.22	,00	1.07	0.47	0.22
4	0.16	,00	0.92	0.42	0.18
5	0.18	,00	0.88	0.40	0.16
6	0.09	,00	0.73	0.34	0.12
7	0.31	,00	1.05	0.46	0.21
8	0.11	,00	0.45	0.22	0.05
9	0.29	,00	1.09	0.48	0.23
10	0.16	,00	0.86	0.39	0.16
11	0.20	,00	0.92	0.42	0.17
12	0.32	,00	1.08	0.47	0.23

### Conclusión

Podemos considerar que todos los ítems de la subescala de combinaciones del MSCEIT poseen elevada capacidad discriminativa y, por ello, nos permitirán clasificar correctamente a la mayoría de los sujetos como personas con alto o bajo nivel de inteligencia emocional. No obstante, los resultados de nuestro estudio sugieren cierta cautela con los ítems 6 y 8 que, a la luz de los hallazgos, deberían ser reformulados o eliminados.

En el caso del ítem 6 parece que el problema está en la muestra normativa de codificación, ya que al parecer no hubo suficiente grado de acuerdo para determinar cuál era la respuesta correcta. Esto implica que elecciones incorrectas llevan aparejadas puntuaciones no muy alejadas de la elección más correcta. De hecho, el rango de respuestas  $0.43-0.02=0.41$  es el más bajo. En resumen, el presente estudio proporciona evidencias de la validez discriminativa de la subescala de combinación del MSCEIT, no obstante, también se sugiere una necesaria reformulación y/o recodificación de este ítem que permitiera incrementar la capacidad discriminativa de la tarea. Futuras investigaciones deberían llevar a cabo tal reformulación y examinar el grado incremental de capacidad discriminativa del instrumento.

### Referencias

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: LEA.
- Extremera, N. y Fernández-Berrocal, P. (2009). Versión española del MSCEIT. Madrid: TEA.
- Fernández-Berrocal, P. y Extremera, N. (2006). Emotional intelligence: A theoretical and empirical review of its first 15 years of history. *Psicothema*, 18, supl. 7-12.

- Matthews, G., Zeidner, M. y Roberts, R. D. (2007). *The science of emotional intelligence: knowns and unknowns*. Oxford: Oxford University Press.
- Mayer, J., Caruso, D. R. y Salovey, P. (2000a). Selecting a measure of emotional intelligence: The case of ability scales. En R. Bar-On y J. Parkes (Eds.). *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 320-342). San Francisco: Jossey-Bass/Pfeiffer.
- Mayer, J.D. y Salovey, P. (1997). What is emotional intelligence. En P. Salovey y D. Sluyter (Eds.): *Emotional development and emotional intelligence: educational implications* (pp. 3-31). New York: Basic Books.
- Mayer, J.D., Salovey, P. y Caruso, D. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). Version 2.0*. Toronto, Canadá: Multi-Health Systems.
- Mayer, J. D., Salovey, P. y Caruso, D. R. (2000) Models of emotional intelligence. En R. J. Sternberg (Ed.). *Handbook of Human Intelligence* (2nd ed), pp 396-420. New York: Cambridge.
- Salovey, P. y Mayer, J.D. (1990): Emotional intelligence. *Imagination, Cognition and Personality*, 9, 185-211.

# EVIDENCIAS DE VALIDEZ RELACIONADAS CON LA ESTRUCTURA INTERNA DE LOS ÍTEMS DE LA SUBESCALA DE COMBINACIÓN DE EMOCIONES DEL *MAYER-SALOVEY-CARUSO EMOTIONAL INTELLIGENCE TEST VERSION* (MSCEIT)

Manuel Sánchez-García<sup>1</sup>, Natalio Extremera<sup>2</sup> y Pablo Fernández-Berrocal<sup>2</sup>

<sup>1</sup>Universidad de Huelva

<sup>2</sup>Universidad de Málaga

Recientemente se considera que la inteligencia emocional, entendida como la capacidad para procesar la información emocional (Mayer y Salovey, 1997), debe ser necesariamente evaluada a través de medidas de ejecución. Para medir las dimensiones propuestas en su teoría, el grupo de IE de Mayer et al., (2002) desarrollan el MSCEIT. Actualmente, el MSCEIT es la prueba de ejecución más utilizada en la literatura científica. Este test se compone de 141 ítems agrupados en 8 subescalas, que a su vez se agrupan en 4 ramas, dos áreas y una puntuación global. Usualmente los análisis psicométricos realizados con este test se realizan siempre a nivel de subescala o superior, pero no llevan a cabo análisis de los ítems. En este trabajo, a través de AFE y AFC, se intenta probar si los 12 ítems que integran la subescala G (Combinación de emociones) del MSCEIT se agrupan en una sola dimensión. El estudio se realizó con 946 sujetos (con edades entre 16 y 58 años). Los resultados aportan evidencias de validez de la escala encontrando una estructura unidimensional que se ajusta aceptablemente a los datos.

La inteligencia emocional (IE) se considera una habilidad cognitiva compuesta por cuatro capacidades emocionales básicas: percepción, asimilación, comprensión y regulación emocional (Mayer y Salovey, 1997). En el campo de la evaluación, las denominadas medidas de habilidad o de ejecución se consideran el formato más aceptado para medir estas capacidades emocionales (Mayer, Roberts y Barsade, 2008). Su principal ventaja es que los resultados obtenidos se basan en la capacidad actual de ejecución de la persona en una tarea y no sólo en su creencia sobre tal capacidad, tal como tradicionalmente recogen las medidas auto-informadas de IE (Extremera y Fernández-Berrocal, 2007).

El equipo de Mayer, Salovey y Caruso ha llevado a cabo una labor amplia de investigación en el campo de las medidas de evaluación cuya más reciente operacionalización de su modelo ha sido la creación del MSCEIT v.2.0 (*Mayer-Salovey-Caruso Emotional Intelligence Test*; Mayer, Salovey y Caruso, 2002). Tanto el MSCEIT v.2.0 como sus predecesores, MSCEIT Research Version 1.1. y el *Multi-Factor Emotional Intelligence Scale* (MEIS) han configurado un acercamiento práctico y novedoso en la medición de la IE en base a su modelo teórico dirigido a medir los niveles actuales de IE (Fernández-Berrocal y Extremera, 2006).

La principal medida de ejecución de IE utilizada actualmente en la literatura científica, el MSCEIT, se compone de 141 ítems que se agrupan en 8 tareas, que a su vez se agrupan en 4 ramas (Percepción, Asimilación o Facilitación, Comprensión y Regulación o Manejo): dos tareas por cada rama. La capacidad para percibir emociones es evaluada mediante tareas de percepción de emociones en rostros faciales y fotografías, el factor de asimilación emocional es medido a través de las tareas de sensación y facilitación, la capacidad de comprensión de emociones es evaluada a través de una tarea de cambios y otra de mezclas de emociones y, finalmente, la capacidad para manejar emociones es evaluada mediante una tarea de manejo emocional y otra tarea de relaciones emocionales. A su vez, las ramas de Percepción y Facilitación se agrupan en el área Experiencial y las ramas de Comprensión y Manejo en el área Estratégica. Por último, las dos áreas citadas se integran en una puntuación global. Así

pues, a partir de las puntuaciones en los ítems de cada sujeto, se generan 15 puntuaciones organizadas jerárquicamente (ver figura 1).

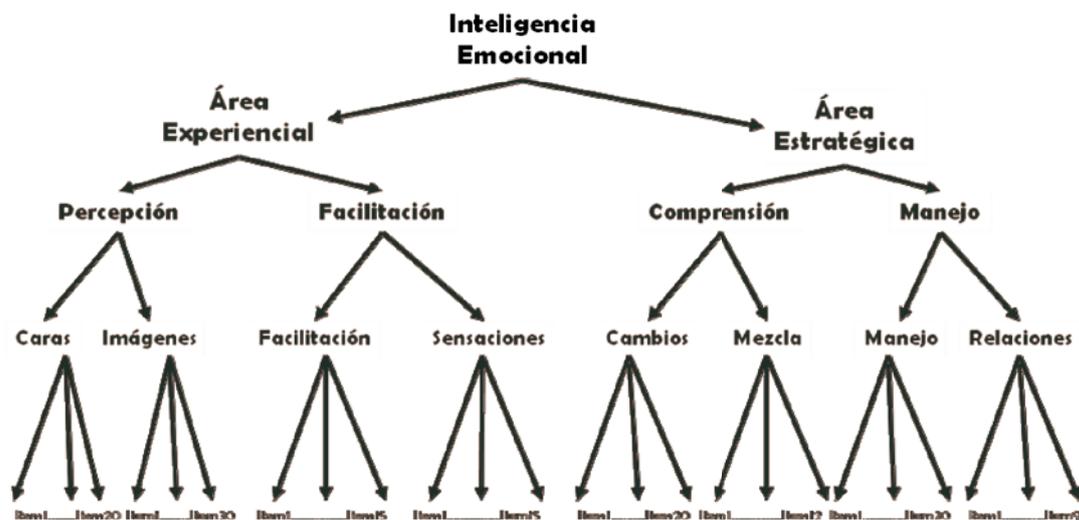


Figura 1. Estructura jerárquica de las puntuaciones MSCEIT

Cada una de las ocho tareas del MSCEIT está compuesta a su vez por un paquete de preguntas, algunas referidas al mismo ítem y otras independientes. De forma resumida, en la tarea de caras los participantes ven una serie de rostros y responden en una escala de cinco opciones, indicando el grado en el cual cada emoción específica está presente. La tarea de imágenes es la misma que la tarea de caras excepto en el contenido de los estímulos que en este caso hacen referencia a paisajes y diseños abstractos. En la tarea de sensaciones, se les pide a los participantes que imaginen y sientan determinadas emociones y emparejen sensaciones a éstas. En las tareas de facilitación los participantes valoran el estado de ánimo que mejor acompañaría o facilitaría tareas cognitivas específicas o comportamientos. En la tarea de combinación de emociones, que es la que examinaremos en este estudio, los participantes identifican emociones que podrían estar combinadas o formar otro tipo de sentimientos más complejos. Así, por ejemplo, los sujetos deben valorar si el rencor es una combinación de las emociones de envidia e ira. En la tarea de transformación o cambio emocional, los participantes deben seleccionar una emoción que resulta de la intensificación de otro sentimiento. En la tarea de manejo de emociones los sujetos deben juzgar qué acción sería más efectiva para obtener un determinado estado emocional final. Por último, en la tarea de relaciones emocionales los sujetos deben juzgar qué acción debería realizar una persona para regular eficazmente las emociones de los demás.

Los ítems originales del test (Mayer et al., 2002) son variables nominales –p.e., distintas emociones o combinación de sentimientos- u ordinales –grado de adecuación de una determinada respuesta ante una situación, en algunos casos- con 5 opciones de respuesta. Las respuestas serán consideradas correctas o incorrectas de acuerdo a ciertos criterios: las puntuaciones cuantitativas de cada uno de los sujetos en cada uno de los ítems corresponden a la proporción de sujetos (de una muestra normativa o de expertos) que han considerado que esa es la respuesta correcta. Por ejemplo, si en el ítem 4 de la tarea de combinación (o mezcla) de emociones el 74.38% de los sujetos de la muestra normativa consideraron que la respuesta correcta era la opción E; a cada uno de los sujetos que eligieron esa opción E en ese ítem se le asignará una puntuación de 0.7438. Las puntuaciones de las tareas, ramas, áreas y total serán las medias de las puntuaciones de rango inferior que la integran.

En la versión original del MSCEIT (Mayer et al., 2002) las puntuaciones cuantitativas de los sujetos se obtenían a partir de dos criterios: consenso general (muestra de 5000 sujetos) y expertos (21 miembros de la *International Society Research on Emotion*). Los datos que se presentan en este trabajo se obtienen después de corregir los ítems a partir de los criterios de una muestra normativa española de 2066 sujetos utilizando la versión española del MSCEIT (Extremera y Fernández-Berrocal, 2009).

El objetivo de este trabajo es aportar evidencias de validez basadas en la estructura interna de la subescala G de combinación o mezcla de emociones, correspondiente a la rama de comprensión. Nuestro objetivo será analizar si los 12 ítems que la integran se agrupan en una sola dimensión, es decir, miden un mismo constructo.

## Método

### *Participantes*

Los análisis se realizaron a partir de las respuestas de 946 personas con edades comprendidas entre los 16 y los 58 años. La media de edad se situó en los 20 años (D.T.=5.55). El 54% de los participantes eran mujeres.

### *Variables*

Las variables de análisis son las puntuaciones cuantitativas de los sujetos en cada uno de los 12 ítems que forman la escala. Las puntuaciones medias en los ítems oscilan entre 0.29 para el ítem 6 y 0.53 para el ítem 12. Las varianzas extremas (0.014 y 0.106) también pertenecen a los ítems 6 y 12, respectivamente.

## Resultados

Para obtener evidencias de validez relacionadas con la estructura interna de la subescala realizaremos distintos análisis factoriales exploratorios (AFE) y confirmatorios (AFC). En primer lugar llevamos a cabo un análisis de componentes principales con SPSS. El programa ofrece una solución con 3 dimensiones. Por defecto la opción que utiliza SPSS para determinar el número de factores a retener es la regla de Kaiser (K1) por la cual se retienen e interpretan todos los factores con autovalor mayor que 1. No obstante, hemos analizado detenidamente los resultados del AFE y realizado otras pruebas descriptivas para demostrar que estos 12 ítems miden una única dimensión.

El gráfico de sedimentación nos permite apreciar claramente la distancia entre el primer y segundo autovalor. Mientras que la distancia entre este segundo autovalor y el último (12º) es sensiblemente menor.

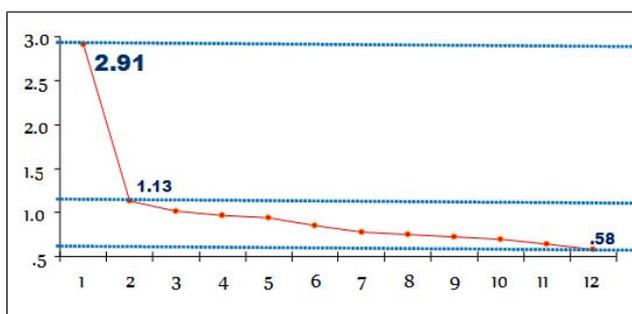


Figura 2. Gráfico de sedimentación

La relación entre los autovalores aporta una prueba más a favor de la unidimensionalidad de estos ítems: calculamos la diferencia entre el primer autovalor y el segundo y la diferencia entre el segundo y el tercer autovalor. La diferencia entre el primer y el segundo autovalor es 15 veces mayor que la diferencia entre el segundo y tercer autovalor.

Por último, realizamos dos pruebas más: análisis paralelo de Horn (1965) y MAP de Velicer (1976) siguiendo las recomendaciones de O'Connor (2000) y utilizando las macros para SPSS que este autor ofrece en su página web (<https://people.ok.ubc.ca/briocconn/nfactors/nfactors.html>). El test MAP indica que el número de componentes que permiten explicar la variabilidad de estos ítems es uno. Por otro lado, el análisis paralelo nos permite comparar (ver tabla 1) los autovalores que se obtendrían en 1000 muestras aleatorias de 946 sujetos y 12 ítems. Podemos apreciar que sólo el primer autovalor empírico es superior a su correspondiente autovalor aleatorio.

*Tabla 1. Autovalores empíricos y aleatorios*

Componente	Autovalor empírico	Autovalor aleatorio	Dimensión relevante
1	2.914	1.186	SI
2	1.131	1.139	NO
3	1.014	1.102	NO
4	.966	1.069	NO
5	.939	1.039	NO
6	.856	1.011	NO
7	.779	.983	NO
8	.750	.956	NO
9	.725	.928	NO
10	.700	.899	NO
11	.641	.865	NO
12	.584	.826	NO

Para obtener un evidencia más de la unidimensionalidad de la escala realizamos un AFC con EQS de un modelo unifactorial. Si nos fijamos detenidamente en los índices de ajuste podemos concluir que el modelo unifactorial es adecuado. Hemos de tener en cuenta que, cuando realizamos AFC, los índices más informativos son los basados en los residuales: nuestro modelo ha de ser capaz de reproducir adecuadamente las relaciones entre los ítems y, en consecuencia, dar lugar a residuales bajos.

*Tabla 2. Índices de ajuste del modelo de AFC de un factor*

FIT INDICES	
BENTLER-BONETT NORMED FIT INDEX (NFI)	= <u>.846</u>
BENTLER-BONETT NON-NORMED FIT INDEX (NNFI)	= <u>.861</u>
COMPARATIVE FIT INDEX (CFI)	= <u>.886</u>
BOLLEN (IFI) FIT INDEX	= .888
MCDONALD (MFI) FIT INDEX	= .936
LISREL GFI FIT INDEX	= .968
LISREL AGFI FIT INDEX	= .954
ROOT MEAN-SQUARE RESIDUAL (RMR)	= .002
STANDARDIZED RMR	= .041
(RMSEA)	= .050
90% CONFIDENCE INTERVAL OF RMSEA	(.042, .058)
Media (en valores absolutos) de los residuales estandarizados	= .030 < .050
Residual estandarizado más elevado	= .141 < .15

No obstante, los índices de ajuste incremental son inferiores a 0.90. Estos índices comparan el valor  $\chi^2$  de nuestro modelo con el valor de  $\chi^2$  de un modelo nulo que prevé la ausencia de relación entre los ítems. Estos primeros índices muestran tan mal ajuste porque, al comparar el valor de  $\chi^2$  del modelo nulo (1161.71) con el valor de  $\chi^2$  de nuestro modelo (178.47), encuentran que no son tan distintos. Así, por ejemplo,  $NFI = (1161.71 - 178.47)/1161.71 = 0.846$ . Es decir, la matriz de varianzas-covarianzas entre ítems con tan bajas varianzas (en torno a 0.04) y covarianzas (en torno a 0.01), no es tan distinta de una matriz de ceros.

Sin embargo, el resto de índices de ajuste y los residuales nos indican que el modelo es adecuado. Por ejemplo, la media de los residuales estandarizados es inferior a 0.05 y que el residual más elevado es también inferior a 0.15 que son los puntos de corte a los que hace referencia la literatura. Por otro lado, los índices de ajuste SRMR y RMSEA también indican un buen ajuste. Los índices ajuste global como GFI y AGFI presentan valores superiores a 0.95. Ambos indican en qué medida la matriz de varianzas-covarianzas reproducida a partir del modelo es igual a la matriz de varianzas-covarianzas empírica. Una sola dimensión permite explicar satisfactoriamente las relaciones entre estos ítems.

### Conclusión

Los datos recogidos aportan evidencias a favor de la hipótesis de unidimensionalidad. Obtenemos pues evidencias de la validez de una de las subescalas que integran el MSCEIT. De cualquier modo, es cierto que hay algunos ítems que tienen saturaciones factoriales muy bajas, aunque estadísticamente significativas, y que habría que replantearse su redacción. Uno de ellos es el ítem 8, con saturaciones factoriales estandarizadas inferiores a 0.4 en el modelo confirmatorio y en el ACP. Los ítems 6 y 1 también tienen valores inferiores a 0.4 en el AFC pero superiores a 0.4 en el AFE.

La causa de las bajas saturaciones factoriales en estos ítems puede ser bien distinta. Para el caso concreto del ítem 8, éste parece no funcionar correctamente en esta muestra ya que presenta las correlaciones más bajas con el resto de ítems y, por otra parte, sugiere ser un ítem de dificultad elevada, al menos para esta muestra, ya que presenta un porcentaje menor de elección de la respuesta más correcta. Una posible explicación de su menor elección pudiera referirse a que el contenido del ítem implica la combinación de cuatro emociones simples donde el participante debe adivinar a qué emoción compleja da lugar. Dado el carácter marcadamente evolutivo de estas habilidades emocionales y su capacidad para ser aprendidas a través de la enseñanza y la experiencia (Mayer et al., 1999), sería necesario realizar este tipo de análisis con muestras cuyas medias de edad fueran superiores a los 20 años. Si el porcentaje de elección de la respuesta mejora, evidentemente este ítem discriminaría bien entre los inexpertos y expertos emocionales de más edad. Si no es así, futuros estudios deberían revisar semánticamente el ítem y tratar de homogeneizar la dificultad de su contenido con el resto de los ítems de la subescala.

Agradecimientos. Queremos agradecer a los participantes y asistentes a la mesa de comunicaciones sus sugerencias y recomendaciones, que han servido para mejorar este trabajo.

## Referencias

- Extremera, N. y Fernández-Berrocal, P. (2009). *Versión española del MSCEIT*. Madrid: TEA.
- Extremera, N. y Fernández-Berrocal, P. (2007). Una guía práctica de los instrumentos actuales de evaluación de la inteligencia emocional. En J. M. Mestre & P. Fernández-Berrocal (coords.), *Manual de Inteligencia Emocional* (pp. 97-120). Madrid: Pirámide.
- Fernández-Berrocal, P. y Extremera, N. (2006). Emotional intelligence: A theoretical and empirical review of its first 15 years of history. *Psicothema*, *18*, 7-12.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.
- Mayer, J. D., Roberts, R. D. y Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507-536.
- Mayer, J.D. y Salovey, P. (1997). What is emotional intelligence. En P. Salovey y D. Sluyter (Eds.): *Emotional development and emotional intelligence: educational implications* (pp. 3-31). New York: Basic Books.
- Mayer, J.D., Caruso, D. & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, *27*, 267-298.
- Mayer, J.D., Salovey, P. y Caruso, D. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). Version 2.0*. Toronto, Canadá: Multi-Health Systems.
- O'Connor, B.P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*, 396-402.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327.

# VALIDACIÓN DE CONSTRUCTO FACTORIAL DE UN INSTRUMENTO OBJETIVO DE TENDENCIA AL RIESGO

Iván Sánchez-Iglesias<sup>1</sup> y Manuel J. Sueiro<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid

<sup>2</sup>Universidad Complutense de Madrid

La *tendencia al riesgo* es un estilo interactivo referido al conservadurismo de los sujetos ante situaciones de elección. Se aborda desde diferentes perspectivas: mediante cuestionarios de autoinforme y pruebas objetivas. El autoinforme ha sido criticado por la interferencia de sesgos y falta de validez predictiva. Las pruebas objetivas requieren recursos que hacen que sean difíciles de aplicar a grandes muestras de sujetos. Se presenta un test de *tendencia al riesgo* de treinta ítems. Cada ítem es una prueba objetiva que presenta al sujeto una situación hipotética en la que tiene que elegir entre tres opciones, cada una con una función de ganancia diferente pero equivalentes en su esperanza matemática. Para valorar sus bondades psicométricas, se aplicó a 222 sujetos y se realizaron análisis de fiabilidad y de validez factorial exploratorio (AFE). Los resultados apoyaron un modelo en tres factores (Deportes y juegos de azar, Planes a largo plazo y Prevención de pérdidas). Tras los ajustes necesarios al adoptar este modelo, e incluyendo un factor global de *tendencia al riesgo*, un Análisis Factorial Confirmatorio (AFC) mostró un ajuste adecuado.

El riesgo se ha abordado dentro de las *decisiones con incertidumbre* (Bernoulli, 1954; Neuman y Politser, 1992; Savage, 1954; Kahneman y Tversky, 1979), también como rasgo (Eyseck, 1967; Costa y McCrae, 1995; Zuckerman, 1979).

La *tendencia al riesgo* es uno de los estilos interactivos propuestos por Ribes y Sánchez (1990). El método fundamental para la evaluación de la personalidad ha sido el cuestionario de autoinforme: fácil de aplicar (con lápiz y papel), clasifica sujetos correctamente, y tiene validez concurrente con otros tests de autoinforme. Sin embargo, pueden darse sesgos de respuesta. Además, hay discrepancia entre los informes del sujeto y lo que realmente hacen (Rachman, 1978, citado por Santacreu et al., 2006).

Las pruebas objetivas evitan esto, pero se necesitan más recursos físicos, y son muy específicos del ámbito en que se evalúa, como juegos de azar (Rubio y Santacreu, 1998).

La conducta es resultado de interacción persona-contexto, que provee de factores disposicionales que modifican la probabilidad de interacción (Kantor, 1959). Tres de estos factores son: competencia, motivación y estilo interactivo (tendencia a comportarse de una forma en una situación; Ribes, 1990).

Pensamos que la *tendencia al riesgo* dependerá del contexto. Buscamos la forma de evaluar esta *tendencia* que permita ver la multidimensionalidad del constructo, de forma que: 1) se mida el rasgo a partir de conductas recogidas en pruebas objetivas con contingencias abiertas, 2) en diversos ámbitos, 3) fácilmente aplicables 4) con papel y lápiz.

Objetivos: 1) Construir un test objetivo de medida de *Tendencia al Riesgo*, que supere inconvenientes asociados a pruebas objetivas y de autoinforme. 2) Valorar la validez de constructo.

Hipótesis: un AFE mostrará una estructura multifactorial con significado sustantivo, dependiente de la morfología de las situaciones.

## Método

### Participantes

Tenemos 222 sujetos de entre 6 y 68 años ( $M= 36.39$ ,  $DT= 12.13$  años); un 60.8% son hombres y un 38.7% mujeres. El 5.5% no tenía estudios, o eran primarios. El 28% tenían estudios secundarios y el 64% universitarios (64%).

### Instrumentos

Para el análisis se usó SPSS 15 y MPlus (Muthén, 2007).

### Procedimiento

Cada ítem es una prueba objetiva. Se presenta al sujeto una situación hipotética, y la hay tres alternativas de respuesta, con distinta ganancia y probabilidad de conseguirla. Las alternativas más arriesgadas tienen una ganancia potencial mayor, pero menor probabilidad. El sujeto no obtiene *feedback* del resultado de su elección, para evitar sesgos posteriores (León y Lopes, 1988). La esperanza matemática de cada opción de respuesta es la misma. La ganancia es una variable aleatoria ( $X$ ) con distintos valores según el ítem, opción de respuesta, y ganancia de cada opción:

$$X_{ia} = \{G_{ia1}, G_{ia2}, \dots, G_{iaj}, \dots, G_{iaJ}\} \quad [1]$$

Donde:

$X_{ia}$  = Ganancia ítem  $i$ , opción  $a$

$G_{iaj}$  = Valor de la ganancia: ítem  $i$ , opción  $a$ , dado que ocurre  $j$

Un ejemplo se puede ver en la figura 1.

- 22) Para seguir siendo competitivo con su cultivo de maíz, decide usar uno de estos tres métodos:
- Continuar con el cultivo tradicional y asegurar 40 toneladas de grano.
  - Usar semillas importadas, que si arraigan (1/4 de posibilidades) darán 160 toneladas de grano.
  - Usar semillas transgénicas, que si arraigan (1/8 de posibilidades) darán 320 toneladas de grano.

Figura 1. Ejemplo de ítem

Cada opción presenta dos valores de ganancia posibles, la asociada a “ganar”, y la asociada a “perder”.

$$X_{22a} = \{40, \mathbf{o}\} \quad [2]$$

$$X_{22b} = \{160, \mathbf{o}\} \quad [3]$$

$$X_{22c} = \{320, \mathbf{o}\} \quad [4]$$

Por ejemplo, la ganancia del ítem 22, opción  $a$  puede tomar los valores 40 (“ganar”) o 0 (“perder”). Existe una función de probabilidad de la ganancia:

$$f(X_{ia}) = \begin{cases} \pi_{ia1} \Leftrightarrow X_{ia} = G_{ia1} \\ \pi_{ia2} \Leftrightarrow X_{ia} = G_{ia2} \\ \cdot \\ \cdot \\ \cdot \\ \pi_{iaj} \Leftrightarrow X_{iaj} = G_{iaj} \\ \cdot \\ \cdot \\ \cdot \\ 1 - \sum_{j=1}^{j-1} \pi_{iaj} \Leftrightarrow X_{iaj} = G_{iaj} \end{cases} \quad [5]$$

La función de probabilidad de ganancia  $f(X_{ia})$ , en una opción de respuesta de un ítem, arrojará una probabilidad determinada según el valor concreto de la ganancia.

En el ejemplo:

$$f(X_{22a}) = \begin{cases} 1 \Leftrightarrow X_{22a} = 40 \\ 1 - 1 \Leftrightarrow X_{22a} = 0 \end{cases} \quad [6]$$

La función de probabilidad de ganancia del ítem 22, opción  $a$ , que representamos con  $f(X_{22a})$ , devuelve probabilidad 1 si y solo si la ganancia  $X_{22a}$  toma el valor 40. La misma función devuelve probabilidad 0 si  $X_{22a}$  toma el valor 0. Ambas opciones suman una probabilidad de 1.

El valor esperado de la ganancia (en una opción de repuesta de un ítem) se calcula:

$$E(X) = \sum \pi X \quad [7]$$

Para el caso general de opciones multinomiales:

$$E(X_{ia}) = G_{ia1}\pi_{ia1} + G_{ia2}\pi_{ia2} + \dots + G_{iaj}(1 - \sum_{j=1}^{j-1} \pi_{iaj}) \quad [8]$$

Aunque la función de probabilidad de cada alternativa sea distinta, el valor esperado de la ganancia es siempre el mismo. Todas las opciones de cada ítem son equivalentes.

En nuestro ejemplo:

$$E(X_{22a}) = 40 \cdot 1 + 0 \cdot 0 = 40 \quad [9]$$

$$E(X_{22b}) = 160 \cdot \frac{1}{4} + 0 \cdot \frac{3}{4} = 40 \quad [10]$$

$$E(X_{22c}) = 320 \cdot \frac{1}{8} + 0 \cdot \frac{7}{8} = 40 \quad [11]$$

Las opciones de cada ítem tienen distinta varianza. La varianza se relaciona con la *tendencia al riesgo*: opciones cuya probabilidad es 1 tienen varianza 0, y cuanto menor la probabilidad de éxito, mayor varianza. El valor de riesgo de una opción suele computarse como el inverso de la probabilidad de éxito de la opción (Arend, Botella, Contreras, Hernández y Santacreu, 2003), pero esto no tiene en cuenta la posibilidad que, de no tener éxito, exista alguna ganancia distinta de 0. Utilizando la varianza para ordenar las opciones según su riesgo sí se contempla esa opción.

Los ítems intentan abarcar diferentes ámbitos. El test tiene 30 ítems con tres opciones de respuesta.

El test se aplicó a la muestra, se grabaron las respuestas y se analizaron los datos.

## Resultados

La puntuación de los sujetos es la suma de las puntuaciones de los ítems. Cada ítem puntúa en función de la probabilidad de la ganancia en la opción elegida. Las puntuaciones se distribuyeron normalmente ( $Z=.868$ ;  $p=.438$ ), con media 53.68 y desviación típica 7.45.

### *Análisis de los elementos*

Hubo variabilidad en todos los ítems, en función de la fuerza de atracción de cada opción presentada.

Veintiséis ítems mostraron un índice de homogeneidad corregido significativo. El ítem 16 es el que menos correlacionó con el total del test ( $H_c = 0.59$ ), y se eliminó.

### *Indicadores de fiabilidad*

El  $\alpha$  de Cronbach para los 30 ítems toma el valor 0.712. El coeficiente de Spearman-Brown fue de 0.746. Ambos valores son adecuados (Nunnally, 1981).

### *Evaluación del nivel de competencia requerido*

Para ser usado en población general, el nivel de competencia necesario no debe ser elevado. Si en el test los sujetos responden al azar, las puntuaciones no reflejan su nivel de rasgo, y las correlaciones entre ítems serán bajas o nulas. Las respuestas al azar disminuyen la fiabilidad del test (Nunnally, 1981).

Se excluyó a los sujetos sin estudios o con estudios primarios (que podían tener dificultades con porcentajes y fracciones), y se comparó  $\alpha$  de la muestra reducida con la total, con el estadístico de Feldt. No se encontraron diferencias ( $F [221,6409]=1031$ ,  $p=.363$ ): el test parece útil para población general.

Se usó AFE para descubrir factores subyacentes. Las puntuaciones de los ítems son categorías ordenadas, por lo que el AFE (ejes principales y rotación Varimax) se calculó con correlaciones policóricas (Olsson, 1979). El modelo más simple con ajuste a los datos fue de tres factores:  $GFI=0.9168$  y  $RMSR=0.0717$ .

Se mantuvo la estructura factorial, ajustando cinco ítems por su contenido, cambiando su factor de pertenencia.

Para comprobar el ajuste de la estructura modificada, realizamos un AFC, que explicita el compromiso teórico con nuestro modelo, para contrastarlo con los datos (Ruiz, 2000). No se recomienda el AFC con los mismos datos del AFE, para evitar capitalización del azar, pero la muestra de este estudio es limitada, y las modificaciones no están realizadas para mejorar el ajuste. De hecho deberían empeorarlo.

Se propone que las dimensiones se derivan de un factor de segundo orden, el factor global de *tendencia al riesgo* (figura 2).

El método de estimación fue WLSMV. El ajuste de este modelo en términos globales fue bueno. (CMIN/DF= 3.694; CFI=0.861; TLI=0.878; SRMR= 0.088; RMSEA= 0.038).

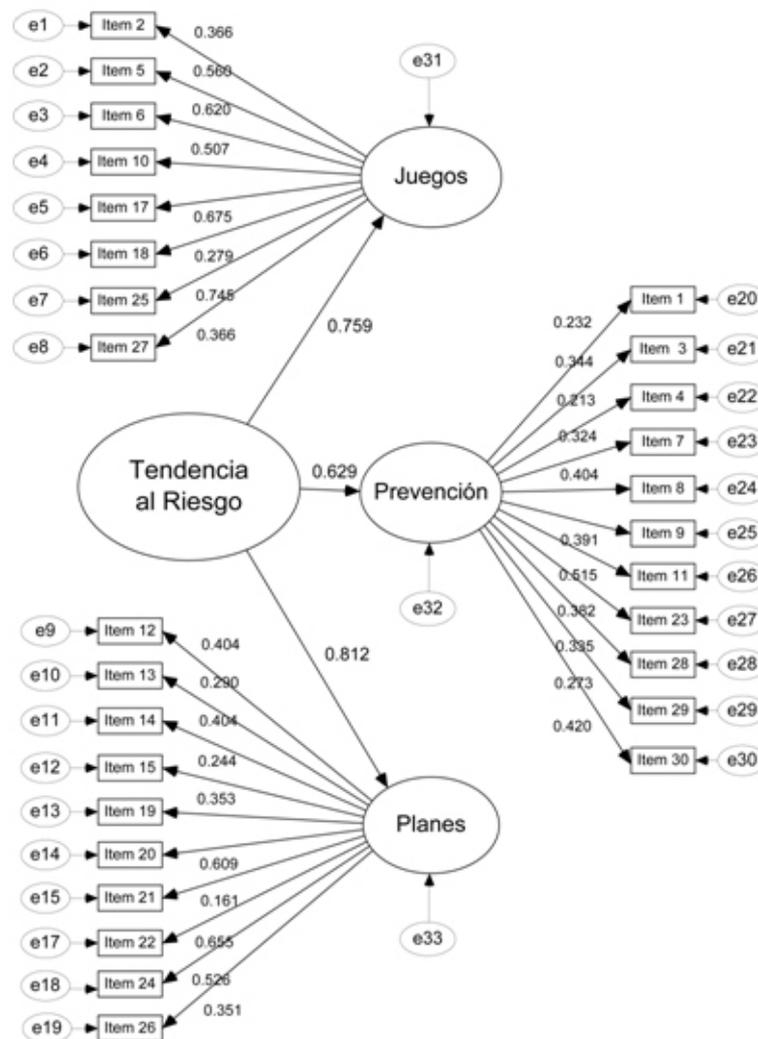


Figura 2. Modelo de un factor de segundo orden. Coeficientes estandarizados

La *tendencia al riesgo* es la “propensión a elegir, cuando hay varias alternativas posibles, la opción con la mayor ganancia aunque la probabilidad de su ocurrencia sea menor”, y depende de las características morfológicas de la situación. Se han encontrado los siguientes factores:

- *Deportes y juegos de azar*: Situaciones de juegos de azar o deportes, y apuestas donde se arriesga dinero.
- *Planes a largo plazo*: Se incluye un elemento temporal. Campañas políticas, hipotecas o cultivos agrícolas, situaciones donde las consecuencias no son evidentes inmediatamente.
- *Prevención de pérdidas y control de daños*: una elección conservadora previene pérdidas económicas o personales. Una elección arriesgada puede salvar todos los recursos, pero lleva a perderlos en mayor cantidad en caso de no “perder”. Se incluye “control de daños”, donde se parte de una situación en que se asumen pérdidas en toda opción, pero se pueden minimizarlas con poca probabilidad de éxito, o asumir pérdidas algo mayores pero controladas. Control de plagas o gestión de vidas humanas ante una crisis son ejemplos.

Existe correlación poco importante entre subescalas. La correlación más alta ocurre entre *Deportes y Planes* ( $r=.288$ ,  $R^2=.083$ ).

Los tres factores podrían utilizarse como subescalas, como instrumentos separados. *Deportes y juegos de azar* mostró un alfa de Cronbach de 0.676, y fiabilidad de 0.688 (dos mitades). *Planes a largo plazo* tuvo un alfa de 0.565, y fiabilidad de 0.627. *Prevención de pérdidas y control de daños*, alfa de 0.505, y fiabilidad de 0.509.

## Discusión

Se presentó una prueba objetiva para evaluar el estilo interactivo *tendencia* al riesgo: “propensión a elegir, cuando hay varias alternativas, la opción con la mayor ganancia aunque la probabilidad de su ocurrencia sea menor”. Este rasgo ha sido evaluado de dos formas: los *autoinformes* tienen sesgos de respuesta y falta de validez predictiva; las *pruebas objetivas* tienen mayor validez predictiva, pero necesitan más recursos para aplicarlas. Además, cada prueba mide un solo ámbito, dificultando la generalización de resultados.

Se construyó un test de medida de *tendencia al riesgo*; cada ítem es una prueba objetiva en sí misma, con distintas situaciones, contingencias abiertas (igualdad de esperanza matemática de las opciones) para una buena discriminación, facilidad de aplicación.

Se interpreta un 75% de la variabilidad en las puntuaciones como variabilidad en el rasgo. Esta prueba que puede ser aplicada a participantes con niveles de estudios bajos, sin merma de fiabilidad ( $F [221,6409]=1031$ ,  $p=.363$ ).

Tres factores latentes subyacen a los datos muestrales (AFE, ejes principales); después de analizar el contenido de los ítems y cambiar la pertenencia de cinco ítems a otros factores, se observó buen ajuste de los datos a un modelo: tres factores de primer orden y un factor de segundo orden (con AFC), que llamamos *tendencia al riesgo*.

Hay que mejorar la varianza explicada por los tres factores (23.18%). No se recomienda un AFC con los mismos datos que el AFE, pero se argumenta que las modificaciones no producen capitalización del azar (puesto que son derivadas de la teoría, y deberían *empeorar* el ajuste). Sería conveniente aplicar AFC a otra muestra.

El modelo tiene un ajuste adecuado en AFE (GFI= 0.9168 y RMSR= 0.0717). Con AFC también encontramos ajuste mediante varios indicadores; seleccionamos un modelo con un factor de segundo orden, la *tendencia al riesgo*, que influye en los tres factores de primer orden. CFI y TLI no fueron adecuados (CFI= 0.861; TLI= 0.856), pero otros índices sí (CMIN/DF=3.694; SRMR=.088; RMSEA=.038). Globalmente, el modelo ajusta a los datos.

Encontramos tres dimensiones que dan cuenta de la relación entre ítems: *Deportes y juegos de azar* sería el ámbito más estudiado: las ganancias surgen de situaciones lúdicas con componente de azar. *Planes a largo plazo* incluye un componente temporal. *Prevención de pérdidas y control de daños* tiene relación con situaciones en las que el riesgo es perder recursos.

De este trabajo se desprende que el estilo interactivo de *tendencia al riesgo* depende de las diferencias individuales de la forma de la situación, en ausencia de *feedback* o elecciones con contingencias reforzantes claras.

Se aporta un método para evaluar la *tendencia al riesgo*, mediante pruebas objetivas fácilmente aplicables, midiendo en distintos contextos de forma rápida y económica.

## Referencias

- Arend, I., Botella, J., Contreras, M. J., Hernández, J. M. y Santacreu, J. (2003). A Betting Dice Test to Study the Interactive Style of Risk-Taking Behavior. *The Psychological Record*, 53, 217-230.
- Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22, 23-36.
- Costa, P. T. y McCrae, R. R. (1995). Primary traits of Eysenck's P-E-N System: Three- and five factor solutions. *Journal of Personality and Social Psychology*, 69, 308-317.
- Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas.
- Kahneman, D. y Tversky, A. (1979). Prospect Theory: An analysis of decision making under risk. *Econometrica*, 47, 263-291.
- Kantor, J. R. (1959). *Interbehavioral Psychology*. Chicago: Principia Press.
- León, O. G. y Lopes, L. L. (1988). Risk preference and feedback. *Bulletin of the Psychonomic Society*, 26 (4), 343-346.
- Lopes, L. L. (1987). Between hope and fear: the psychology of risk. *Advances in Experimental Social Psychology*, 20, 255-295.
- Muthén, B. O. (2007). Mplus. Statistical Analysis With Latent Variables. [Software]. Los Angeles, CA: Muthén & Muthén.
- Neuman, P. J. y Politser, P. E. (1992). Risk and optimality. En J. F. Yates (Ed.), *Risk-taking behavior*. Chichester: Wiley.
- Nunnally, J. C. (1981). *Psychometric theory*. New York, NY: McGraw-Hill.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.
- Rachman, S. (1978). Human fears: a three systems analysis. *Scandinavian Journal of Behaviour Therapy*, 7, 237-245.
- Ribes, E. y Sánchez, S. (1990). *Problemas conceptuales en el análisis del comportamiento humano*. México: Trillas.
- Rubio, V. J. y Santacreu, J. (1998). *Test de riesgo asumido mediante dados*. N° R. P. I.: M-70573.
- Ruiz, M. A. (2000). *Introducción a los modelos de ecuaciones estructurales*. Madrid: UNED Ediciones.
- Santacreu, J., Rubio, V. J. y Hernández, J. M. (2006). The objective assessment of personality: Cattell's T-data revisited and more. *Psychology Science*, 48 (1), 53-68.
- Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley.
- Zuckerman, M. (1979). Sensation seeking and risk taking. En C. E. Izard (Ed.), *Emotions in personality and psychopathology*. New York, NY: Plenum.

# VALIDEZ Y FIABILIDAD DE UNA VERSIÓN ESPAÑOLA DEL CUESTIONARIO SA-45: APOYO EMPÍRICO PRELIMINAR

Andrés Sánchez Prada y Beatriz Palacios  
Universidad Pontificia de Salamanca

El presente trabajo analiza las propiedades psicométricas básicas del *Symptom Assessment-45 Questionnaire* (SA-45) a partir de una muestra de estudiantes universitarios, con el propósito de contribuir al proceso de adaptación de dicho instrumento a la población española. Los distintos análisis realizados muestran las inconsistencias encontradas en otros estudios sobre la validez y fiabilidad. Así mismo, en nuestro propósito de contribuir al estudio de este instrumento, nuestros análisis sugieren la posibilidad de tener en cuenta otras posibilidades sobre el SA-45 que exponemos a continuación.

Cada vez más en el ámbito de la psicoterapia, los instrumentos de recogida de datos más utilizados son los autoinformes, y ello es debido a la necesidad de una perspectiva multi-método en el estudio de los cambios que experimentan los usuarios a lo largo del proceso terapéutico.

El *Symptom Checklist-90, SCL-90*, elaborado por Derogatis, Lipman y Covi en 1973 es uno de los más utilizados. Es un autorregistro de 90 ítems que evalúa al usuario atendiendo a 9 dimensiones psicopatológicas (*Somatización, Ansiedad, Psicoticismo, Obsesividad-compulsividad, Hostilidad, Sensibilidad interpersonal, Ansiedad fóbica, Ideación paranoide y Depresión*) y dos índices globales (Índice de Severidad Global (GSI) y Total de Síntomas Positivos (PST)) Existe una versión abreviada de 45 ítems elaborada por Davison y colaboradores, en 1977, *El Symptom Assessment-45 Questionnaire, SA-45*, con una escala de 1 a 5.

Centrándonos en el SA-45, objeto de nuestro estudio, se han realizado numerosos estudios acerca de su fiabilidad y validez.

Sobre la fiabilidad del instrumento, Davison et al. (1997) encontraron valores para el Alfa de Cronbach en cada escala del SA-45, entre 0.70 y 0.90, menos en *psicoticismo e ideación paranoide*. Además, estos autores, comprobaron que había mayor fiabilidad en adultos que en adolescentes, y en muestra clínica frente a la no clínica. Más tarde, Maruish, Bershinsky y Goldstein (1998), evaluaron la fiabilidad de la escala con los procedimientos test-retest utilizando la escala SA-24, y obteniendo correlaciones moderadas en el GSI y en *depresión*, aunque con valores significativos (0.57 en el GSI y 0.56 en *depresión*,  $p < .001$ ). Maruish (1999b), evaluó la correlación ítem-escala en dos estudios, en ambos obtuvo valores de correlación muy altos entre los ítems, 93% y 96% respectivamente. En cuanto a la validez, Davison y colaboradores (1997), confirmaron la existencia de las nueve escalas y además, demostraron la posibilidad del instrumento para discriminar entre sujetos con diferentes tipos y grados de patología. También, encontraron correlación entre las escalas e índices globales del SA-45 y el SCL-90. Maruish, Bershinsky y Goldstein (1998) obtuvieron resultados positivos y significativos en la correlación inter-escala, y correlaciones significativas entre el SA-45 y el Componente físico y mental del SF-12 Health Survey (Ware, Kosinski & Keller, 1995). En la adaptación española, Sandin y colaboradores (2008) encontraron, en general, y en relación a la fiabilidad de la escala, niveles altos de consistencia interna (Cfr. Davison et al., 1997). Y específicamente, un coeficiente alfa global de 0.95 para la escala, y coeficientes para cada sub-escalas de entre 0.63 y 0.85, así como moderadas correlaciones ítem-escala corregidas. En sus análisis de la validez de la escala, los resultados mostraron valores de

0.73 hasta 0.39 entre escalas, y buenos índices de correlación con otras escalas como la PANAS y ASI. Además, por otro lado, demuestran la existencia de una estructura multidimensional del SA-45, con buenos índices de bondad de ajuste, destacando los componentes de *depresión* y *hostilidad*.

Sin embargo, se continúa encontrando algunas inconsistencias y solapamientos entre escalas (e.g. *sensibilidad interpersonal* e *ideación paranoide*), que hacen necesario seguir estudiando este instrumento.

### *Objetivos del estudio*

Objetivo general: Aportación de nuevos datos al proceso de validación iniciado por los autores españoles.

Objetivos específicos:

1. Contrastar la validez de constructo del SA-45 en una muestra no clínica.
2. Determinar cuantitativamente la fiabilidad del SA-45 en dicha muestra.

## Método

### *Instrumentos*

El instrumento utilizado ha sido el SA-45, versión española de Sandín et al. (2008). Es una escala de autoinforme de 45 ítems a través de los cuales el sujeto ha de valorar su sintomatología indicando en qué medida los síntomas descritos se ajustan a su experiencia cotidiana, en una escala de 0 a 4, siendo 0 “nada en absoluto” y 4 “mucho o extremadamente”. Este instrumento, al igual que sus predecesores, recoge nueve subescalas: *Somatización*, *Ansiedad*, *Psicoticismo*, *Obsesividad-Compulsividad*, *Hostilidad*, *Sensibilidad Interpersonal*, *Ansiedad fóbica*, *Depresión* e *Ideación Paranoide*.

### *Participantes*

La muestra seleccionada fue de 478 estudiantes universitarios/as, de Salamanca (69%), Santander (23.9%) y Barcelona (7.1%), distribuidos según la variable “sexo” en un 23.2% de hombres y un 76.8% de mujeres. La media de edad para los hombres es de 21.82 y DT= 4.28 y de 20.96 y DT= 4.11 para las mujeres, sin diferencias significativas en la variable edad ( $t_{(edad)} = 1.879$   $p = .061$   $\alpha = .05$ ).

Tras un estudio de los valores perdidos encontrados, de los 478 sujetos iniciales, se eliminó el único caso con más de un valor perdido dentro de una misma subescala. Además, se imputaron por regresión los valores perdidos en los restantes 477 casos. Y posteriormente se eliminaron 13 de los valores imputados, correspondientes a un total de 12 casos, por encontrarse éstos fuera del rango de la escala (0-4).

De esta manera, la muestra de 477 sujetos, fue utilizada para la realización del estudio de fiabilidad, mientras que para el análisis de componentes principales y el análisis factorial se empleó una muestra total de 465 sujetos.

## Resultados

Al igual que en el estudio de Sandín et al. (2008), se realizó un Análisis de Componentes Principales (ACP) sobre la matriz de correlaciones obtenida con la muestra total, forzando una solución final de nueve componentes con rotación Promax. ( $KMO = 0.918$ ;  $\chi^2(990) = 8514,97$  ( $p < .001$ )).

Los nueve componentes extraídos explican un 56,44% de la varianza total de la matriz de correlaciones, distribuida según los correspondientes autovalores del modo siguiente: 12.55 (27.9%), 2.43 (5.41%), 2.08 (4.63%), 1.88 (4.18%), 1.60 (3.56%), 1.37 (3.05%), 1.22 (2.71%), 1.18 (2.63%), 1.07 (2.37%). Una vez rotados dichos componentes se obtuvo la matriz en la que, los datos no apoyaban la estructura inicialmente esperada, encontrándose las siguientes inconsistencias:

- En la subescala *sensibilidad interpersonal*, los ítems se reparten entre el primer componente (*depresión*) y el tercero (*ideación paranoide*).
- Aparece un componente de sólo dos ítems correspondientes a la escala de *obsesividad-compulsividad*, con una supuesta inestabilidad (menos de 3 variables por factor) (Costello & Osborne, 2005; Velicer & Fava, 1998).
- En la subescala *ansiedad*, sus ítems se agrupan, por un lado, en una terna bien diferenciada (componente 8), mientras los dos ítems restantes se entremezclan con los propios del cuarto componente (*ansiedad fóbica*).
- Los componentes segundo (*somatización*), quinto (*hostilidad*) y séptimo (*psicoticismo*) no cambian su estructura.
- El ítem 39 resultó en cierto modo ambiguo, ya que, además de no ser suficientemente explicado por el quinto componente ( $r_{xy} = 0.48$ ; patrón= 0.26) podría estar en el primero ( $r_{xy} = 0.47$ ; patrón= 0.22).

Finalmente, en general, las correlaciones entre componentes resultaron ser moderadas.

Para estudiar la estructura latente del SA-45, se optó por el Análisis de Ejes Principales (AEP), ( $p < .001$  en la prueba de Kolmogorov-Smirnov en todos los ítems).

Teniendo en cuenta las restricciones propias del AEP al estimar tanto las comunalidades como los autovalores y pesos factoriales (Guttman, 1956; Thurstone, 1947), la solución factorial final resultó básicamente idéntica a la obtenida mediante ACP, al igual que en el estudio de Sandín et al. (2008). El AEP corrigió a la baja las comunalidades y saturaciones obtenidas en el ACP previo, y al alza las correlaciones entre factores. Sin embargo, el porcentaje de varianza explicada por los nueve factores (45.92% de la varianza total) superaba el máximo de varianza común introducida inicialmente en la matriz de correlaciones (45.58% del total). En buena lógica (Harman, 1976), nos llevó a realizar una nueva factorialización de la matriz de correlaciones, extrayendo un número menor de factores.

Para este nuevo análisis, se empleó como criterio para la extracción de factores el *Parallel Analysis* (PA; Horn, 1965). Se compararon los autovalores observados con los estimados mediante PA, y se decidió retener los ocho primeros factores cuyos autovalores resultaban superiores a al punto de corte.

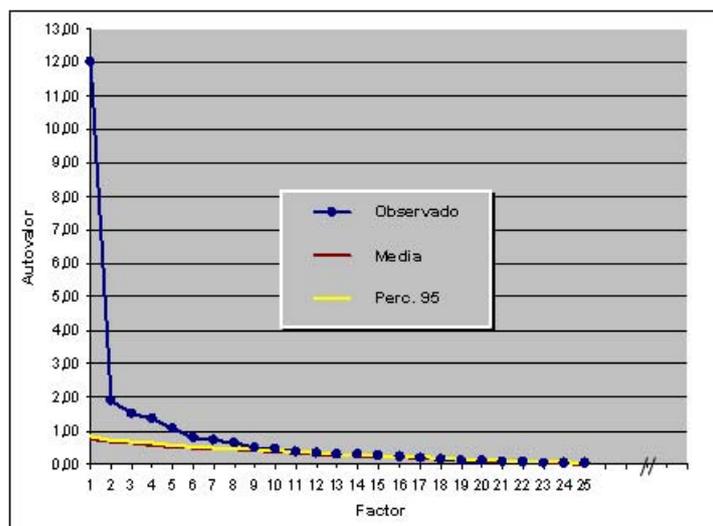


Gráfico 1. AP: Autovalores observados y simulados

Los ocho factores extraídos explican el 97.89% de la varianza común, y el 44.62% de la varianza total.

Los resultados finalmente obtenidos conservan las fortalezas y debilidades de las soluciones con nueve factores anteriormente expuestas y en general, la estructura factorial resultante parece aproximarse más a la estructura simple que las anteriores.

El primer factor mantiene los ítems y la denominación originales de “*depresión*”, más tres ítems pertenecientes inicialmente a la escala de *sensibilidad interpersonal*.

A su vez, el segundo factor conservaría su denominación como “*ideación paranoide*”, considerando tres ítems, en principio pertenecientes a otra escala pero que podrían ser considerados como parte de la sintomatología paranoide.

Por otra parte, los factores tercero (“*somatización*”), quinto (“*hostilidad*”), sexto (“*obsesividad-compulsividad*”) y séptimo (“*psicoticismo*”) resultaron acordes a las expectativas teóricas, manteniendo los ítems y etiquetas originales.

Finalmente, se decidió modificar la denominación de los factores cuarto y octavo intentando representar más adecuadamente dimensiones latentes diferentes. Así, al cuarto factor se le asignó la etiqueta “*ansiedad aguda-situacional*”, ya que, además de incluir los síntomas agorafóbicos asociados a determinados contextos (escala original de *ansiedad fóbica*), incorpora dos ítems teóricamente pertenecientes a la escala *ansiedad*, pero diferenciados de los tres restantes en cuanto a lo abrupto e intenso de la sintomatología. Estos tres últimos ítems conformarían el octavo factor, que ha sido denominado “*ansiedad general*”, que se refieren a estados de ansiedad, nerviosismo, intranquilidad o tensión más globales y, en cierto modo, menos intensos que los anteriores.

Las correlaciones más altas se observaron en el primer factor (*depresión*; desde 0.459 con *ansiedad aguda-situacional* hasta 0.616 con *obsesividad-compulsividad*), y las más bajas en el octavo (*ansiedad general*; desde 0.245 con *psicoticismo* hasta 0.528 con *depresión*) (Tabla 1).

Tabla 1. Correlaciones entre factores (8)

	1	2	3	4	5	6	7	8
1	1,000							
2	,573	1,000						
3	,569	,445	1,000					
4	,459	,418	,440	1,000				
5	,563	,449	,462	,411	1,000			
6	,616	,508	,549	,399	,381	1,000		
7	,478	,427	,388	,516	,563	,349	1,000	
8	,528	,413	,445	,315	,387	,400	,245	1,000

Para completar el estudio psicométrico, con la solución de ocho factores se llevó a cabo el estudio de fiabilidad, tanto de las subescalas obtenidas mediante AFE como de la escala total, teniendo como referente las nueve subescalas teóricas de cinco ítems cada una (véase tabla 2).

Las subescalas de *hostilidad* ( $\alpha= 0.791$ ), *obsesividad-compulsividad* ( $\alpha= 0.744$ ), *psicoticismo* ( $\alpha= 0.687$ ) y *somatización* ( $\alpha= 0.780$ ), mantienen iguales a los índices de las subescalas teóricas. En el resto de subescalas se obtuvieron mejores resultados con la nueva estructura.

Tabla 2. Subescalas teóricas vs. Solución AFE

Subescalas teóricas	$\alpha$	Subescalas AFE	$\alpha$
<i>Depresión</i>	0.798	<i>Depresión</i>	0.849
<i>Ideación paranoide</i>	0.691	<i>Ideación paranoide</i>	0.778
<i>Somatización</i>	0.780	<i>Somatización</i>	0.780
<i>Ansiedad fóbica</i>	0.742	<i>Ansiedad aguda-situacional</i>	0.794
<i>Hostilidad</i>	0.791	<i>Hostilidad</i>	0.791
<i>Obsesividad-compulsividad</i>	0.744	<i>Obsesividad-compulsividad</i>	0.744
<i>Psicoticismo</i>	0.687	<i>Psicoticismo</i>	0.687
<i>Ansiedad</i>	0.778	<i>Ansiedad general</i>	0.831
<i>Sensibilidad interpersonal</i>	0.783		

Total escala SA-45:  $\alpha= 0.939$

## Discusión

Los resultados aportados por nuestro análisis son concordantes con los estudios anteriores en múltiples aspectos, como en la predominancia en términos factoriales y de consistencia interna de la escala de *depresión*, la mayor inestabilidad hallada en las escalas teóricas de *ideación paranoide* y *psicoticismo*, y la relativamente alta diferenciación de esta última escala junto con la de *somatización* (e. g. Maruish et al., 1998).

Sin embargo, se sugiere una estructura factorial compuesta por ocho factores correlacionados, en los que la dimensión teórica *sensibilidad interpersonal*, que principio mostraba una consistencia interna aceptable, termina diluyéndose entre las dimensiones *depresión* e *ideación paranoide*. Esta aparente anomalía estructural, no resultaría en todo caso sorprendente a tenor de hallazgos previos, que ya habían aludido a posibles solapamientos o al menos una significativa relación entre estas

tres dimensiones teóricas (Davison et al., 1997; Maruish et al., 1998; Maruish, 1999b). Del mismo modo, este dato apunta en la dirección señalada por Sandín y sus colaboradores, quienes aventuraban esta posibilidad aduciendo un posible problema de validez de contenido en los ítems de la escala de *sensibilidad interpersonal*. Algo similar ocurriría con la escala teórica de *ansiedad*, la cual en nuestro estudio aparece dividida entre un factor de *ansiedad general* claramente diferenciado, y el factor originalmente denominado *ansiedad fóbica*. De hecho, y en la línea anticipada por Davison et al. (1997) acerca de las diferentes “intensidades” a las que se podrían estar refiriendo los ítems de *ansiedad*, en nuestra solución de ocho factores puede observarse cómo la consistencia interna de las escalas redefinidas como *ansiedad general* y *ansiedad aguda-situacional* mejora en comparación con sus homólogas teóricas.

Finalmente, sería conveniente contrastar estos resultados y conclusiones en estudios con muestras más amplias. En consonancia con Sandín et al. (2008), pensamos que sería de especial interés comenzar a implementar estudios en ámbito clínico, ya que éste es el propósito fundamental para el que fue desarrollado el SA-45. Teniendo en cuenta que posiblemente los indicadores de fiabilidad, estabilidad factorial y capacidad discriminativa mejoren en esta población (Davison et al., 1997), merecería la pena seguir indagando en pos de la optimización de un instrumento prometedor, tanto en la investigación de resultados en psicoterapia como en la eficacia del instrumento, en la praxis clínica cotidiana.

#### Referencias

- Costello, A.B. & Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation, 10*(7). Recuperado el 10 de marzo de 2009, de <http://pareonline.net/pdf/v10n7.pdf>
- Davison, M.K., Bershinsky, B., Bieber, J., Silversmith, D., Maruish, M.E. & Kane, R.L. (1997). Development of a brief, multidimensional, self-report instrument for treatment outcomes assessment in psychiatric settings: Preliminary findings. *Assessment, 4*, 259-276.
- Derogatis, L.R. & Cleary, P.A. (1977). Confirmation of the dimensional structure of the SCL-90: A study in construct validation. *Journal of Clinical Psychology, 33*, 981-989.
- Derogatis, L.R., Lipman, R.S. & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale - Preliminary report. *Psychopharmacology Bulletin, 9*, 13-27.
- Guttman, L. (1956). Best possible systematic estimates of communalities. *Psychometrika, 21*, 273-285.
- Harman, H.H. (1976). *Análisis factorial moderno*. Madrid: Saltés.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Maruish, M.E. (1999b). Symptom Assessment-45 Questionnaire (SA-45). En M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (2ª ed., pp. 725-757). Mahwah: Lawrence Erlbaum Associates.

- Maruish, M.E., Bershady, B. & Goldstein, L. (1998). Reliability and validity of the SA-45: Further evidence from a primary care setting. *Assessment*, 5, 407-419.
- Sandín, B., Valiente, R.M., Chorot, P., Santed, M.A. y Lostao, L. (2008). SA-45: forma abreviada del SCL-90. *Psicothema*, 20, 290-296.
- Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Velicer, W.F. & Fava, J.L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3, 231-251.
- Ware, J.E., Kosinski, M. & Keller, S.D. (1995). *SF-12: How to score the SF-12 physical and mental health summary scales* (2ª ed.). Boston: The Health Institute, New England Medical Center.

# LVE: LABORATORIO VIRTUAL DE ESTADÍSTICA

José Chacón<sup>1</sup>, Víctor de la Fuente<sup>2</sup>, Trinidad Ruiz Gallego-Largo<sup>1</sup>, Francisco Serradilla<sup>2</sup>,  
María José Hernández-Lloreda<sup>1</sup> y M. Rosario Martínez-Arias<sup>1</sup>

<sup>1</sup>Universidad Complutense de Madrid

<sup>2</sup>Universidad Politécnica de Madrid

En el marco del Espacio Europeo de Educación Superior (EEES), el proyecto Laboratorio Virtual de Estadística (LVE) es un sistema integrado multiplataforma accesible vía Web y que sirve de apoyo al alumno en el aprendizaje de la estadística inferencial. El sistema permite una aproximación visual a diferentes conceptos y procedimientos estadísticos y los presenta tanto teórica como numérica y gráficamente. Además, mediante simulaciones dinámicas (*applets* Java), permite al usuario interactuar con él, manipulando los parámetros de cada módulo. Los aspectos más relevantes que ofrece el LVE son: (1) La posibilidad de visualizar lo que la teoría afirma y manipular sus elementos constituyentes. (2) La existencia de un “tutor” interactivo, que proporciona una explicación contextual de los elementos seleccionados. (3) La posibilidad de utilizar los ejemplos del sistema o los datos que el propio alumno introduzca. (4) Una gestión autónoma por parte del alumno, tanto en relación a la distribución temporal como en el grado de profundidad. (5) La facilidad de acceso para los estudiantes, vía web (aunque también puede descargarse para su uso en modo local).

No es una novedad indicar que en la enseñanza de la estadística, especialmente en los procedimientos de inferencia, se observa una dificultad por parte de los alumnos para entender algunos conceptos centrales. Ejemplo paradigmático es la distribución muestral de un estadístico, eje de cualquier procedimiento de inferencia pero del que muy pocos alumnos podrían decir cómo se obtiene, qué es (qué representa) y, sólo entonces, qué podemos hacer con ello. Si podemos responder a estas tres preguntas, el resultado de cualquier contraste deviene evidente. Pero esto no es lo más habitual.

Aceptemos que el conjunto de ideas a desarrollar no son de fácil comprensión. Precisamente por ello, los que enseñamos esta materia hemos utilizado procedimientos varios: desde la omnipresente pizarra a las transparencias “animadas”, presentaciones, estáticas o dinámicas, o simulaciones. La simulación quizás sea lo más fructífero; después de manipular una y otra vez los parámetros de una simulación, parece alcanzarse un nivel de comprensión suficiente como para responder a las preguntas anteriores. Pero al cabo de una o dos clases, aquella comprensión ha desaparecido, o no parece haber la misma claridad.

Quizás una solución, o al menos ayuda, puede ser que los propios alumnos lleven a cabo la simulación por ellos mismos, de forma interactiva; que sean ellos quienes modifiquen los parámetros y vean los resultados; que los lleven a los límites y comprueben qué ocurre. Y que puedan volver a hacerlo si lo creen necesario. Con esto en mente hemos construido el Laboratorio Virtual de Estadística (LVE).

## *Objetivos y características*

El LVE consiste en un conjunto de páginas web en las que los alumnos disponen básicamente de (1) una breve introducción donde aparece una presentación teórica que aglutina los aspectos principales sobre el concepto tratado y (2) un programa empotrado (*applet* Java) que permite la interacción con el alumno y que muestra gráfica y numéricamente varias aplicaciones de algún concepto estadístico.

Aunque existen algunas experiencias a este respecto, tanto en inglés (por ejemplo, Lane et al., 2006) como en español (Área docente de estadística, Universidad de Córdoba, sin fecha; Departamento de Matemática Aplicada, UCM, sin fecha), ninguna de esas opciones permitía contemplar de forma integrada los objetivos que teníamos en mente. En este sentido, si queremos exponer las características del sistema, quizás la mejor manera sea enunciando los objetivos que perseguíamos.

El primer objetivo era permitir “ver”. Eso nos ha llevado al conjunto de programas que muestran visualmente algún proceso estadístico y su resultado. Y, para evitar las ambigüedades de una foto fija, permiten manipular esa vista hasta entender lo que tenemos frente a nosotros, por lo que los programas responden interactivamente a la manipulación de los parámetros implicados en cada caso. Previamente se dispone de una introducción teórica, breve, donde se expone el concepto a tratar y los elementos que participan, lo que en ningún caso pretende ser un sustituto de algún manual o de las clases presenciales.

Un segundo objetivo era evitar la posible desconexión entre la manipulación de los *applets* y los conceptos estadísticos implicados. La solución fue lo que llamamos *tutor*: una pequeña ventana emergente que, a petición del usuario, informa de qué es y/o qué función tiene el elemento sobre el que está situado el ratón.

En tercer lugar, creíamos crucial que los estudiantes pudiesen utilizar el sistema a su propio ritmo, y este formato lo permite: acudir directamente al lugar de interés, repetir las simulaciones tantas veces como se quiera o decidir el grado de detalle en la simulación.

También queríamos ofrecer una serie de ejemplos que ilustrasen los conceptos estadísticos, pero a la vez parecía interesante que los estudiantes pudiesen introducir sus propios datos. La solución fue ofrecer ambas opciones.

Y en quinto pero imprescindible lugar, la facilidad de acceso. Actualmente está disponible a través de Internet (<http://aicu.eui.upm.es/lve>), lo que permite el acceso desde dentro o fuera de la facultad, y también puede descargarse para ejecutarlo posteriormente en un ordenador no conectado a Internet. El único requisito es disponer de un navegador que tenga instalada la máquina virtual de Java (gratuita, autoinstalable y descargable desde Internet).

### *Contenidos*

La página de presentación (<http://aicu.eui.upm.es/lve>) da acceso a los contenidos principales, que podemos agrupar en tres categorías: conceptos previos, distribución muestral y aplicaciones de la distribución muestral.

### *Conceptos previos*

Los dos primeros módulos tratan de la construcción de histogramas y del muestreo. Entender estos conceptos es imprescindible para entender otros posteriores, y nuestra experiencia apunta a que no siempre se consigue adecuadamente. El primero está dedicado a tratar en profundidad el elemento gráfico que será utilizado en los demás programas: los histogramas, y su construcción (fig. 1). El programa permite modificar el número de intervalos, su anchura o los límites de la distribución, permitiendo apreciar, por ejemplo, las formas tan diferentes que puede adoptarse el histograma a partir de unos mismos datos. Todos los elementos (incluyendo la tabla de frecuencias) responden interactivamente a cualquier modificación.

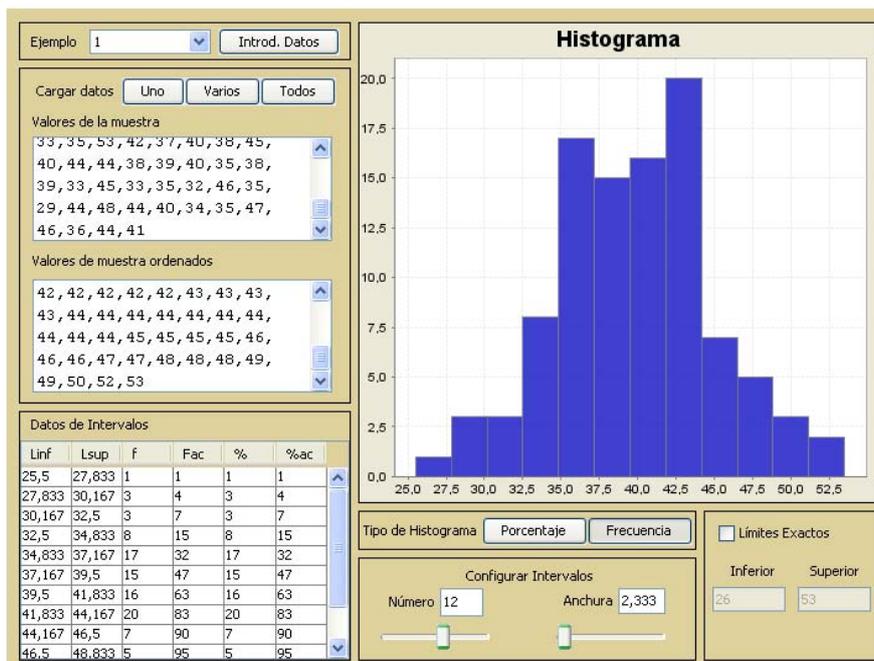


Figura 1. Programa para la construcción de histogramas

Su funcionamiento parte de la “extracción” de los datos, que puede realizarse dato a dato, extrayendo varios de una vez o todos con una sola pulsación. Los datos se van ordenando y representando conforme van siendo extraídos, y las características del histograma pueden ser modificadas en cualquier momento.

El siguiente programa trata el muestreo; otro concepto que, creemos, no siempre es bien entendido o, por expresarlo más exactamente: lo que habitualmente tenemos en mente al pensar en un muestreo no siempre corresponde con los resultados reales (o posibles) de tal proceso. El programa sobre el muestreo permite observar los resultados de un muestreo aleatorio simple, extrayendo seis muestras de tamaño arbitrario de una de dos poblaciones disponibles (normal o uniforme). El usuario determina el tamaño muestral, y lleva a cabo las extracciones con el nivel de detalle que desee (desde dato a dato hasta todas las muestras de una vez). Los propios gráficos, así como las medias y desviaciones típicas (indicadas para la población y cada muestra), permiten observar las relaciones entre la población, el tamaño muestral, y las propiedades de las muestras así seleccionadas (figura 2).

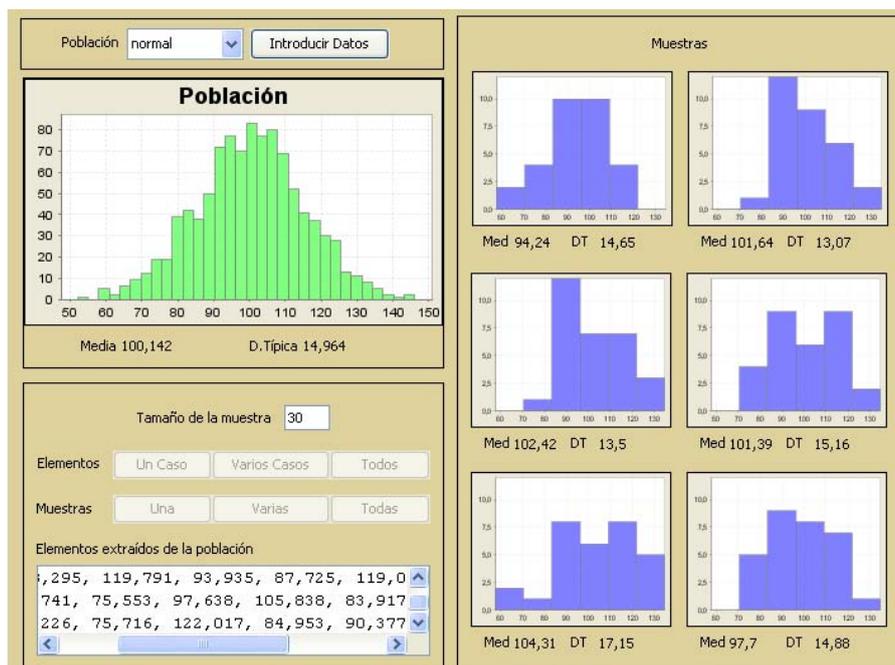


Figura 2. Programa para la realización del muestreo, con seis muestras extraídas de una población normal

### Distribución muestral (DM)

No es exagerado decir que éste es el concepto central de la estadística inferencial, ni que no es comprendido completamente por nuestros alumnos. Por ello contamos con dos programas, para atender el caso de la DM de una media, y el de una proporción.

El primero de ellos (figura 3) permite visualizar la construcción, mediante simulación, de la DM de la media. El alumno, después de seleccionar la población de ejemplo, debe indicar el tamaño de cada muestra y el número de muestras a extraer. Al igual que en programas anteriores, puede extraer las muestras una a una, varias, o todas de una vez. Para cada muestra extraída se ofrece su media y desviación típica, y la primera es “anotada” en el histograma correspondiente a la DM de la media. El programa permite, además, extraer de forma inmediata 1000 muestras, así como superponer al histograma obtenido la curva correspondiente a la DM teórica para su comparación.

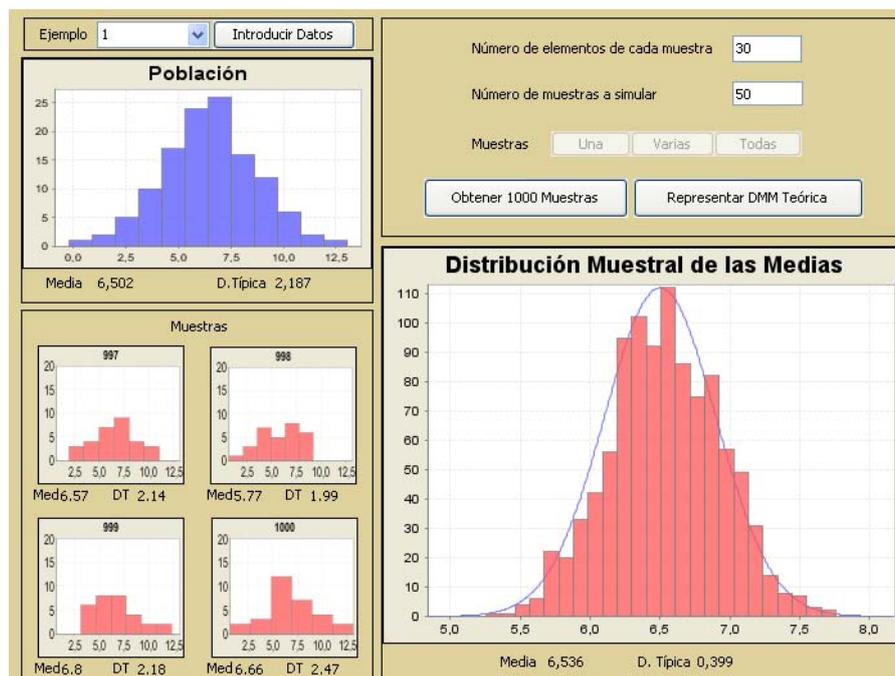


Figura 3. Programa para la distribución muestral de la media

El programa para la DM de la proporción (figura 4) permite hacer algo similar al anterior, sólo que ahora partimos de una variable dicotómica, y el usuario ha de indicar la probabilidad de éxito y el número de ensayos (o extracciones) realizadas, así como el número de muestras a extraer. Ello permite generar la DM del estadístico “nº de éxitos”. De igual forma, el programa permite comprobar el porcentaje de casos que coincide con un valor dado, así como el porcentaje de casos con un número de éxitos inferior y superior.

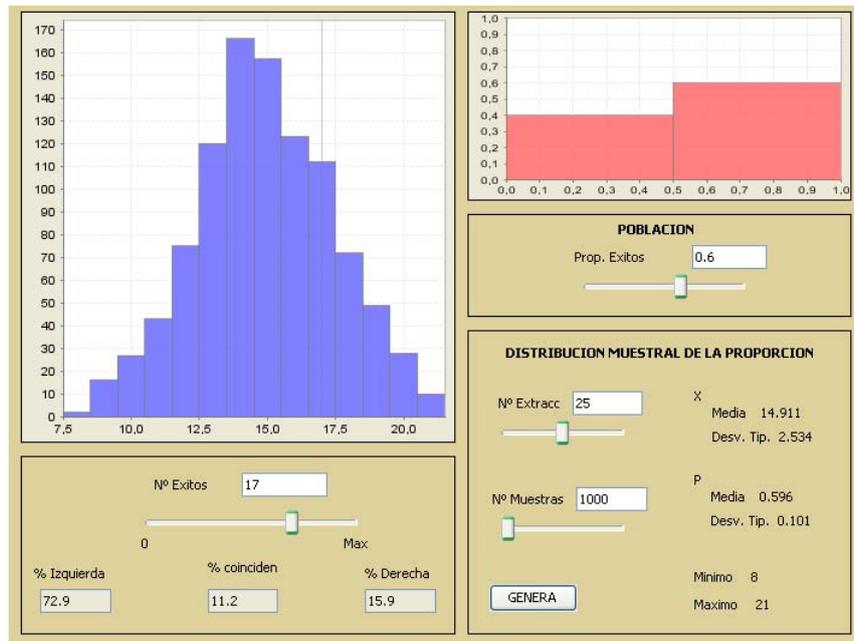


Figura 4. Programa para la distribución muestral de la proporción

### Aplicaciones de la distribución muestral

Hay cuatro módulos que hacen uso de la distribución muestral para realizar otros análisis. El primero de ellos permite la construcción y manipulación de intervalos de confianza para el caso de una media, pudiendo comenzar con alguno de los ejemplos propuestos o introduciendo datos propios. El programa proporciona unos descriptivos básicos y ofrece el intervalo a un determinado NC, indicando sus límites en la escala de la distribución  $t$  o en la escala de la variable ( $X$ ). Los controles bajo la representación del intervalo (ver figura 5) permiten modificar parámetros (media, desviación típica, tamaño muestral o el nivel de confianza) y ver cómo afecta al intervalo obtenido.

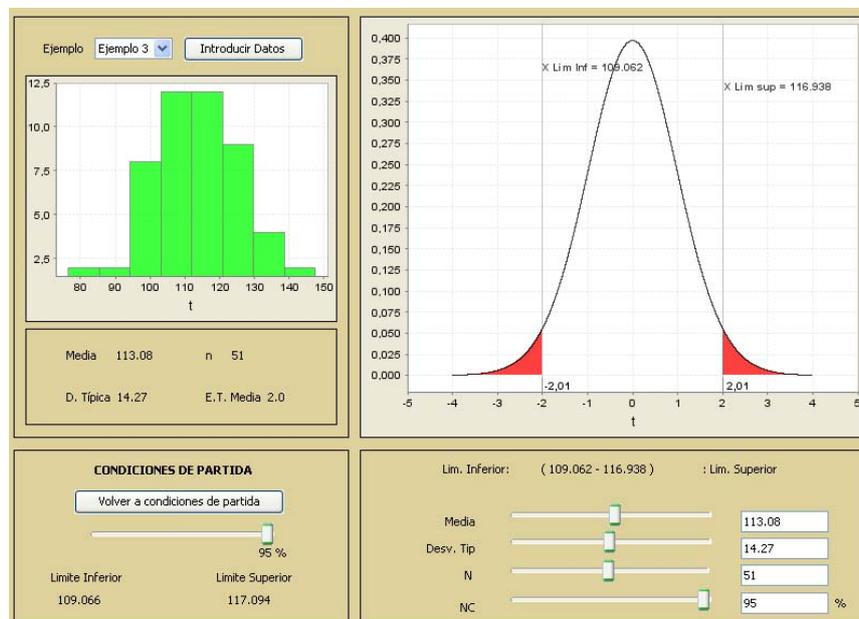


Figura 5. Programa para el intervalo de confianza para una media

Con respecto a los contrastes de hipótesis, contamos con tres módulos. El primero de ellos (figura 6) permite manipular los parámetros que influyen en el contraste para una media ( $\alpha$ , distancia a la media, desviación típica y tamaño muestral), y visualizar su efecto en las distribuciones de probabilidad correspondientes a las hipótesis estadísticas (errores tipo I y II, nivel de significación, potencia, nivel crítico y tamaño del efecto).

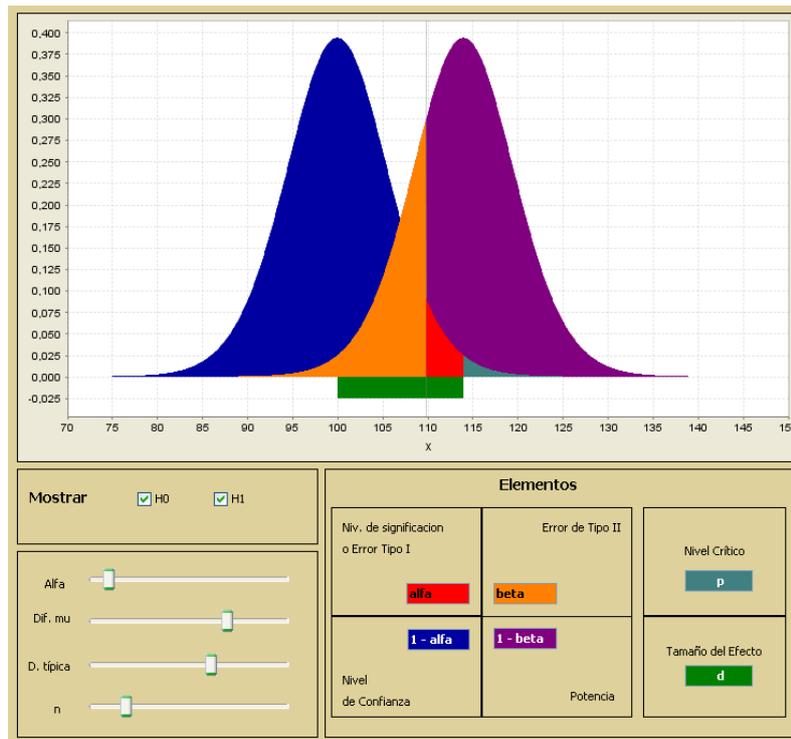


Figura 6. Programa para los elementos del contraste para una media

El siguiente programa realiza el contraste para una media respecto a un valor propuesto (figura 7). Ofrece un histograma de los datos junto a unos descriptivos básicos. En este caso, el programa permite seleccionar la direccionalidad del contraste, el valor de la media poblacional propuesta y el nivel de confianza para el contraste. El programa ofrece los valores para el estadístico de contraste,  $T$ , el límite de la región crítica, y el nivel crítico.

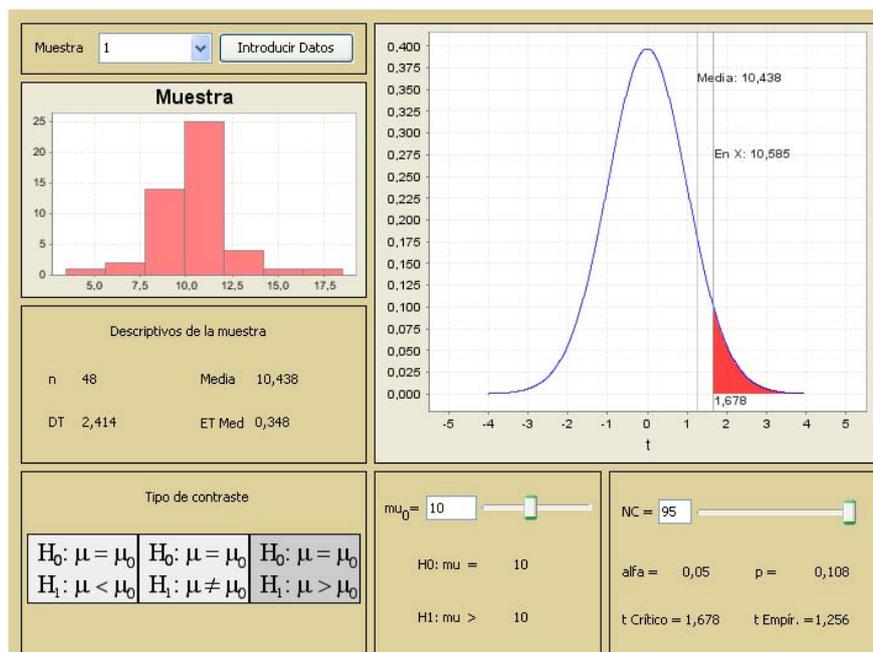


Figura 7. Programa para el contraste de una media

Y por último, contamos con un programa para el contraste de dos medias independientes. Es similar en todo al anterior, pero atendiendo a dos conjuntos de datos. En este caso incorpora el test de Levene, que comprueba el supuesto de igualdad de varianzas, y que es necesario realizar antes de realizar el contraste de medias para determinar la distribución del estadístico de contraste.

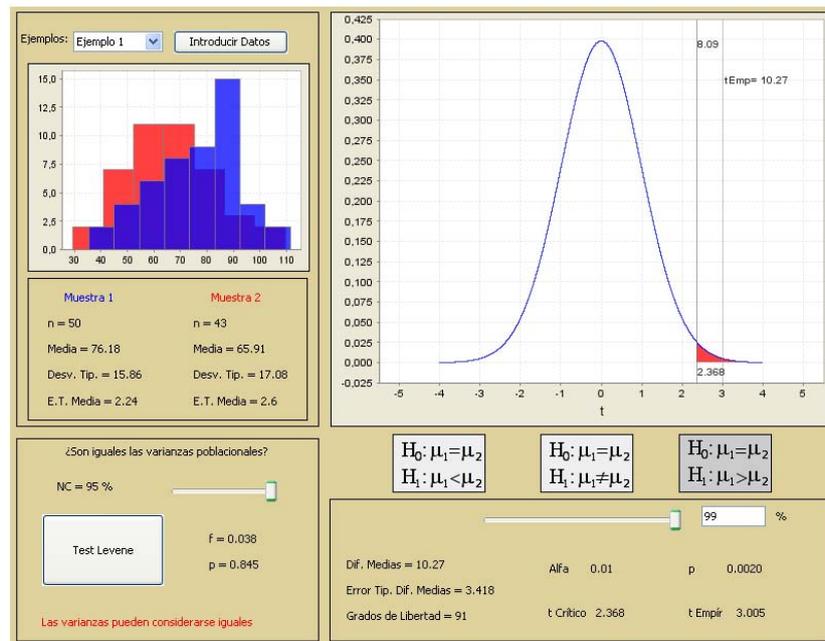


Figura 8. Programa para el contraste de dos medias independientes.

## Conclusión

El LVE está orientado a dos objetivos básicos: el aprendizaje a través de la interactividad de conceptos estadísticos tradicionalmente difíciles, y la autogestión por parte del alumno tanto en contenidos como en dedicación. Y creemos que el formato y las características del sistema permiten estos objetivos. No obstante, también creemos que los ejemplos y sugerencias que se proporcionan no siempre garantizan que el alumno capte la estructura de la información con una coherencia suficiente, lo que podría derivar en una mera manipulación de números y gráficos sin sentido. En este sentido, creemos que este sistema puede mostrar su máximo rendimiento si lo usamos proporcionando a nuestros alumnos datos o situaciones y preguntas “inteligentes”, que (idealmente, sólo) puedan responderse manipulando los programas.

Así, quizá con una batería de problemas adecuados y la esperable cooperación por parte de cada alumno, algunos conceptos difíciles puedan serlo un poco menos.

## Nota de los autores

El Laboratorio Virtual de Estadística es un proyecto conjunto de los departamentos Metodología de las Ciencias del Comportamiento de la UCM y Sistemas Inteligentes Aplicados de la UPM, realizado gracias a la financiación del Vicerrectorado de Desarrollo y Calidad de la Docencia de la UCM, a través de los Proyectos de Innovación y Mejora de la Calidad Docente nº 515 (2006-2007) y nº 127 (2007-2008). En el desarrollo de las aplicaciones contenidas en el Laboratorio Virtual de Estadística se han empleado las librerías JFreeChart y JSci, ambas con licencia LGPL.

## Referencias

- Área docente de estadística, Universidad de Córdoba. (sin fecha). *Simulaciones Estadísticas Telemáticas*. Accedido el 7 de febrero de 2010 desde [http://www.uco.es/simulaciones\\_estadisticas/index.php?menu=inicio](http://www.uco.es/simulaciones_estadisticas/index.php?menu=inicio).
- Departamento de Matemática Aplicada, Universidad Complutense de Madrid. (sin fecha). *Aula Virtual de Bioestadística*. Accedido el 7 de febrero de 2010 desde [http://e-estadistica.bio.ucm.es/cont\\_mod\\_1.html](http://e-estadistica.bio.ucm.es/cont_mod_1.html).
- Lane, D. M. et al. (2006). *Rice Virtual Lab in Statistics*. Accedido el 7 de febrero de 2010 desde <http://onlinestatbook.com/rvls/index.html>.

# IDENTIFICACIÓN DE PERFILES PROAMBIENTALES EN EL ALUMNADO DEL CERTIFICADO DE APTITUD PEDAGÓGICA EN LA UNIVERSIDAD DE MÁLAGA

Ligia I. Estrada y Antonio Matas  
Universidad de Málaga

La presente investigación tiene como objetivo la identificación de las diferentes formas de entender la Educación Ambiental que tiene el alumnado del curso CAP (Certificado de Aptitud Pedagógica) de la Universidad de Málaga. Para ello, se creó un cuestionario “*ad hoc*” partiendo de las distintas propuestas recogidas en la literatura sobre educación ambiental, siendo administrado a un total de 412 alumnos. Posteriormente se realizó un análisis de las características psicométricas del instrumento, así como el análisis descriptivo de los datos. Los resultados muestran la existencia de distintas formas de entender la educación ambiental, así como el predominio de la muestra hacia interacciones ambientalmente responsables, si bien esta situación está mediada por algunas variables sociodemográficas.

Esta investigación se encuentra inmersa dentro de una línea de trabajo más amplio, centrado en el análisis de las diferentes formas que tiene la ciudadanía de concebir y de afrontar su interacción con el medioambiente. En este caso, la investigación que se presenta se centra en el análisis de las posibles formas de conceptualizar la Educación Ambiental que posee el alumnado que estudió el curso para la obtención del Certificado de Aptitud Pedagógica (C.A.P.) de la Universidad de Málaga (promoción 2008-09).

La Educación Ambiental es tratada a lo largo de esta comunicación como una parte importante del bagaje cognitivo, actitudinal y procedimental del docente, siendo de gran interés por ser parte del material que usará el futuro profesor para llevar a cabo la labor de formar y educar a jóvenes ciudadanos.

Se puede considerar que la Educación Ambiental es «*consecuencia del cambio de lectura que el hombre empieza a realizar, a fines de la década de los sesenta, del escenario de su vida*» (Sureda, Cañellas y Colom, 1989, p. 90). En el mundo contemporáneo, muchas instituciones vienen diseñando estrategias para la formación en educación ambiental, valorando la actitud y el comportamiento que el ser humano dirige hacia el medioambiente que le rodea. Por ello, desde instituciones y organismos, tanto internacionales como nacionales, como UNESCO, Parlamento Europeo y del Consejo, Agenda 21, Ministerio de Medio Ambiente y Medio Rural y Marino, etc., se impulsa el fomento de actuaciones responsables hacia el medioambiente a través de tratados, convenios y acuerdos internacionales, así como la legislación comunitaria, nacional, regional o local sobre el medio ambiente o relacionados con el mismo (La Carta de Belgrado, 1975; Declaración de Salónica, 1997; etc.). Se trata de promover la implicación de todos, incluyendo el compromiso de los gobiernos, a través de distintas conferencias, así como de las organizaciones gubernamentales, intergubernamentales y no gubernamentales (ONGs).

En nuestro país son varias las actuaciones que se están llevando a cabo en educación ambiental, como las desarrolladas por el “Centro Nacional de Educación Ambiental (CENEAM)”<sup>[1]</sup>, el programa de integración “Aulas de la naturaleza”<sup>[2]</sup>, “Campaña Municipal de Educación Ambiental”<sup>[3]</sup>, “Programa de Educación Medioambiental”<sup>[4]</sup>, o “Recursos sobre educación Ambiental”<sup>[5]</sup>.

[1] [http://www.mma.es/portal/secciones/formacion\\_educacion/ceneam01/](http://www.mma.es/portal/secciones/formacion_educacion/ceneam01/)

[2] <http://www.educarm.es>

[3] <http://www.diarioinformacion.com>

[4] <http://www.motril.es/index>

[5] <http://www.educa.madrid.org/>

También son varios los estudios que se realizan desde instituciones públicas y privadas (como los llevados a cabo el área de trabajo de Sostenibilidad y Medio Ambiente perteneciente a la Cátedra UNESCO de Desarrollo sostenible y educación ambiental<sup>[6]</sup> desde el Programa MaB, o el Ecobarómetro Andaluz, para determinar ciertas actitudes y conductas hacia el medioambiente, pero que investigan de manera local y con objetivos muy específicos que les ayuden a solventar los problemas que se encuentran.

En la Declaración de Tbilisi de 1977 se considera que *“La Educación Ambiental debe impartirse a personas de todas las edades, a todos los niveles y en el marco de la educación formal y no formal”* (UNESCO-PNUMA, 1977, p. 1), y el Libro Blanco de la Educación Ambiental en España afirma que la Educación Ambiental trata de *“promover la acción pro-ambiental entre individuos y grupos sociales; (...), realizada en los contextos vitales de las personas: hogar, trabajo, escuela, ocio y comunidad”* (Ministerio de Medio Ambiente, 1999, p.1). Teniendo en cuenta estas manifestaciones, se infiere la importancia que pueden tener las distintas perspectivas que sobre la Educación Ambiental pueden existir entre los educadores.

Dentro del ámbito escolar español, la Educación Ambiental fue tratada como un eje transversal del curriculum con el desarrollo de la Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo. En su título preliminar, artículo 2, contempla como uno de los principios de la actividad educativa: *“la relación con el entorno social, económico y cultural”*, así como *“la formación en el respeto y defensa del medio ambiente”*. Posteriormente en la Ley Orgánica 2/2006, de 3 de mayo, de Educación, se reestructura el sistema educativo, si bien se sigue haciendo hincapié en la importancia de los contenidos medioambientales en la formación de futuros ciudadanos. De ahí la consideración de estudiar al alumnado del C.A.P., puesto que como futuros docentes, desempeñarán un papel básico en el desarrollo de valores, actitudes, creencias y comportamientos en su alumnado. Como expone Sauvé (2004, p. 2):

*“La formación de profesores, animadores y otros dinamizadores en Educación Ambiental es fundamental para el desarrollo de este campo y responde a una triple problemática: ambiental, social y educativa.”*

La formación y capacitación de docentes para la Educación Ambiental es necesaria, pero ¿se conoce el perfil que tienen estos profesionales en relación a la forma de entender la educación ambiental?

Aunque esta pregunta podría parecer básica para el desarrollo de Educación Ambiental en las aulas, a la hora de realizar una revisión de los estudios realizados con relación a la misma, se puso de manifiesto el reducido número de investigaciones que tratan este tema. Esto puede deberse a varios motivos, uno de ellos que se asume que la conceptualización de la Educación Ambiental se desarrolla al mismo tiempo que los docentes “aprenden” sobre cómo ejercer como docentes. Otra posibilidad es que la Educación Ambiental sea considerada como una materia que el profesor transmite a sus alumnos de forma similar a como se hace con las matemáticas, la literatura, o el dibujo. Es decir, a través de unos contenidos “pre-establecidos” que son explicados al alumnado, y éste posteriormente, realizará una serie de actividades evaluativas para demostrar que ha adquirido dichos conocimientos. En cualquier caso, y no cerrando las puertas a otras muchas posibilidades, la literatura sobre formación en Educación Ambiental pone de manifiesto la necesidad de analizar posibles formas diferentes de entender este campo formativo, así como el hecho de que esta cuestión no está suficientemente estudiada desde nuestro punto de vista. En

---

[6] [http://www.unescoeh.org/base/ingurumena.php?id\\_atala=5&id\\_azpiatala=11000&hizk=es&id\\_kont=588](http://www.unescoeh.org/base/ingurumena.php?id_atala=5&id_azpiatala=11000&hizk=es&id_kont=588)

este sentido cabe señalar el estudio de Sauv  (2004) que no solamente pone de manifiesto esta situaci n, sino que se aventura proponiendo hasta 15 posibles perfiles, a partir de un amplio estudio documental (procedente de distintos eventos, actuaciones, instituciones, investigaciones y estudios).

Justific ndonos en esta situaci n, y considerando que esta investigaci n es parte de una l nea de trabajo m s amplia, como se ha dicho anteriormente, el objetivo de este estudio est  dirigido a la identificaci n de las distintas formas de conceptualizar (perfiles) la Educaci n Ambiental que presenta el alumnado del CAP de la Universidad de M laga.

## M todo

Se consider  el alumnado del curso para la obtenci n de la Certificaci n de Adaptaci n Pedag gica (C.A.P) de la Universidad de M laga como poblaci n objetivo.

Se utiliz  un cuestionario para recoger la informaci n sobre las variables del estudio, cuyo proceso de construcci n del instrumento fue *ad hoc*, con un total de 75  tems. Para ello se recurri  a un grupo de expertos en Educaci n Ambiental quienes propusieron un total de 60  tems relativos a las distintas corrientes propuestas por Sauv  (2005), al que se denomin  *Cuestionario de Identificaci n de Perfiles en Educaci n Ambiental* (CIPEA). Se a nadi  el Cuestionario del Nuevo Paradigma Ecol gico (NPE) (Dunlap et al., 2000), con la intenci n de tener un instrumento que permitiese recoger datos que *a priori* podr an estar relacionados con los perfiles. Estos  tems a nadi dos, permitieron adem s una validaci n parcial del constructo. De esta forma, el instrumento qued  conformado por 60  tems sobre perfiles m s 15 del NPE.

El instrumento de medida se configur  como una de estimaci n que var a entre el valor 1 y 5 (teniendo el valor 1 = Totalmente en desacuerdo, 2= En desacuerdo, 3= Indeciso, 4= De acuerdo, y 5= Totalmente de acuerdo) en cada uno de los enunciados.

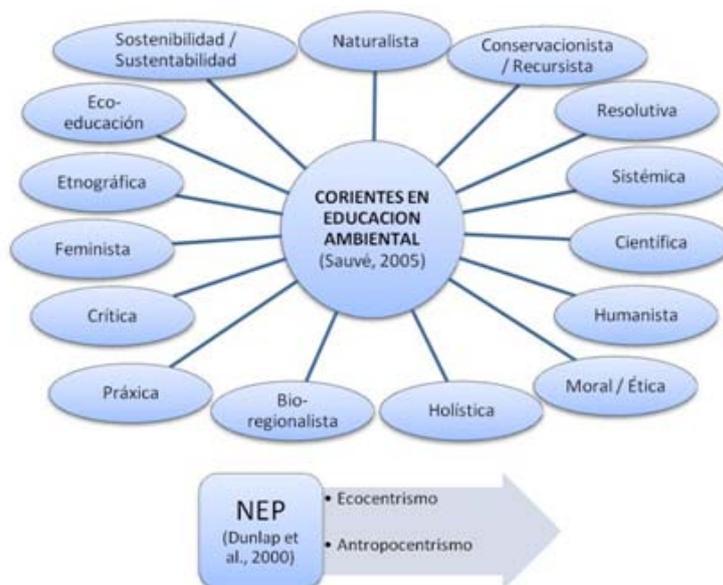


Figura 1. Las quince corrientes propuestas por Sauv  (2005) y el ecocentrismo ante el antropocentrismo propuesto desde el Nuevo Paradigma Ecol gico (2000)

El instrumento se administr  entre los meses de enero a febrero de 2009, en las instalaciones donde se imparten las materias pertenecientes al C.A.P., en la Universidad de M laga, siempre a la misma hora y por la misma encuestadora, durante las sesiones normales de clase.

Inicialmente, el diseño de encuesta incluye un estudio psicométrico del cuestionario, un análisis descriptivo de la información y un análisis de componentes principales que permita identificar distintos grupos, segmentos o patrones.

## Resultados

### *Características psicométricas*

Se ha obtenido que el Cuestionario de Identificación de Perfiles en Educación Ambiental (CIPEA) tiene una fiabilidad bastante alta, cuya consistencia interna alcanza un valor de 0,894 ( $\alpha$  de Cronbach). Al consultar la tabla 1 se podrá apreciar que estas mismas características psicométricas en el cuestionario del Nuevo Paradigma Ecológico (NEP) son inferiores que en el CIPEA, con un valor de 0,428 ( $\alpha$  de Cronbach).

*Tabla 1. Coeficientes Cronbach's Alpha (CIPEA y NEP)*

Cuestionario	Cronbach's Alpha	Nº de Items
CIPEA	0,893	60
NEP	0,428	15

Se realizó un análisis de componentes principal con rotación Normalización Varimax sobre el CIPEA. Los resultados sugieren la existencia de 15 componentes (varianza explicada del 59%). Este mismo análisis de componentes principales, aplicado sobre los datos procedentes del NPE sugiere la existencia de cuatro componentes (varianza explicada superior al 54%) Estos últimos resultados son coherentes con la literatura consultada (Dunlap et al., *opus cit.*).

Al consultar la tabla de Matriz de Componentes Rotados del CIPEA, se observó que al relacionar los nuevos perfiles con los propuestos por Sauv e no se corresponde  ntegramente ninguno, surgiendo as  una redistribuci n de las caracter sticas que se seleccionaron en cada una de las mismas. Por tanto, los perfiles identificados en la muestra no coinciden con los propuestos por Sauv e.

### *An lisis descriptivo de los datos*

La muestra inicial de 412 participantes se redujo a 123 personas para la realizaci n de este an lisis al eliminar todos los cuestionarios incompletos. Dicha muestra est  compuesta por 35 hombres (28,5%) y 88 mujeres (71,5%), procedentes de diversas titulaciones universitarias o de grado medio (diplomados, licenciados o con una titulaci n considerada con el mismo valor, como es el caso de los Conservatorios Superiores), cuya edad oscila entre los 21 y los 50 a os (la moda de la edad es de 24 a os). En cuanto al per odo de comienzo y finalizaci n de las titulaciones, estas fluctuaban entre el a o 1983 y el 2008.

La distribuci n por titulaciones fue de 28 sujetos (22,8 %) Licenciados en Biolog a o Bioqu mica, 14 sujetos (11,4 %) Licenciados en Ciencias Ambientales, 3 sujetos (2,4 %) Licenciados en Ciencias del Mar, 2 sujetos (1,6 %) Licenciados en Geolog a, 13 sujetos (10,6 %) Licenciados en Qu mica, 6 sujetos (4,9 %) Licenciados en Ingenier a Qu mica, 1 sujeto (0,8 %) Licenciado en Veterinaria, 1 sujeto (0,8 %) Ingeniero Agr nomo, 8 sujetos (6,5 %) Diplomados en Turismo, 3 sujetos (2,4 %) Diplomados en Magisterio, 16 sujetos (13,0 %) Licenciados en Psicolog a, 5 sujetos (4,1%) Licenciados en Comunicaci n Audiovisual, 13 sujetos (10,6 %) Licenciados en Periodismo y 10 sujetos (8,1 %) Licenciados en Geograf a.

La titulaci n que mayor representatividad manifest  fue Licenciado en Biolog a o Bioqu mica con una moda de 5.

### *Análisis descriptivo de las respuestas a los ítems en el CIPEA*

El rango de respuesta oscila entre 1 y 5, por lo que se puede considerar que la muestra manifiesta una diversidad en el mayor o menor acuerdo con el enunciado al que se responde. El valor 4 es el más elegido por la muestra, lo cual indicaría que la muestra tiende a estar “de acuerdo” con la mayoría de las acciones y características que se proponen desde las corrientes de intervención propuestas por Sauv . No presentan moda 4, trece ítems del total que definen los perfiles identificados en Educaci n Ambiental (uno con  $Mo=2$ , tres con  $Mo=3$  y nueve con  $Mo=5$ ).

A grandes rasgos, se puede decir que de las corrientes propuestas por Sauv  los participantes est n m s a favor de las siguientes: a) conservacionista o recursista (dos de sus cuatro enunciados tiene  $Mo=5$  y otros dos  $Mo=4$ ), b) hol stica (dos enunciados  $Mo=5$  y otros dos  $Mo=4$ ), y c) sostenibilidad o sustentabilidad (dos enunciados  $Mo=5$  y otros dos  $Mo=4$ ). Tambi n est n muy a favor, aunque en menor grado que las anteriores por presentar moda uno de sus enunciados y moda 4 los otros tres, la corriente naturalista y la bio-regionalista.

### Conclusi n

El Cuestionario para la Identificaci n de Perfiles en Educaci n Ambiental presenta unas caracter sticas psicom tricas aceptables.

El Cuestionario para la Identificaci n de Perfiles en Educaci n Ambiental (CIPEA) y la Escala del Nuevo Paradigma Ecol gico (NEP) mantienen relaciones significativas, sin embargo la relaci n debe estudiarse m s. Se intuye que existen estructuras esp reas en el CIPEA por su comportamiento en el ACP.

Se han identificado quince perfiles en la forma de concebir la idea de “Educaci n Ambiental”, que siguiendo las caracter sticas que definen las corrientes propuestas por Sauv , podr an llamarse as : Perfil 1 “Ecol gico”, Perfil 2 “Naturalista pr ctico con eco-educaci n”, Perfil 3 “Moral conservacionista y de sostenibilidad”, Factor o Perfil 4 “Cient fico bioregionalista cr tico-pr ctico”, Perfil 5 “Igualdad de g nero”, Perfil 6: “Biorregionalista moral”, Perfil 7 “Hol stico cr tico”, Perfil 8 “Sist mico resolutivo”, Perfil 9 “Sist mico naturalista-sostenible”, Perfil 10 “Hol stico”, Perfil 11 “Resolutivo sist mico”, Perfil 12 “Humanista sist mico”, Perfil 13 “Etnogr fico naturalista-resolutivo”, Perfil 14 “Sostenible” y Perfil 15 “Cient fico humanistico”.

La muestra presenta bastante acuerdo hacia las iniciativas llevadas a cabo en Educaci n Ambiental, pudi ndose decir que la muestra tiende a estar “de acuerdo” con la mayor a de los enunciados (la mayor a de los enunciados tienen  $Mo=4$ ).

Existen una serie de enunciados con los cuales, la muestra se encuentra totalmente de acuerdo, por lo que se podr a decir que dichos enunciados son las iniciativas que con mayor tendencia se pueden dar en Educaci n Ambiental en la muestra objeto de estudio, como puede ser: a) el llevar a cabo programas de gesti n sostenible para el desarrollo econ mico en una regi n, b) aprender a utilizar los recursos con el fin de que haya suficientes para todos, asegurando las necesidades de hoy y las del ma ana, o c) desarrollar habilidades de conservaci n sobre los recursos ambientales.

## Referencias

- Dunlap, R.E., Van Liere, K.D., Mertig, A.G. y Jones, R.E. (2000). Measuring endorsement of the new ecological paradigm: A revised NEP Scale. *Journal of Social Issues*, 56 (3), 425-442.
- Sacristán, J. G. y Gómez, Á. I. P. (1999). *Comprender y transformar la enseñanza*. Madrid: Morata.
- Sureda, J., Cañellas, A. J. C. y Colom, A. J. (1989). *Pedagogía ambiental*. Madrid: CEAC.
- Ministerio de Medio Ambiente (1999). *Libro Blanco de la Educación Ambiental en España en pocas palabras*. Madrid.
- Sauvé, L. (2004). Perspectivas curriculares para la formación de formadores en Educación Ambiental. *I Foro Nacional sobre la Incorporación de la Perspectiva Ambiental en la Formación Técnica y Profesional*. México.
- Sauvé, L. (2005). Uma cartografia das correntes em educação ambiental. En Sato, Sr. y Carvalho, I. (Dir.). *Educação ambiental - Pesquisa e desafios*. Oporto Animado: Artmed, 17-46.
- UNESCO-PNUMA (1977). *Conferencia intergubernamental sobre Educación Ambiental en Tbilisi, Georgia*. Madrid, España: UNESCO.

# CONTENIDOS METODOLÓGICOS EN EL GRADO EN PSICOLOGÍA EN LA COMUNIDAD AUTÓNOMA DE ANDALUCÍA. DESCRIPCIÓN DEL CASO DE LA UNIVERSIDAD DE ALMERÍA

Juan García García y Leticia de la Fuente  
Universidad de Almería

En este trabajo se presenta el proceso seguido por la Comunidad Autónoma Andaluza en la elaboración de los nuevos títulos universitarios de Grado a raíz del proceso de construcción del Espacio Europeo de Educación Superior. En el marco de dicho proceso, el Consejo Andaluz de Universidades ha homogeneizado el reconocimiento automático de módulos y materias con un porcentaje del 75% del título entre las universidades andaluzas que lo imparten, incluyendo las prácticas externas y el trabajo fin de grado. En este trabajo, se explicita el proceso seguido en el caso del Grado en Psicología, presentándose un resumen de la estructura de las fichas elaboradas por la comisión andaluza para la elaboración del Grado en Psicología, ejemplificadas a través del Módulo de Métodos, Diseños y Técnicas de investigación en Psicología y estableciendo la comparativa entre universidades. Se expone también la concreción del proceso realizado en la Universidad de Almería y se presentan la estructura de las materias y asignaturas que corresponden al entorno del citado módulo.

La titulación de Grado en Psicología atiende a la formación universitaria inicial de un ámbito de conocimiento científico y académico y de un campo de ejercicio profesional consolidado y ampliamente extendido, tanto en el Estado Español como en otros países de referencia, europeos y americanos.

En España, los estudios de Psicología comienzan a aparecer en las universidades, en asignaturas o cátedras, a finales del siglo XIX, y adquieren su consolidación, estructurados en titulación universitaria a finales de los años sesenta del siglo XX. En 1973, se fijan las directrices para los planes de estudio de Psicología y el desarrollo normativo y reglamentario de La Ley de Reforma Universitaria (LRU), recogió y asentó esta consolidación en el ámbito universitario, con el establecimiento de seis áreas de conocimiento directamente vinculadas a esta disciplina, a saber, Metodología de las Ciencias del Comportamiento; Personalidad, Evaluación y Tratamientos Psicológicos; Psicobiología; Psicología Básica; Psicología Evolutiva y de la Educación y Psicología Social. Esta estructura se ha mantenido en los sucesivos planes estudio hasta la actualidad, que en cierta forma se pone en cuestión con el inicio de la reforma derivada de la creación del Espacio Europeo de Educación Superior (EEES).

## *El Espacio Andaluz de Educación Superior*

En la Comunidad de Andalucía, y como consecuencia de la adaptación al EEES, el Consejo Andaluz de Universidades, aprobó en 2008 un documento sobre “Líneas Generales, Protocolos y Metodología de trabajo para la solicitud de autorización de Titulaciones Oficiales en el Sistema Universitario Andaluz, en el que, entre otras cuestiones, se reglamentan las Comisiones de Rama de Conocimiento y las Comisiones del Título. Tras un posicionamiento inicial del título de Psicología a la rama de Ciencias Sociales, finalmente se adscribió a CC. de la Salud. Entre las consideraciones a establecer en los nuevos títulos emanados de las universidades andaluzas, se establecía el reconocer los módulos y en su caso las materias y de todos los títulos de Psicología en un 75% de las enseñanzas comunes en el que se incluye incluyen las Prácticas Externas y los seis créditos que, como máximo, se podrá reconocer a los

estudiantes por la participación en actividades universitarias culturales, deportivas, de representación estudiantil, solidarias y de cooperación, y que generan una oferta de optatividad para aquellos alumnos que no realicen dichas actividades.

Para el caso de los contenidos metodológicos se establece la siguiente referencia (ver cuadro 1):

*Cuadro 1. Módulo y competencias de los contenidos metodológicos del 75% común en Andalucía*

Denominación del módulo	18 créditos ECTS:
MÉTODOS, DISEÑOS Y TÉCNICAS DE INVESTIGACIÓN EN PSICOLOGÍA	12 créditos de Rama (Ciencias de la Salud) + 6 créditos obligatorios
Duración y ubicación temporal dentro del plan de estudios	
COMPETENCIAS Y RESULTADOS DEL APRENDIZAJE QUE EL ESTUDIANTE ADQUIERE CON DICHO MÓDULO	
Conocer distintos diseños de investigación, los procedimientos de formulación y contrastación de hipótesis y la interpretación de los resultados.	
Saber seleccionar y administrar los instrumentos, productos y servicios y ser capaz de identificar a las personas y grupos interesados	
Saber planificar la evaluación de los programas y las intervenciones	
Ser capaz de medir y obtener datos relevantes para la evaluación de las intervenciones	
Saber analizar e interpretar los resultados de la evaluación	
Ser capaz de describir y medir variables y procesos psicológicos.	
MATERIAS	
ESTADÍSTICA	6 créditos ECTS de materia obligatoria a determinar por cada Universidad
Materia básica de Rama	
12 créditos ECTS	
Rama: Ciencias de la Salud	
BREVE RESUMEN DE CONTENIDOS	
Métodos cuantitativos y cualitativos. Diseños de investigación. Técnicas de análisis de datos.	
Psicometría.	

La propuesta completa de la Comisión Andaluza del Título puede consultarse en: <http://www.uco.es/organizacion/eees/documentos/nuevastitulaciones/comisiones/titulacion/salud/PSICOLOGIA-definitiva.pdf>

La traslación que cada universidad andaluza hizo de este módulo a sus planes de estudios puede verse en la tabla 1. Algunas interpretaciones particulares a destacar pueden ser que la Universidad de Sevilla situó los contenidos de Fundamentos Metodológicos en un módulo de Fundamentos de la Psicología, y algunas universidades optaron por introducir la mayoría de los contenidos de Psicometría en materias/asignaturas fuera de esta troncalidad andaluza.

*Tabla 1. Materias/asignaturas con contenidos metodológicos derivados del 75% común en las universidades andaluzas*

	Métodos, Diseños y Técnicas de Investigación en Psicología		
<b>UAL</b>	- Fundamentos Metodológicos en Psicología	- Análisis de Datos en Psicología	- Psicometría
<b>UGR</b>	- Descripción en Psicología	- Técnicas de Análisis de Datos en Psicología - Métodos y diseños de investigación psicológica	
<b>UHU</b>	- Análisis de Datos en Psicología I	- Análisis de Datos en Psicología II - Fundamentos Metodológicos en Psicología	
<b>UJAEN</b>	-Análisis de datos I	-Análisis de datos II. -Métodos y diseños de investigación psicológica	
<b>UMA</b>	-Metodología de la investigación y Estadística I	-Metodología de la investigación y Estadística II	- Psicometría
<b>US</b>	- Diseño y Análisis de Datos en Psicología I	- Diseño y análisis de Datos en Psicología II	- Psicometría

#### *Plan de estudios de Psicología de la Universidad de Almería*

El plan de estudios del Grado en Psicología por la Universidad de Almería tiene un total de 240 créditos, distribuidos en 4 cursos (8 cuatrimestres) de 60 créditos cada uno (30 ECTS por cuatrimestre). El plan de estudios se estructura en materias, integradas en módulos de mayor amplitud siguiendo un criterio de contenido. Todas las materias son cuatrimestrales, incluidas las Prácticas Externas y el Trabajo fin de Grado. A excepción de las Prácticas Externas (18 créditos ECTS), todas las materias tienen una carga de 6 ó 12 créditos ECTS.

Los 60 créditos ECTS correspondientes a materias de formación básica conforman 9 materias de rama (1 materia de 12 créditos y 8 materias de 6 créditos cada una), y se sitúan en los dos primeros años de los estudios de Grado. Estas materias proceden de las materias de rama que se contemplan en el Anexo II del RD 1393/2007 y se vinculan a Psicología (12 créditos), Biología (6 créditos), Fisiología (6 créditos) y Estadística (12 créditos) de la Rama de Ciencias de la Salud (un total de 36 créditos); y Psicología (12 créditos), Sociología (6 créditos) y Educación (6 créditos) de la Rama de Ciencias Sociales y Jurídicas (un total de 24 créditos). Esta distribución fue acordada por la Comisión Andaluza de la Titulación ya que estos 60 créditos forman parte de los 180 ECTS (75% de acuerdo) que, según la normativa citada más arriba, deben ser comunes a todas las Facultades de Psicología de Andalucía.

Las materias obligatorias restantes se subdividen en 18 materias que proceden de los módulos comunes a las Facultades de Psicología Andaluzas y el porcentaje restante son materias obligatorias propias de la Facultad de Psicología de la Universidad de Almería. De esta forma se configura un Grado compuesto, casi exclusivamente por materias obligatorias, por decisión expresa del Rectorado de la Universidad de Almería.

Centrándonos en los contenidos metodológicos del plan de estudios de la Universidad de Almería, estos se concentran en un módulo de 30 créditos denominado Métodos, Diseños y Técnicas de Investigación en Psicología. Este módulo proporciona formación básica de carácter metodológico relacionada con la Psicología, como ciencia y como profesión, y en las técnicas y procedimientos de análisis de datos e interpretación de resultados obtenidos a partir de la investigación. Así como, los

fundamentos de la medida del comportamiento y sus aplicaciones. Es un módulo clave para proporcionar la formación metodológica básica para que el estudiante pueda interpretar y valorar adecuadamente la información metodológica a la que tengan acceso durante el resto del Grado y en su inicio profesional, así como durante su proceso de aprendizaje a lo largo de toda la vida.

Los contenidos de este módulo se corresponden con las materias/asignaturas:

Materia 1. Fundamentos de metodología y análisis de datos en Psicología.

Asignaturas:

*Fundamentos metodológicos de la Psicología:* Cuatrimestre 1º

*Fundamentos del análisis de datos en Psicología:* Cuatrimestre 2º

Materia 2. Medición en Psicología.

Asignaturas:

*Psicometría:* Cuatrimestre 4º

Materia 3. Métodos de investigación Aplicados.

Asignaturas:

*Diseños de investigación y análisis de datos:* Cuatrimestre 5º

*Metodología de Encuestas:* Cuatrimestre 7º

Esta estructura supone un peso similar al que estos contenidos tenían en la troncalidad actual de la Licenciatura (13,7%), y junto con su carácter obligatorio aseguran una formación base homogénea al futuro graduado/a en Psicología por la Universidad de Almería.

# LA INNOVACIÓN EDUCATIVA DESDE LA METODOLOGÍA COMO DISCIPLINA TRANSVERSAL: UN SISTEMA PARA CONTRASTAR MODELOS TEÓRICOS EN LAS CLASES DE PRÁCTICAS

José Fernando García Pérez, Enrique Gracia, M. Soledad Lila y M. del Castillo Fuentes  
Universidad de Valencia

Los programas de innovación educativa requieren emplear sistemas simples, pero sistemáticos, para contrastar los supuestos teóricos de los modelos psicológicos con los datos reales de la investigación empírica. Siguiendo la idea de que la psicología es una ciencia con enunciados ordinales (Frick, 1996) este trabajo plantea un sistema con el objetivo de: 1) relacionar los componentes explicativos de un sistema teórico con dos ejes ortogonales; y 2) contrastar el orden que seguirán en diferentes criterios de ajuste psicológico las cuatro tipologías básicas que se derivan de este planteamiento teórico con dos ejes. Sobre el principio de igualdad o diferencia se ordenan los componentes de los ejes ortogonales en ocho regiones: sobre un eje o en medio de dos. Con el mismo principio de igualdad o diferencia se ordena el ajuste de las cuatro tipologías: igual, mayor o menor. Se evalúa el rendimiento académico y los cambios de actitudes hacia la metodología. Los resultados indican que: 1) la realización de esta experiencia de investigación contrastada con datos reales mejora la formación del alumno, y 2) supone una toma de contacto con la realidad metodológica que favorece sus actitudes hacia la misma.

En múltiples ocasiones se ha criticado que la psicología sea una ciencia que se caracteriza por una desorganización teórica importante (e.g., García, Frías y Pascual, 1999). Una de las frases que en un tono humorístico mejor representa esta situación es que cada psicólogo de la Tierra tiene su propia teoría, y que éstas llegan a ser de un uso tan personal como los cepillos dentales, de tal manera que cuando uno mismo desaparece o pierde su interés por la teoría, la teoría también muere (Watkins, 1981, 1984). Los metodólogos de las ciencias del comportamiento han trabajado incesantemente para aportar soluciones metodológicas que contribuyan a la acumulación consistente de conocimientos.

Por una parte, siguiendo la línea de Cohen (1962, 1994), quien advertía insistentemente de que las conclusiones de un estudio no dependían solamente de que fuera real la relación entre la variable independiente y la dependiente, sino también del tamaño de la muestra, varios autores han coincidido en destacar la importancia del mejor control de la validez estadística (véase García, Pascual, Frías, Van Krunkelsven y Murgui, 2008). Muchos trabajos han culpado directamente de estos problemas a la estadística y su uso psicológico que se hace de la hipótesis nula (e.g., Bakan, 1966; Cohen, 1994; Nickerson, 2000), proponiendo que se introdujese sistemáticamente en la psicología la necesidad de los cálculos de los intervalos de confianza y las medidas del tamaño del efecto (e.g., Cumming y Finch, 2005; Nickerson, 2000).

Por otra parte, destaca la aportación de Fritz (1996), que desde una perspectiva opuesta (1º, modelo teórico; 2º, hipótesis; y 3º, validez estadística) defiende que los enunciados que se derivan de las teorías psicológicas son en muchas ocasiones ordinales. Es decir, enuncian algún tipo de relación ordinal entre variables como, por ejemplo, que a mayor motivación se producirá un mayor rendimiento (el grupo con mayor motivación tendrá mayor rendimiento, etc.). No es muy frecuente, sin embargo, que se relacionen cuantitativamente las variables (mediante cantidades concretas y una fórmula matemática). Siguiendo el ejemplo anterior, qué cantidad de incremento en el grado de motivación produciría qué cantidad de incremento del rendimiento académico. Desde la perspectiva de Fritz (1996) la solución es analizar si los supuestos teóricos que plantean los modelos teóricos son de naturaleza

ordinal o cuantitativa, para definir sin ambigüedades los sistemas de comprobación de las hipótesis, teniendo en consideración los supuestos críticos que la hipótesis teórica plantea (Kukla, 1989).

El *presente estudio* es un proyecto de innovación educativa que tiene como objetivo la mejora de la formación teórica y metodológica de los alumnos (promocionando una actitud positiva hacia la metodología). Para ello se propone un sistema que permite analizar las relaciones ordinales que establece un modelo teórico clásico de la socialización familiar, pero que actualmente aun es fuente de continuas polémicas en la literatura especializada (véase García y Gracia, 2009a, en prensa). El modelo se compone de dos ejes ortogonales, aspectos teóricamente independientes y directrices de la socialización familiar, que organizan la relación que existe entre los tipos de prácticas socializadoras de los padres y permiten definir para cada padre o familia un estilo de socialización familiar. A partir de la lectura de una selección de la literatura especializada que relaciona teóricamente varias prácticas con los dos ejes se determina que la precisión que determinan los autores de los trabajos se reduce, como mucho, a determinar ocho puntos en el cuadrante de los dos ejes ortogonales sobre la base de la relación que tiene que existir entre cada eje, cada medida de cada práctica y la coherencia implícita entre las relaciones de todos los componentes del modelo. Con este análisis, se definen ocho posiciones para las prácticas entre los dos ejes: cada práctica sobre un eje o en medio de dos (Figura 1).

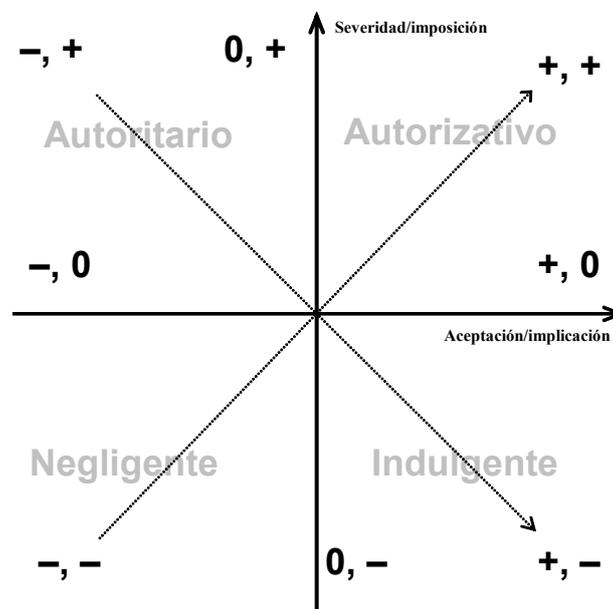


Figura 1. Representación de los ocho patrones de relación entre los ejes y las prácticas

Teóricamente se predicen diferencias entre los cuatro estilos de socialización y los diferentes criterios de ajuste en los hijos. En las predicciones acerca del orden que seguirán las cuatro tipologías en los criterios de ajuste y desajuste la predicción varía en la posición relativa entre las cuatro tipologías, si bien son posibles hasta cuatro posiciones diferentes de orden en muchos casos no alcanza a más de dos, o a lo sumo tres niveles, quedando varios de los cuatro estilos indistinguibles dentro de un mismo grado (Figura 2). Con el mismo principio de igualdad o diferencia se ordena el ajuste de las cuatro tipologías: igual, mayor o menor. También se midió en este primer apartado de hipótesis si la relación entre las prácticas familiares y los grados de ajuste de los hijos interactuaban con el contexto social (favorable o desfavorable) en el que se producía la socialización mediante los patrones de medias.

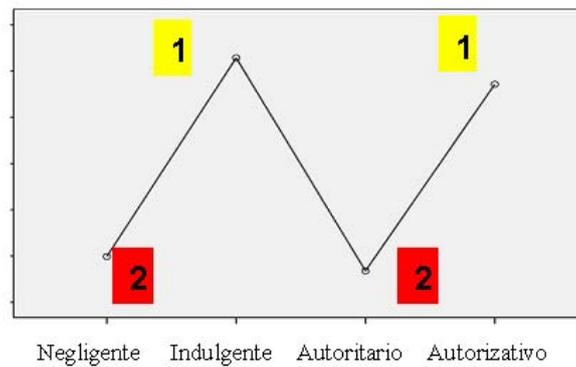


Figura 2. Ejemplo de un patrón de medias de dos niveles: el 2-1-2-1

Además, como han indicado varios estudios, se analiza el efecto del tamaño muestral en la consistente replicación inter-muestral de los resultados de los tres conjuntos de hipótesis planteadas realizando primero los análisis de datos con muestras muy pequeñas ( $N = 30$ ). Los detalles sobre los contenidos específicos del programa están publicados (García y Gracia, 2009b; García, Gracia y Lila, 2009ab).

El objetivo de este programa de innovación educativa es que aplicando los dos sistemas metodológicos recomendados (utilizar muestras de un tamaño suficiente y definiendo hipótesis precisas consonantes con los enunciados teóricos de un modelo) en el tratamiento transversal del conocimiento psicológico que propone el programa mejorarán la comprensión, la formación y las actitudes del alumno respecto de las bases metodológicas de la psicología.

## Método

### Participantes

La muestra fueron 84 alumnos que siguieron el programa, y que también cursaron la asignatura de Métodos y Diseños de Investigación en la Facultad de Psicología de la Universidad de Valencia.

Las medidas fueron el rendimiento académico, la nota conseguida en el examen de la asignatura de junio (entre 0 y 10), el grado de rendimiento en las actividades del programa, y la precisión conseguida durante los pronósticos (medida a partir de la proporción de aciertos sobre el total de preguntas, entre 0 y 1) de las tres hipótesis contrastadas en las fases 1ª (después de examinar el planteamiento teórico) y 3ª del programa (después de analizar los resultados de una muestra de 30 sujetos, que fue denominada fase 2ª). Por último, se midieron las actitudes hacia los cuatro componentes estadísticos de la metodología (García y Gracia, 2009b; Márquez, Cervantes y Gutiérrez, 2003) evaluados con el *Survey of Attitudes Toward Statistics* (SATS; Schau, 2003) en septiembre (*Pre*) y junio (*Post*). La consistencia interna en Afecto fue de 0,70 en el *Pre* y de 0,78 en el *Post*; en Competencia Cognitiva, 0,80 y 0,74; en Valoración, 0,79 y 0,75; y en Facilidad, 0,54 y 0,51.

## Resultados

Primero, se analizó la precisión en las predicciones de las tres hipótesis durante las fases primera y tercera (Tabla 1). Constatando que se producía un mayor grado de acierto en la primera hipótesis (ubicación de las prácticas en los dos ejes ortogonales) que en las dos últimas (predecir el orden del patrón de medias). Sin embargo los patrones de respuesta no varían en la fase 3ª respecto de la fase 1ª (reconocen que la potencia estadística del estudio de la práctica era muy baja con 30 observaciones).

Tabla 1. Proporción de aciertos entre las fases 1ª y 3ª en las 3 hipótesis formuladas

Hipótesis	Fase	N	Mínimo	Máximo	M	DT	r	t	gl
1ª	1ª	34	0,12	0,90	0,61	0,19	0,57**	-0,54	28
	3ª	29	0,14	0,88	0,63	0,20			
2ª	1ª	52	0,12	0,77	0,34	0,12	0,44**	1,91	44
	3ª	45	0,15	0,54	0,31	0,10			
3ª	1ª	33	0,08	0,50	0,30	0,11	0,43**	-0,96	30
	3ª	31	0,08	0,52	0,31	0,10			

\*  $p < 0,05$ ; \*\*  $p < 0,01$ .

Segundo, se analizó la relación entre el rendimiento en el programa de innovación y la calificación en el examen de Junio. Cuando el rendimiento en el programa fue alto ( $n = 16$ ,  $M = 7,17$ ,  $DT = 1,03$  vs.  $n = 29$ ,  $M = 5,96$ ,  $DT = 1,44$ ) la nota del examen fue mayor,  $F(1, 43) = 8,73$ ,  $p < 0,01$ . Además, la distribución cruzada de la calificación en el acta (no presentados, NP, y suspendidos vs. aprobados) con el rendimiento en el programa de innovación (bajo vs. alto) indicaba que ninguno de los alumnos con alto rendimiento en el programa suspendió la asignatura (Tabla 2).

Tabla 2. Distribución del rendimiento en el acta final de Junio por el rendimiento en el programa de innovación educativa

		Rendimiento en el programa		
		Bajo	Alto	Total
Nota final de junio	NP ó suspenso	42	0	42
	Aprobado ó más	26	16	42
	Total	68	16	84

Tercero, se analizó el cambio de actitudes hacia los componentes estadísticos de la metodología con cuatro diseños factoriales  $2 \times 2$  (*pre vs. post*, en cada componente del SATS; rendimiento alto vs. bajo, en las actividades de programa de innovación) cuya interacción se representa en la Figura 3. Los efectos principales del tiempo siempre fueron significativos ( $\alpha = 0,05$ ) indicando constantemente una mejora positiva asociada con el transcurso del programa, incrementándose las medias tanto de los de alto rendimiento en el programa como en los de bajo.

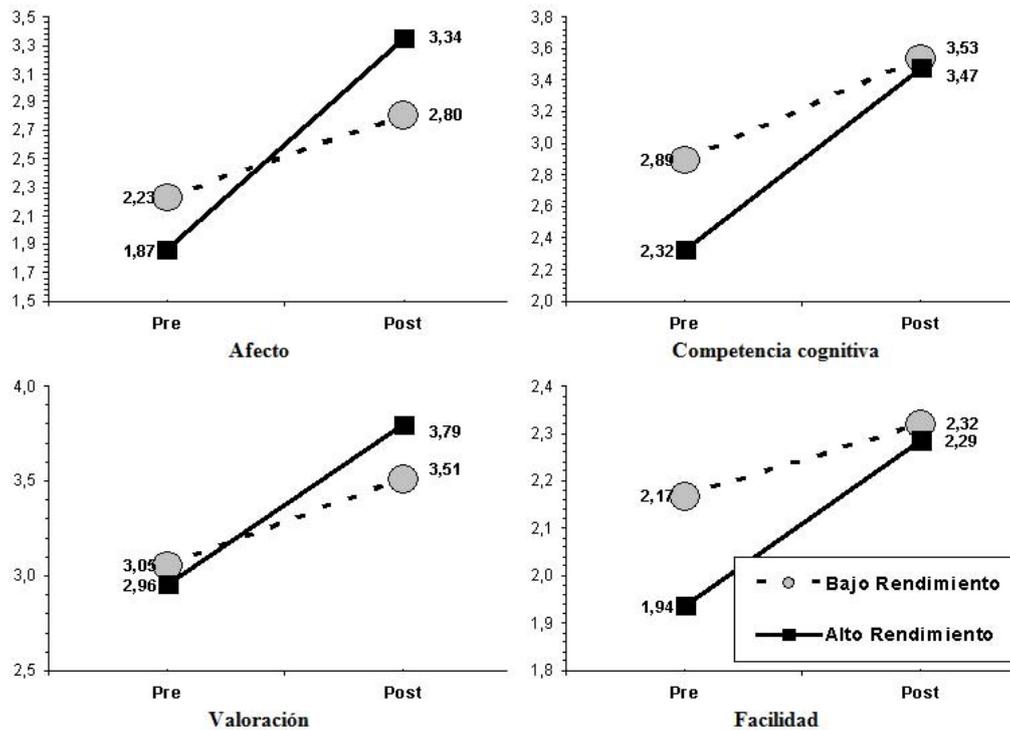


Figura 3. Rendimiento en el programa por cambio de actitudes de las cuatro escalas del SATS

La tendencia de la interacción de Afecto,  $\Lambda = 0,842$ ,  $F(1, 82) = 15,34$ ,  $p < 0,001$ , indicaba que el grupo con alto rendimiento, aunque partió de posiciones más bajas en el *pre*, superó en la nota del *post* a los de bajo rendimiento. La actitud de Competencia Cognitiva inferior del grupo con alto rendimiento,  $\Lambda = 0,920$ ,  $F(1, 82) = 7,13$ ,  $p < 0,01$ , se compensó al final del programa. La Valoración se cruzaba con el rendimiento,  $\Lambda = 0,949$ ,  $F(1, 82) = 4,42$ ,  $p < 0,05$ , indicando que las posiciones iniciales de relativa desventaja del grupo con mayor rendimiento en el programa se transformaba en una mayor ventaja relativa en el *post*. Finalmente, en la Facilidad no se constataron efectos de interacción,  $\Lambda = 0,955$ ,  $F(1, 82) = 3,83$ ,  $p > 0,05$ , únicamente los efectos principales asociados con el tiempo,  $\Lambda = 0,769$ ,  $F(1, 82) = 24,60$ ,  $p < 0,001$ , que indicaban que independientemente del grado de rendimiento en el programa, la tendencia era que el final les resultaba más sencillo que el principio.

## Discusión

El programa de innovación educativa analizado en este trabajo se ha relacionado positivamente con la formación de los alumnos. En el contexto real de este programa, que va más allá de las simulaciones controladas de los clásicos ejercicios, los alumnos no variaron las predicciones de sus hipótesis con muestras muy pequeñas (Cohen, 1962, 1994; García et al., 2008). Formularon sus hipótesis de manera que fueran contrastables sin ambigüedad (Fritz, 1996; Kukla, 1989), partiendo de los modelos teóricos de la literatura psicológica, no precisamente convergentes, coherentes y claros (Watkins, 1981, 1984; García y Gracia, 2009a). En este contexto hay que destacar que el rendimiento de los alumnos con mejores indicios de calidad en el programa fue claramente superior en los criterios de formación académica. También se han constatado avances satisfactorios en todos los alumnos en criterios mediadores muy importantes como son las actitudes hacia las bases estadísticas de estas asignaturas (García y Gracia, 2009b). Todos estos resultados favorables están limitados por la naturaleza cuasiexperimental de la metodología, también puede contribuir el mero desarrollo de las clases. Pese a esta importante limitación hay que valorar, como advierte Fritz (1996) al considerar la dirección en las relaciones ordinales, que no se ha constatado ningún indicador de que la participación en este programa haya supuesto algún retroceso en la formación. Por lo que la experiencia que supone un contacto real y

directo con la investigación psicológica y su metodología, de la manera tan radicalmente directa que plantea este programa, parece bastante fructífera para la formación del estudiante.

Agradecimientos. Este trabajo forma parte del proyecto de investigación “Las prácticas con datos reales: Contrastando en las clases de prácticas los supuestos de los modelos teóricos” financiado por el Vicerrectorado de Convergencia Europea y Calidad de la Universidad de Valencia, código: 72/DT/26.

### Referencias

- Cohen, J. (1962). Statistical power of abnormal-social psychological-research: A review. *Journal of Abnormal Psychology, 65*, 145-153.
- Cohen, J. (1994). The earth is round (p-less-than .05). *American Psychologist, 49*(12), 997-1003.
- Cumming, G. & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170-180.
- Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods, 1*, 379-390.
- García, F. & Gracia, E. (2009a). Is always authoritative the optimum parenting style? Evidence from Spanish families. *Adolescence, 44*(173), 101-131.
- García, F. & Gracia, E. (2009b). Attitudes toward research methodology among psychology students: Assessment of an educative intervention programme. In L. G. Chova, D. M. Belenguer & I. C. Torres (Eds.), *INTED2009 Proceedings CD* (pp. 4830-4836). Valencia, Spain: IATED.
- García, F. & Gracia, E. (en prensa). What is the optimum parenting style in Spain? A study with children and adolescents aged 10-14. *Infancia y Aprendizaje*.
- García, F. Gracia, E. & Lila, M. (2009a). Comprobando en las clases prácticas un modelo teórico. En F. F. Cerdá, A. F. March & J. M. M. March (Eds.), *Actas del V Congreso Iberoamericano de Docencia Universitaria* (pp. 2351-2356). Valencia, Spain: Universidad Politécnica de Valencia.
- García, F., Gracia, E. & Lila, M. (2009b). Integration of theory and practice among different areas of psychology: Testing in classroom practices a theoretical model of parenting. In L. G. Chova, D. M. Belenguer & I. C. Torres (Eds.), *INTED2009 Proceedings CD* (pp. 3672-3681). Valencia, Spain: IATED.
- García, J. F., Frías, M. D., & Pascual, J. (1999). *Los diseños de la investigación experimental: Comprobación de las hipótesis*. Valencia, Spain: Cristóbal Serrano Villalba.
- García, J. F., Pascual, J., Frías, M. D., Van Krunckelsven, D. & Murgui, S. (2008). Design and power analysis: n and confidence intervals of means. *Psicothema, 20*, 933-938.
- Kukla, A. (1983). Non empirical issues in psychology. *American Psychologist, 44*, 785-734.
- Márquez, R. J. M., Cervantes, R. J. M. y Rodríguez, R. M. (2003, septiembre). *Actitudes y rendimiento en asignaturas estadísticas*. Comunicación presentada en el Congreso de Metodología de las Ciencias Sociales y de la Salud. Valencia, Spain.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.
- Watkins, M. J. (1981). Human memory and the information processing metaphor. *Cognition, 10*, 331-336.
- Watkins, M. J. (1984). Models and toothbrushes. *Behavioral and Brain Sciences, 7*, 86.

# DIFICULTADES Y TEMORES DEL ALUMNADO UNIVERSITARIO EN SU SALIDA AL MUNDO LABORAL

Pilar García Soidán<sup>1</sup> y Herminia Dios<sup>2</sup>

<sup>1</sup>Universidad de Vigo

<sup>2</sup>Asociación de Desenvolvemento Rural Proder II. Comarca de Pontevedra

Durante los últimos años de sus estudios, surgen en el alumnado universitario importantes dudas y temores de cara a su salida al mundo laboral, que apuntan fundamentalmente a la falta de formación y recursos para afrontar las situaciones con las que se van a encontrar. Por ello, este trabajo se ha centrado en conocer las cuestiones que se plantean los estudiantes en relación a su futuro ejercicio profesional, analizar su grado de información sobre los servicios de empleo, los recursos y medios disponibles, así como su nivel de valoración de los mismos. Para ello se combinó metodología cuantitativa y cualitativa, con objeto de elaborar un diagnóstico de la situación actual. En la primera parte del estudio, se realizó una encuesta sobre una muestra representativa de este colectivo, según cuotas de titulación proporcionales a su distribución en el Campus. Seguidamente se trató de complementar este estudio con un análisis cualitativo de la situación, mediante la observación participante de los hechos, a través de las aportaciones que realizó un grupo reducido de alumnos que accedieron a compartir su perspectiva con mayor detalle.

El compromiso adquirido por las distintas Universidades para su adaptación al Espacio Europeo de Educación Superior (EEES) ha requerido un proceso de estudio y adecuación importante para converger a un sistema de titulaciones comprensible y comparable con dos niveles principales: grado y postgrado. En España, el Documento Marco de Integración, elaborado por el Ministerio de Educación, Cultura e Deporte en Febrero de 2003, establecía que los estudios de grado deberían proporcionar una formación universitaria en la que se integrasen armónicamente las competencias con la formación integral de las personas y deberían facilitar a los futuros titulados su inserción en el mercado de trabajo. Por lo tanto, el diseño de los nuevos planes de estudio, así como de las actividades y formación complementaria que ofertasen las Universidades, deberían estar orientados a promover e impulsar las oportunidades de trabajo de los titulados.

Desde esta perspectiva, parecía importante conocer y recopilar temas de preocupación común a las personas que, al finalizar sus estudios, afrontarán por vez primera su incorporación al ámbito laboral. De hecho, en los últimos cursos de las diferentes titulaciones surgen, en el alumnado universitario, dudas y temores de cara a su salida al mundo laboral, que no sólo tienen que ver con la formación recibida en la carrera elegida. En ocasiones apuntan, más bien, a la falta de conocimiento, preparación y recursos para afrontar las situaciones con las que se van a encontrar que no siempre sabrán cómo resolver, dónde encontrar respuesta o el asesoramiento preciso.

Teniendo en cuenta lo anterior, este trabajo se ha centrado en conocer las cuestiones que se plantean los estudiantes en relación a su futuro ejercicio profesional, analizar su grado de información sobre los servicios de asesoramiento de empleo, los recursos y medios disponibles, así como su nivel de valoración de los mismos.

## Método

### *Diseño del estudio*

En la primera parte de esta investigación, se diseñó una encuesta en la que se entrevistaron a 262 estudiantes, cuyas características se resumen en la ficha técnica (Tabla 1).

*Tabla 1. Ficha técnica del estudio cuantitativo*

Ámbito:	Campus de Pontevedra.
Universo:	Alumnado de los 2 últimos cursos de las 11 titulaciones oficiales de la Universidad de Vigo en el Campus de Pontevedra en el curso 2008/2009.
Tamaño de la muestra:	Diseñada: 254 personas. Realizada: 262 personas.
Procedimiento de muestreo:	Muestreo aleatorio simple. El diseño se llevó a cabo según cuotas de titulación proporcionales a su distribución en el Campus.
Error de muestreo:	Para un nivel de confianza del 90% y $P=Q$ , el grado máximo de error es del 5,7% para el conjunto do Campus de Pontevedra.
Método de entrevista:	Encuestas realizadas mediante entrevista personal.
Período de realización:	Febrero-Junio de 2008.

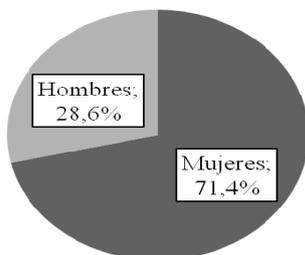
En la segunda parte del estudio se trató de complementar el análisis anterior, añadiendo la valoración y aportaciones que realizó un grupo más reducido de alumnos (119 en total), cuya distribución por titulaciones se resume en la tabla 2. Este grupo de estudiantes respondió a diversas cuestiones abiertas, diseñadas con objeto de complementar las ideas descritas anteriormente.

*Tabla 2. Distribución del alumnado participante en el estudio cualitativo*

Titulación	Nº alumnos/as
Lic. Bellas Artes	21
Dip. Fisioterapia	12
Lic. Comunicación Audiovisual	4
Lic. Publicidad y Relaciones Públicas	21
Dip. Gestión y Administración Pública	4
Ing. Técnica Forestal	8
Maestro/a Educación Musical	6
Maestro/a Educación Física	7
Maestro/a Educación Infantil	10
Maestro/a Educación Primaria	5
Lic. Ciencias Actividad Física y Deporte	21

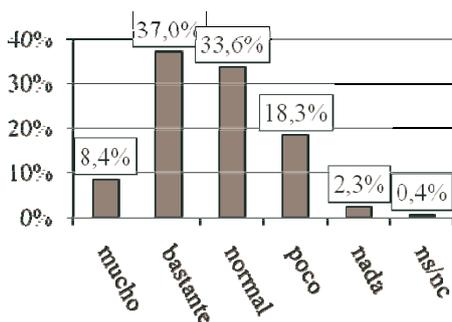
*Análisis cuantitativo del grado de preparación del alumnado*

Para realizar la primera parte del estudio, la selección de los encuestados fue realizada aleatoriamente, con cuotas por titulación, resultando la distribución por género en la muestra, reflejada en la figura 1 (71,4% mujeres y 28,6% hombres), que es próxima a la correspondiente a la población objeto de estudio (65% mujeres y 35% hombres).



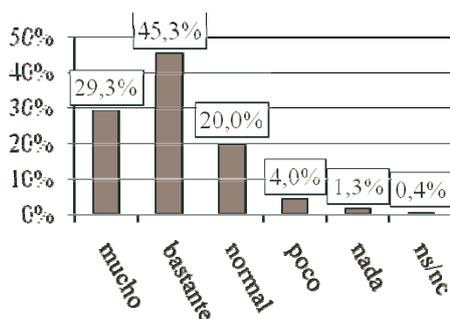
*Figura 1. Distribución del alumnado encuestado por género*

El grado de motivación de los estudios realizados podría revelarse como un condicionante importante en la valoración del alumnado y por ello se analizó este aspecto. Interesaba conocer la prioridad otorgada por el alumnado a la titulación cursada, obteniéndose que la gran mayoría de los entrevistados había tenido acceso a la carrera elegida en primer lugar (72,1%). Además, en general, los estudiantes manifiestan una opinión satisfactoria sobre la docencia recibida, de acuerdo con la información presentada en la figura 2.



*Figura 2. Grado de satisfacción del alumnado sobre la docencia recibida*

La realización de prácticas de empresa proporciona a los alumnos un primer acercamiento al mercado laboral. No obstante, no se trata de una actividad obligatoria, de ahí que tan sólo una pequeña proporción del alumnado encuestado las haya cursado (28,7%). La experiencia de las prácticas ha sido claramente satisfactoria para los alumnos, de acuerdo con la información del gráfico siguiente.



*Figura 3. Grado de satisfacción del alumnado sobre las prácticas realizadas*

Las prácticas de empresa proporcionan a los alumnos un referente importante del mercado de trabajo, las demandas que requiere y su grado de preparación. Es más, aunque su objetivo no es propiamente concluir con una oferta laboral, ha ocurrido así para los alumnos que las han cursado en casi un 20% de los casos.

Una vez analizado el perfil de la población observada, se procedió a determinar el grado de conocimiento y opinión del alumnado sobre los recursos y servicios de empleo disponibles. El hecho de considerar en este trabajo al alumnado matriculado en los dos últimos cursos de la titulación elegida supone que han finalizado ya un año de sus estudios y disponen de mayor información sobre el medio universitario. Aún así, menos de la tercera parte de las personas encuestadas (30,5%) conoce alguna oficina o servicio de empleo dependiente de la Universidad.

Desde la Universidad y a distintos niveles se organizan actividades de orientación al empleo dirigidas al alumnado de las diferentes titulaciones. Preguntados los alumnos por su conocimiento de este tema, resultó que menos de la quinta parte tenía noticias de la organización de tales actividades (18,4%). De entre las personas que sí sabían de la celebración de alguna de ellas, casi la mitad tenían como referencia las jornadas que organiza anualmente la Universidad bajo el título de Forumempleo (52,1%), como se refleja en la figura 4, si bien las charlas informativas que se desarrollan en los propios centros son para el 16,7% de estos estudiantes la única actividad conocida.

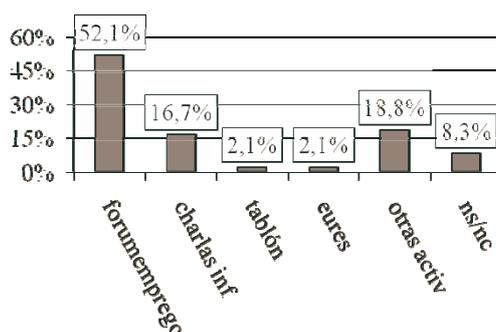


Figura 4. Actividades de empleo conocidas por el alumnado

Para el alumnado, mejorar su conocimiento del mercado laboral y de los recursos disponibles, supone un importante beneficio, además de la buena acogida que le dispensan a este tipo de actividades. En este sentido, casi la totalidad (95,8%) coinciden en que se debería ofertar una materia específica de libre elección sobre empleo que les preparase para su salida de la Universidad.

Por otra parte, se ha abordado el estudio de los recursos *on-line* que están a disposición del alumno sobre temas de empleo y a los que puede acceder a través de la página WEB de la Universidad. Como dato indicativo, la práctica totalidad de los encuestados conoce la dirección WEB de la Universidad (96,2%); sin embargo, el número de estudiantes que sabe que la página WEB de la Universidad ofrece información sobre empleo es más reducido (63,4%).

Profundizando en los procedimientos *on-line* a disposición de los futuros titulados, comenzamos por el que proporciona herramientas de apoyo para la elaboración del currículum, que es conocido por un porcentaje mínimo del alumnado (11,1%). No obstante, para los estudiantes que han accedido a este recurso, se trata de una herramienta útil y aplicable a la que otorgan una valoración positiva, calificándola de normal (40%) o buena (60%).

Mayor desconocimiento muestra el alumnado sobre la posibilidad de acceder a una bolsa de trabajo *on-line*, alcanzando tan sólo el 8,9% el porcentaje de estudiantes que conocen su existencia. En este sentido, deberían incrementarse los esfuerzos de cara a la promoción de los recursos, que han de ser accesibles y simples para ser de utilidad, y que son tan demandados por todos hoy en día puesto que suponen una economía de tiempo y dinero, a la vez que se adaptan a los horarios y disponibilidad de los usuarios.

### *Valoración y aportaciones del alumnado desde un enfoque cualitativo*

En la última parte del estudio se trató de complementar el análisis anterior, añadiendo la valoración y aportaciones que realizó un grupo más reducido de alumnos a diversas cuestiones abiertas que se les plantearon. Las diferentes opiniones e impresiones se resumen a continuación, donde se hace una descripción de tipo cualitativo de la información recogida.

La principal referencia que tienen los alumnos en los centros, de cara a su preparación para desarrollar una actividad laboral, son las prácticas de empresa. No todos los centros las ofertan, no son obligatorias y no hay suficientes plazas para atender a la demanda. Aún así, los alumnos que las han cursado tienen en general una impresión positiva sobre las mismas y consideran que les preparan para ser mejores profesionales, si bien deben concretarse los objetivos de las mismas y estar orientados a la preparación para desarrollar un trabajo. Se ha repetido bastante la petición a los equipos directivos y decanales de firmas de convenios con empresas para aumentar la oferta, ya que su desconocimiento del mundo laboral suele ser nulo por la falta de prácticas y esta experiencia les ayudaría a saber cómo desenvolverse, especialmente al principio.

Un comentario extendido entre los estudiantes es su desconocimiento de la existencia de oficina alguna de orientación al empleo en la Universidad. De ahí que consideren que debería ser objetivo prioritario de la institución académica el hecho de trabajar para dar a conocer los recursos existentes al alumnado ya que, a la vista de los resultados, los mecanismos de difusión que se utilizan son insuficientes.

Las nuevas tecnologías ya se aplican para este cometido y, así, algunos estudiantes confirman que reciben mensajes (por correo electrónico o SMS) sobre ofertas de prácticas, trabajo, etc. No obstante, debería hacerse un replanteamiento del sistema de difusión de la información, puesto que se ha llegado un punto en el que existe saturación por el volumen de información recibida y en ocasiones deja de prestársele la atención requerida. En cualquier caso, la página WEB debe proporcionar información útil y detallada sobre el tema del empleo, la existencia de oficinas y la organización de actividades, pero debe estar actualizada.

Por otra parte, aunque se organizan actividades de preparación a la búsqueda de empleo, apenas se conocen. De ahí que también se deba potenciar la difusión de estos recursos e incrementar el número de plazas y actividades, ya que no siempre son suficientes para la demanda existente.

A pesar del creciente interés actual por las actividades *on-line*, algunos de los entrevistados insisten en la necesidad de que se combine formación a distancia con actividades presenciales. Es cierto que los recursos *on-line* posibilitan el acceso a estas actividades al alumnado que trabaja o al que le resulta incompatible con su horario de clases y también que la mayoría de la gente dispone de acceso a internet, pero suponen una reducción de la atención e interés. En este sentido, la formación presencial podría aportar una motivación y cercanía para la preparación en la búsqueda de empleo, mientras que las actividades *on-line* podrían servir para mantener el contacto, para el desarrollo de la parte técnica (herramientas y recursos de apoyo) y para el intercambio de información.

## Conclusión

La posibilidad de cursar las prácticas de empresa supone para muchos de los alumnos universitarios su primer acercamiento al mercado laboral y, aunque no es el objetivo de las mismas, en algunos casos se traducen en contratos laborales. Para aquellos que han cursado las prácticas, se trata en general de una experiencia satisfactoria, aunque es muy reducido el porcentaje de alumnos que ha accedido a ellas. En este sentido, se demanda una ampliación de la oferta de las prácticas de empresa a los diferentes centros universitarios, lo cual redundaría en beneficio del alumnado.

Por otra parte, los equipos decanales/directivos y los propios docentes podrían tener una mayor implicación, en opinión de los estudiantes, en el proceso de formación para el empleo. Incluso casi todos los encuestados coinciden en el hecho de que se debería ofertar una materia específica de formación para afrontar la búsqueda de empleo.

Asimismo, los estudiantes consideran que se debería trabajar más la difusión de los servicios y las actividades de empleo que ofrece la Universidad, porque existe una gran desinformación al respecto. Sobre estas últimas, muestran un gran interés en participar y la valoración de aquellos que han cursado alguna de esas actividades es positiva, aunque se demanda mayor oferta y mayor frecuencia.

Se aprecia también un grado importante de desconocimiento de los recursos *on-line* orientados a su salida al mundo laboral. De hecho aunque la mayoría de los alumnos intuyen que la página WEB pone a su disposición material y herramientas ligados a la búsqueda de empleo, algo más de la tercera parte de los estudiantes no la ha consultado. En consecuencia, pocos de los futuros titulados conocen los recursos que se ofrecen a través de la página WEB para elaborar el currículum o para acceder a una bolsa de trabajo, entre otras opciones que se ofertan.

Agradecimientos. El presente trabajo ha sido desarrollado en el marco del Proyecto García 01/CFV 07 “Dificultades del alumnado universitario del Campus de Pontevedra en su salida al mundo laboral: detección de posibles carencias de formación extraacadémica o fallos en los canales de comunicación y propuestas de mejora para su implantación”, subvencionado por la Cátedra Filgueira Valverde, financiada por Caixanova y la Universidad de Vigo (Campus de Pontevedra) en la convocatoria 2007.

## Referencias

### *Direcciones WEB:*

- <http://europa.eu.int/eures/home.jsp?lang=es>
- <http://traballo.xunta.es/index.jsp?lang=gal&content=0003>
- <http://www.emancipacionjoven.es/empleajoven/index.php>
- <http://www.fundacionuvigo.es/>
- <http://www.inem.es/>
- <http://www.infojobs.net/>
- <http://www.injuve.mtas.es/injuve/portal.portal.action>
- <http://www.integrants.es/>
- <http://www.monster.es/>
- <http://www.oficinaempleo.com/>
- <http://www.primerempleo.com/>
- <http://www.universia.es/>
- [http://www.060.es/todo\\_sobre/buscar\\_trabajo/index-ides-idweb.jsp](http://www.060.es/todo_sobre/buscar_trabajo/index-ides-idweb.jsp)

# INNOVACIÓN DOCENTE EN “DISEÑOS EXPERIMENTALES Y APLICADOS”: ECTS, CAMPUS VIRTUAL Y EVALUACIÓN CONTINUA

Olatz López Fenández, Manel Viader, Antoni Coscolluela, M. Luisa Honrubia y Joan M. Malapeira  
Universidad de Barcelona

Se describe la progresiva adaptación de la asignatura troncal de “Diseños experimentales y aplicados” de la Facultad de Psicología de la Universidad de Barcelona en el marco de las directrices del Espacio Europeo de Educación Superior. Se abarcan las tres principales innovaciones docentes realizadas por el equipo docente entre 2005-2009: la adaptación del nuevo plan docente en base al ECTS, la implementación del campus virtual (*Moodle*) y la incorporación de una metodología de evaluación basada en un sistema de pruebas de validación, que incorporan el trabajo del estudiante individual y grupal a lo largo del semestre. Esta adaptación ha sido evaluada mediante la administración de cuestionarios a los estudiantes, con el propósito de iniciar un estudio empírico para conocer su opinión y percepción acerca de los cambios que se han ido introduciendo en esta asignatura. En concreto, se revisan los resultados del impacto inicial del campus virtual y del nuevo sistema de evaluación basado en pruebas de validación con un trabajo del estudiante organizado con una semi-presencial asimétrica fundamentada en las sesiones presenciales teórico-prácticas. Finalmente, entre las conclusiones halladas destaca una mejora global del rendimiento, un decremento del absentismo y una satisfacción general con el sistema de evaluación continua.

La asignatura troncal “Diseños experimentales y aplicados” de la Facultad de Psicología de la Universidad de Barcelona se ha ido adaptando progresivamente para encajar en el marco de las directrices del Espacio Europeo de Educación Superior (EEES) según la Universidad de Barcelona (UB) (UB, 2006). Entre 2005-2009 se han abarcado tres principales innovaciones docentes realizadas por el equipo de profesores: la adaptación del nuevo plan docente en base al ECTS, la implementación del campus virtual (*Moodle*) y la incorporación de una metodología de evaluación basada en un sistema de pruebas de validación, que incorporan el trabajo del estudiante individual y grupal a lo largo del semestre. A continuación se describe el desarrollo cronológico del trabajo de adaptación realizado.

## *Desarrollo de la innovación docente (2005-2009)*

- Curso académico 2005-2006

Se establece un sistema de prácticas individualizadas que consiste en la resolución de problemas mediante diferentes diseños de investigación, donde se trabajan los conceptos metodológicos y las técnicas estadísticas aplicadas a cada diseño con temáticas diversas enmarcadas en el ámbito Psicológico. Los estudiantes disponen de un entorno virtual de aprendizaje propio de la UB, conocido como *Dossiers electrònics*, que era altamente jerarquizado y sólo permitía publicar directorios con documentos, un foro y la opción de envío, a parte del calendario y avisos del profesorado (ver figura 1). En éste los estudiantes de forma individualizada envían sus problemas resueltos en fechas previamente indicadas para su evaluación continua de las prácticas. Estos problemas eran resueltos en las clases de prácticas, organizadas en función del avance de las clases teóricas. El trabajo propio del estudiante, que se organizaba en grupos que asistían a seminarios mensualmente, se organizaba a parte de esta evaluación teórico práctica, donde la vertiente de las prácticas estaba en evaluación continua, mientras que la teórica en evaluación sumativa con una prueba tipo test al final del semestre.

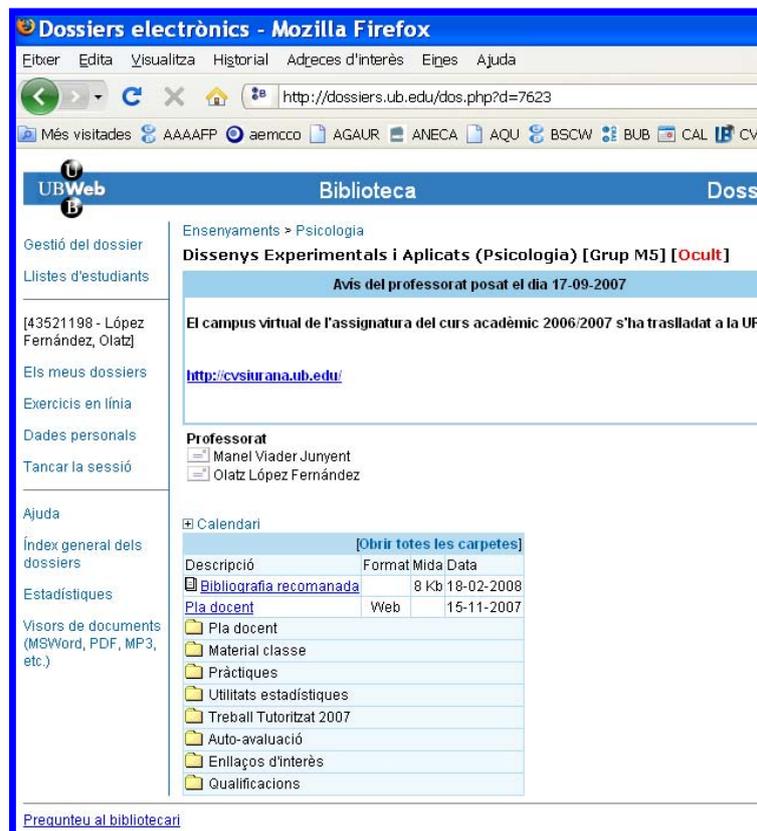


Figura 1. El Dossier electrònic de DEIA (2005-2006)

- Curso académico 2006-2007

Se inició el diseño del campus virtual con la plataforma *Moodle* (v.1.7) en base a la actividad del estudiante: clases de teoría, clases de prácticas y trabajo propio (ahora llamado “trabajo dirigido”), lo que hace precisar tanto en el plan docente como en la metodología los tres ejes de la asignatura y las modalidades en las que se impartirá (ver figura 3). La metodología de enseñanza, evaluación y aprendizaje es similar a la del curso anterior, pero se innova en emigrar del *dossier electrònic* al campus virtual *Moodle*, lo que supone una formación de todo el equipo docente, un trabajo cooperativo en el diseño de la estructura y funcionalidad del campus virtual y actualización de materiales (ver figura 3). Ello permitió iniciar una investigación piloto con un cuestionario *on-line* implementado en el propio campus virtual y estudiar el cambio que percibían los estudiantes respecto a los dos entornos utilizados, pues formamos parte de la prueba piloto de *Moodle* en la UB (Viader, López, Rifà, Cifré, Cosculluela y Malapeira, 2007; López Fernández, Viader Junyent, Cosculluela Mas, Malapeira Gas, Rifà Ros y Cifre León, 2008).

<p>Teoría: Incorporación de material y actividades de aprendizaje individual (documentación básica, complementaria, auto-avaluaciones) y grupal (foro) – PRESENCIAL Y APRENDIZAJE AUTÓNOMO</p> <p>Prácticas: evaluación continuada a través de tareas en línea, escala cuantitativa para calificar el trabajo de cada estudiante. PRESENCIAL Y APRENDIZAJE AUTÓNOMO</p> <p>Trabajo Dirigido: Incorporación de actividades dirigidas al soporte del aprendizaje grupal (wikis y tareas) e individual (cuestionarios) – NO PRESENCIAL Y TUTORIZADO</p>
--

Figura 2. Estructura de la enseñanza presencial y no presencial de DEIA (2006-2007)

ANTES de 2005-2006	AHORA: desde 2006-2009
<ul style="list-style-type: none"> <li>● “Dossiers electrònics” (aplicativo propio de la UB)</li> <li>● Sistema de carpetas con documentos digitales, un calendario, un foro, un espacio de envío de trabajos común (sin posibilidad de gestión) y posibilidad de crear cuestionarios</li> <li>● Uso del correo electrónico UB</li> <li>● Roles: <ul style="list-style-type: none"> <li>● Docente: publicación de parte del material, gestión calendario y foro</li> <li>● Discente: envío de problemas por el aplicativo o email</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>● Moodle: Campus Virtual UB (fase piloto: v. 1.7.)</li> <li>● Recursos y actividades dirigidas al soporte del aprendizaje y la evaluación continuada, así como de bloques de gestión (calendario, etc.)</li> <li>● Uso del correo electrónico UB</li> <li>● Roles: <ul style="list-style-type: none"> <li>● Docente: publicación de todo el material digitalizado, gestión y evaluación en el campus, se comunica también a través del campus</li> <li>● Discente: descarga materiales, envía problemas y recibe su evaluación a través del campus, realiza actividades complementarias</li> </ul> </li> </ul>

Figura 3. Comparación entre propiedades del Dossier electrònic y el Campus virtual Moodle de DEIA (2005-2009)

Se diseñó un campus virtual estructurado, con temas centrales numerados, donde el primero (no numerado) correspondía a los “Aspectos generales” y contenía: plan docente, bibliografía, foro de noticias, enlaces web didácticos (en la Facultad de Psicología y a la BUB-CRAI). Los cuatro siguientes (numerados) correspondían a los cuatro bloques de DEIA, de forma que todos tenían la misma distribución (ver figura 4).

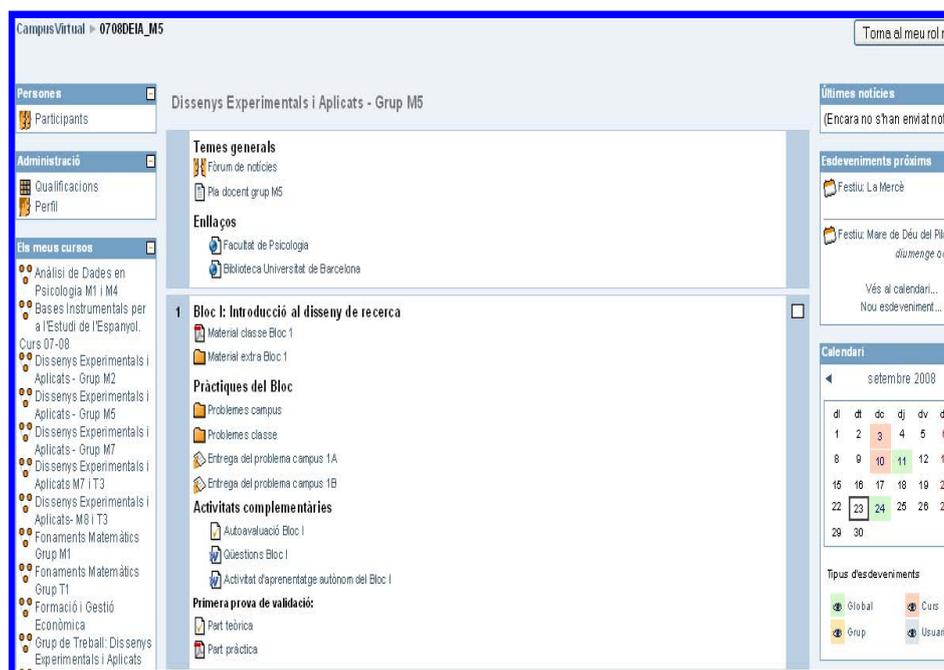


Figura 4. Pantalla principal del campus virtual DEIA M5 (2006-2007)

- Curso académico 2007-2008 y 2008-2009

Se diseñó un sistema de evaluación continua en base a pruebas de validación del conocimiento teórico-práctico de la asignatura. Para ello se desarrolló un nuevo sistema de actividades prácticas, donde todo el planteamiento se basaba en tres temáticas de investigación de nuestra disciplina abordables desde todos los diseños de investigación estudiados en la asignatura. Estas temáticas eran el hilo conductor de dos tipos de problemas a resolver:

*Prácticas presenciales de clase:* eran grupales y no evaluables directamente. Consistían en unas fichas que se facilitaban la semana antes el problema en resolver en la clase de prácticas, donde se trataba uno de los tres temas expuestos. Estaban en el directorio de “Problemas clase” de cada bloque y los estudiantes los llevaban impresos en papel y dedicaban la mitad de clase a trabajarlos en pequeño grupo con la ayuda puntual del docente de prácticas y después se les explicaban las soluciones con la participación de los estudiantes (ver figura 5).

	Nom	Mida	Modificat
	DEIA_M5_BII_ProbClasse_2.doc	135.5Kb	27 octubre 2009, 12:53
	DEIA_M5_BII_ProbClasse_2.odt	93.3Kb	27 octubre 2009, 12:53
	DEIA_M5_BII_ProbClasse_3.doc	131.5Kb	27 octubre 2009, 12:53
	DEIA_M5_BII_ProbClasse_3.odt	92.9Kb	27 octubre 2009, 12:53

Figura 5. Directorio “Problemas clase” del campus virtual DEIA M5 (2007-2008)

*Prácticas no presenciales autónomas:* eran individuales y evaluables directamente. Consistían en unas fichas que trataban sobre uno de los temas planteados e indicaba la puntuación de cada pregunta (por ejemplo: [Puntuación de 1,25]). Se realizaban individualmente y enviaban a través de la opción de “tarea cuelga un fichero” (por ejemplo: “Entrega del problema campus 2”). A la semana siguiente eran evaluados y su nota se publicaba a través del campus virtual, en el área de la tarea o a la opción de “Calificaciones”.

A parte se continuaba con el Trabajo dirigido, no presencial y autorizado que se adaptó al sistema: se realizaba inicialmente en pequeño grupo (cuatro miembros) y la última parte de forma individual. Se evaluaba aparte (20% de la nota final de DEIA). Se incluyó en la parte inferior de la estructura del campus virtual, tema número 5. Se estructuraba en tres actividades: búsqueda de resúmenes a través de las bases de datos científicas respecto a un tema asignado (en grupo), análisis metodológico del tema de investigación (en grupo) y análisis metodológico de un artículo científico original en inglés (individual). Trabajaban no presencialmente bajo la tutorización del docente encargado de esta tarea, a través de dos entrevistas presenciales tipo seminario (obligatorias) convocadas a través de la wiki (ver figura 6).

**5 Treball dirigit**

-  INSTRUCCIONS BÀSIQUES PER AL TREBALL DIRIGIT. CURS 2008/09
-  Llistat dels temes escollits pel grup M5 al curs 2008-2009
-  Llistat de grups i temes (13-03-2009)
-  CALENDARI
-  Fòrum de treball dirigit per cada grup

---

**Activitat 1:**

-  Activitat 1 - Cerca bibliogràfica (material de suport)
-  ACTIVITAT 1 DEL GRUP
-  ACTIVITAT 1 DEL GRUP (revisada)
-  CALENDARI PRIMERES ENTREVISTES

---

**Activitat 2:**

-  ACTIVITAT 2 DEL GRUP
-  ACTIVITAT 2 DEL GRUP (revisada)
-  CALENDARI SEGONES ENTREVISTES

---

**Activitat 3:**

-  ACTIVITAT 3 INDIVIDUAL: Article assignat
-  Instruccions ACTIVITAT 3
-  ACTIVITAT 3 INDIVIDUAL

---

**Enllaços a bases de dades de Psicologia:**

-  PsycINFO
-  PsycARTICLES

**Traductors online:**

-  Traductor UB
-  Translendum

Figura 6. Trabajo dirigido del campus virtual DEIA M5 (2008-2009)

Los criterios de evaluación del sistema de evaluación continua eran comunicados desde la presentación de la asignatura. Consistían en tres pruebas de validación de conocimientos teórico-prácticos, donde cada parte tenía un valor del 50%. El sistema tenía que garantizar el rendimiento del estudiante y la acumulación de conocimientos, por el que se optó por el siguiente (la primera puntuación del paréntesis es la obtenida por la prueba de validación y la segunda la que se daba al estudiante por los problemas campus realizados, siempre y cuando aprobase la prueba, si no la puntuación de las prácticas no le era sumada): bloque I (1.5+0.5) + bloques I+II (2+1) + bloques I+II+III+IV (2+1); de forma que el total de las tres pruebas era 8/10 (sólo faltaban los 2 puntos restantes del trabajo dirigido en incluir al final). Se elaborarán los nuevos cuestionarios de la innovación docente para evaluarla bajo el punto de vista de los estudiantes. Para el estudio de esta nueva innovación, se administró un cuestionario adaptado en parte del EMID (Bono, Arnau, Blanca 2006), que se re-diseñó con el fin de estudiar la innovación docente realizada en el nuevo sistema de evaluación de DEIA y el uso del campus virtual para el logro de los objetivos de aprendizaje (López Fernández, Viader Junyent, Cosculluela Mas, Honrubia Serrano, Malapeira Gas, en prensa).

### Conclusión

La adaptación al nuevo plan docente en base al EEES se ha llevado de forma gradual en cursos académicos (2005-2009). En el 2006-2007 se introdujo la innovación docente del campus virtual de la asignatura por cada grupo, cuya estructura representaba los 3 ejes citados y daba soporte personalizado a la evaluación continuada de las prácticas. El diseño e implementación del campus virtual ofreció al profesorado una gestión dinámica de los contenidos por parte del equipo docente, que ha representado un paso positivo hacia la docencia en el marco de la convergencia europea. En general los estudiantes han valorado positivamente la innovación docente (exceptuando, en los primeros años descritos algunas herramientas de *Moodle*: como el editor, la wiki de grupo y algunas herramientas de comunicación que se eliminaron o modificaron en 2007-2008 y mejoraron a partir de la versión 1.9), así como en el tiempo invertido en la realización de actividades prácticas y de trabajo dirigido, medidas con ECTS (una percepción compartida en estudios similares que puede ser debida a múltiples causas).

Los resultados obtenidos a través de la encuesta con *Moodle* fueron rápidos de analizar descriptivamente, que en ese momento era el objetivo, aunque no permitieron elaborar una matriz de datos para analizarlos de forma bivariable. Por ello, se prosiguió con la adaptación de un cuestionario similar en formato papel y con hojas de lectora óptica para facilitar la continua evaluación de los cambios que se han ido implementando hasta lograr la adaptación a las directrices del EEES según la UB (2006), de forma óptima para profesorado y estudiantes. Los últimos resultados indicaron datos objetivos en relación a la estimación del tiempo que dedican los estudiantes a cada elemento del modelo, así como una valoración del sistema de evaluación continuada en base a las pruebas de validación, las prácticas de aprendizaje autónomo (problemas campus), etc.

En general, los estudiantes valoran el grado de utilidad de los elementos del modelo presentado, así como el soporte del campus virtual y el sistema de evaluación asistido por el equipo docente a través dicho entorno virtual de aprendizaje. No obstante, aunque la conclusión de los resultados preliminares parece positiva desde la perspectiva de los estudiantes, que han mostrado una mejoría del rendimiento en general, así como satisfacción global con el sistema evaluativo, todavía se están analizando algunos de los resultados en detalle, pues en ocasiones sigue habiendo una percepción de sobrecarga de trabajo, el incremento de estudiantes que superan la asignatura parece ir asociado al decremento de detección de estudiantes excelentes y la aparición de nuevas versiones y softwares produce una revisión y estudio anual para facilitar el trabajo con soporte TIC a los estudiantes (como el uso de Open Office, etc.).

Por último, se considera que este modelo propuesto puede ser extensible a otras asignaturas metodológicas similares que se desarrollen con un equipo docente, soporte de campus virtual (o entorno similar) y posibilidades y soporte en innovación docente, tanto en los materiales didácticos como en la organización del sistema de evaluación.

### Referencias

- Bono, R., Arnau, J. y Blanca, M.J. (2006). Tecnologías de la información y la comunicación en la enseñanza de diseños experimentales y aplicados. *Psicothema*, 18(3), 646-651.
- López Fernández, O., Viader Junyent, M., Cosculluela Mas, A., Malapeira Gas, J.M., Rifà Ros, X. y Cifre León, I. (2008). Encuesta de innovación docente universitaria en Diseños experimentales y aplicados implementada en un campus virtual moodle. Ponencia presentada en el Congreso de Metodología de Encuestas, IESA, Universidad de Córdoba, Septiembre de 2008.
- López Fernández, O., Viader Junyent, M., Cosculluela Mas, A., Honrubia Serrano, M.L. y Malapeira Gas, J.M. (en prensa). Innovación docente en diseños experimentales y aplicados: ECTS, campus virtual y evaluación continua. Ponencia presentada en el XI Congreso de Metodología de las Ciencias Sociales y de la Salud, Málaga, Septiembre 2009.
- Viader, M., López, O., Rifà, X., Cifré, I., Cosculluela, A. y Malapeira, J.M. (2007). Incorporació de la plataforma Moodle com a eina de suport a la docència del marc europeu en l'assignatura Disseny experimental i aplicats. Ponència presentada a la Quarta Trobada de Professorat de Ciències de la Salut, Facultat de Medicina de la UB, juny de 2007.
- Universitat de Barcelona (2006). *Normes reguladores dels plans docents de les assignatures per als ensenyaments de la Universitat de Barcelona segons les directrius de l'Espai Europeu d'Educació Superior*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.

# ESTRATEGIAS DE APRENDIZAJE Y SU RELACIÓN CON EL RENDIMIENTO ACADÉMICO PREVIO DE LOS ESTUDIANTES DE NUEVO INGRESO EN LA TITULACIÓN DE PSICOLOGÍA

Dolores López-Montiel, María J. Blanca, Gema López-Montiel, Rafael Alarcón y Jesús Miranda  
Universidad de Málaga

Actualmente, en el contexto universitario, debido al desarrollo de nuevos paradigmas que explican los procesos de aprendizaje del alumno y a la influencia de la sociedad del conocimiento, se exige al estudiante universitario que trabaje autónoma y comprometidamente sus aprendizajes, estando claramente aceptado que las habilidades del alumno para autorregular su propio proceso de aprendizaje son definitivas para alcanzar el éxito académico. El trabajo que aquí se presenta se enmarca en una investigación cuyo objetivo general es conocer y comprender las características del alumno que accede a la titulación de Psicología para construir un punto de partida desde el que diseñar situaciones de aprendizaje eficaces que ayuden a los alumnos a ser aprendices independientes y activos. Como primer paso de esta investigación, se describe la relación entre las características académicas previas (rendimiento y conocimientos previos) y las estrategias cognitivas que son capaces de poner en práctica para el control de su propio aprendizaje los alumnos que inician sus estudios universitarios en la titulación de Psicología.

Uno de los temas centrales de la investigación actual en el campo de la educación es el concepto de aprendizaje autorregulado, que se desarrolla fundamentalmente a partir de los trabajos de Zimmerman y Schunk (1989). En los últimos años, en el contexto del estudio de los procesos y estrategias de aprendizaje, se ha desarrollado un ingente cuerpo de investigaciones sobre los procesos que los alumnos activan a la hora de aprender (Boekaerts y Corno, 2005; Castejón, Gilar y Pérez, 2006; Núñez, González-Pienda, Solano y Rosario, 2006). La autorregulación está relacionada con los pensamientos, sentimientos y acciones que son planeadas y sistemáticamente adaptadas, cuando sea necesario, para incrementar la motivación y el aprendizaje (Pintrich, 2004; Rosário, Soares, Núñez, González-Pienda y Rúbio, 2004; Zimmerman, 2000, 2002), de manera que el núcleo de los procesos de autorregulación reside en la posibilidad de elección y en el control (Rosário, Mourão, Núñez, González-Pienda, Solano y Valle, 2007).

En este sentido, y desde una perspectiva sociocognitiva, destaca el modelo de aprendizaje autorregulado de Pintrich (2000), que tiene por objeto la clasificación y análisis de los distintos procesos que, según la literatura científica, están implicados en el aprendizaje autorregulado. En este modelo los procesos reguladores se organizan en cuatro fases: planificación, auto-observación, control y evaluación. Y en cada una de ellas, las actividades se enmarcan dentro de cuatro áreas: cognitiva, motivacional/afectiva, comportamental y contextual (Tabla 1).

*Tabla 1. Fases, áreas y procesos implicados en el aprendizaje autorregulado (Pintrich, 2000b, p.454).  
Traducción de Torrano y González-Torres (2004, p. 5)*

FASES	COGNICIÓN	MOTIVACIÓN/ AFECTO	COMPORTAMIENTO	CONTEXTO
1 PREPARACIÓN PLANIFICACIÓN ACTIVACIÓN	Establecimiento de metas. Activación del conocimiento previo. Activación del conocimiento metacognitivo.	Adopción de metas. Juicios de autoeficacia. Activación de las creencias sobre el valor de la tarea. Activación del interés personal. Afectos (emociones).	Planificación del tiempo y del esfuerzo.	Percepción de la tarea. Percepción del contexto.
2 AUTO- OBSERVACIÓN	Conciencia y autoobservación de la cognición.	Conciencia y autoobservación de la motivación y del afecto.	Conciencia y autoobservación del esfuerzo, del empleo del tiempo y de la necesidad de ayuda.	Conciencia y autoobservación de las condiciones de la tarea y del contexto.
3 CONTROL REGULACIÓN	Uso de estrategias cognitivas y metacognitivas.	Uso de estrategias de control de la motivación y del afecto.	Incremento/ disminución del esfuerzo. Persistencia. Búsqueda de ayuda.	Cambios en los requerimientos de la tarea y en las condiciones del contexto.
4 EVALUACIÓN	Juicios cognitivos. Atribuciones.	Reacciones afectivas. Atribuciones.	Elección del comportamiento.	Evaluación de la tarea y del contexto.

Las estrategias de aprendizaje abarcan tanto estrategias cognitivas, como metacognitivas y de regulación de recursos (Weinstein, Husman y Dierking, 2000). Así, por ejemplo, estrategias cognitivas podrían ser las de repaso, elaboración y organización (Pintrich et al., 1991; Pintrich y García, 1993). Metacognitivas serían la planificación, el control y la regulación de las actividades (Pintrich et al., 1991). Y entre las de regulación de recursos destacarían, por ejemplo, la organización del tiempo, la regulación del esfuerzo o la búsqueda de ayuda.

En el contexto universitario actual, debido al desarrollo de estos paradigmas que explican los procesos de aprendizaje del alumno y a la influencia de la sociedad del conocimiento, se demanda al estudiante universitario que trabaje autónoma y comprometidamente sus aprendizajes, de manera que en la transición de la Secundaria a la Universidad se exige mucho de los nuevos universitarios (Chemers, Hu y García, 2001; Cook y Leckey, 1999; Delaney, 2002; Lawrence, 2001; Lowe y Cook, 2003). Sin embargo, aunque está claramente aceptado que las habilidades del alumno para autorregular su propio proceso de aprendizaje son definitivas para alcanzar el éxito académico (Nota, Soresi y Zimmerman, 2004), en muchos estudios se comprueba que los estudiantes llegan a la Universidad sin ser capaces de emplear adecuadamente estas capacidades de autorregulación (Allgood, Risko, Álvarez y Fairbanks, 2000). En este sentido, como afirman Rosário et al. (2007), los niveles de complejidad estructural de la comprensión de los alumnos universitarios se sitúan, en muchos casos, por debajo de los pretendidos, con la consiguiente repercusión en los resultados académicos (Rosário, Núñez, González-Pienda, Almeida, Soares, y Rúbio, 2005; Rosário, Mourão, Soares et al., 2005).

Por todo ello, y de acuerdo con Sander (2005), para proporcionar una educación eficiente y eficaz, que facilite la transición del alumno de Secundaria a la Universidad, es imprescindible que los profesores conozcan las características de sus alumnos como aprendices.

El objetivo de este trabajo es conocer y comprender algunas de las características del alumno que accede a la titulación de Psicología, así como la relación entre el rendimiento académico previo de estos alumnos y las estrategias de aprendizaje que de manera general son capaces de poner en práctica.

## Método

### *Participantes*

Participaron en el estudio 135 sujetos, con una media de edad de 18,82 años (D.T.= 1,884), de los cuales 99 eran mujeres y 36 hombres. Todos los participantes eran estudiantes de primer curso de Psicología del curso académico 2008-2009 de la Universidad de Málaga.

### *Instrumentos*

Las variables y los instrumentos de evaluación correspondientes se resumen en la Tabla 2.

*Cuestionario de Datos de Acceso a la Universidad.* Se ha elaborado un cuestionario que consta de 10 ítems con cuestiones referidas a los datos personales, detección del alumnado que cursa por primera vez la carrera de psicología, si eran repetidores, la modalidad e itinerario cursado durante el bachillerato y la nota media alcanzada durante el bachillerato y selectividad, entre otras cuestiones. Los formatos de respuesta elegidos para la medición de los diferentes ítems fueron: de verdadero-falso, de opción múltiple, con tres o más alternativas de respuesta y, por último abiertas.

De este cuestionario, como indicador del rendimiento académico en el bachillerato se eligieron las variables *nota media obtenida en el Bachillerato* y *nota alcanzada en la prueba de Selectividad*.

Y como indicador de los conocimientos previos de los alumnos se eligió la variable *Opción elegida para realizar la Selectividad* (Científico-Tecnológica, Ciencias de la Salud, Ciencias Sociales y Humanidades), por cuanto refleja la especialización elegida en el bachillerato (Tecnológico, Ciencias de la Naturaleza y de la Salud, Ciencias Sociales y Humanidades) y, por tanto, el tipo de conocimientos mayoritariamente adquiridos.

*Cuestionario de Estrategias de Aprendizaje y Motivación (CEAM II)* (Roces, Tourón y González, 1995), adaptación al castellano del *Motivated Strategies for Learning Questionnaire (MSQL)* de Pintrich, Smith y McKeachie (1991). Es un instrumento que mide la motivación y las estrategias de aprendizaje de los estudiantes y consiste en una escala tipo Likert (1-7) compuesta por 81 ítems, 31 de motivación y 50 de estrategias de aprendizaje. Se ha considerado aquí la escala de estrategias de aprendizaje que mide los aspectos: *uso de estrategias de elaboración, manejo del tiempo y concentración, uso de estrategias de organización, búsqueda de ayuda, constancia, uso de estrategias de autorregulación metacognitiva y uso de estrategias de autointerrogación.*

Tabla 2. Variables e Instrumentos

Variabes	Instrumentos de Evaluación
<i>V. Sociodemográficas</i>	
Edad y género	Cuestionario de Datos de Acceso a la Universidad
<i>Rendimiento académico y conocimientos previos</i>	Cuestionario de Datos de Acceso a la Universidad
Nota Bachillerato	
Nota Selectividad	
Opción Selectividad	
<i>Estrategias de aprendizaje</i>	
Elaboración	
Manejo del tiempo y concentración	Escala de Estrategias de Aprendizaje del Cuestionario de Estrategias de Aprendizaje y Motivación (CEAM II) (Roces, Tourón y González, 1995). Adaptación del MSLQ (Pintrich, Smith y McKeachie, 1991)
Organización	
Búsqueda de ayuda	
Constancia	
Metacognición	
Autointerrogación	

### Procedimiento

Los cuestionarios se administraron al comienzo del curso académico, durante dos sesiones de una hora cada una, respetando los grupos naturales de clase y en el horario habitual, con el consentimiento del Centro, el profesorado responsable en ese momento y el alumnado.

### Resultados y discusión

En la tabla 3, se presenta la descripción de las variables incluidas en el estudio.

Tabla 3. Resultados del análisis descriptivo

<i>Sociodemográficas</i>	<i>Rendimiento académico en el bachillerato</i>	<i>Estrategias de aprendizaje</i>	
Edad: 18,82 años (D.T.= 1,884).	Nota Bachillerato: 6,84 puntos (D.T.= 0,952)	N=135	
Género:	Nota Selectividad: 5,93 puntos (D.T.= 1,10)	M	DT
<i>Mujeres: 73,3%</i>	Opción Selectividad:	E	41,06 7,789
<i>Hombres: 26,7%</i>	CC. Sociales: 39,3%	TC	40,69 8,359
	CC. Salud: 29,6%	O	41,84 8,967
	Humanidades: 14,8%	A	32,86 5,79
	Científico-Tec.: 3,7%	C	15,75 3,023
	Artes: 5,2%	M	36,74 6,16
		AU	14,14 3,788

(E: Elaboración; TC: Tiempo y Concentración; O: Organización; A: Ayuda; C: Constancia; M: Metacognición; AU: Autointerrogación)

Como se aprecia en la tabla 3, en el curso 2008-09, la mayoría de los alumnos evaluados que se matricularon por primera vez en Psicología fueron mujeres y accedieron a la licenciatura por las vías Ciencias Sociales y Ciencias de la Salud en su mayoría, con una nota media de acceso de aprobado. Esto indica que la mayoría (el 72,6%, si se suman también a los alumnos del bachillerato de Tecnología) recibieron una formación continuada en Matemáticas y otras asignaturas consideradas de Ciencias, frente al 20% de los bachilleratos de Arte y de Humanidades, que no la recibió.

Para el análisis de la relación entre el rendimiento académico previo de los alumnos y las estrategias de aprendizaje empleadas, se ha llevado a cabo un análisis de correlación canónica. El primer conjunto de datos está formado por las puntuaciones directas de las variables *Nota en Bachillerato* y *Nota en Selectividad*. En el segundo se incluyen como variables las estrategias de aprendizaje siguientes: *elaboración, tiempo y concentración, organización, búsqueda de ayuda, constancia, autorregulación metacognitiva y autointerrogación*.

El análisis, empleando el criterio de Wilks, arroja sólo una variable canónica significativa [ $F(14, 214) = 2,03; p < 0,05$ ], con una correlación canónica de 0,44, lo que supone la existencia de intercorrelaciones significativas entre los dos conjuntos de variables resumidas en una dimensión.

En la Tabla 4 se exponen las correlaciones de las variables canónicas con las variables del primer y segundo conjunto.

Tabla 4. Correlaciones de las variables canónicas con las variables originales del primer y segundo conjunto

Variables	Saturaciones
V. Primer conjunto	
<i>Nota en Bachillerato</i>	0,79
<i>Nota en Selectividad</i>	0,93
V. Segundo conjunto	
<i>Estrategias de elaboración</i>	0,22
<i>Manejo del tiempo y concentración</i>	0,65
<i>Estrategias de organización</i>	0,07
<i>Búsqueda de ayuda</i>	0,45
<i>Constancia</i>	0,55
<i>Estrategias de autorregulación metacognitiva</i>	0,71
<i>Estrategias de autointerrogación</i>	0,63

Eligiendo una correlación de corte de 0,35 para la interpretación, las variables que en el conjunto segundo se correlacionan con la variable canónica son, de forma positiva, *manejo del tiempo y concentración, búsqueda de ayuda, constancia, estrategias de autorregulación metacognitiva y estrategias de autointerrogación*. Y no correlacionan con la variable canónica las variables *estrategias de elaboración y estrategias de organización*.

Por tanto, los resultados muestran que los estudiantes que acceden a la Titulación de Psicología por primera vez con rendimiento alto en Bachillerato y en Selectividad, también tienden a puntuar alto en la mayoría de las variables relativas a las estrategias de aprendizaje estudiadas. Así, alumnos con notas altas, emplean estrategias de regulación de recursos como son el manejo del tiempo y la búsqueda de ayuda entre los profesores o sus iguales durante la realización de las tareas académicas, siendo capaces de centrarse en aquello que están realizando en cada momento; demuestran diligencia y esfuerzo para llevar al día las actividades y trabajos de las diferentes asignaturas y alcanzar las metas establecidas; son capaces de establecer de antemano metas y regular el estudio y la propia comprensión; así como de hacerse preguntas durante el estudio con el objeto de centrarse en el objetivo, autoevaluar su nivel de comprensión y cuestionar la veracidad de lo estudiado.

Por otro lado, no se ha encontrado relación con las puntuaciones obtenidas en la subescala *elaboración*, que valora si el alumno aplica conocimientos anteriores a situaciones nuevas para resolver problemas, tomar decisiones o hacer evaluaciones críticas, y si es capaz de establecer conexiones de la información nueva con la que ya disponía para recordarla mejor. Tampoco se ha encontrado relación con la variable *organización*, que recoge estrategias que emplea el alumnado para estudiar el contenido y seleccionar la información relevante como esquemas, resúmenes, subrayado, etc.

Por tanto, aunque los resultados obtenidos en este estudio preliminar indican que el alumnado de alto rendimiento emplea estrategias de aprendizaje autorreguladoras, no se encuentran entre ellas el uso de estrategias de aprendizaje tan importantes como las de *elaboración* y *organización*. Parece que sería necesario entonces, instruir a los alumnos de nuevo ingreso, junto a la enseñanza de los contenidos disciplinares, en el uso de estas estrategias, de manera que se favorezca una adecuada transición desde el Bachillerato a la Universidad.

Agradecimientos. Este trabajo ha sido subvencionado por el Vicerrectorado de Profesorado, Formación y Coordinación de la Universidad de Málaga, proyecto PIE08-042.

### Referencias

- Allgood, W. P., Risko, V. J., Álvarez, M. C. y Fairbanks, M. M. (2000). Factors that influence study. En Flippo, R. F. y Caverly, D. C. (Coord.), *Handbook of college reading and study strategy research*. NJ: LEA.
- Boekaerts, M. y Corno, L. (2005). Self regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An international review*, 54(2), 199-231.
- Castejón, J.L., Gilar, R. y Pérez, A.M. (2006). Aprendizaje complejo: el papel del conocimiento, la inteligencia, motivación y estrategias de aprendizaje. *Psicothema*, 18(4), 679-685.
- Chemers, M.M., Hu, L. and Garcia, B., (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology*, 93 (1), 55-65.
- Cook, A. y Leckey, J. (1999). "Do expectations meet reality? A survey of changes in first-year student opinion". *Journal of Further and Higher Education*, 23 (2), 157-71.
- Delaney, A.M. (2002). Enhancing support for student diversity through research. *Tertiary Education and Management* (8) 12, 146-166.
- Lawrence, J. (2001). Academics and first years students: collaborating to access success in an unfamiliar university culture. *Widening Participation and Lifelong Learning* (3) 3, 4-14.
- Lowe, H. y Cook, A. (2003). Mind the Gap; are the students prepared for high education? *Journal of Further and Higher Education* (27) 1, 61-76.
- Nota, L., Soresi, S. y Zimmerman, B.J. (2004). Self-regulation and academic achievement and resilience: a longitudinal study. *International Journal of Educational Research*, 41(3), 198-251.
- Núñez, J.C., Solano, P., González-Pienda, J. y Rosário, P. (2006). El aprendizaje autorregulado como medio y meta de la educación. *Papeles del Psicólogo*, 27(3), 139-146.

- Pintrich, P.R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- Pintrich, P.R. (2000). The role of goal orientation in self-regulated learning. En Boekaerts, M. y Pintrich, P. R. (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press.
- Pintrich, P.R. y Garcia, T. (1993). Intraindividual differences in students' motivation and selfregulated learning. En *German Journal of Educational Psychology* (7), 3, 99-107.
- Pintrich, P.R., Smith, D.A.F., García, T. y McKeachie, W.J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. National Center for Research to Improve PostSecondary Teaching and Learning. Ann Arbor, MI.
- Roces, C., Tourón, J. y González, M.C. (1995). Validación preliminar del CEAM II. *Psicológica*, 16, 347-366.
- Rosário, P., Mourão, R., Núñez, J. C., González-Pienda, J., Solano, P. y Valle, A. (2007). Eficacia de un programa instruccional para la mejora de procesos y estrategias de aprendizaje en la enseñanza superior. *Psicothema*, 19 (3), 422-427.
- Rosário, P., Núñez, J.C., González-Pienda, J.A., Almeida, L., Soares, S. y Rúbio, M. (2005). El aprendizaje escolar examinado desde la perspectiva del Modelo 3P de J. Biggs. *Psicothema*, 17, 20-30.
- Rosário, P., Soares, S., Núñez, J.C., González-Pienda, J. y Rúbio, M. (2004). Processos de auto-regulação da aprendizagem e realização escolar no Ensino Básico. *Psicologia, Educação e Cultura*, 8(1), 141-157.
- Sander, P. (2005). La investigación sobre nuestros alumnos, en pro de una mayor eficacia en la enseñanza universitaria. *Revista electrónica de Investigación Psicoeducativa y Psicopedagógica*, 5-3 (1), 113-130.
- Torrano Montalvo, F. y González Torres, M.C. (2004). Self-regulated learning: current and future directions. *Electronic Journal of Research in Educational Psychology*, 2(1), 1-34.
- Weinstein, C.J., Husman, Y. y Dierking, D. (2000). Self regulation interventions with a focus on learning strategies. En Boekaerts, M., Pintrich, P. y Zeidner, M. (2000) *Handbook of Self-regulation*. San Diego, CA: Academic Press.
- Zimmerman, B.J. (2000). Attaining self-regulation. A social cognitive perspective. En M. Boekaerts, P. Pintrich y M. Zeidner (Eds.), *Handbook of self-regulation*. New York: Academic Press.
- Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2), 64-70.
- Zimmerman, B.J. y Schunk, D.H. (Eds). (1989). *Self-regulated Learning academic achievement: Theory, research and practice*. New York: Springer-Verlag.

# DIFERENCIAS INDIVIDUALES EN LA AUTORREGULACIÓN DE LOS APRENDIZAJES EN ESTUDIANTES UNIVERSITARIOS: ESTRATEGIAS MOTIVACIONALES Y ASPECTOS VOLITIVOS DE LA PERSONALIDAD

Dolores López-Montiel, María J. Blanca, Gema López-Montiel, Rafael Alarcón y Jesús Miranda  
Universidad de Málaga

El modelo de aprendizaje autorregulado de Pintrich, desde la perspectiva de las características del alumno, implica una autorregulación de componentes comportamentales, motivacionales, cognitivos y contextuales que permiten llevar a cabo aprendizajes eficaces, logrando los resultados académicos deseados. Una de las estrategias comportamentales de control que los alumnos son capaces de poner en marcha son las que conciernen a la voluntad, orientadas a evitar las distracciones externas e internas, mantener la concentración, su esfuerzo y su motivación durante la realización de las tareas académicas. En este trabajo se estudia la relación entre aspectos volitivos de la personalidad, como la escrupulosidad y la perseverancia, y las estrategias motivacionales en los alumnos que inician los estudios de Psicología. El análisis de correlación canónica llevado a cabo muestra que las estrategias motivacionales creencias de control y autoeficacia para el aprendizaje, autoeficacia para el rendimiento, orientación hacia metas intrínsecas y extrínsecas y valor de la tarea, correlacionan positivamente con las variables escrupulosidad y perseverancia.

En la actualidad, en la Educación Superior Europea, para la que se pretende una mejora significativa de los procesos de enseñanza-aprendizaje, evolucionando desde una perspectiva centrada en el profesor a otra centrada en el aprendizaje y en el estudiante (Fernández y Fernández, 2006), se acepta que las habilidades del alumno para autorregular su propio proceso de aprendizaje son definitivas para alcanzar el éxito académico (Nota, Soresi y Zimmerman, 2004). De acuerdo con Torrano y González-Torres (2004) y en la línea de Zimmerman (2002), lo que caracteriza a los alumnos que autorregulan su aprendizaje es que se sienten agentes de su conducta, creen que el aprendizaje es un proceso proactivo, están automotivados y usan las estrategias que les permiten lograr los resultados académicos deseados. Según el modelo de aprendizaje autorregulado de Pintrich (2000), los procesos reguladores se organizan en cuatro fases: planificación, auto-observación, control y evaluación. Y en cada una de ellas, las actividades se enmarcan dentro de cuatro áreas: cognitiva, motivacional/afectiva, comportamental y contextual, de manera que la autorregulación de estos componentes permite llevar a cabo aprendizajes eficaces, logrando los resultados académicos deseados. Torrano y González-Torres (2004) recomiendan algunas direcciones para la investigación en el campo del aprendizaje autorregulado, como el estudio del papel que juegan las diferencias individuales en el mismo. Así, por ejemplo, Paris y Paris (2001), sugieren el estudio de la influencia de características de la personalidad como la impulsividad, la paciencia o la resistencia ante las distracciones en el proceso de autorregulación. También se sugiere el estudio de cómo los alumnos controlan su propia motivación, emociones y comportamiento, para una mayor implicación y desempeño académicos (Pintrich, 1999). Esta capacidad de control ha sido conceptualizado como volitivo, y cobra mayor relevancia cuando los objetivos a alcanzar requieren de concentración y esfuerzo durante largos períodos de tiempo (Corno, 1993; Heckhausen y Kuhl, 1985).

El objetivo de este trabajo es el estudio de la relación de dimensiones de la personalidad, como la escrupulosidad y la perseverancia, con las estrategias motivacionales empleadas por los alumnos que inician los estudios de Psicología.

La escrupulosidad se asocia con la meticulosidad y el orden, y la perseverancia está relacionada con la persistencia y tenacidad, características de la personalidad que de forma global hacen referencia a la capacidad de autorregulación o autocontrol con la que se llevan a cabo las tareas emprendidas.

## Método

### *Participantes*

Participaron en el estudio 135 sujetos, con una media de edad de 18,82 años (D.T.= 1,884), de los cuales 99 eran mujeres y 36 hombres. Todos los participantes eran estudiantes de primer curso de Psicología del curso académico 2008-2009 de la Universidad de Málaga.

### *Instrumentos*

Para la evaluación de las variables mencionadas de los alumnos que acceden a la Facultad de Psicología por vez primera se utilizaron los cuestionarios siguientes:

- *Cuestionario “Big Five” (BFQ)* (Caprara, Barbaranelli y Borgogni, 1993). Test de personalidad compuesto por 132 ítems que se miden en una escala tipo Likert (1-5). Consta de 5 dimensiones (energía, afabilidad, tesón, estabilidad emocional, apertura mental), y diez subdimensiones (dinamismo, dominancia, cooperación/empatía, cordialidad/amabilidad, escrupulosidad, perseverancia, control de las emociones, control de los impulsos, apertura a la cultura, apertura a la experiencia). También incluye una escala de distorsión (D) que mide la tendencia de ofrecer un perfil falseado.

La dimensión *Tesón*, hace referencia a la capacidad de autorregulación o autocontrol. Las personas que obtienen una puntuación alta en esta dimensión tienden a describirse como muy reflexivas, escrupulosas, ordenadas, diligentes y perseverantes. Sin embargo, aquellas otras que obtienen puntuaciones baja, tienden a describirse como poco reflexivas, poco escrupulosas, poco ordenadas, poco diligentes y poco perseverantes. Consta de las siguientes subescalas:

- *Escrupulosidad*. Mide aspectos relacionados con la meticulosidad y gusto por el orden.
  - *Perseverancia*. Se refiere a aspectos relacionados con la persistencia y tenacidad con que se llevan a cabo las tareas y actividades emprendidas.
- *Cuestionario de Estrategias de Aprendizaje y Motivación (CEAM II)* (Roces, Tourón y González, 1995), adaptación al castellano del MSQI de Pintrich, Smith y McKeachie (1991). Es un instrumento que mide la motivación y las estrategias de aprendizaje de los estudiantes y consiste en una escala tipo Likert (1-7) compuesta por 81 ítems, 31 de motivación y 50 de estrategias de aprendizaje. Se ha empleado la escala de *motivación*, que mide los factores: *creencia de control, autoeficacia para el rendimiento, orientación hacia metas intrínsecas, orientación hacia metas extrínsecas, valor de la tarea y ansiedad durante la realización de tareas escolares*.

Tabla 1. Variables e Instrumentos

Variabes	Instrumentos de Evaluación
<i>V. Sociodemográficas</i> Edad y género	Cuestionario de Datos de Acceso a la Universidad.
<i>Variables de personalidad</i> Escrupulosidad Perseverancia	Subdimensiones Escrupulosidad y Perseverancia del Cuestionario "Big Five", BFQ, (Caprara, Barbaranelli y Borgogni, 1995).
<i>Estrategias motivacionales</i> Creencia de control Autoeficacia para el rendimiento Orientación hacia metas intrínsecas Orientación hacia metas extrínsecas Valor de la tarea Ansiedad	Escala de Motivación del Cuestionario de Estrategias de Aprendizaje y Motivación (CEAM II) (Roces, Tourón y González, 1995). Adaptación del MSLQ (Pintrich, Smith y McKeachie, 1991).

### Procedimiento

Los cuestionarios se administraron al comienzo del curso académico, durante dos sesiones de una hora cada una, respetando los grupos naturales de clase y en el horario habitual, con el consentimiento del Centro, el profesorado responsable en ese momento y el alumnado.

### Resultados y discusión

En la tabla 2 se describen las variables utilizadas en el estudio.

Tabla 2. Resultados del análisis descriptivo

<i>Sociodemográficas</i>	<i>Personalidad</i>		<i>Estrategias motivacionales</i>			
	N=135		N=135			
	M	DT	M	DT		
Edad: 18,82 años (DT= 1,884)	ES	40,12	6,91	CC	41,11	5,82
Género: <i>Mujeres: 73,3%</i> <i>Hombres: 26,7%</i>	PER	45,20	6,10	AR	18,69	3,60
				MI	19,34	3,61
				ME	20,43	4,49
				VT	33,46	4,78
				A	20,09	6,49

(ES: Escrupulosidad; PER: Perseverancia; CC: Creencia de Control; AR: Autoeficacia Rendimiento; MI: Metas Intrínsecas; ME: Metas Extrínsecas; VT: Valor de la Tarea; A: Ansiedad).

Se ha llevado a cabo un análisis de correlación canónica entre las variables de personalidad escrupulosidad y perseverancia y las estrategias motivacionales de los alumnos. El primer conjunto de datos está formado por las puntuaciones directas de las variables *perseverancia* y *escrupulosidad*. En el segundo se incluyen las variables de motivación para el estudio del alumno: *creencias de control* y *autoeficacia para el aprendizaje*, *autoeficacia para el rendimiento*, *orientación hacia metas intrínsecas*, *orientación hacia metas extrínsecas*, *valor de la tarea* y *ansiedad* durante la realización de los exámenes.

El análisis, empleando el criterio de Wilks, arroja sólo una variable canónica significativa [F (12, 238)=3,9; p<0,001], con una correlación canónica de 0,49, lo que supone la existencia de intercorrelaciones significativas entre los dos conjuntos de variables resumidas en una dimensión.

En la Tabla 3 se exponen las correlaciones de las variables canónicas con las variables del primer y segundo conjunto.

Tabla 3. Correlaciones de las variables canónicas con las variables originales del primer y segundo conjunto

Variabes	Saturación
V. Primer conjunto	
<i>Escrupulosidad</i>	0,84
<i>Perseverancia</i>	0,83
V. Segundo conjunto	
<i>Creencias de control y autoeficacia para el aprendizaje</i>	0,35
<i>Autoeficacia para el rendimiento</i>	0,73
<i>Orientación hacia metas intrínsecas</i>	0,63
<i>Orientación hacia metas extrínsecas</i>	0,51
<i>Valor de la tarea</i>	0,69
<i>Ansiedad</i>	0,27

Eligiendo una correlación de corte de 0,35 para la interpretación, las variables del segundo conjunto que correlacionan con la variable canónica son, de forma positiva, *creencias de control* y *autoeficacia para el aprendizaje*, *autoeficacia para el rendimiento*, *orientación hacia metas intrínsecas*, *orientación hacia metas extrínsecas* y *valor de la tarea*, no correlacionando la variable *ansiedad*.

Por tanto, sujetos con puntuaciones altas en escrupulosidad y perseverancia también tienden a puntuar alto en las variables relativas a los aspectos motivacionales de las estrategias de aprendizaje empleadas por los alumnos. Así, estudiantes que se describen como reflexivos, meticolosos y ordenados, diligentes y perseverantes en el desempeño de las actividades emprendidas, creen que el dominio en una asignatura depende del esfuerzo y modo de estudiar, se consideran capaces de aprender lo que se requiere en las diferentes materias, confían positivamente en sus capacidades para alcanzar un buen rendimiento académico y consideran importantes las asignaturas académicas para su propia formación y útiles para entender otras asignaturas. No encontrándose relación con las puntuaciones obtenidas en la subescala ansiedad que recoge pensamientos negativos y reacciones fisiológicas manifestadas durante la realización de los exámenes.

También se ha encontrado que estos sujetos se implican en las tareas de aprendizaje tanto por razones intrínsecas (de reto, maestría o dominio), como extrínsecas (notas, recompensas u opinión de los otros). Es decir, que se orientan tanto a la tarea, como a los resultados.

En definitiva, los resultados obtenidos en este estudio preliminar indican que características de la personalidad como la escrupulosidad y la perseverancia, aspectos de la dimensión tesón que

hace referencia a la capacidad de autorregulación o autocontrol, se asocian con los componentes motivacionales implicados en el aprendizaje.

Agradecimientos. Este trabajo ha sido subvencionado por el Vicerrectorado de Profesorado, Formación y Coordinación de la Universidad de Málaga, proyecto PIE08-042.

### Referencias

- Caprara, G.V., Barbaranelli, C., Borgogni, L. y Perugini, M. (1993). The “Big Five Questionnaire”: a new questionnaire to assess the five factor model. *Personality and Individual Differences*, 15, 281-288.
- Corno, L. (1993). The best-laid plans: Modern conceptions and educational research. *Educational Research*, 22(2), 14-22.
- Fernández, E. y Fernández, M. (2006). El Espacio Europeo de Educación Superior. En J.L. Benítez, A.B.G. Berbén, F. Justicia y J. de la Fuente (Coords.): *La universidad ante el reto del espacio europeo de educación superior: investigaciones recientes*. Madrid: EOS.
- Heckhausen, H. y Kuhl, J. (1985). From wishes to action: The dead ends and short cuts on the long way to action. En M. Frese y J. Sabini (Eds.), *Goal directed behavior*. Hillsdale, NJ: Erlbaum.
- Nota, L., Soresi, S. y Zimmerman, B.J. (2004). Self-regulation and academic achievement and resilience: a longitudinal study. *International Journal of Educational Research*, 41(3), 198–251.
- Paris, S.G. y Paris, A.H. (2001). Classroom application of research on self-regulated learning. *Educational Psychologist*, 36(2), 89-101.
- Pintrich, P.R. (1999). The Role of Motivation in Promoting and Sustaining Self-regulated Learning. *International Journal of Educational Research*, 31 (6), 459-470.
- Pintrich, P.R. (2000). The role of goal orientation in self-regulated learning. En Boekaerts, M. y Pintrich, P. R. (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press.
- Pintrich, P.R., Smith, D.A.F., García, T. y McKeachie, W.J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. National Center for Research to Improve PostSecondary Teaching and Learning. Ann Arbor, MI.
- Roces, C., Tourón, J. y González, M.C. (1995). Validación preliminar del CEAM II. *Psicológica*, 16, 347-366.
- Torrano, F. y González Torres, M.C. (2004). Self-regulated learning: current and future directions. *Electronic Journal of Research in Educational Psychology*, 2(1), 1-34.
- Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2), 64-70.

# LA RELACIÓN INTERPERSONAL DOCENTE-ESTUDIANTE EN LA EDUCACIÓN SUPERIOR. ADAPTACIÓN DEL QTI Y RESULTADOS POR RAMAS DE CONOCIMIENTO

Esther Mena y Juan Carlos Tójar  
Universidad de Málaga

Este trabajo resume una investigación más amplia realizada sobre las relaciones interpersonales entre profesorado y alumnado universitario. La investigación utiliza el modelo original MITB (*Model for Interpersonal Teacher Behaviour*) materializado en el instrumento QTI (*Questionnaire on Teacher Interaction*) de Wubbels, Créton y Hooymayers (1985). Este instrumento continúa siendo un referente en la investigación sobre relaciones interpersonales en el aula (p. ej. Telli, Brok y Cakiroglu, 2007). El QTI posee ocho escalas (liderazgo, apoyo, comprensión, responsabilidad/libertad, incertidumbre, insatisfacción, sanción e inflexibilidad), que según el modelo MITB se organizan en dos ejes principales (cooperación-oposición, y dominancia-sumisión). En este trabajo se presenta la traducción y adaptación del QTI a la Educación Superior, validando el instrumento (expertos, análisis factorial,...), y mostrando algunas conclusiones sobre las relaciones interpersonales docente-estudiante en las diferentes ramas y especialidades universitarias. Para esto último, el instrumento adaptado se aplicó a una muestra de alumnado universitario en la provincia de Málaga de diferentes ramas de conocimiento y titulaciones ( $n=509$ ).

Este trabajo de investigación trata de la percepción del estudiante sobre el comportamiento interpersonal entre éste y el docente en la Educación Superior. Existen varias formas de evaluar las percepciones de la interacción entre docentes y estudiantes. En las siguientes páginas se resume una investigación más amplia realizada sobre las relaciones interpersonales entre el profesorado y el alumnado universitario. La investigación utiliza el modelo original MITB (*Model for Interpersonal Teacher Behaviour*) materializado en el instrumento QTI (*Questionnaire on Teacher Interaction*) de Wubbels, Créton y Hooymayers (1985). Para esta investigación se ha elegido uno de los instrumentos más internacionalmente utilizados. El QTI original poseía 77 ítems en una escala Likert de 5 puntos. Versiones posteriores han ido reduciendo el número de ítems hasta un total de 48 (den Brok, Fisher y Koul, 2005). Ésta última es la versión utilizada en este estudio. El QTI se compone de 8 escalas o sectores interpersonales (liderazgo, apoyo, comprensión, libertad/responsabilidad, incertidumbre, insatisfacción, sanción e inflexibilidad), organizados a su vez en dos dimensiones (o ejes independientes): Influencia (dominancia-sumisión) y Proximidad (oposición-cooperación).

El QTI ha sido utilizado en multitud de estudios para analizar las relaciones interpersonales docente-estudiante en ambientes de aprendizaje eficaces (den Brok et al., 2004; den Brok, Fisher y Scott, 2005). Disponer de un instrumento que permita conocer exhaustivamente las formas de interacción de los mejores docentes podría ser de mucha utilidad para tratar de favorecer las relaciones interpersonales que conduzcan a mejorar los ambientes de aprendizaje.

Los objetivos que persigue esta investigación son los siguientes:

1. Traducir, adaptar y validar el QTI a una muestra española de estudiantes de Educación Superior.
2. Conocer las percepciones que tienen los estudiantes sobre las relaciones interpersonales con sus mejores docentes.
3. Analizar si existen diferencias entre las percepciones de los estudiantes, sobre las relaciones interpersonales con sus docentes, según la rama de conocimiento de procedencia.

La hipótesis principal de este trabajo es que, teniendo en cuenta la percepción de los estudiantes sobre los mejores profesores que han tenido durante su carrera universitaria, debería haber diferencias en cuanto a las relaciones interpersonales que se dan en el aula dependiendo de la rama de conocimiento de pertenencia. Para ello se ha aplicado el QTI a una muestra de alumnado universitario en la provincia de Málaga de diferentes ramas de conocimiento y titulaciones ( $n=509$ ).

### Método

Para esta investigación, se partió de la versión del QTI (48 ítems) utilizada por den Brok, Fisher y Koul (2005). Esta versión se tradujo al castellano, fue revisada por varios expertos y se aplicó a una muestra inicial (experiencia piloto) de 62 estudiantes titulados universitarios (matriculados en el curso para la obtención del Certificado de Aptitud Pedagógica, CAP). En esta aplicación piloto se realizó el mismo protocolo que se emplearía en el futuro con la muestra total de sujetos. En particular se les indicó, de acuerdo con lo objetivos de la investigación, que rellenaran la escala pensando en el/los mejor/es docentes que habían tenido durante su carrera de procedencia. Los resultados de las consideraciones de los expertos y de la aplicación piloto consiguieron afinar en la definición de algunos de los ítems de la escala.

Existen numerosos estudios sobre la fiabilidad y la validez del QTI. Por ejemplo, en los Países bajos (Wubbels et al.1985); en los Estados Unidos (Wubbels y Levy, 1991) y en Australia (Fisher, Fraser y Wubbels, 1992). En todos ellos se obtuvo una fiabilidad aceptable.

Los análisis realizados con la muestra inicial de esta investigación mostraron resultados similares a los señalados anteriormente. La consistencia interna, medida con el coeficiente alpha de Cronbach, obtuvo un valor de 0,92. Además de los análisis factoriales tradicionales para datos cuantitativos, debido a la naturaleza ordinal de los datos, se optó por realizar un análisis factorial para datos categoriales (CATPCA). Dicho análisis mostró una solución satisfactoria con una estructura de dos factores que representaba adecuadamente los ocho conjuntos de ítems correspondientes a las escalas que componen el QTI. Ambos factores, que pueden perfectamente ser interpretados como las dimensiones de proximidad e *influencia* del modelo teórico, explican el 60,54% de la varianza de las escalas del QTI.

### Participantes

Para la realización de esta investigación se han elegido a alumnos que ya habían terminado su carrera universitaria, en concreto a  $N=509$  alumnos que representan a las 5 ramas de conocimiento en las que se agrupan las diferentes carreras que se estudian en la UMA (Humanidades, Ciencias, Ciencias de la salud, Ciencias sociales y jurídicas e Ingenierías y arquitectura).

Tabla 1. Distribución de la muestra por ramas de conocimiento

Ramas de conocimiento	Porcentaje
Humanidades	44.0
Ciencias	13.1
Ciencias de la Salud	7.0
Ciencias Sociales y Jurídicas	25.4
Ingenierías y Arquitectura	7.0
Otros	3.6

La edad de estos estudiantes oscilaba entre 22 años mínimo y 65 máximo, con una media de edad de 27.82 y una desviación típica de 6.12. Dada la diferencia de edades anteriormente descrita también los años de inicio y término de la carrera universitaria es diverso, oscila entre 1970 y el 2007, con una duración media de 5.39 años y una desviación típica de 1.97. De esta muestra hay que destacar que 170 personas eran hombres y 338 mujeres, representando unos porcentajes de 33.5 y 66.5 respectivamente.

### Análisis

Para intentar comprobar la existencia de diferencias en la percepción de los estudiantes sobre los mejores docentes dependiendo de la rama de conocimiento de pertenencia, se planteó un diseño multivariante. Como variable independiente se tomaron las ramas de conocimiento y como dependientes cada una de las escalas del QTI. El diseño se resolvió a través de un análisis multivariante de la varianza (MANOVA). Para comprobar la igualdad de las matrices de covarianzas se utilizó la prueba de Box que contrasta la hipótesis nula de que las matrices de covarianza observadas son iguales en todos los grupos.

Se utilizó también el contraste de Levene sobre la igualdad de las varianzas error para contrastar la hipótesis de que la varianza de error de la variable dependiente es igual a lo largo de todos los grupos. Tras la realización de los contrastes multivariados y univariados se procedió a la realización de comparaciones múltiples (pruebas *post hoc*) mediante la DHS de Tukey para determinar en qué escalas concretas se producían diferencias en las ramas de conocimiento.

### Resultados

Para realizar los contrastes multivariados se calcularon la traza de Pillai, la Lambda de Wilks, la traza de Hotelling y la raíz mayor de Roy. Las pruebas *F* contrastan el efecto multivariado de la Rama de conocimiento, y todas ellas se basan en comparaciones por pares linealmente independientes entre las medias marginales estimadas. Como se observa en la tabla 2, todos los contrastes son significativos ( $p < 0.0005$ ), por lo que se rechaza la hipótesis nula de independencia entre las escalas o factores del QTI y las ramas de conocimiento. En la tabla 2 se incluye además el cálculo del tamaño del efecto a partir del coeficiente eta al cuadrado parcial.

Tabla 2. Contrastes multivariados realizados en el MANOVA

Efecto		Valor	<i>F</i>	Gl de la hipótesis	Gl del error	<i>p</i>	Tamaño del efecto
Ramas De Conocimiento	Traza de Pillai	1.208	12.510	48	2382	0.000	0.201
	Lambda de Wilks	0.004	83.587	48	1932	0.000	0.602
	Traza de Hotelling	200.592	1631.206	48	2342	0.000	0.971
	Raíz mayor de Roy	200.358	9942.765	8	397	0.000	0.995

En las pruebas de los efectos inter-sujetos realizadas, los contrastes en todos los casos son significativos ( $p < 0.0005$ ) lo que implica rechazar la hipótesis de independencia entre los factores del QTI y las ramas de conocimiento. Esto es, en las distintas ramas de conocimiento las percepciones sobre los diversos factores del QTI no es la misma en ningún caso.

Para tratar de determinar en qué escalas concretas se producían diferencias en las ramas de conocimiento, se procedió a la realización de comparaciones múltiples (pruebas *post hoc*) mediante la DHS de Tukey. En las siguientes líneas se comentan sólo las relaciones significativas entre las ramas de conocimiento y las escalas del QTI.

- En cuanto al liderazgo, la percepción de los estudiantes de humanidades sobre sus mejores docentes es mayor que los estudiantes de Ingenierías y Arquitectura ( $p < 0.0005$ ). Lo mismo ocurre con la percepción sobre el liderazgo que tienen los estudiantes de Ciencias y Ciencias sociales y jurídicas. En ambos casos son mayores que en Ingeniería y Arquitectura ( $p = 0,004$  y  $p = 0,008$ , respectivamente).
- En relación a la variable incertidumbre, la percepción de los estudiantes de Humanidades sobre sus mejores docentes es menor que la de los estudiantes de Ingeniería y Arquitectura ( $p = 0,001$ ). La percepción sobre la incertidumbre en los mejores docentes de Ciencias también es menor que la que manifiestan los estudiantes de Ingeniería y Arquitectura ( $p = 0,019$ ).
- En la variable apoyo, que brindan los mejores docentes, la percepción de los estudiantes de Humanidades es mayor que la de los de Ciencias ( $p = 0,012$ ) y que los de Ingeniería y Arquitectura ( $p = 0,001$ ).
- En relación a la variable insatisfacción, los mejores docentes son percibidos con menor valoración en Humanidades ( $p = 0,002$ ) y en Ciencias Sociales y Jurídicas ( $p = 0,001$ ) que en Ingeniería y Arquitectura.
- En cuanto a la variable comprensión, los estudiantes de Humanidades ( $p < 0,0005$ ) y de Ciencias ( $p = 0,023$ ) valoran más a sus mejores docentes que los estudiantes de Ingeniería y Arquitectura.
- En la variable libertad, los estudiantes de Humanidades perciben que sus mejores docentes les dan más margen de libertad que los docentes de Ciencias ( $p = 0,001$ ) y que los de Ingeniería y Arquitectura ( $p = 0,003$ ).

No existen evidencias de relaciones significativas en relación a las variables de sanción y rigidez con respecto a ninguna de las ramas de conocimiento.

### Conclusión

Se ha conseguido traducir, adaptar y validar el QTI a una muestra española de estudiantes de Educación Superior. El QTI es un test muy validado y utilizado internacionalmente para evaluar las relaciones interpersonales entre docentes y estudiantes. Tanto las aplicaciones como las validaciones se han realizado en diferentes niveles educativos. Es la primera vez que se realiza una aproximación a la validación de este instrumento para la educación superior. La adaptación y aplicación del instrumento a una muestra de estudiantes de educación superior es al mismo tiempo una evidencia de validez del modelo de Wubbels et al. (1985), el modelo circunplejo de comportamiento interpersonal del docente, para la educación superior.

La validación del instrumento permitió aplicarlo para conocer las percepciones que tienen los estudiantes sobre las relaciones interpersonales con sus mejores docentes. Para ello se elaboró

un cuestionario que solicitó a los estudiantes que pensarán, a la hora de rellenar el cuestionario, en las relaciones interpersonales que mantuvieron con su/s mejor/es docente/s. Los resultados de esta parte de la investigación (que no se ha incluido exhaustivamente en este trabajo por cuestiones de espacio y tiempo), muestran un perfil de las relaciones interpersonales entre los mejores docentes y sus estudiantes, desde el punto de vista de estos últimos, en una muestra obtenida de estudiantes licenciados en la Universidad de Málaga.

Se han analizado y comprobado, aplicando un MANOVA sobre los resultados obtenidos tras la aplicación del QTI, las diferencias entre las percepciones de los estudiantes, sobre las relaciones interpersonales con sus docentes, según la rama de conocimiento de procedencia. En este sentido se han podido detectar algunas relaciones significativas que muestran diferente percepción de las relaciones interpersonales en los estudiantes de unas ramas de conocimiento con respecto a otras. En principio se podría pensar que las relaciones interpersonales son, por ejemplo, más “próximas” (en el eje proximidad, más cercano al polo de cooperación). Pero se ha de tener en cuenta que no sólo los docentes considerados son diferentes, también lo son los evaluadores (estudiantes, en este caso), y, por supuesto, las titulaciones. Algunas variables relativas a los diferentes contextos docentes, y de la titulación, incluso a las características de los estudiantes (por ejemplo, diferente nivel de exigencia), hacen bastante complicada la comparación sin más.

Los resultados han probado la hipótesis principal de este trabajo. En este sentido se puede afirmar que existen diferencias en cuanto a las relaciones interpersonales que se dan en el aula dependiendo de la rama de conocimiento de pertenencia. Estas diferencias, como ya se tuvo en cuenta en la redacción de la hipótesis principal, hace referencia a la percepción de los estudiantes sobre los mejores profesores que han tenido durante su carrera universitaria.

Sin dejar de tener en cuenta las consideraciones expresadas en este apartado en relación al cumplimiento del tercer objetivo, se puede afirmar que con los resultados se intuye el perfil de los mejores docentes, percibido desde la perspectiva del alumnado, que predomina en cada una de las ramas de conocimiento.

Los mejores docentes (que se mueven en un rango del 19,5% al 28,27%, en la muestra considerada), no tienen por qué ser similares en cada una de las ramas de conocimiento. Los resultados de investigaciones de estas características pueden profundizar en las relaciones interpersonales docente-estudiante más eficaces en las diferentes ramas de conocimiento.

## Referencias

- den Brok, P., Brekelmans, M. & Wubbels, T. (2004). Interpersonal Teacher Behaviour and Student Outcomes. *School Effectiveness and School Improvement*, 15(3), 407-442.
- den Brok, P., Fisher, D. & Koul, R. (2005). The Importance of Teacher Interpersonal Behaviour for Secondary Science Students' Attitudes in Kashmir. *Journal of Classroom Interaction*, 40(2), 5-19.
- den Brok, P., Fisher, D. & Scott, R. (2005). The Importance of Teacher Interpersonal Behaviour for Student Attitudes in Brunei Primary Science Classes. Research Report. *International Journal of Science Education*, 27(7), 765-779.
- Fisher, B. J., Fraser, D. L. y Wubbels, T. (1992). *Teacher communication style and school environment*. Paper presentado en la European Conference on Educational Research. Enschede.

- Telli, S., den Brok, P. & Cakiroglu, J. (2007). Students' Perceptions of Science Teachers' Interpersonal Behaviour in Secondary Schools: Development of a Turkish Version of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 10(2), 115-129.
- Wubbels, T. y Levy, J. (1991). A comparison of interpersonal behavior of Dutch and American teachers. *International Journal of Intercultural Relationships*, 15, 1-18.
- Wubbels, T., Créton, H. A. y Hooymayers, H. P. (1985). Discipline Problems of Beginning Teachers, Interactional Teacher Behaviour Mapped Out. Recuperado febrero 4, 2009, a partir de <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED260040>.

# INTEGRACIÓN DEL INGLÉS COMO SEGUNDA LENGUA EN EL APRENDIZAJE DE DISEÑOS EXPERIMENTALES

Juan Carlos Oliver y Patricia Salazar  
Universidad Jaume I de Castellón

El objetivo del trabajo fue el desarrollo y evaluación de un contexto bilingüe en el aprendizaje de diseños experimentales. Se realizó durante dos cursos en coordinación con la asignatura *Inglés para Psicólogos*. Participantes voluntarios realizaron pretests y postests en esta lengua con contenidos específicos de diseño. La proporción de participantes que completaron satisfactoriamente el curso aumentó desde un 33% a un 66% tras la introducción el segundo año de un contrato de aprendizaje y un seguimiento semanal. Se obtuvieron ganancias variables en los postests de inglés con máximos sobre el 20% y dependientes del tipo de tarea. Hubo una alta reincidencia de errores en pruebas personalizadas. Se observó relación entre la distribución temporal del esfuerzo y el rendimiento. Se sugiere la utilización de modelos mixtos con muestras mayores. En la evaluación se expresó dificultad pero satisfacción por el trabajo realizado, el ambiente estimulante de la clase e interés por una mayor oferta de cursos en inglés. El control de la complejidad de las tareas, el entrenamiento sostenido y la articulación de un mayor número de refuerzos como créditos de idioma de libre configuración o puntos para becas erasmus pueden ser útiles en reducir el abandono y mejorar los resultados.

El conocimiento de lenguas extranjeras y el aprendizaje intercultural constituyen demandas crecientes para la docencia universitaria en el contexto europeo. Tanto para estudiantes propios como extranjeros los idiomas facilitan tanto la realización de estancias en otros países como un ejercicio profesional satisfactorio. Como respuesta, hay un interés creciente en el fomento del inglés como lengua académica en distintas titulaciones (Fortanet, 2008).

El objetivo del curso es el aprendizaje de la metodología de elaboración de un trabajo de investigación personal en ciencias humanas. Se aprende a producir conocimientos propios mediante destrezas de innovación con escasez y demanda creciente en la sociedad actual. Se trata así de un curso práctico en donde el estudiante plantea un pequeño proyecto y lo ejecuta en todas sus fases: identificación del problema, planteamiento de la hipótesis y de un diseño de recogida de datos, análisis, y comunicación de resultados de forma oral y escrita según normas de publicación. El trabajo con contenidos en inglés es básico puesto que la mayoría de fuentes documentales están en esta lengua.

El objetivo del estudio fue en primer lugar desarrollar contextos bilingües en castellano e inglés para la asignatura. Esto se realizó en mediante (a) la elaboración y organización en red de materiales y tareas en ambos idiomas, y (b) la integración de la docencia con los departamentos de estudios ingleses y traducción ya que el aprendizaje de habilidades de lengua extranjera constituye un objetivo distinto a los propios de psicología. En segundo lugar se evaluó los efectos del contexto bilingüe sobre (a) las ganancias en inglés específico para psicología y la asignatura, (b) el rendimiento en psicología, y (c) la satisfacción y dificultad percibida por el alumno.

## ESTUDIO 1

### Método

#### *Participantes*

Tres estudiantes voluntarios de la asignatura diseños experimentales durante el curso 2006-2007. Ninguno de ellos era extranjero. Como incentivo podían obtener hasta dos puntos adicionales en la asignatura *Inglés para Psicólogos*.

#### *Materiales Docentes y Medidas*

Se elaboró los siguientes materiales en ambas lenguas: programa de la asignatura, diapositivas, prácticas de diseño y análisis de datos y guías para la redacción del proyecto tanto en contenidos como en formatos. La bibliografía recomendada en clase estaba también en ambos idiomas.

Como medidas de inglés general en el pretest y postest se utilizó dos versiones del *Quick Placement Test* (University of Cambridge Local Examination Syndicate, 2001). Como medidas de inglés específico para psicología se empleó a) dos pruebas análogas de comprensión y expresión escrita con textos de contenido clínico, y b) dos pruebas análogas de vocabulario específico de la asignatura extraídas del glosario.

Adicionalmente, se administró junto con el postest pruebas a medida que contenían diez errores cometidos por el estudiante durante el semestre. Se pretendía así evaluar el grado de reincidencia y el efecto de las correcciones sobre el aprendizaje.

Se elaboró también un cuestionario sobre la satisfacción y la dificultad experimentada por el estudiante en una escala de 7 puntos. Este incluía preguntas abiertas.

#### *Procedimiento*

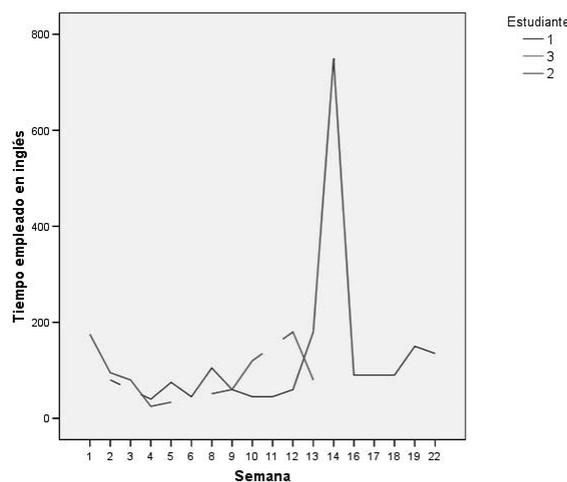
Al comienzo del curso se dio información sobre incentivos y se realizaron los tres pretests. Se les informó sobre la posibilidad de elegir el inglés de forma optativa en cualquiera de las actividades con comunicación oral y escrita durante el curso. Las tareas eran enviadas por correo electrónico a una profesora del departamento de inglés quien hacía las correcciones de idioma antes de que se evaluaran los contenidos de psicología.

Dado el carácter exploratorio de la clase bilingüe y el nivel de los estudiantes, la teoría se impartió en castellano durante la primera parte del semestre, y en inglés durante la segunda parte. Todo el material de transparencias, lectura de apuntes, y prácticas se realizó en inglés. Se instruyó a los participantes a que contabilizaran mediante un cuestionario el tiempo que invertían en el trabajo en lengua inglesa para tres tipos de actividades: teoría, prácticas de aula y proyecto. Al final del curso se realizaron los tres postests. La organización del estudio siguió así la de un diseño pretest-postest sin grupo de control. Finalmente contestaron el cuestionario de evaluación.

## Resultados

### *Indicadores de Motivación*

La media semanal de dedicación al trabajo en lengua inglesa durante el período de clases fue de 63 minutos, con una desviación típica de 42,65 minutos. La continuidad media del trabajo en inglés fue de 6,67, definida como el número de semanas en que el estudiante informa haber realizado tareas. Una descripción gráfica individualizada de este esfuerzo se encuentra en la Figura 1 en donde aparecen tres períodos claramente diferenciados: a) semanas 1 y 13, b) semana 14 anterior al examen, y c) tiempo posterior al examen, dedicado fundamentalmente a la elaboración del trabajo. Sólo un estudiante se presentó al examen y completó el trabajo, con tiempos medios de dedicación ( $M_D = 106.1$  minutos) y continuidad de 11 semanas mayores que el segundo ( $M_D = 62.7$  m. y Continuidad=8 semanas) o el tercer estudiante ( $M_D = 20.7$  m. y Continuidad=1 semana). Un análisis de regresión polinómica con modelo autorregresivo sobre el tiempo dedicado al inglés durante el primer período reveló que el segundo estudiante que no terminó mostraba una continuidad irregular con un componente cuadrático significativo  $\hat{\beta} = 4.37$ ,  $t(1) = 3.96$ ,  $p < .008$ , presentando una curvatura en forma de U con tiempos mínimos sobre la mitad del semestre (SAS Institute, 2007; Ato y Vallejo, 2007). El primer estudiante no mostró tendencias irregulares y para el tercero no se contó con una secuencia de datos suficiente para el análisis.



*Figura 1. Tiempos individuales empleados en el trabajo en inglés*  
 Nota. El tercer estudiante no aparece representado como trazo puesto que sólo se dispone de un único registro semanal

### *Rendimiento en Inglés*

Los resultados en las tres pruebas administradas se muestran en la Figura 2 con datos un solo estudiante. En éste se observaron ganancias del 10 y 20 % en las pruebas de inglés específico para psicología pero un ligero retroceso del 4% en la prueba de inglés general. Tomando como referencia la variabilidad en los pretests esto corresponde a efectos de .40, .40 y .10 desviaciones típicas respectivamente. Los dos primeros efectos son cercanos al tamaño medio y el tercero muy bajo según el criterio de Cohen (Olejnik & Algina, 2000). En la prueba individualizada con errores propios se produjo una reincidencia de un 80% respecto a la ejecución inicial en prácticas y trabajo.

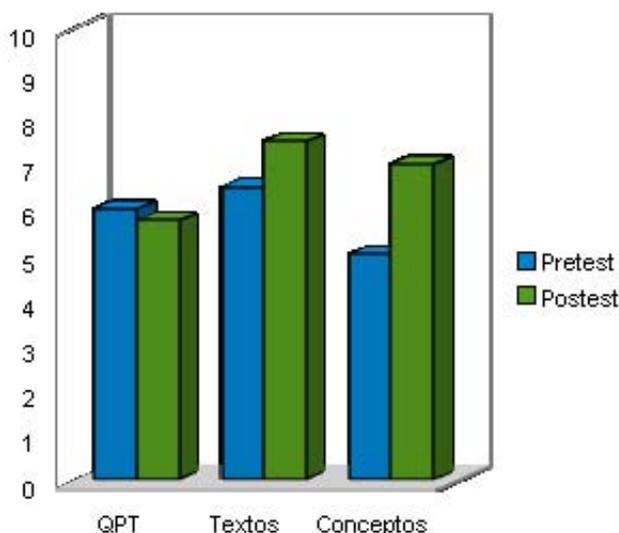


Figura 2. Rendimiento en las medidas de inglés durante el curso 2006-2007

### *Rendimiento en Psicología*

Las notas de práctica de aula fueron de un 97,3% sobre el total para el primer estudiante, un 60% para el segundo, y un 25% para el tercero. El primer estudiante obtuvo una nota en el examen de un 88% y en el proyecto de un 95% con una calificación global ponderada de 9.2. El segundo y tercer estudiante no se presentaron al examen ni entregaron el proyecto final.

### Discusión

La alta tasa de abandono de dos tercios se ha visto asociada a la irregularidad del tiempo de dedicación semanal al inglés. Esto puede ser debido a que se hayan visto desbordados por la dificultad combinada de comprender los contenidos de la clase y realizar el proyecto en una lengua no materna que no conocen bien. Para prevenirlo se ensayó en el siguiente curso el control de la dificultad de las tareas mediante un contrato de aprendizaje. Se intentó así adaptar el uso del inglés a la motivación y niveles del estudiante tanto en inglés como en psicología al ser la asignatura de libre configuración. Se contrastó también un seguimiento semanal para estimular la motivación y continuidad. Se incluyó además tareas y entrenamiento en expresión y comprensión oral.

## ESTUDIO 2

### Método

#### *Participantes*

Tres estudiantes de la clase diseños experimentales durante el curso 2007-2008 con los mismos incentivos que el curso anterior. Ninguno de ellos era extranjero.

### *Materiales Docentes y Medidas*

Se utilizó los mismos materiales que el curso anterior sustituyendo la tarea de comprensión y expresión de textos con contenido clínico. En su lugar se elaboró dos tareas de comprensión y expresión escrita y dos tareas de comprensión y expresión oral con resúmenes de artículos de revista.

### *Procedimiento*

Fue análogo al del curso anterior. La complejidad de las tareas se controló mediante un contrato de aprendizaje individualizado. Las propuestas para el proyecto fueron esta vez expuestas oralmente en inglés y en formato *powerpoint* a toda la clase con el fin de fomentar el debate y la crítica cooperativa. Estas fueron grabadas digitalmente, corregidas por la profesora posteriormente en presencia del alumno, y vueltas a grabar hasta que fueron correctas. El seguimiento se realizó mediante información verbal sobre la participación y el rendimiento durante el curso.

## Resultados

### *Indicadores de motivación*

La media semanal de dedicación al trabajo en lengua inglesa durante el período de clases fue de 65 minutos con una desviación típica de 39.37 minutos. La continuidad media fue de 8,67 semanas. Los tiempos individuales en dedicación fueron de 111.36, 52.27 y 31.36 minutos y la continuidad de 11, 10 y 5 semanas para los tres estudiantes respectivamente. Hubo omisiones en la realización de postests, por lo que los resultados de mejora están basados en un solo estudiante diferente para cada tipo de prueba.

### *Rendimiento en inglés*

Los resultados en las pruebas pueden consultarse en la Figura 3. En el *QPA* se produjeron mejoras del 3.6%, en la prueba de conceptos una mejora promedio de 17.5% y en la de resúmenes un retroceso del 8.25%. En la prueba individualizada con errores propios se produjo una reincidencia de un 80% respecto a la ejecución inicial en prácticas y trabajo.

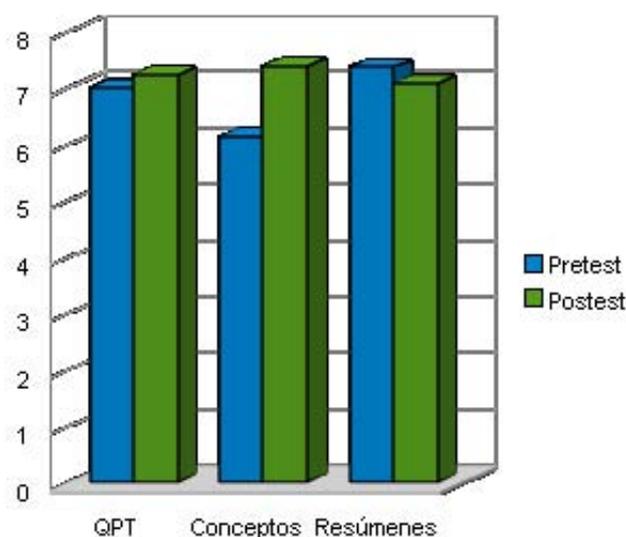


Figura 3. Rendimiento en las medidas de inglés durante el curso 2007-2008

## *Rendimiento en psicología*

Las notas de práctica de aula fueron de un 53% sobre el total para el primer estudiante, un 74.8% para el segundo, y un 30.4% para el tercero. El primer estudiante obtuvo una nota en el examen de un 63%, el segundo de un 52% y el tercero del 14%. Las notas globales ponderadas fueron del 56.2%, 63.4% y 29.2% respectivamente.

## Conclusión

El control de la complejidad de las tareas mediante el contrato de aprendizaje y el seguimiento han sido útiles: a) en la reducción del abandono que ha pasado de dos tercios a un tercio y b) en el aumento de la continuidad semanal media del esfuerzo en un treinta por ciento. En ambos cursos se ha producido un cumplimiento satisfactorio cuando la dedicación al inglés ha sido mayor de 50 minutos y la continuidad mayor de 10 semanas. Al ser resultados descriptivos y con valores faltantes hay que interpretarlos con cautela porque pueden contener sesgo. Sería de interés utilizar muestras mayores y procedimientos de inferencia que contrastaran la relación entre el rendimiento y los niveles y distribución temporal del esfuerzo.

El rendimiento en inglés ha variado en función de la tarea. Las pruebas de conceptos de la asignatura han sido donde se han replicado mejoras de rendimiento entre un quince y un veinte por ciento en ambos estudios. Las pruebas de resúmenes no han mostrado en cambio esas mejoras. Aunque el retroceso en rendimiento sea posiblemente debido a sesgos metodológicos al realizar en un caso los postests por teléfono, en los otros casos tampoco hubo mejoras. Aunque la expresión y comprensión verbal de resúmenes sea una habilidad básica de la asignatura, es posible que el lenguaje utilizado sea muy heterogéneo en las diversas áreas de especialidad en psicología. Las tareas de inglés general tampoco mostraron mejoras en promedio, probablemente porque no han recibido durante el curso entrenamiento en habilidades básicas como la gramática. Asimismo, la alta reincidencia en errores del 80% se ha replicado en ambos estudios.

Como estrategias de mejora podría ser útil acotar con más detalle las competencias lingüísticas asociadas a los diferentes contenidos de la asignatura a fin de definir y escalar mejor los objetivos de aprendizaje. Otras posibilidades serían introducir controles por temas que motivaran un esfuerzo más continuado y la utilización de un entrenamiento individualizado mediante tutoriales informatizados.

En la evaluación de ambos estudios se ha replicado la valoración positiva de la corrección coordinada, la satisfacción por el trabajo realizado, la demanda de más cursos de inglés y más práctica oral en clase así como la satisfacción global buena. Como aspectos negativos se replicó la dificultad percibida de las tareas. En moderado desacuerdo entre ambos estudios estuvo el nivel de flexibilidad en el uso del idioma y la preferencia de un curso bilingüe frente a uno monolingüe en inglés. Convendría conocer sus opiniones con más detalle en este tema. Una posibilidad a contrastar es ofertar una clase mayoritariamente en inglés con flexibilidad y disponibilidad de materiales y recursos. El estudiante podría así recibir recompensa acorde a su nivel de esfuerzo sin comprometer los objetivos de aprendizaje de la asignatura. Para incentivar la participación se podría articular un mayor número de refuerzos como créditos de idioma de libre configuración, puntos para becas erasmus o la obtención de la aptitud en inglés para psicología en el título europeo de grado.

Agradecimientos. Se agradece a la Unitat de Suport Educatiu de la Universitat Jaume I la concesión de dos proyectos de innovación educativa para la realización de este trabajo.

## Referencias

- Ato, M. y Vallejo, G. (2007). *Diseños experimentales en psicología*. Madrid: Pirámide.
- Fortanet, I. (2008). *Hablar inglés en la universidad: Docencia e investigación*. Oviedo: Septem Ediciones.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. & Li, W. (2005). *Applied linear statistical models* (5th ed.). London: Irwin.
- Olejnik, S. & Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations and limitations. *Contemporary educational psychology*, 25, 241-286.
- SAS Institute (2007). The autoreg procedure. In *Sas Online Documentation Version 9.1.3*. Acceso el 1 de Septiembre de 2007 de la pagina <http://support.sas.com/onlinedoc/913/docMainpage.jsp>
- University of Cambridge Local Examination Syndicate (2001). *Quick placement test* (v. 1 y 2). Oxford: Oxford University Press.

# APLICACIÓN DE LA METODOLOGÍA DE LA ENCUESTA PARA DETERMINAR LA ACTITUD DEL PROFESOR UNIVERSITARIO ANTE EL USO DEL ORDENADOR. CASO: UNIVERSIDADES PÚBLICAS VENEZOLANAS

Hugo Padrón<sup>1</sup> y Marcos Ruiz<sup>2</sup>

<sup>1</sup>Universidad Centroccidental Lisandro Alvarado, Venezuela

<sup>2</sup>Universidad de Málaga

El objetivo de esta investigación fue aplicar la metodología de la encuesta para determinar la actitud del profesor universitario ante el uso del ordenador, en las universidades venezolanas. Para la elaboración de la encuesta se utilizó una escala de Likert con 45 ítems y cuatro opciones de respuesta referidas a las tres dimensiones de toda actitud: cognoscitiva, emocional y conductual. Para seleccionar la muestra se aplicó un muestreo estratificado y proporcional, según la cantidad de profesores en las universidades participantes. Se calculó el tamaño de la muestra para un error máximo entre el 3% y el 4%, para una población total de 8.571 profesores; 563 personas eran suficientes, no obstante se incrementó a 756 aumentando el tamaño de las submuestras de las universidades con menos profesorado, para asegurar la representatividad de los estratos menores. El estudio evidencia que el profesor universitario asume diversas actitudes respecto al uso del ordenador, dependiendo de ciertos factores presentes en el proceso.

El desarrollo vertiginoso que han tenido las TIC está presente en todos los ámbitos del quehacer humano. Sin embargo este desarrollo tecnológico, por diversas razones no se ha dado a la misma velocidad en el ámbito educacional, incluyendo el ambiente universitario.

Cuando se habla del tema, las universidades reconocen la importancia del desarrollo tecnológico, además están conscientes de las exigencias del sector empresarial respecto a los profesionales universitarios egresados de los Centros de Educación Superior, sin embargo no se toman decisiones acordes con la necesidad de respuesta esperada por las empresas que esperan a los egresados.

Las universidades de los países desarrollados iniciaron hace varios años proyectos que han facilitado la introducción de las TIC en el campo educativo y, a través de planes estratégicos, han podido experimentar, establecer y evaluar proyectos de largo alcance que finalmente se han fortalecido en el tiempo. En Venezuela todas las universidades, tanto públicas como privadas, han hecho un gran esfuerzo para introducir las TIC en sus respectivos ambientes. El gobierno, a través del Ministerio de Ciencia y Tecnología, dictó el Decreto 825, según el cual se señalan los objetivos y la normativa relacionada con el desarrollo tecnológico, sobre todo para difundir su uso en las instituciones de educación superior, liceos y escuelas. Pero, a pesar de existir interés por la integración de las TIC en el proceso de enseñanza-aprendizaje, podemos afirmar que a veces no existe una mentalidad o actitud de abierta aceptación de las TIC (Power, 2002).

Para llevar adelante estos proyectos es indispensable que exista una actitud de aceptación, una alfabetización informática que, a su vez, exige un cambio de mentalidad y la disposición a asumir los retos que implican estos cambios; de ello dependerá el éxito o el fracaso de los planes y proyectos (Tomas et al., 1999; Aparici, 2000).

Los alumnos de nuevo ingreso son hijos de las TIC, razón por la cual los profesores deben formarse y familiarizarse con las TIC para poder atender y motivar en el proceso de aprendizaje y dar respuestas adecuadas a las exigencias actuales (Diesbach, 2002).

### *La investigación*

Nuestra realidad que se debate entre el ser y el deber ser, entre lo que se tiene y lo que se quisiera tener, nos ha llevado a realizar esta investigación, a fin de conocer la actitud del profesor universitario ante el uso del ordenador.

### *Objetivos*

1. Caracterizar al profesor universitario en sus características personales, académicas y relacionadas con el uso del ordenador.
2. Determinar la actitud del profesor universitario ante el uso del ordenador para fines docentes.
3. Conocer la necesidad de formación tecnológica de los profesores universitarios referentes al uso del ordenador.

### *La muestra*

Para que la muestra fuese representativa y los resultados pudiesen generalizarse al *Sistema de Educación Superior* venezolano, se optó por escoger una muestra de siete universidades de las diecisiete universidades públicas existentes en Venezuela. Las universidades seleccionadas están ubicadas en tres regiones: Central (Región Capital), Centroccidental y Región de Los Llanos. Para seleccionar la muestra se aplicó un muestreo estratificado y proporcional, según la cantidad de profesores en las universidades participantes y, a su vez, dentro de éstas por departamentos. Para ello se dividió la población en estratos (ver Tabla 1).

*Tabla 1. Distribución de profesores por universidad (población y muestra)*

<i>Universidad</i>	<i>CARRERA</i>			<i>Total</i>	<i>POBLACIÓN</i>		<i>MUESTRA</i>	
	<i>Científ.</i>	<i>Humanist.</i>	<i>Técni</i>		<i>Profesoras</i>	<i>Profesores</i>	<i>Profesoras</i>	<i>Profesores</i>
UCV	✓	✓	✓	5951	2573	3378	173	227
UPEL	✓	✓	✓	236	148	88	63	37
USB	✓	✓	✓	666	256	410	20	30
UCLA	✓	X	✓	1198	566	632	43	47
UNEXPO	✓	X	✓	252	87	165	18	32
UNELLEZ	X	X	✓	174	66	108	16	24
UNEY	X	X	✓	94	28	66	9	17
<b>Total</b>				<b>8571</b>	<b>3724</b>	<b>4847</b>	<b>342</b>	<b>414</b>

Nota: las siglas corresponden a las siguientes universidades

UCV: Universidad Central de Venezuela

UPEL: Universidad Pedagógica Experimental Libertador

USB: Universidad Simón Bolívar

UCLA: Universidad Centroccidental Lisandro Alvarado

UNEXPO: Universidad Experimental Politécnico

UNELLEZ: Universidad Experimental de Los Llanos Ezequiel Zamora

UNEY: Universidad Experimental del Yacuy

Se calculó el tamaño de la muestra para un error máximo entre el 3% y el 4%. Para una población total de 8.571 profesores era suficiente una muestra de 563 personas y para un error del 3% se hubieran necesitado 954 personas.

El aumento de 563 a los 756 de la muestra definitiva se obtuvo aumentando el tamaño de las submuestras de las universidades con menos profesorado (práctica habitual para asegurar la representatividad de los estratos menores).

### *Determinación de las dimensiones de la escala*

Para la elaboración de la escala nos basamos en el *Cuestionario de actitudes de los profesores hacia las computadoras* de Knezek y Christensen (1994), por ser un cuestionario utilizado en diversas investigaciones en los Estados Unidos y en México. Se elaboró una escala Likert con 45 ítems de cuatro puntos y se tuvieron en cuenta las tres dimensiones de toda actitud: cognoscitiva, emocional y conductual.

### *Los ítems*

Se elaboró una escala Likert con 45 ítems de tres dimensiones de toda actitud: cognoscitiva, emocional y conductual. Se formularon 45 ítems enunciados como afirmaciones (positivas o negativas) ante las cuales se debía escoger una opción única que concordase con el sentir de la persona en uno de estos cuatro valores: (1) Totalmente en desacuerdo (TD); (2) En desacuerdo (D); (3) De acuerdo (A); (4) Totalmente de acuerdo (TA). Se procedió a redactar el conjunto de los ítems de manera que cada una de las afirmaciones se refiriese a actitudes positivas o negativas del profesor universitario hacia el uso del ordenador.

### *Análisis*

Se realizaron cuatro estudios. Estudios 1: Análisis descriptivo del perfil del profesor. Estudio 2 Análisis Factorial de la Escala de Actitudes. Estudio 3: Aplicación de la regresión logística para la predicción de actitudes y Estudio 4: Análisis de conglomerados para buscar una tipología de profesores. Para efectos de esta comunicación se hará referencia a los Estudio 3 y 4: Aplicación de la regresión logística para la predicción de actitudes y Análisis de conglomerados, respectivamente. Se trata de detectar aquellas variables que permiten predecir las actitudes de los profesores respecto al uso docente del ordenador. Estos predictores se señalan en la Tabla 2.

Tabla 2. Predictores iniciales de la actitud de los docentes

<i>Variables</i>	<i>Categorías</i>
Sexo	(1) Masculino (2) Femenino
Edad	(3) 28 años (4) 33 (5) 38 (6) 43 (7) 48 (8) 53 (9) Más de 53
Años de docencia	(1) Instructor (4 años) (2) Asistente (4 años) (3) Agregado (4 años) (4) Asociado (4 años) (5) Titular (5 años)
Asignatura	(1) Científica (2) Humanística (3) Técnica
Tiempo usando el ordenador	(1) 10 (2) 15 (3) 20 (4) 25
Formación	(1) Autodidacta (2) Universidad (3) Academia
¿Tiene ordenador en casa?	(1) Sí (2) No
Lugar de uso del ordenador	(1) Sólo en casa (2) Sólo en la universidad (3) En casa, en la universidad y ciber

### *Variables relevantes en el análisis*

Sexo (masculino, femenino);

¿Qué asignatura imparte? (científica, humanística, técnica);

¿Tiene ordenador en casa? (sí, no);

Lugar donde usa el ordenador (sólo en casa, sólo en la universidad, en casa y en la universidad, en casa en la universidad y ciber);

Tipo de formación (alguna, ninguna, curso de hoja de cálculo, elaboración de presentaciones, otros cursos);

Universidad a la que pertenece (UPEL, UNEXPO, USB, UNELLEZ, UCLA, UNEY, UCV);

Lugar donde recibió la formación (autodidacta, cursos en la universidad, cursos en academias, autodidacta y universidad, autodidacta y academias, academias y universidad).

### *Análisis de conglomerados para buscar una tipología de profesores*

Se aplicó el análisis de clusters mediante el procedimiento de análisis de conglomerados en dos fases, con el objeto de generar grupos basados tanto en variables categóricas como continuas. La distribución

de conglomerados se realizó teniendo en cuenta el puntaje general de las respuestas dadas a los ítems, de modo que una mayor puntuación indicaba una actitud más positiva hacia el ordenador y una menor puntuación indicaba una actitud más negativa hacia el ordenador. (Ver Tabla 3)

*Tabla 3. Clasificación de los profesores según el uso que hacen del ordenador*

TIPO	SEXO	ASIGNAT.	ORD. CASA	LUGAR USO CASA - UNIV	FORMAC.
	M 48%	CIENT. 22%	94% tiene		
TECNÓFILO		HUM. 26%	45% no	48%	10% y 13%
	H 55%	TEC. 52%			
	M 46%	CIENT. 24%			
TECNOMODERADO.		HUM. 27%		55%	17% y 27%
	H 55%	TEC. 22%			
	M 52%	CIENT. 34%			
TECNÓFOBO		HUM. 38%	15	46%	4% y 9%
	H 48%	TEC. 38%			

### Conclusión

En esta investigación, sobre las actitudes de los profesores universitarios respecto al uso del ordenador en Venezuela, se seleccionó una muestra representativa que permitiera proyectar los resultados al conglomerado de las universidades venezolanas. En este sentido, se han presentado algunos resultados que confirman, en todo o en parte, hallazgos anteriores. Los aspectos más relevantes que ofrecen los resultados de este estudio tienen que ver con la actitud global de los profesores universitarios frente al uso del ordenador y de las TIC, en el ámbito de la enseñanza-aprendizaje, con el significado de las tecnologías en el ámbito universitario y con la necesidad de formación del profesor en esta era tecnológica. Después de comparar conceptos, analizar datos y relacionar resultados, podemos formular las conclusiones siguientes:

En general, se observa que la actitud de los profesores hacia el ordenador es positiva y, además, se aprecia la importancia de esta herramienta en la docencia; Agualeles (1995) considera esta actitud positiva como primer y principal elemento para facilitar la autorrenovación. El grupo de profesores con actitud positiva alcanza el 92%, aunque en la práctica no la usen sistemáticamente.

Teniendo en cuenta la actitud de los profesores con respecto al uso del ordenador y de las TIC, se han establecido tres grupos que, según Aparici (2001), definen el desempeño en esta área: los tecnófobos (quienes están lejos de las tecnologías, no se sienten a gusto al utilizarlas y se expresan con muchos recelos respecto a ellas); los tecnomoderados (quienes conocen las tecnologías y las usan adecuadamente, sin entusiasmarse excesivamente por ellas) y los tecnófilos (quienes conocen las tecnologías, las usan permanentemente y están al día en lo último de las tecnologías tratando de resolver casi todos los problemas mediante las mismas).

También constatamos que la edad es un factor importante en el uso del ordenador; los profesores de mayor edad presentan cierta reticencia al cambio y utilizan poco el ordenador en su práctica docente; coincide esta conclusión con el estudio de Christense (1999) y Meredirh et al. (1999).

Finalmente se considera que la implantación de las TIC en el ambiente de aprendizaje suele llevar consigo dificultades e inconvenientes los cuales son perfectamente superables cuando se busca una estrategia que permita integrar intereses pedagógicos, tecnológicos y profesionales. Esta tríada de intereses son avalados por Karsenti, Savoie y Larose (2001), cuando señalan la importancia de programar la formación de los profesores universitarios.

### Referencias

- Aguareles, M. (1995). Análisis de las posibilidades educativas en internet: las conferencias mundiales de Pangea para los curricula universitarios de ciencias de la educación. *Edutec 95*. <http://www.uib.es/depart/gte/aguarca.html> (Consulta: junio 2005).
- Aparici, R. (2001). *Mitos de la Educación a Distancia y de las nuevas tecnologías* <http://www.uned.es/ntedu/espanol/temas.dedebate/mitos/nuevastecnos.htm> (Consulta: 18 julio 2006).
- Christensen, R. y Knezek (1999). *Construct validity toward computer. A metanalysis*. 10th International Conference of the Society for Information Technology and teacher Education USA, Sn. AntonioTexas.
- Diesbach, N. (2002). *Los retos de la educación en el amanecer del nuevo milenio*. Vitoria-Gasteiz, España: La Llave.
- Hernández, A., Espejo, B., González, V. y Gómez, J. (2001). Escalas tipo Likert: ¿Es relevante la alternativa “indiferente”? *Metodología de Encuestas*. Vol. 3, Nº 2, 2001.
- Karsenti, T., Savoie, L. y Larose, F. (2001). Les futurs enseignants confrontés aux TIC : Changements dans l'attitude, la motivation et les pratiques pédagogiques. *Éducation et francophonie*, 29. <http://www.acelf.ca/revue> (Consulta : 15 Febrero 2005).
- Knezek y Chirstensen, (1994). *Computer Attitude Questionnaire*, Texas Center for Educational Technology, University of North Texas.
- Meredyth, D., Russell, N., Blackwood, L., Thomas, J. & Wise. P. (1999). Real Time: Computers, Change and Schooling. National Sample Study of the Information Technology Skills of Australian School Students. Department of Education, Training and Youth Affairs, Canberra. <http://www.dtya.gov.au/schools/publications/realtime.pdf>. (Consulta: Febrero 2000).
- Power, I. (2002) *Senderos hacia la sociedad del conocimiento: imágenes y acciones de políticas públicas en Venezuela*. Foro ciencia y tecnología en la sociedad de la información. [http://www.funredes.org/mistica/castellano/ciberoteca/participantes/docupart/esp\\_doc\\_66/](http://www.funredes.org/mistica/castellano/ciberoteca/participantes/docupart/esp_doc_66/) (Consulta: junio 2006).
- Tomas, M., Feixas, M. y Marqués, P. (1999). *La Universidad ante los retos que plantea la sociedad de la información. El papel de las TIC*. Actas de las Jornadas EDUTEK-99”.

# ¿QUÉ DEBERÍAMOS CAMBIAR EN LA DOCENCIA DE ANÁLISIS DE DATOS?

Marcos Ruiz Soler  
Universidad de Málaga

Los desarrollos acontecidos durante los últimos decenios tanto en los modelos estadísticos como en los programas de ordenador han generado unas circunstancias que deberían, cuando menos, llevarnos a reflexionar sobre cómo podríamos mejorar la transmisión de los conocimientos necesarios para el análisis de los datos empíricos de nuestras investigaciones. En consecuencia, se presentan aquí siete propuestas concretas acerca de los modos en que podríamos aumentar la capacidad de nuestros alumnos para comprender cómo abordar el análisis estadístico de datos de cualquier situación de investigación acorde con las posibilidades de los modelos actuales y sin una exigencia significativamente mayor en cuanto a conocimientos matemáticos. Aunque el modo tradicional de enseñanza basado en la explicación de pruebas estadísticas sigue siendo el dominante y comprensible por razones históricas, no puede omitirse que existen buenas razones para justificar el cambio hacia un enfoque unificado del análisis de datos desde el modelo lineal generalizado capaz de dar respuesta más acorde con la naturaleza de las variables estudiadas.

## *¿Qué problemas tienen nuestros alumnos con la estadística?*

Cualquier profesor que haya impartido alguna asignatura que en sus contenidos incluya el análisis de datos es muy probable que haya comprobado que el proceso de enseñanza-aprendizaje es más dificultoso de lo que inicialmente había supuesto. Aunque en el diagnóstico de esta situación deberían incluirse muy diversos factores (motivacionales, contextuales, etc.), vamos a centrarnos ahora en aquellos que hemos observado repetidamente en la docencia en ciencias del comportamiento, pero que posiblemente sean generalizables a otras áreas de las ciencias sociales. Algunos de los problemas son los siguientes:

- (1) Ausencia de una visión de conjunto de las técnicas disponibles para el análisis de datos
- (2) Dificultades para relacionar los problemas de investigación con los modelos de análisis de datos correspondientes
- (3) Carencia de una visión unificada del proceso de análisis estadístico (la representación mental del alumno parece ser un simple listado incompleto de pruebas para determinadas situaciones)
- (4) Los datos recogidos para realizar trabajos en distintas asignaturas son frecuentemente de naturaleza categórica (elección de categorías o escalas ordinales) y se encuentran con que los modelos de análisis que deberían aplicar no les han sido explicados (por ej., regresión logística o regresión de Poisson)
- (5) Inversión de un tiempo excesivo en la resolución de problemas con calculadora, los cuales no siempre garantizan una comprensión profunda de los conceptos
- (6) Falta de una representación visual explícita de los resultados obtenidos, lo cual dificulta notoriamente la comprensión de conceptos abstractos
- (7) Limitaciones prácticas a la hora de disponer de todos los materiales necesarios para el trabajo personal, por encontrarse en fuentes muy dispersas o de difícil acceso (por ejemplo, casos ilustrativos)

de las situaciones que tienen que resolver, manuales del programa informático utilizado que se ajuste a sus necesidades, etc.).

La gravedad de estos problemas es, desde una perspectiva docente, mayor si se considera que disponemos de elementos que podrían facilitar que fueran subsanados introduciendo ciertos cambios en nuestro modo de transmitir los conocimientos. Es por ello que ofrecemos un conjunto de sugerencias que podrían ayudar a cambiar la situación.

### *Algunas sugerencias prácticas para cambiar nuestra enseñanza*

*Sugerencia 1: Desarrollo desde lo general a lo particular.* En la transmisión de conocimientos es posible seguir dos estrategias principales (Ruiz-Soler y Luna, 1997): (1) partir de un enfoque general que muestre el “mapa” del territorio estudiado (de arriba a abajo o *top-down*) o (2) partir de un enfoque particular que desde las distintas “zonas” llega finalmente a mostrar el mapa completo (de abajo a arriba o *bottom-up*). Aunque cada una tiene sus ventajas e inconvenientes asociados, pensamos que sería muy útil para el alumno disponer ya desde el primer momento de ese panorama general de los modelos de análisis existentes (pese a que no sea posible explicarlos con detalle). Resulta lamentable, por ejemplo, que gran número de alumnos desconozca por completo la amplia gama de modelos de análisis disponibles para datos categóricos. No es de extrañar entonces que se apliquen siempre unos mismos análisis, aun no siendo los adecuados a la naturaleza de las variables, pues “cuando sólo se tiene un martillo, todos los problemas tienen forma de clavo”.

*Sugerencia 2: Más énfasis en la conexión diseño – análisis.* Tradicionalmente se ha explicado la estadística (o el análisis de datos) de modo independiente a las metodologías que emplean tales técnicas analíticas. Así, por un lado existen asignaturas de introducción a la metodología de investigación que exponen tipos de problemas e hipótesis, técnicas de investigación social, etc., pero con poco énfasis en el análisis estadístico, ya que éste se trata en otras asignaturas. Una notable excepción al respecto lo constituye la metodología experimental, ya que desde siempre se han explicado los diseños experimentales conjuntamente con el análisis estadístico correspondiente. Aunque para el profesor la conexión entre diseño y análisis resulta evidente, no siempre es así para el alumno (es por ello que en los nuevos planes de grado de la Universidad de Málaga hemos optado por unir ambos aspectos, de modo que los distintos diseños [experimentales, de encuesta u observacionales] se expliquen conjuntamente con los conceptos estadísticos más frecuentes).

*Sugerencia 3: Empleo del enfoque del modelado estadístico.* La idea de la explicación como construcción de modelos sobre la realidad no es algo nuevo (para una revisión histórica, véase Ruiz-Soler y Pelegrina, 1996). Sin embargo, durante las dos últimas décadas la idea parece haberse consolidado, particularmente al ofrecer un marco de trabajo que unifica la inmensa mayoría de los modelos estadístico-matemáticos. La ecuación fundamental, como es bien sabido, es la siguiente:

$$\text{Datos} = \text{Modelo} + \text{Error}$$

Gracias al esfuerzo pionero de diversos autores (McCullagh y Nelder, 1989, entre otros) se ha logrado integrar todos los modelos en familias con características similares (modelos lineales generalizados, modelos aditivos generalizados, etc.), consiguiendo así una simplicidad conceptual derivada de la unificación de modelos en apariencia distintos, pero similares desde una perspectiva matemática. Una sólida comprensión del modelo de regresión lineal se convierte así en el camino más rápido y seguro para comprender la variedad de modelos con los que tratar de describir/explicar conjuntos de datos. Y aunque inicialmente la entrada puede parecer más costosa, las ventajas que se derivan son notables, pues este enfoque posibilita una enorme flexibilidad para adaptarse a todo tipo de situaciones y una

más rica interpretación de los parámetros del modelo, reflejando la tarea del investigador como una tarea de construcción de modelos sobre la realidad que tiene muy presentes las características de esa realidad (tanto la naturaleza de las variables criterio o dependientes como los tipos de relación posibles entre las variables predictoras o independientes).

*Sugerencia 4: Modelar toda clase de datos categóricos.* En la actualidad, en la que existen muy buenos programas que implementan modelos de análisis que en épocas pasadas eran de difícil acceso, no parece muy justificado el continuar aplicando modelos de análisis que no son los más apropiados a la naturaleza de las variables de respuesta. Esto sucede particularmente –aunque no exclusivamente– en ciencias humanas y sociales, en las que el uso de variables categóricas es muy habitual. Por ello, en nuestra docencia deberían incorporarse aquellos modelos de análisis apropiados para frecuencias, recuentos, así como variables binarias y ordinales, máxime cuando los resultados de la aplicación de unos y otros no resulta equivalente (López-González y Ruiz-Soler, 2002, Ruiz-Soler y Pisano, 2005).

*Sugerencia 5: Realizar menos cálculos estadísticos manualmente.* Todos los profesores estaríamos de acuerdo en que lo importante es que el alumno realmente comprenda los conceptos estadísticos que forman parte de su aprendizaje. Sin embargo, no es infrecuente observar que lo que muchos alumnos aprenden es simplemente ciertos conjuntos de operaciones para resolver determinados tipos de problemas. Es decir, saben realizar ciertas cosas sin entender lo que hacen. Expresado de un modo sintético y metafórico: conocen la sintaxis, pero no la semántica (una situación muy similar a lo que sucedía en la habitación china ideada por J. Searle). El problema radica en el modo de enseñar: parece que el realizar los cálculos manualmente conducirá progresivamente a una mejor comprensión del análisis estadístico (por ejemplo, realizando numerosos ANOVAs). Sin embargo, me parece que esto es un error por dos motivos: (1) las operaciones algebraicas no conducen necesariamente a la comprensión y (2) se pierde una gran cantidad de horas que podrían dedicarse a actividades más fructíferas. Afortunadamente hay autores que han abierto el camino hacia este enfoque más comprensivo, proponiendo ejercicios ingeniosos que llevan a un aprendizaje intuitivo –pero real– de los conceptos estadísticos (son significativos al respecto los de Hurlburt, 1994, denominados “eye-ball estimation/calibration”, cuya traducción libre podría ser la de *Estadística a ojímetro*).

*Sugerencia 6: Emplear más representaciones visuales para ilustrar los conceptos.* Como escuché decir a un profesor (Botella, 1999), lo que deseo es que mis alumnos sean “consumidores responsables en el mercado estadístico”, es decir, que sepan hacer un uso adecuado de los análisis estadísticos que emplean. En este sentido, un elemento de gran ayuda y que se ha desarrollado espectacularmente durante los últimos años es el de las representaciones gráficas (Friendly y Ruiz-Soler, 2000; Ruiz-Soler, Anguera y Wallace, 2002). De ahí, que servirse de esta útil herramienta podría ser de un gran provecho. No se trata tan sólo de ver las medias aritméticas o los porcentajes de distintos grupos en un gráfico o de mostrar la forma de una distribución para ciertas situaciones sino de alcanzar una comprensión auténtica de lo que sucede cuando aparecen patrones más complejos, tales como interacciones entre diversos factores, visualizando los patrones de cambio en la variable de respuesta o dependiente empleando técnicas tan útiles como son los gráficos interactivos (Young, Valero-Mora y Friendly, 2006; Ruiz-Soler, 2009).

*Sugerencia 7: Diseñar entornos integrados de aprendizaje.* Para resultar efectivas las sugerencias anteriores parece preciso disponer de herramientas que faciliten la tarea tanto de docentes como discentes (Ruiz-Soler y Pelegrina, 1997). En este sentido, convendría facilitar al máximo la disponibilidad de todo aquello que precisa el alumno para aprender a analizar e interpretar datos (desde el programa de ordenador con el que analizar los datos hasta la documentación necesaria para los análisis, pasando por todo tipo de ejemplos que abarquen la pluralidad de situaciones reales con las que se puede encontrar). Es por ello que la disponibilidad de entornos webs amigables con ejemplos ilustrativos y documentación abundante bien estructurada es un medio que puede servir a este fin (Ruiz-Soler y Wallace, en prensa).

## Conclusión

Las sugerencias que acabamos de ofrecer son tan sólo algunas ideas sobre ciertos cambios que probablemente mejorarían el modo de presentar el análisis de datos a nuestros alumnos. Algunos de ellos son meramente instrumentales, pero otros suponen un cambio profundo en la concepción de las asignaturas y en los objetivos que éstas deberían perseguir derivados de una reflexión sobre la metodología docente seguida, siempre orientada a mejorar nuestra labor como facilitadores de aprendizajes significativos aplicables a situaciones reales.

## Referencias

- Friendly, M. y Ruiz-Soler, M. (2000). Nuevos procedimientos *gráficos* para datos categóricos: de la representación a la cognición. En E. Oñate, F. García-Sicilia y L. Ramallo (Coords.), *Métodos numéricos en Ciencias Sociales* (pp. 83-96). Barcelona: CIMNE.
- Hurlburt, R.T. (1994). *Comprehending Behavioral Statistics*. California: Brooks/Cole.
- López-González, E. y Ruiz-Soler, M. (2002). Regresión normal vs. regresión ordinal: un análisis comparativo. *Metodología de las Ciencias del Comportamiento*, vol. especial, 355-358.
- McCullagh, P. y Nelder, J.A. (1989). *Generalized Linear Models*. London, Chapman and Hall.
- Ruiz-Soler, M. (2009). *Gráficos interactivos: nuevas herramientas para la interpretación de datos*. Comunicación presentada en el XI Congreso de Metodología de las Ciencias Sociales y de la Salud. Málaga, 15-18 septiembre.
- Ruiz Soler, M., Anguera, M.T. y Wallace, A. (2002). Diagramas de Bangdiwala: representaciones para el acuerdo. *Metodología de las Ciencias del Comportamiento*, vol. especial, 492-495.
- Ruiz-Soler, M. y Luna, R. (1997). *Metodología de la investigación en las ciencias del comportamiento: razones para investigar lo que los metodólogos enseñamos*. Actas del V Congreso de Metodología de las Ciencias Humanas y Sociales (pp. 695-699). Sevilla.
- Ruiz-Soler, M. y Pelegrina, M. (1996). Modelización matemática en Psicología: Un recorrido histórico hasta nuestros días. *Revista de Historia de la Psicología*, 17 (3-4), 317-324.
- Ruiz Soler, M. y Pelegrina, M. (1997). Enseñanza de metodología experimental con nuevas tecnologías: Hacia un entorno informático integrado. En M. Cebrián (Coord.), *Creación de materiales para la innovación educativa con nuevas tecnologías* (pp. 346-349). Málaga: Universidad de Málaga, Instituto de Ciencias de la Educación.
- Ruiz-Soler, M. e Pisano, I. (2005). *Un experimento y tres formas de análisis: Un estudio comparativo de cómo evaluar los aciertos*. Póster presentado en el IX Congreso de Metodología de las Ciencias Sociales y de la Salud. Granada, 14-16 de septiembre.
- Ruiz-Soler, M. y Wallace, A. (en prensa). *Modelado estadístico con R: un entorno web para la docencia*.
- Young, F.W., Valero-Mora, P.M. y Friendly, M. (2006). *Visual Statistics. Seeing Data with Dynamic Interactive Graphics*. New Jersey: John Wiley & Sons.

# MODELADO ESTADÍSTICO CON R: UN ENTORNO WEB PARA LA DOCENCIA

Marcos Ruiz Soler y Agustín Wallace  
Universidad de Málaga

El enfoque del modelado estadístico es una aproximación creciente en las ciencias sociales y del comportamiento, pero en ocasiones se hace difícil su enseñanza dada la escasez de recursos didácticos en nuestro idioma y su adaptación a las necesidades prácticas de ciertas investigaciones psicológicas. Dada esta situación, por una parte, y las crecientes posibilidades ofrecidas por Internet, por otra, consideramos que resultaría de gran utilidad diseñar un entorno integrado de aprendizaje accesible desde Internet. Con este propósito se ha elaborado un entorno CSS (*cascading style sheets*) en el que se integran la mayoría de las operaciones/transformaciones realizables con datos, los diferentes diseños de investigación experimental y sus correspondientes posibilidades de análisis así como diversas clases de representación gráfica. Se intenta superar así algunas dicotomías convencionales en estadística, presentando de una forma simple y eficaz los comandos necesarios en R para realizarlos. La perspectiva del modelado estadístico aporta una visión unificada de todo el proceso que permite una mejor comprensión de la labor del investigador. La Web se complementa con manuales en línea y pdf, enlaces de interés y algunos ejercicios. Su diseño abierto ofrece asimismo la posibilidad de incorporar nuevos materiales ajustables a necesidades docentes específicas.

Desde hace ya años, es posible encontrar obras que presentan la tarea del analista de datos como la tarea de encontrar un modelo estadístico-matemático capaz de ajustarse suficientemente bien a los datos (entre ellos cabe destacar, las de Lunneborg, 1994; Judd y McClelland, 1989 y en el ámbito de las ciencias del comportamiento en nuestro país las del Grupo ModEst, 2000a, 2000b, Ramos, Catena y Trujillo, 2004 y, más recientemente, Ato y Vallejo, 2007). Es bien sabido que esta forma de trabajo se conoce como el *enfoque del modelado estadístico* y se presenta como alternativa al enfoque convencional (también denominado *enfoque del recetario estadístico*); el propósito de este último consiste -simplificando- en encontrar para cada situación la prueba estadística apropiada según la naturaleza de los datos. Sobre las ventajas de este relativamente nuevo enfoque no vamos a entrar ahora, dado que ha sido objeto de tratamiento en otros trabajos (v.g. López-González y Ruiz-Soler, en prensa).

Por otra parte, respecto al software estadístico que el analista de datos puede seleccionar para su tarea existe una amplia gama de posibilidades (SPSS, MNITAB, SYSTAT, ViSta, S-Plus, etc.), aunque dentro del ámbito de las Ciencias Humanas y Sociales, parece que el mayoritariamente utilizado, tanto en la docencia como en la investigación, es el programa SPSS™. No obstante, desde hace algunos años, varios profesores hemos optado por el empleo del programa R, por sus innumerables ventajas frente a los paquetes estadísticos convencionales (para saber más sobre ellas, pueden consultarse Ruiz-Soler y López-González, 2009).

## *De las necesidades discentes a las soluciones docentes*

Convencidos como estamos de las ventajas del enfoque del modelado estadístico, por una parte, y de su implementación mediante el programa R, por otra parte, pensamos que sería conveniente incluir ambos elementos en nuestra docencia. Sin embargo, tras varios años de enseñanza, hemos podido comprobar que su empleo origina ciertos problemas a la hora de realizar los ejercicios prácticos; entre

éstos encontramos: (a) errores menores en la escritura de los comandos, pero que impiden proseguir el ejercicio, (b) dificultades en aplicar los comandos a modelos con más variables que las de los ejemplos simples, (c) ocasional olvido del material de apoyo con las instrucciones correspondientes. Como consecuencia de todo ello se dedica un tiempo excesivo a cuestiones que nada tienen que ver con la comprensión del modelo de análisis sino que se relacionan únicamente con cuestiones –digámoslo así– “sintáctico-procedimentales”, máxime considerando que el tiempo de prácticas es tan sólo de una hora semanal y resulta imprescindible maximizar su aprovechamiento.

Por consiguiente, pensamos que sería conveniente crear un entorno de trabajo que subsanara tales problemas y además posibilitara cubrir nuestros objetivos docentes, por lo que éste debería cumplir una serie de requisitos, a saber:

- (1) Válido para el análisis de datos tanto experimentales como no experimentales
- (2) Con información sobre todos los aspectos relativos a la preparación de los datos (recodificación de valores, selección y transformación de variables, etc.)
- (3) Situado en un nivel intermedio de complejidad, ni demasiado básico ni con tantas opciones que fácilmente lleve al alumno a perderse entre las mismas
- (4) Posibilidad de ampliar los conocimientos adquiridos mediante enlaces a diversos documentos de trabajo disponibles en la red
- (5) Resultar intuitivo en su navegación, con una estructura clara y fácilmente accesible
- (6) Resultar estéticamente agradable, buscando la homogeneidad entre contenidos y formas
- (7) Ampliable con facilidad por parte de otros docentes para adecuarlo a sus necesidades.

### *Un entorno de trabajo para el análisis de datos*

Con objeto de satisfacer todos estos requisitos se diseñó un entorno de trabajo basado en hojas de estilo en cascada (Cascading Style Sheets o CSS). Las CSS constituyen una forma de presentación de un documento escrito en HTML o XML que permite separar los contenidos de su modo de presentación. De esta manera, la información de estilo puede ser adjuntada como un documento independiente o como en el mismo documento HTML.

Las ventajas de usar las hojas de estilo son: (a) rapidez de actualización, ya que podemos cambiar en cualquier momento alguna parte o la totalidad del diseño de nuestras páginas con sólo modificar nuestra hoja de estilo; (b) fácil accesibilidad, ya que el usuario puede configurar su propia hoja de estilo; (c) diversidad de formatos, ya que un página puede disponer de diferentes hojas de estilo según el dispositivo de salida o incluso según la preferencia del usuario (dispositivo móvil, sintetizador de voz, etc.); (d) optimiza el tiempo de presentación, ya que al dividir contenido y apariencia obtenemos archivos más ligeros.

El entorno diseñado se encuentra estructurado siguiendo un nivel de dificultad progresivo. En la página principal se ofrece una condensada introducción al modelado estadístico, contrastando sus diferencias con el enfoque clásico de análisis de datos.

La página está organizada de forma que se pueden copiar tal cual están escritos los comandos para luego pegarlos directamente en el intérprete de comandos de *R*, teniendo que cambiar únicamente el nombre de los archivos de datos.

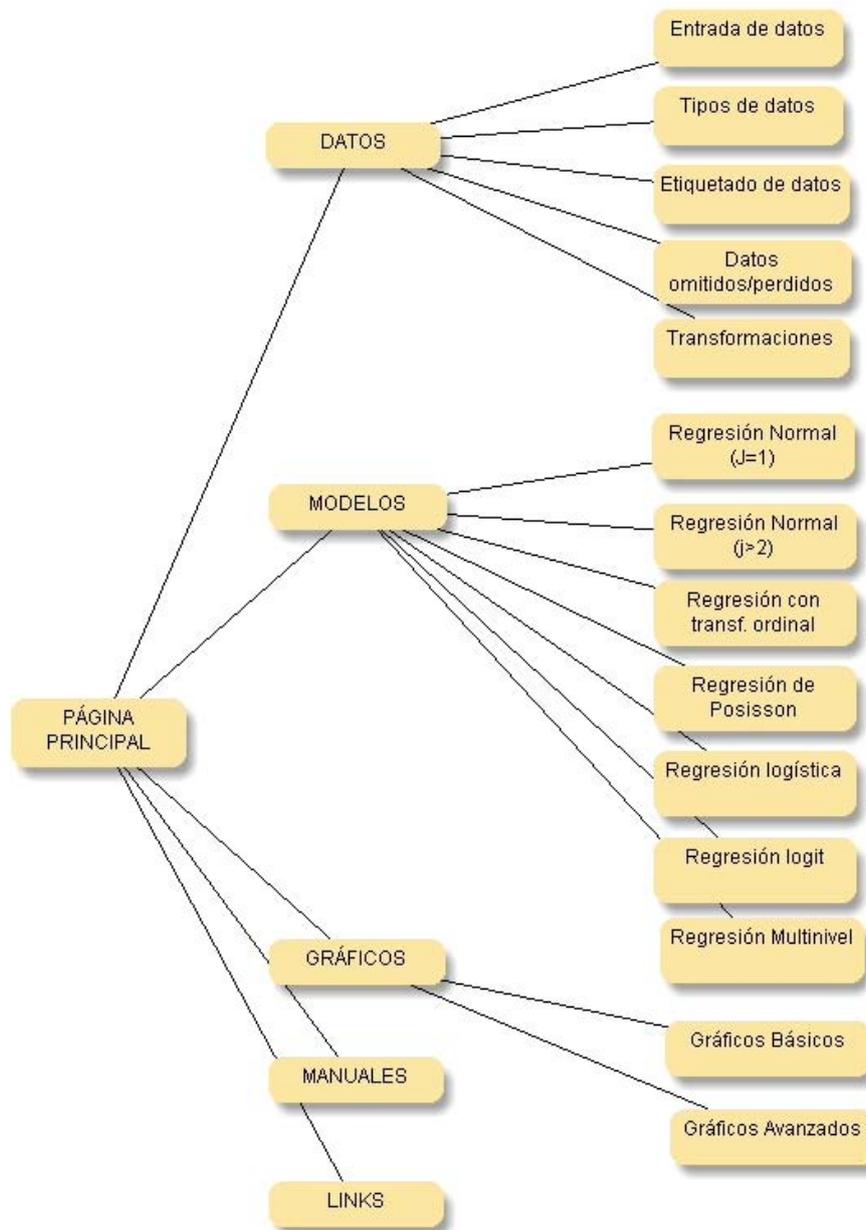


Figura 1. Estructura de las páginas Web de modelado con *R*

Desde la página principal, tenemos acceso a una serie de pestañas que indican las principales secciones de la Web. Al pulsar sobre ellas tendremos una breve descripción de los contenidos que vamos a encontrarnos. Las diferentes secciones (pestañas) son las siguientes: Datos (donde se expone la forma de introducir, transformar, codificar, eliminar y etiquetar los datos), Modelos (donde se presentan los diferentes modelos de regresión, lineal simple, lineal múltiple, con transformación ordinal, de Poisson, logístico, logit y multinivel), Gráficos (donde se muestran gráficos básicos [diagramas de barras, de sectores, tallos y hojas, histogramas y gráficos de cajas] y gráficos avanzados [gráficos lineales, de Trellis, tridimensionales y combinados]), Manuales (donde podemos encontrar documentos para aprender a utilizar *R* y ampliar la información proporcionada por la Web), *Links* (donde encontramos enlaces a la principales páginas que tratan y desarrollan el programa *R*).

Estos diferentes submenús nos permiten navegar por la Web de una manera intuitiva, adentrándonos en cada uno de los aspectos señalados, con el valor añadido de incluir la sintaxis específica para realizar cada uno de los análisis o las modificaciones que se indican.

### *Aportaciones del entorno integrado de aprendizaje*

Atendiendo a la clasificación tripartita de los objetivos docentes (en cognitivos, actitudinales y procedimentales), creemos que esta Web es capaz de satisfacer algunos aspectos de ellos. Desde una perspectiva cognitiva, la Web permite aprender a aplicar modelos de análisis para una gran variedad de datos, algunos de los cuales han sido considerados tradicionalmente como avanzados (por ejemplo, modelos de regresión logística o de regresión de Poisson). Posiblemente ya es hora de incorporar tales modelos desde los niveles iniciales de la docencia, aunque la comprensión sea únicamente la de mero usuario que sabe cuándo puede aplicar un modelo determinado y sabe cómo debe interpretar los resultados del modelo, sin tener que entrar necesariamente en los pormenores analíticos que lo sustentan o los algoritmos para la estimación de los parámetros.

Desde una perspectiva actitudinal, el ofrecer un entorno estéticamente agradable, bien estructurado, sin tecnicismos innecesarios y plenamente ajustado a las necesidades prácticas planteadas en las asignaturas, probablemente genere una predisposición más positiva en el alumno, que ya no concibe la tarea del análisis como un angustioso obstáculo que hay que salvar de cualquier manera.

Desde una perspectiva procedimental, posiblemente es donde se vea la mayor ventaja de una Web docente como ésta, al posibilitar que el alumno concentre todo su esfuerzo intelectual en la comprensión de los modelos, tanto en sus fases de especificación como de interpretación y evaluación, sin tener que perder tiempo en los aspectos procedimentales más estrechamente ligados con la sintaxis del programa. Es cierto que existen interfaces orientados a facilitar también esta labor, como *R-Commander* (Fox, 2005), pero aun tratándose de estupendas iniciativas precisan todavía de depuración, dado que hoy por hoy generan ciertos problemas con frecuencia que pueden hacer perder toda una sesión de trabajo.

Sobre la realidad de estas aportaciones esperamos tener evaluaciones futuras de los propios alumnos en próximos cursos y así poder incorporar aquellos elementos que hagan de esta herramienta un elemento facilitador del aprendizaje de esas asignaturas metodológicas que requieren siempre un continuado esfuerzo para su transmisión didáctica.

### Referencias

- Ato, M. y Vallejo, G. (2007). *Diseños experimentales en Psicología*. Madrid: Pirámide.
- Fox, J. (2005). The R Commander: A Basic-Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14 (9), 1-42.
- Grupo ModEst (2000a). *Análisis de datos. Del contraste de hipótesis al modelado estadístico*. Barcelona: CBS.
- Grupo ModEst (2000b). *Análisis de datos: Modelo Lineal Generalizado*. Barcelona: CBS.
- Judd, C. M. & McClelland, G. H. (1989). *Data analysis: A model comparison approach*. New York: Harcourt Brace Jovanovich.

- Lunneborg, C. E. (1994). *Modeling Experimental and Observational Data*. Belmont, Ca.: Duxbury Press.
- Ramos, M. M., Catena, A. y Trujillo, H. A. (2004). *Manual de métodos y técnicas de investigación en ciencias del comportamiento*. Madrid: Biblioteca Nueva.
- Ruiz-Soler, M. y López-González, E. (2009). El entorno estadístico R: ventajas de su uso en la docencia y la investigación. *Revista Española de Pedagogía*, 67 (243), 255-274.

# INTENCIONES DEL PROFESORADO DE EDUCACIÓN FÍSICA ESPAÑOL HACIA LA SALUD Y LA CALIDAD DE VIDA

Francisco Salinas, M. Teresa Miranda, Armando Cocca, Kamal Mohamed y Jesús Viciano  
Universidad de Granada

El objetivo de esta investigación fue comprobar la intención que los profesores en formación inicial (PTs) de Educación Física (EF) de Educación Secundaria Obligatoria (ESO) de España, tenían hacia el desarrollo curricular de la salud. Se seleccionaron a 1225 PTs de EF de las 22 Facultades de Ciencias de la Actividad Física y el Deporte (FCCAFD) de España. Empleamos un cuestionario tipo Likert (38 ítems, preguntas cerradas), graduado en una escala de 0 (totalmente en desacuerdo) a 100 (totalmente de acuerdo) y validado anteriormente (*alfa de Cronbach* .884, y coeficientes de *Spearman* en repetibilidad de  $p < .001$ ). El análisis de los resultados se llevó a cabo a través del software estadístico SPSS 13.0. Los resultados mostraron que los PTs españoles priorizan el desarrollo de la salud (85.64) sobre el aprendizaje motor (75.19) o la recreación (74.90) en la EF, con diferencias altamente significativas ( $p = .000$ ). La creación de hábitos de práctica física y actitudes positivas (84.87) es la meta prioritaria, existiendo diferencias altamente significativas ( $p = .000$ ) con respecto al aprendizaje de conceptos (73.10) y procedimientos (69.15). Los PTs prefieren los juegos y deportes (79.31), junto a la condición física-salud (78.18) frente a otros contenidos de EF.

En la actualidad, existen diferencias entre lo plasmado en el curriculum oficial y el tratamiento que realmente hace el profesor de EF en sus planificaciones respecto a los diferentes bloques de contenidos (Salinas y Viciano, 2006). A pesar de que el profesorado de EF, independientemente de la etapa de formación en que se encuentre, debe atender en sus planificaciones a todos los bloques de contenidos (Viciano, 2002), éste hace especial énfasis en el bloque de juegos y deportes (Matanin y Collier, 2003; Zabala, Viciano y Lozano, 2002).

Al mismo tiempo, la expresión corporal y las actividades en el medio natural no se encuentran definitivamente insertadas en las clases de EF (Méndez, 1996), ya que son los bloques menos valorados por los profesores (Castejón et al., 2001; Napper-Owen et al., 1999), aunque hoy día están adquiriendo mayor fuerza y protagonismo (Zabala et al., 2001).

Atendiendo a la dimensión referente a los objetivos, los PTs de EF muestran un mayor grado de preferencia por los objetivos encaminados a la salud, seguido de los objetivos orientados al aprendizaje y los objetivos encauzados a la vivencia y la recreación (Xiang, Lowy y McBride, 2002; Corbin, 2002).

En ocasiones para los PTs la vivencia y motivación es un objetivo que debe prevalecer sobre el resto (Viciano et al., 2004; Matanin y Collier, 2003), no siempre se cumple (Zabala, Viciano y Lozano, 2002).

Finalmente, atendiendo a los tipos de aprendizaje, los docentes de EF en formación inicial orientan sus planificaciones en un aprendizaje encaminado a la adquisición de hábitos y valores, por encima del aprendizaje de procedimientos y la adquisición de conceptos, siendo desvalorizado éste último en todas las ocasiones (Viciano et al., 2004; Burrows, Wright y Jungersen-Smith, 2002; Corbin, 2002); aunque no siempre sucede esta tendencia (Zabala, Viciano y Lozano, 2002; Placek, 1984).

## Método

El diseño de la investigación ha sido de tipo sociológico descriptivo, registrando la opinión que una muestra representativa de profesores de EF, de las distintas FCCAFD de España, tenían sobre diferentes aspectos de la salud y calidad de vida.

La muestra fue establecida entre la población total de profesores en formación inicial de todas las FCCAFD de las Universidades españolas (un total de 22) que cursaron las asignaturas correspondientes al “Diseño del currículo”, “Enseñanza”, “Didáctica” o “Planificación de la EF”, y fueron un total de 2504 sujetos. Dicha muestra fue estratificada y proporcional. El procedimiento de muestreo fue de tipo probabilístico, aleatorio simple, siendo el nivel de confianza del 95% y la precisión del 2%, habiendo aplicado la corrección por población finita. El resultado fue una muestra total de 1225 profesores (874 profesores y 351 profesoras), con una edad media de 23.39 años, que supone un 48.9% de la población total. 240 profesores no habían tenido ninguna experiencia docente, mientras que 985 sí habían tenido algún contacto con la docencia en la actividad física fuera del ámbito educativo, por lo que podemos decir que la valoración sobre las opiniones del cuestionario sobre la planificación de la enseñanza de la EF se debió a la formación recibida en las facultades donde se desarrollaba la formación inicial.

El instrumento de recogida de datos fue un cuestionario tipo Likert con una escala nominal de 0 (totalmente en desacuerdo) a 100 (totalmente de acuerdo), validado estadísticamente con una muestra piloto de 150 profesores y diferentes a los participantes en el estudio.

El cuestionario poseía 38 ítems divididos en 13 dimensiones, de las cuales aquí mostramos la dimensión 3 (objetivos del profesor de EF con sus alumnos), dimensión 4 (tipos de aprendizaje) y dimensión 5 (bloques de contenidos de la EF).

El análisis de datos del cuestionario se realizó mediante una estadística descriptiva. Dado que los datos no siguen la distribución normal, aplicamos el *Test no paramétrico de Friedman*, seguido de sus correspondientes comparaciones múltiples dos a dos, teniendo en cuenta la *corrección de Bonferroni*.

## Resultados y Discusión

El bloque de contenidos de las cualidades motrices coordinativas es el más valorado por los profesores de EF en formación inicial de España (80.68 puntos). Prácticamente al mismo nivel se sitúa el bloque de juegos y deportes (79.31 puntos) y condición física-salud (78.18 puntos). La expresión corporal (72.91 puntos) y las actividades en el medio natural (72.19 puntos) son los que menor puntuación reciben, aunque la diferencia con respecto al resto de bloques de contenidos es escasa. En el estudio de Salinas, Miranda y Viciano (2006), realizado con 112 profesores de EF en formación inicial de la FCCAFD de Galicia se obtuvieron resultados prácticamente similares. Por contra, en la investigación de Matanin y Collier (2003), donde se estudió la evolución del pensamiento de tres profesores de EF en formación inicial tras su paso por los cuatro años de Universidad, los docentes mostraron una preferencia por el bloque de juegos y deportes, seguido de la condición física-salud, debido fundamentalmente a que fueron este tipo de actividades las que realizaron cuando se encontraban en el sistema educativo. Igualmente, Zabala, Viciano y Lozano (2002), comprobaron que los docentes destinaban más de 25 sesiones por curso a impartir contenidos de deportes. Según el estudio de Napper-Owen et al. (1999), esta tendencia hacia este tipo de actividades es aún mayor en años anteriores.

A pesar de que cada etapa educativa posee unos contenidos propios y diferentes a Secundaria, las planificaciones de los profesores de EF de Educación Primaria Obligatoria, también conceden una

elevada importancia a los deportes (Faulkner y Reeves, 2000) y la condición física (Castejón et al., 2001), mostrando una actitud muy positiva a la enseñanza de éstos en sus clases (Faulkner y Reeves, 2000). Asimismo, la expresión corporal ocupa nuevamente un lugar poco relevante (Castejón et al., 2001). A pesar de todo, se ha producido una disminución del impacto sportivista existente hasta ahora en el pensamiento de los profesores de EF en formación inicial de España (Vizueté, 2002).

En todos estos resultados los agentes del contexto juegan un papel vital (Matanin y Collier, 2003), así como el dominio que muestren los profesores hacia los bloques de contenidos (Castejón et al., 2001). Además, está demostrado que las actividades que llevan a cabo los profesores en su infancia y adolescencia a su paso por el sistema educativo, inciden directamente en las planificaciones que realizan cuando son adultos (Faulkner y Reeves, 2000; Matanin y Collier, 2003).

Parece ser que hemos pasado de una etapa, en la cual la expresión corporal y las actividades en la naturaleza estaban poco insertadas en los centros educativos (Méndez, 1996), a una época actual en la que este tipo de actividades pueden estar mucho más presentes en las clases de EF en un futuro más próximo. De cualquier manera, es necesario que los profesores Universitarios de España continúen concienciando a los profesores en formación inicial de la importancia y trascendencia que tiene el diseño de unas planificaciones que incluya todo tipo de actividades.

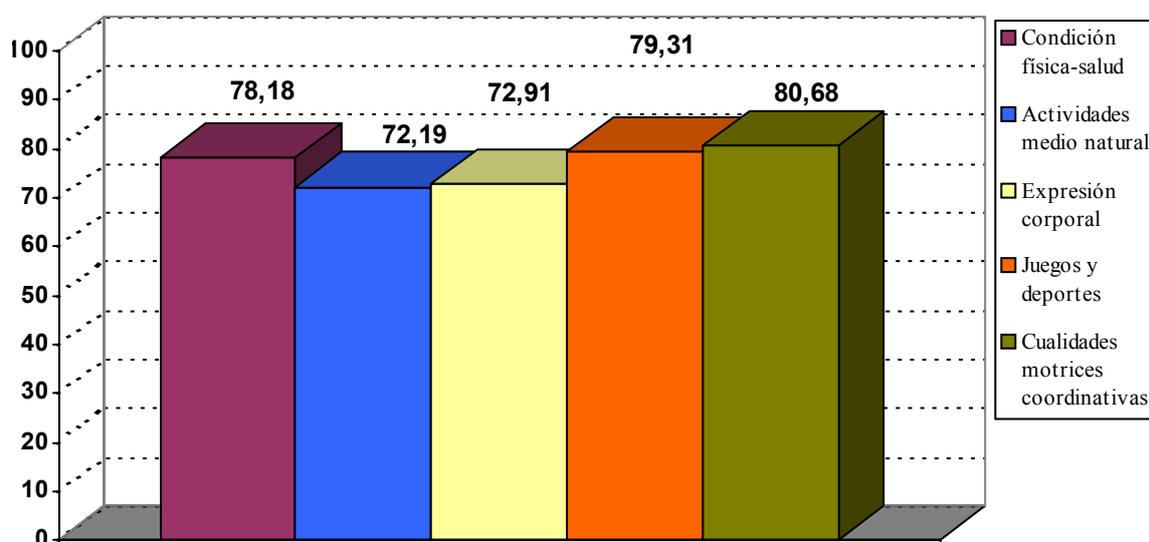


Figura 1. Preferencia hacia los bloques de contenidos de la EF

Atendiendo a los objetivos que los profesores pretenden conseguir con sus alumnos (figura 2), la concienciación de la relación EF-salud ocupa el primer lugar (85.64 puntos), lo que ayudará a disminuir en un futuro el número de abandonos de la práctica física (Pérez y Delgado, 2003) y a aumentar el número de adultos implicados en un estilo de vida saludable, según los resultados de estudios longitudinales (Pieron, 2005; Viciano, 2005). En el estudio de Viciano y Salinas (2005), la concienciación de la relación EF-salud fue el principal objetivo que los profesores de EF en formación inicial pretendieron conseguir con sus alumnos, tanto antes (80.51 puntos), como después de realizar el practicum (85.13 puntos). Además, está constatado que los profesores y los alumnos de otros países, también están familiarizados y concienciados de la relación EF-salud (Burrows, Wright y Jungersen-Smith, 2002).

Para los profesores de España, el aprendizaje de contenidos relacionados con la EF (75.19 puntos) es un objetivo de importancia similar a la vivencia, recreación y satisfacción del alumnado (74.9 puntos). El hecho de los alumnos se lo pasen bien y terminen la clase de EF con un gran sentimiento de

satisfacción, es un objetivo que los profesores de EF en formación inicial han priorizado sobre el resto en varias ocasiones (Matanin y Collier, 2003; Viciano et al., 2004), aunque ello no siempre ha sido así (Viciano y Salinas, 2005).

Los objetivos de vivencia y experimentación, han sido en ocasiones valorados en mayor medida que los de aprendizaje (Viciano et al., 2004). En el estudio de Viciano y Salinas (2005b), el practicum repercutió de forma significativa en la puntuación que los profesores en formación inicial de Granada concedieron a este ítem (72.05 puntos antes del practicum y 61.54 puntos después). El efecto pendular de los procesos (entendiéndose una EF centrada en la motivación, la satisfacción y la vivencia) y los productos (entendiéndose los contenidos y aprendizajes), resaltado por Viciano (2002), se sitúa en la actualidad en un lugar central, como muestran los resultados recogidos de la opinión de los profesores españoles.

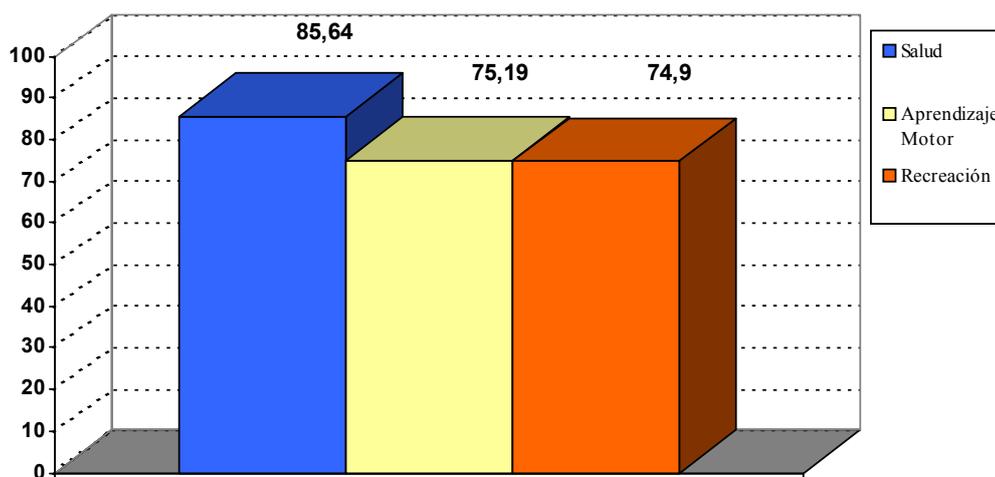


Figura 2. Objetivos de PTs con sus alumnos

Respecto a los tipos de aprendizaje, la adquisición de hábitos de práctica física y actitudes positivas es la meta prioritaria que los PTs de España pretenden conseguir con sus alumnos en sus clases (84.87 puntos). Un lugar intermedio ocupa la adquisición de conceptos relacionados con la EF (73.10 puntos), mientras que la consecución de conductas motrices y habilidades es el ítem menos valorado (69.15 puntos).

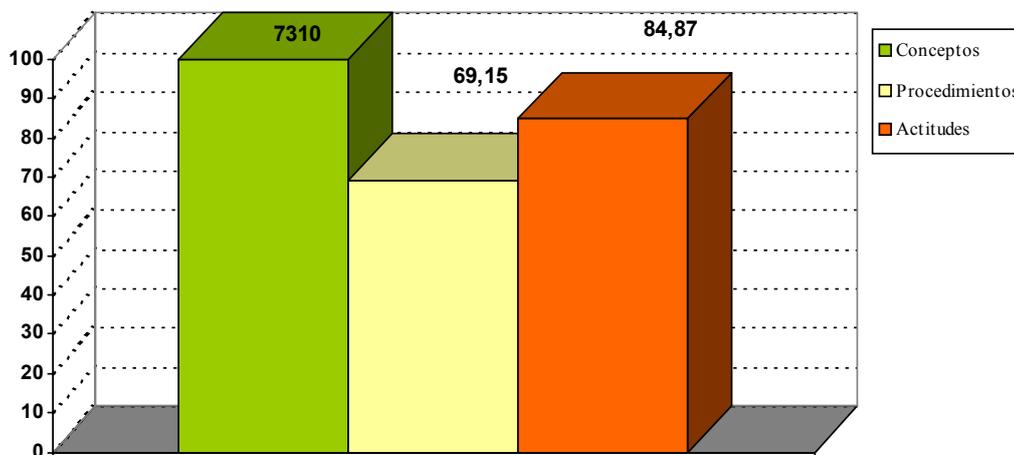


Figura 3. Tipos de Aprendizaje de PTs con sus alumnos

Igualmente, los resultados de las investigaciones precedentes indican que, la creación de hábitos y valores positivos es la preferencia mostrada por los profesores de EF en sus planificaciones

(Burrows, Wright y Jungersen-Smith, 2002; Corbin, 2002). Se considera por tanto, uno de los aspectos que justifican la presencia de la EF en el currículo de la ESO, es decir, que los alumnos desarrollen hábitos y actitudes que les permitan realizar actividad física de manera autónoma durante toda su vida (Vizueté, 2002).

### Conclusión

Los PTs españoles priorizan el desarrollo de la salud sobre el aprendizaje motor o la recreación en la EF. La creación de hábitos de práctica física y actitudes positivas es la meta prioritaria que los PTs de España pretenden lograr con sus alumnos en las clases de EF, lo cual, favorecerá el estado de salud y calidad de vida posterior de los alumnos. Los PTs prefieren los juegos y deportes, junto a la condición física-salud frente a otros contenidos de EF. Podemos decir, que la formación actual de los PTs en España es adecuada desde el punto de vista de su concienciación y preparación en este campo, ya que conseguir niveles saludables en los escolares y crear hábitos de práctica que garanticen una calidad de vida adecuada en el futuro son los principales objetivos.

### Referencias

- Burrows, L., Wright, J. y Jungersen-Smith, J. (2002). "Measure your belly". New Zealand children's constructions of health and fitness. *Journal of teaching in physical education*, 22, 39-48.
- Castejón, F.J., Caro, D., Gamarra, A., Hernando, A., López, I., Nieto, V., Rol, M., Ruiz, A., Sánchez de la Rosa A.M. y De la Torre, A.B. (2001). La evaluación de la programación del profesor y la influencia de su pensamiento en el diseño de la misma. *Revista de Educación Física: renovación de teoría y práctica*, 81, 5-12.
- Corbin C.B. (2002). Physical activity everyone: what every physical educator should know about promoting lifelong physical activity. *Journal of teaching in physical education*, 21, 128-144.
- Curtner-Smith, M. (1996). The impact of an early field experience on preservice physical education teacher's conceptions of teaching. *Journal of teaching in physical education*, 15, 224-250.
- Faulkner, G. y Reeves, C. (2000). Primary school student teachers' physical self-perceptions and attitudes toward teaching physical education. *Journal of teaching in physical education*, 19, 311-325.
- Matanin, M. y Collier, C. (2003). Longitudinal analysis of preservice teachers' beliefs about teaching physical education. *Journal of teaching in physical education*, 22, 153-168.
- Méndez, A. (1996). Motivación para la práctica: de la obligatoriedad en el Sistema Educativo a la necesidad de movimiento. *Revista de Educación Física: renovación de teoría y práctica*, (63), 19-23.
- Napper-Owen, G.E., Kovar, S.K., Ermler, K.L. y Mehrhof, J.H. (1999). Curricula equity in required ninth-grade physical education. *Journal of teaching in physical education*, 19, 2-21.
- Salinas, F., Miranda, M.T. y Viciano, J. (2006). ¿Hacia dónde orientan los profesores de EF en formación inicial sus planificaciones? Comparación entre las Facultades de España y Galicia. *Revista de Educación Física: renovación de teoría y práctica*, 101, 5-14 (en prensa).

- Salinas, F. y Viciano, J. (2006). La planificación de los bloques de contenidos de la Educación Física en Educación Secundaria Obligatoria. *Revista Digital* [www.cienciaydeporte.net](http://www.cienciaydeporte.net).
- Viciano, J. (2000). Principales tendencias innovadoras en la educación física actual. El avance del conocimiento curricular en educación física. *Lecturas de Educación Física: Revista Digital* [www.efdeportes.com](http://www.efdeportes.com), 19, Internet, Argentina.
- Viciano, J. (2002). *Planificar en Educación Física* (1ª edición). INDE. Barcelona.
- Viciano, J. y Delgado, M.A. (1999). La programación e intervención didáctica en el deporte escolar (II). Aportaciones de los diferentes estilos de enseñanza. *Apunts, Educación Física y Deportes*, 56, 17-24.
- Viciano, J. y Salinas, F. (2005). Modifications in the purposes of Physical Education planning because of effect of internship with preservice teachers of Granada. *Abstracts book AIESEP 2005, World Congress. Active Lifestyles: the impact of Education and Sport* (pp. 77). Lisboa: Faculdade de Motricidade Humana.
- Vizuet, M. (2002). La didáctica de la educación física y el “área de conocimiento de expresión corporal: profesores y currículum”. *Revista de Educación*, 328, 137-154.
- Zabala, M., Viciano, J. y Lozano, L. (2002). La planificación de los deportes e la educación física de E.S.O. *Lecturas de Educación Física: Revista Digital* [www.efdeportes.com](http://www.efdeportes.com), 48, Internet, Buenos Aires.

# UNA EXPERIENCIA DE INNOVACIÓN DOCENTE EN MÉTODOS DE INVESTIGACIÓN EN EDUCACIÓN

José Sánchez Santamaría  
Universidad de Castilla-La Mancha

El objetivo de este trabajo es describir y valorar una experiencia de innovación docente (ECTS), puesta en marcha en el curso 2008/09, en la asignatura de “métodos de investigación en educación” del plan de estudios de psicopedagogía (Universidad de Castilla-La Mancha). Con un diseño de un solo grupo y una muestra no aleatoria (N=23), se han analizado los efectos del uso de los esquemas científicos (finalidad exploratoria), mediante el estudio de su dificultad y del componente emocional de aprendizaje y, la valoración de su utilidad. Para dar cuenta de ello, se aportan los motivos que avalan el cambio metodológico, a modo de contextualización, a la vez que se caracteriza el tipo de metodología basada en esquemas científicos; posteriormente, se describe el proceso de implantación; y, por último, se muestran los resultados y conclusiones obtenidas, con intención de servir a la discusión. Un avance de los resultados revela un grado elevado de dificultad en el aprendizaje inicial de la estrategia y, una valoración muy positiva por parte del alumnado. Los resultados nos animan abrir una línea de trabajo caracterizada por incorporar un número mayor de alumnos (generalización) y un grupo de control para determinar la eficacia frente a otras estrategias.

## Introducción

El presente texto recoge los aspectos más relevantes de un proceso exploratorio de innovación docente, que tuvo su origen en el curso académico 2008/09 y que se proyectará hasta el 2010, con la intención de poner en práctica y evaluar el impacto de la estrategia didáctica basada en el uso de esquemas científicos, adaptada a las exigencias de trabajo de la organización establecida por los ECTS, dentro de la asignatura de métodos de investigación en educación del plan de estudios de psicopedagogía de la Universidad de Castilla-La Mancha.

Fruto de ese proceso, en esta comunicación se describe y analiza, a partir de un solo grupo y una muestra no aleatoria (N=23), los efectos del uso de los esquemas científicos (finalidad exploratoria), prestando atención al grado de dificultad y al componente emocional y a la valoración de su utilidad en el proceso de formación del estudiante.

## Marco teórico y contexto de actuación

### *Los esquemas científicos*

Se trata de un recurso esquemático basado en la significatividad lógica del contenido científico complejo. Una herramienta que permite resumir, sintetizar, tratar y relacionar contenidos complejos de forma gráfica, obteniendo un producto; y, al igual que los mapas conceptuales están formados por “conceptos y expresiones denominadas de enlace... que permiten unir frases que tienen significado y que se denominan proposiciones” (González García, 2000, p. 87), enfatizando el principio del aprender a aprender.

Los esquemas científicos se caracterizan por (Sánchez Santamaría y Ballester, 2009):

- a) Es un potente mediador en el proceso de aprendizaje del alumno, ya que permite la significación lógica y, en cierto modo psicológica, orientada al aprender a aprender, donde los nuevos conceptos son asimilados a las estructuras existentes.
- b) Es un método de enseñanza basado en competencias que sirve para representar la estructura conceptual y las relaciones lógicas del contenido de una disciplina, de lo general a lo específico.
- c) Establece una jerarquía conceptual: dos o más expresiones se integran bajo uno más inclusivo.
- d) Los conectores lógicos pueden ser cruzados, lo que permite relacionar distintas ramas jerárquicas entre sí.

Los componentes de los esquemas científicos son (Sánchez Santamaría y Ballester, 2009):

- a) Los cuadros en los que se recogerán las proposiciones y que servirán de marcos para el posterior establecimiento de relaciones lógicas.
- b) Los conectores lógicos que son los que nos permitirán establecer relaciones entre las proposiciones expresadas en los cuadros, que previamente se han extraído de un texto de trabajo.

El procedimiento de elaboración de un esquema científico se compone de una serie de tareas intelectuales vinculadas con el uso de competencias instrumentales de tipo cognitivo y lingüístico. Los pasos de este procedimiento son cinco (Sánchez Santamaría y Ballester, 2009):

- a) Lectura, subrayado y resumen del texto.
- b) Extracción de descriptores primarios y secundarios.
- c) Síntesis de las dependencias entre descriptores.
- d) Elaboración de cuadros de contenidos.
- e) Aplicación de conectores lógicos entre los cuadros de descriptores generales y específicos.

### Contexto de actuación

En la siguiente imagen se identifica el contexto de actuación en el que se ha llevado a cabo el proceso exploratorio de innovación docente (imagen 1).

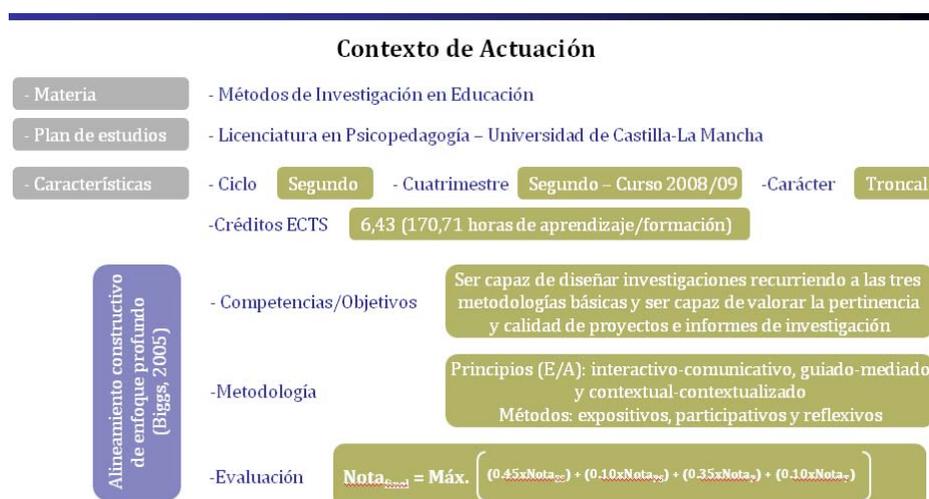


Imagen 1. Contexto de actuación del proceso exploratorio de innovación docente

Fuente: Elaboración propia, 2009

## Planteamiento del problema y objetivos

El problema de investigación propuesto se centró en establecer la valoración del impacto sobre el uso de los esquemas científicos, basado en las percepciones de los estudiantes y en las calificaciones obtenidas, en una asignatura de corte metodológico en educación, para poder extraer evidencias descriptivo-exploratorias que nos permitieran avanzar en un estudio más exhaustivo -evidencias para trabajos posteriores de carácter explicativo y confirmatorio- de este método didáctico e introducir posibles mejoras en su aplicación.

Los objetivos propuestos fueron:

1. Conocer las calificaciones medias obtenidas en la materia analizada por el alumnado.
2. Conocer en qué medida este método de trabajo ha representado una ayuda en términos de mejora de la comprensión de la materia; aplicación a casos prácticos; y, aumento de interés en la materia y del apoyo recibido.
3. Identificar el impacto emocional y sus implicaciones en el proceso de aprendizaje del alumno.
4. Obtener una valoración global de los esquemas científicos.

La falta de un grupo de control o de comparación no nos permite determinar la existencia de diferencias significativas, a nivel estadístico, sobre las cuestiones objeto de nuestra investigación. Algo muy deseable en experiencias futuras.

## Método

### *Muestra*

La muestra del estudio responde a un muestreo no probabilístico de tipo causal por accesibilidad (Aliaga, 2000, p. 89), ya que los sujetos que forman parte de la muestra son los alumnos que en el curso académico 2008/09 estaban matriculados en la asignatura de métodos de investigación en educación y se acogieron a la evaluación continua, es decir, de un total de 52 alumnos matriculados, la muestra invitada final fue de 23 alumnos, lo que representa el 44,2% del total.

### *Procedimiento*

El proceso de mejora se ha desarrollado a lo largo de toda la asignatura mediante fases progresivas. La siguiente tabla (1) muestra las características de este proceso de mejora, con los hitos principales, incluyendo las estrategias para la obtención y tratamiento de la información.

Tabla 1. Procedimiento y estrategias metodológicas de valoración del proceso de mejora

	MÓDULO I	MÓDULO II	MÓDULO III	MÓDULO IV	Global	
Contenido	EPISTEME	METODOLOGÍA				
	Conceptos y paradigmas	Orientación cuantitativa	Orientación cualitativa	Toma de decisiones y cambio		
Método	Esquemas científicos combinado con pruebas de seguimiento personalizado y tutorías					
Temporalización	Febrero	Febrero-Marzo	Marzo	Mayo	Cuestionario	Entrevistas
Diseño Metodológico	Calificaciones, emociones experimentadas y percepción global del método				Análisis estadístico	Análisis contenido

Fuente: Elaboración propia, 2009

### Instrumentos de recogida de datos

La instrumentación empleada para recoger evidencias sobre el impacto de este proceso de mejora consistió en:

- Sistema de indicadores. La evolución del alumno respecto a las calificaciones en el proceso de elaboración de los esquemas científicos. Asimismo, se va contemplar la frecuencia de consulta (tutorías y *on-line*) vinculada con el contenido de la misma. Además, también se apuntaron evidencias registradas mediante los momentos de revisión y evaluación de los esquemas, destacando los aspectos más complejos.
- Adaptación de 2 Cuestionarios
  - Cuestionario sobre técnicas de enseñanza y aprendizaje, compuesto por dos escalas tipo Likert (rango de 1-4), de Giménez Bertomeu y otros (2009): aprendizaje obtenido ( $\alpha = .80$ ) y apoyo recibido ( $\alpha = .70$ ).
  - Cuestionario tipo Likert (rango de 0-4) de bienestar emocional de Rebollo y otros (2008). Análisis descriptivo de los índices de bienestar ( $\alpha = .92$ ) y malestar emocional ( $\alpha = .93$ ).
- Entrevista semiestructurada individual para todos los participantes, con la intención de identificar puntos fuertes y débiles del método utilizado. Las dimensiones preguntadas se centraron en el potencial del método para el aprendizaje de la materia (aspectos positivos y susceptibles de mejora) y propuestas de mejora.

### Análisis y discusión

#### Rendimiento académico y consultas de los estudiantes

Las calificaciones medias de los participantes fue de 6,31, con una moda de 5 y una desviación típica de 1,284.

Sobre las consultas realizadas a lo largo de este proceso, bien en tutorías como a través del correo electrónico, se extrae que de las 232 consultas (Media = 10,08 por alumno), 176 fueron presenciales en horario de tutorías (Media = 7,65 por alumno) y 56 *on-line* (2,43 por alumno). Las consultas se centraron en tres cuestiones: a) uso de conectores lógicos (73,3%), organización de la información (17,4%) y uso de la herramienta informática (9,3%).

### Valoración de los estudiantes del aprendizaje obtenido, apoyo recibido y valoración global del método de trabajo

De la valoración de los estudiantes sobre el aprendizaje obtenido a partir de los esquemas científicos hemos extraído los siguientes datos (tabla 2):

Tabla 2. Aprendizaje obtenido por los estudiantes mediante la aplicación de los esquemas científicos

	M	DT	Mo	Rango	
				Mín.	Máx.
COMATERIA	2,43	.99	2	1	4
APMATERIA	2,87	.69	3	1	4
IMATERIA	2,65	.83	3	1	4

Fuente: Elaboración propia, a partir de CEA de Giménez Bertomeu y otros, 2009.

Leyenda: COMATERIA (comprensión); APMATERIA (aplicación teórica); y, IMATERIA (interés)

Del apoyo recibido para la realización de los esquemas científicos, los estudiantes manifestaron un grado elevado de satisfacción (tabla 3).

Tabla 3. Tipo y grado de apoyo recibido en la realización de los esquemas científicos

	M	DT	Mo
IRACTIVIDAD (a)	3,22	.902	3
MEACTIVIDAD (b)	3,00	.674	3
TACTIVIDAD (c)	2,78	.795	3
APACTIVIDAD (d)	3,04	.976	4
ACACTIVIDAD (e)	3,09	.900	4

Fuente: Elaboración propia, a partir de CEA de Giménez Bertomeu y otros, 2009.

Leyenda: IRACTIVIDAD (las instrucciones recibidas facilitan la actividad); MEACTIVIDAD (el material entregado facilita la actividad); y, TACTIVIDAD (la cantidad de tiempo ha sido la adecuada); APACTIVIDAD (el apoyo del profesor ha sido el necesario); y, ACACTIVIDAD (la interacción con los compañeros facilita la actividad)

Sobre la valoración global, en términos de aprendizaje logrado, dificultad encontrada y apoyo global recibido (tabla 4):

Tabla 4. Valoración global de los alumnos sobre los esquemas científicos

	M	DT	Mo	Rango	
				Mín.	Máx.
ALOGRADO	2,65	.647	3	2	4
ARECIBIDO	3,30	.635	3	2	4
DENCONTRADA	3,35	.487	3	1	3

Fuente: Elaboración propia, a partir de CEA de Giménez Bertomeu y otros, 2009.

Leyenda: ALOGRADO (grado de aprendizaje logrado); ARECIBIDO (grado de apoyo recibido); y, DENCONTRADA (grado de dificultad encontrada)

En definitiva, se aprecia una valoración muy positiva por parte de los alumnos para trabajar con los esquemas científicos, no obstante se pone de relieve la necesidad de mejorar los aspectos vinculados con el grado de aprendizaje logrado.

### *Percepción sobre el impacto emocional del uso de los esquemas científicos*

El análisis descriptivo de los índices (0-1) de bienestar y malestar emocional en el aprendizaje mediante esquemas científicos, revela que el bienestar emocional obtiene un índice medio más alto (media = 0,55, desviación típica de 0,23 y una moda de 0,63) que el malestar (media = 0,27, desviación típica de 0,32 y una moda de 0,26).

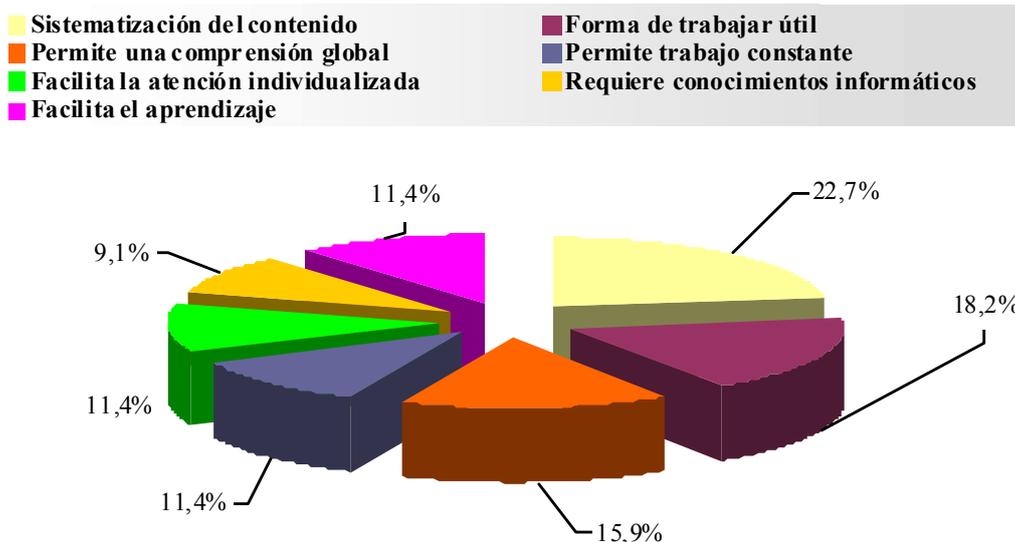
En un análisis pormenorizado de la presencia de cada emoción en el proceso de elaboración de los esquemas científicos, los resultados muestran que los valores más altos se dan en las emociones positivas con una media de 1,7 (cerca del valor 2 = en bastantes ocasiones), mientras que en las emociones negativas asume una media de 0,4 (prácticamente nunca).

Las emociones positivas han estado más presentes que las negativas, ya que todas sus medias se sitúan por encima de 1, siendo la perseverancia (2,08), la orientación (2,04) y el acompañamiento (1,96) las que presentan medias más altas. En el caso de las emociones negativas, a excepción del estrés-cansancio (1,88), la tensión-preocupación (1,33) y la angustia-ansiedad (1,21), el resto apenas se acercan al 1. Esto último preocupa, debido a que se constata el impacto negativo que en ciertos momentos tiene este método y que habrá que tener en cuenta en el futuro.

### *Ventajas e inconvenientes del uso de los esquemas científicos*

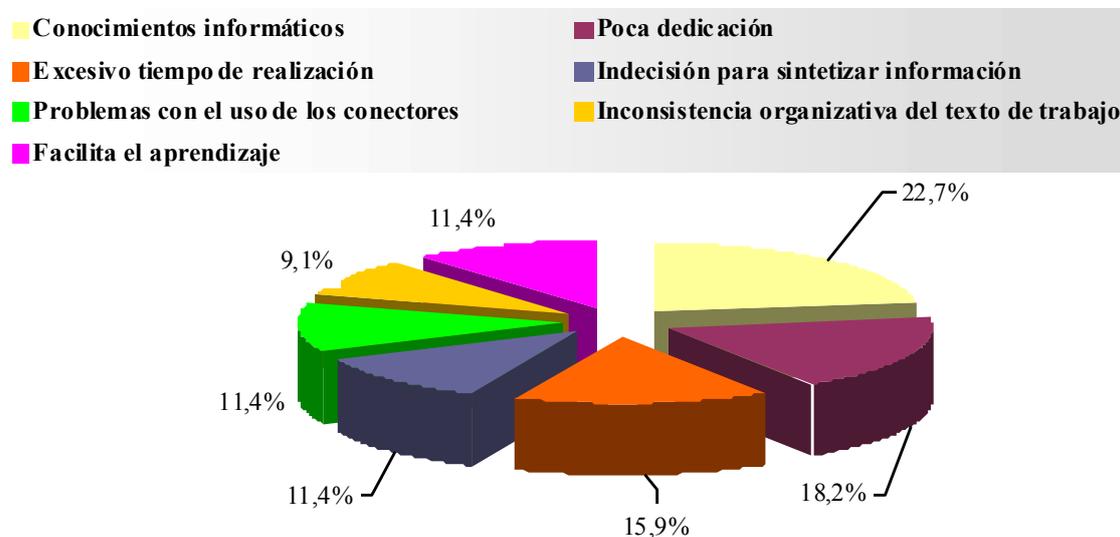
En nuestro caso, simplemente vamos a realizar una identificación de la valoración general que los alumnos participantes en el proceso de mejora junto con una relación de las dos ventajas y dos desventajas que los estudiantes han destacado del método de esquemas científicos.

La valoración global es muy positiva, aunque esta percepción se asume prácticamente al final del proceso, debido a que al principio el alumno tiene que hacer un esfuerzo importante.



Gráfica 1. Ventajas que los participantes del proceso de mejora han destacado (dos ventajas)  
 Fuente: Elaboración propia, 2009

Sobre las dos principales ventajas sobresalen: la sistematización del contenido que proporciona este método, el cual puede ser transferido a otras asignaturas, así como el hecho de que permite una comprensión global y representa una forma útil para abordar el estudio de gran cantidad de conocimientos.



Gráfica 2. Inconvenientes que los participantes del proceso de mejora han destacado  
 Fuente: Elaboración propia, 2009

Se observa que el principal problema se vincula con el uso apropiado de los conectores lógicos, lo que implica un escaso dominio del razonamiento lógico. Del mismo modo, se destaca la existencia de problemas a la hora de decidir qué información es la importante y cómo hay que sintetizarla. Esto último preocupa ya que estamos hablando de alumnos que tienen una dilatada formación universitaria (al menos de 3 a 4 años), por lo que se hace necesario identificar los motivos exactos que expliquen de forma adecuada este problema, debido a que el tipo de conocimiento puede estar influyendo.

## Conclusión

Recordando que se trata de un proceso de mejora que se ha analizado, más desde una perspectiva exploratoria que explicativa, se puede concluir que los resultados obtenidos ponen de manifiesto argumentos para continuar trabajando en el potencial que los esquemas científicos muestran en el aprendizaje universitario, en concreto, en materias de corte metodológico. Aunque parece evidente que tal y como nos advierte Carbonell (2001, p.22-23) “de una forma más o menos intencionada y planificada se ponen en movimiento ideas, estrategias y actividades pero lo importante es que estas confluyan, se interrelacionen y hasta se confundan en un todo indivisible”, así como contemplar los aspectos vinculados con la dedicación, el esfuerzo y los resultados reales que este método tiene para el proceso de aprendizaje del alumno.

Los resultados, aunque de naturaleza exploratorio, nos animan a profundizar en el potencial de este método. Una línea de trabajo a afianzar y evidenciar empíricamente es la eficacia de este método en el aprendizaje del alumno, sobre todo vinculado al éxito académico.

## Referencias

- Aliaga, F. (2000). *Bases epistemológicas y proceso de investigación psicoeducativa*. Valencia: CSV.
- Carbonell, J. (2001). *La aventura de innovar*. Madrid: Morata.
- Giménez Bertomeu, V. M. y otros (2009). El mapa conceptual desde la perspectiva del estudiante en los estudios de trabajo social. *VII Jornadas de Redes de Investigación en Docencia Universitaria*. Recuperado el 12 de junio de 2009, de <http://www.eduonline.ua.es/jornadas2009/comunicaciones/2C2.pdf?PHPSESSID=a65e926453e17a8ffbfd35e962cf31e2>
- González García, M<sup>a</sup>. F. (2000). *Una aportación a la mejora de la calidad de la docencia universitaria: los mapas conceptuales*. Navarra: Universidad Pública de Navarra.
- Rebollo Catalán, M<sup>a</sup> A. y otros (2008). Las emociones en el aprendizaje online. *Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE)*, 14 (1). Recuperado el 16 de junio de 2009, de [http://www.uv.es/RELIEVE/v14n1/RELIEVEv14n1\\_2.htm](http://www.uv.es/RELIEVE/v14n1/RELIEVEv14n1_2.htm).
- Sánchez Santamaría, J. y Ballester, M<sup>a</sup>.G. (2009). *Protocolo para la elaboración de esquemas científicos en asignaturas de corte metodológico*. Cuenca (documento policopiado).

# BASES TEÓRICAS PARA UNA INNOVACIÓN DOCENTE: LOS ESQUEMAS CIENTÍFICOS EN ASIGNATURAS METODOLÓGICAS

José Sánchez Santamaría y María Gracia Ballester  
Universidad de Castilla-La Mancha

“Es evidente que de una forma más o menos intencionada y planificada se ponen en movimiento ideas, estrategias y actividades pero lo importante es que estas confluyan, se interrelacionen y hasta se confundan en un todo indivisible” Carbonell, 2001, p. 22-23[1]

La metodología docente universitaria basada en créditos ECTS, en el marco del desarrollo de competencias, se caracteriza por: a) otorgar un mayor protagonismo al trabajo autónomo del alumno universitario; b) la creación de diversos escenarios de aprendizaje a partir de la diversificación de tareas que permitan el abordaje didáctico de las competencias; c) el impulso de las tutorías, dotándolas con mayor relevancia; d) el aprovechamiento de los recursos tecnológicos; e) un cambio profundo en las relaciones de trabajo y comunicación entre profesor-alumno; y, f) un mayor peso del aprendizaje cooperativo, entre otros. Unido a las dificultades de aprendizaje que, en no pocas ocasiones, tienen que hacer frente los estudiantes universitarios en las asignaturas de metodología.

Por estos motivos, se propone una aproximación teórica a la estrategia didáctica de esquema científico en el marco de las asignaturas metodológicas en ciencias sociales. En concreto, se intenta dar respuesta, con vocación de ser sometido a crítica, a las siguientes cuestiones: ¿por qué y para qué (naturaleza/sentido)?; ¿qué es y cuáles son sus características (aspectos básicos)?; ¿qué elementos la configuran (estructura interna)?; ¿cómo se elabora (criterios/procedimiento)?; ¿complementariedad/combinación didáctica con otras estrategias?; ¿cómo se evalúa (sistema)?; y, ¿cuáles son sus limitaciones?

## Introducción

En los últimos años son muchos los trabajos (Goñi, 2005; De Miguel, 2006) que vienen centrando su interés en las implicaciones educativas, tanto a nivel organizativo como metodológico, del proceso de armonización de la educación superior en Europa. En este escenario de cambios tan profundos de la realidad universitaria, también asistimos a la proliferación de experiencias de innovación docente basadas en las metodologías activas (Díez y Ayala, 2006; Rodríguez y Hernández, 2006).

En este sentido, este trabajo pretende una aproximación teórica al método<sup>[2]</sup> de esquemas científicos para su aplicación en las asignaturas metodológicas de Ciencias Sociales. Para lo cual se abordan siete cuestiones que tienen como objetivo servir al debate, a saber:

- ¿Por qué y para qué los esquemas científicos? (Naturaleza/sentido)
- ¿Qué es y cuáles son sus características de los esquemas científicos? (Delimitación y caracterización)
- ¿Qué elementos configuran los esquemas científicos? (Estructura interna)
- ¿Cómo se elaboran los esquemas científicos? (Criterios/procedimiento)

[1] Carbonell, J. (2001). La aventura de innovar. Madrid: Morata.

- ¿Los esquemas científicos y otras estrategias? (Complementariedad/combinación didáctica)
- ¿Cómo se evalúan los esquemas científicos? (Sistema/criterios)
- ¿Cuáles son las limitaciones y posibilidades de los esquemas científicos? (Críticas/revisiones)

Este trabajo forma parte de un proyecto de innovación docente llevado a cabo en el curso académico 2008/09, dentro de la asignatura de Métodos de Investigación en Educación del plan de estudios de psicopedagogía de la Facultad de Ciencias de la Educación y Humanidades de la Universidad de Castilla-La Mancha.

### Los esquemas científicos: Una aproximación desde los mapas conceptuales

#### *Fundamento y sentido*

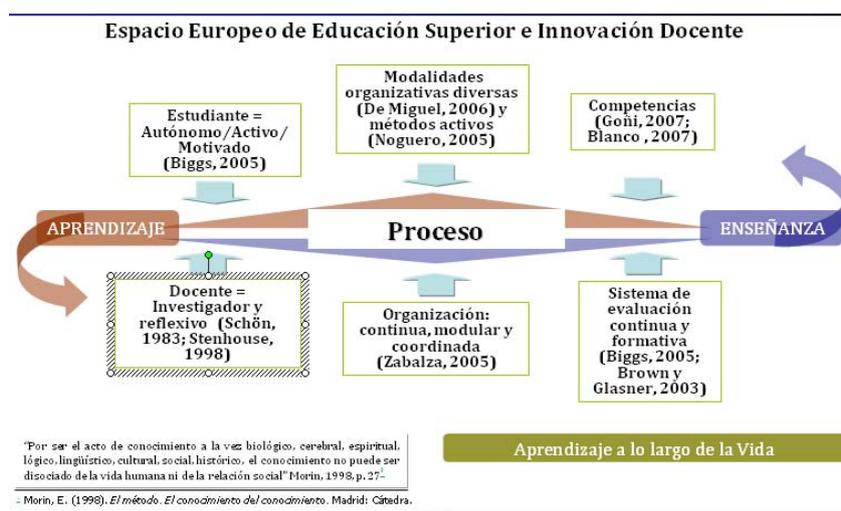
Desde el punto de vista de las teorías del aprendizaje, los esquemas científicos se asientan en los principios constitutivos del aprendizaje significativo propuesto por Ausubel, Novak y Hanesian (1978), al destacar la importancia de la diferenciación progresiva (inclusividad y transformación) y de la reconciliación integradora (vínculos conceptuales).

Si nos centramos en las teorías de la enseñanza, entendemos que la aplicación de los esquemas científicos implica una concepción del profesor anclada en los planteamientos del profesor como investigador y reflexivo sobre su propia práctica docente, tomando conciencia de su papel como mediador y generador de escenarios de formación, orientando sus esfuerzos y recursos a potenciar la formación crítica e integral del alumno.

Así, el sentido que inspira este método parte de la consideración de cambio intrínseco del paradigma educativo de los actuales sistemas de educación superior, los cuales deben desarrollar las habilidades metodológicas, sociales, participativas y los conocimientos de carácter práctico favorecedores de la aplicación de conocimientos teóricos (Mora, 2004). Por ello, los esquemas científicos, en combinación con otras técnicas didácticas, vienen a dar respuesta a las demandas de aplicación de metodologías activas en la docencia universitaria, propias de los retos metodológicos del espacio europeo de educación superior que reclaman un mayor protagonismo del alumno en su proceso de formación y el fortalecimiento de métodos didácticos acordes con el abordaje didáctico de las competencias generales, específicas y transversales que un alumno debe alcanzar tras finalizar un grado, un master o un doctorado, sabiendo que tenemos que tener presente el referente del aprendizaje a lo largo de la vida (esquema 1).

---

[2] Aunque puede haber quien vea excesiva la calificación de “método”. A nosotros nos parece acertada ya que se trata de un instrumento de búsqueda, organización, guía y creación en el desarrollo del proceso de aprendizaje del estudiante universitario en base a unos propósitos y objetivos de enseñanza.



Esquema 1. Referentes implicados en el proceso de enseñanza-aprendizaje universitario

Fuente: Elaboración propia, 2009

El para qué de los esquemas científicos queda resumido en cinco aspectos, a saber: a) facilitar orientaciones y herramientas que permitan al profesorado universitario planificar procesos de enseñanza-aprendizaje de acuerdo al enfoque de competencias (De Miguel, 2006); b) mejora de la calidad de la docencia universitaria (González García, 2000); c) mejora del desarrollo y aplicación de competencias instrumentales y sistemáticas, así como la mejora de dominios propios de la competencia emocional como la valoración adecuada de sí mismo y autogestión (logro, adaptabilidad, iniciativa y optimismo); d) contar con “productos” que conecten con la formación académica y el desarrollo profesional; y, e) que los estudiantes aprendan a cómo pensar estadísticamente y adquieran interés por las aplicaciones estadísticas (Schau, 2003).

### Delimitación y características

Se trata de un recurso esquemático basado en la significatividad lógica del contenido científico complejo. Una herramienta que permite resumir, sintetizar, tratar y relacionar contenidos complejos de forma gráfica, obteniendo un producto; y, al igual que los mapas científicos están formados por “conceptos y expresiones denominadas de enlace... que permiten unir frases que tienen significado y que se denominan proposiciones” (González García, 2000, p. 87), enfatizando el principio del aprender a aprender.

Además, entre otras similitudes con los mapas conceptuales, los esquemas científicos articulan conceptos entre sí, lo que permite “desarrollar la capacidad de abordaje global de fenómenos, diagnóstico y pronóstico a partir de representaciones gráficas de los elementos clave que los configuran y sus interacciones” (González García, 2008, p. 3). Esto presenta dos ventajas: primero, el estudiante universitario puede estructurar los conceptos mediante una representación gráfica relativamente sencilla; y, segundo, el profesor puede identificar posibles inconsistencias, lagunas o distorsiones en la forma de interpretar y comprender el concepto y sus relaciones con otros términos.

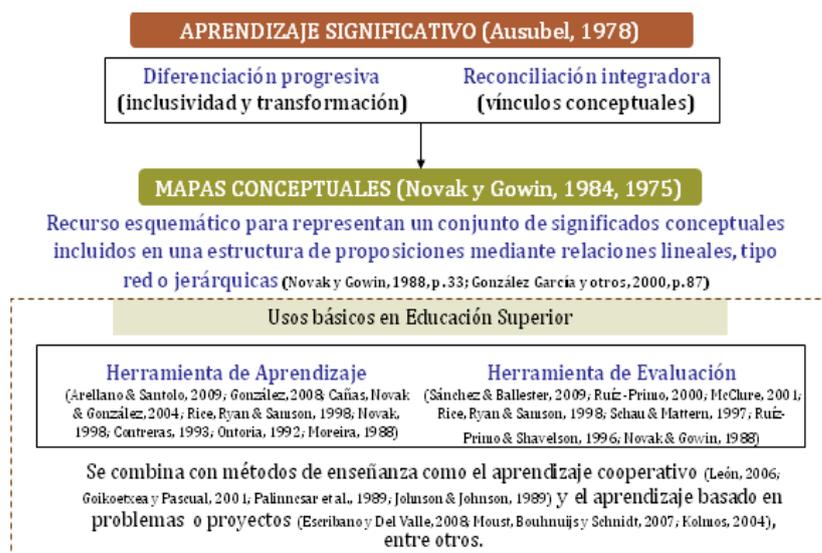
Pero, los mapas conceptuales, según nuestro juicio, presentan algunas limitaciones respecto a los esquemas científicos: a) no permiten el pensamiento circular basado en las predicciones, es decir, la relación lineal sobre la que se configuran representa un déficit para un proceso cognitivo complejo; b) se alinea con una forma de discurrir deductiva, no contempla un pensamiento inductivo; c) su elaboración

se centra en la configuración de proposiciones, no obstante la simplificación y fragmentación del texto complica un pensamiento basado en relaciones lógicas, más allá del concepto; y, d) es complicado su uso extensivo, es decir, aplicarlo a todo un contenido de una materia, lo que hace que se restrinja a determinados contenidos excesivamente conceptuales.

Por ello, proponemos la idea de un “método de enseñanza basada en esquemas científicos”, con la intención de dar respuesta a las limitaciones presentadas. Esto no quiere decir que el esquema científico no contemple aspectos útiles de un mapa conceptual, tal y como hemos expuesto, pero no la totalidad de sus componentes.

Algunas de las características de los esquemas científicos son:

- Es un potente mediador en el proceso de aprendizaje del alumno, ya que permite la significación lógica y, en cierto modo psicológica, orientada al aprender a aprender, donde los nuevos conceptos son asimilados a las estructuras existentes.
- Es un método de enseñanza basado en competencias que sirve para representar la estructura conceptual y las relaciones lógicas del contenido de una disciplina, de lo general a lo específico.
- Establece una jerarquía conceptual: dos o más expresiones se integran bajo uno más inclusivo.
- Los conectores lógicos pueden ser cruzados, lo que permite relacionar distintas ramas jerárquicas entre sí.



Esquema 2. Referentes teóricos y metodológicos de los esquemas científicos

Fuente: Elaboración propia, 2009

### Elementos de los esquemas científicos

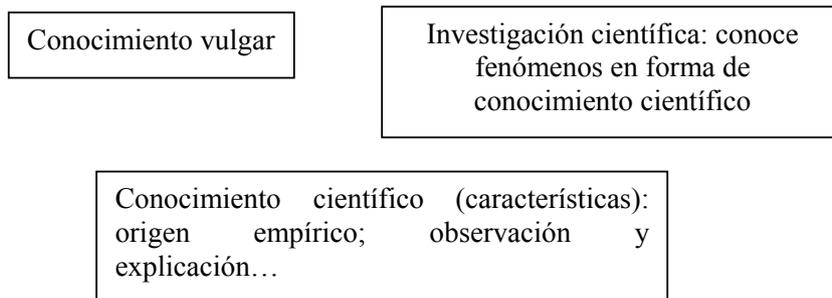
Los esquemas científicos se basan en proposiciones generales y específicas que conforman el contenido de los mismos. Por ejemplo, si el alumno tiene ante sí un texto vinculado con la definición de la investigación científica, estableciendo una diferenciación con el conocimiento vulgar, a la vez que aborda la delimitación y caracterización del conocimiento científico, de aquí obtendrá tres tipos de proposiciones generales conectadas a proposiciones específicas, tal y como se indica:

**PROPOSICIONES: GENERALES (PG) Y ESPECÍFICAS (PE)**

- a. Investigación científica (PG)+definición de investigación científica (PE)
- b. Conocimiento vulgar: en oposición a la investigación científica (PG).
- c. Conocimiento científico (PG)+características (PE).

Tras ello, podemos identificar dos elementos más de carácter gráfico de los esquemas científicos, a saber:

- Los cuadros en los que se recogerán las proposiciones y que servirán de marcos para el posterior establecimiento de relaciones lógicas.



- Los conectores lógicos que son los que nos permitirán establecer relaciones entre las proposiciones expresadas en los cuadros, que previamente se han extraído de un texto de trabajo. Tenemos cuatro tipos:

*Tabla 1. Identificación de los conectores lógicos en los esquemas científicos*

Conector	Descripción	Representación
Relación	A se relaciona con B	←-----→
Influencia	A influye sobre B.	→
Oposición	A se opone a B	→ /
Determinación	A determina a B; determina la forma o la existencia de B	→

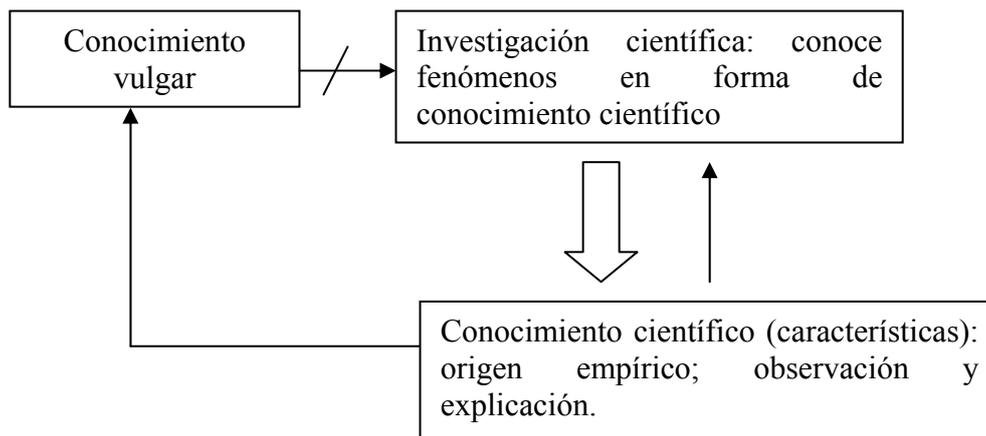
Fuente: Elaboración propia, 2009.

*¿Cómo se elaboran los esquemas científicos? (Criterios/procedimiento)*

El procedimiento de elaboración de un esquema científico es prácticamente igual que el de un mapa conceptual. Se compone de una serie de tareas intelectuales vinculadas con el uso de competencias instrumentales de tipo cognitivo y lingüístico. Los pasos de este procedimiento son cinco:

1. Lectura, subrayado y resumen del texto.
2. Extracción de descriptores primarios y secundarios.
3. Síntesis de las dependencias entre descriptores.
4. Elaboración de cuadros de contenidos.
5. Aplicación de conectores lógicos entre los cuadros de descriptores generales y específicos.

El resultado final que se obtendría siguiendo el ejemplo anterior sería el siguiente:



Si tuviéramos que explicar este esquema lo haríamos así:

“El conocimiento vulgar está en oposición a la investigación científica. Bien los fenómenos o bien el conocimiento que tengamos de esos fenómenos determinan el tipo de conocimiento científico que extraigamos. Este conocimiento científico tendrá unas características que influirán sobre la investigación científica y, por tanto, sobre el conocimiento de los fenómenos. Además, el conocimiento científico por el tiempo ejercerá una influencia sobre el conocimiento vulgar”.

*¿Los esquemas científicos y otras estrategias? (Complementariedad/combinación didáctica)*

Los esquemas científicos se pueden combinar con otros métodos de enseñanza-aprendizaje, tales como: expositivos, estudios de casos, resolución de ejercicios y problemas, aprendizaje basado en problemas y/o proyectos, aprendizaje cooperativo, etc. (De Miguel, 2006). Tendremos que tener presente las competencias y objetivos que se pretenden, así como la dedicación del alumno a las tareas en las que se organice la combinación entre esquemas y otros métodos.

*¿Cómo se evalúan los esquemas científicos? (Sistema/criterios)*

Una de las principales limitaciones en el uso de los esquemas científicos, y aquí ocurre lo mismo que con los mapas conceptuales, es quizá el grado de complejidad que puede representar el proceso de evaluación. En este sentido, y teniendo presente que toda evaluación debe estar organizada en un sistema y responder a criterios de objetividad, justicia y veracidad de los datos o información que se obtiene, hay dos cuestiones que abordar sobre la evaluación de los esquemas científicos:

¿Qué evaluar?:

- Grado de jerarquización: descriptores generales y específicos.
- Diferenciación progresiva.
- Integración conceptual.

¿Cómo evaluar? Asignar un valor numérico a los diferentes aspectos de evaluación:

- Identificación del contenido (primario-secundario): 2 puntos.
- Jerarquías válidas: 1 punto.

- Relaciones válidas: 5 puntos.
- Conexiones cruzadas válidas y significativas: 1 punto.
- Grado de visibilidad de la información: 1 punto.

*¿Cuáles son las limitaciones y posibilidades de los esquemas científicos? (Críticas/revisiones)*

En la siguiente tabla se muestran las principales ventajas e inconvenientes de los esquemas científicos:

Posibilidades	Límites
Enfatizar la estructura conceptual de la materia	Falta de significado para los estudiantes (aprendizaje memorístico)
Proporciona una visión integrada en términos de inclusividad y generalidad y presenta los conceptos de forma jerárquica de modo que facilita el aprendizaje y la recuperación inmediata	Exceso de complejidad y dedicación temporal: el estudiante puede dedicar más esfuerzo a la elaboración (estructura) que a la comprensión de las relaciones y su sentido
Obtener un producto que puede ser utilizado para su aplicación en prácticas y consultado en su ejercicio profesional	Distorsiones en la elaboración debido a las preferencias y percepciones de los estudiantes
Adaptación a personas con discapacidad visual, aplicando un sistema simbólico comprensible	Puede presentar complicaciones respecto a cómo evaluarlo
Herramienta con múltiples posibilidades para el aprendizaje, organizador previo, estudio, cooperación, evaluación y resolución de problemas	Requiere de cierto período de entrenamiento por parte del estudiante

Fuente: Elaboración propia, a partir Cañas, Novak y González, 2004

### Conclusión

La intención que ha inspirado este trabajo ha sido la de mostrar los principales aspectos de los esquemas científicos, que como se ha podido comprobar mantienen una estrecha relación con los mapas conceptuales, pero cuya principal diferencia es que los esquemas científicos van más allá del concepto y trabajan con contenido exclusivamente científico.

Esta propuesta debe ser sometida a una revisión mucho más exhaustiva, cuyo impacto educativo se encuentra en proceso de exploración. A pesar de ello, consideramos que los esquemas científicos en las asignaturas de corte metodológico de ciencias sociales puede abrir un proceso de consolidación de las técnicas de trabajo intelectual, incluidas en el tratamiento pedagógico de las competencias en la formación universitaria, así como pueden reforzar un modo de hacer donde la capacidad de ir más allá del texto se vea reforzada.

## Referencias

- Ausubel, D., Novak, J. & Hanesian, H. (1978). *Educational Psychology: A Cognitive View* (2nd Ed.). New York: Holt, Rinehart & Winston.
- Cañas, A., Novak, J.D. & González, F. (Eds.) (2004). *Concept maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping*. Pamplona, España: Universidad Pública de Navarra.
- De Miguel, M. (Coord.) (2006). *Metodologías de enseñanza y aprendizaje para el desarrollo de competencias*. Madrid: Alianza Editorial.
- Díaz Fondón, M. y Riesco, M. (2006). Desarrollo de técnicas de aprendizaje aplicables a grandes grupos, en la línea del modelo propuesto por el EEEs. R. Rodríguez y J. Hernández (Coords.). *Docencia universitaria: Proyectos de innovación docente* (pp. 59-72), Oviedo: Documentos ICE.
- Díez, M<sup>a</sup>. C. y Ayala, I. M<sup>a</sup>. (2006). La distribución de las actividades académicas dirigidas en la franja horaria: un reto para la convergencia. *Actas de las Jornadas de Trabajo sobre Experiencias Piloto de Implantación del Crédito Europeo en las Universidades Andaluzas*. Cádiz: Universidad de Cádiz. 19-21 septiembre de 2006. Recuperado el 13 de junio de 2009, de <http://www2.uca.es/orgobierno/rector/jornadas/documentos/097.pdf>.
- González García, M<sup>a</sup>. F. (2000). *Una aportación a la mejora de la calidad de la docencia universitaria: los mapas conceptuales*. Navarra: Universidad Pública de Navarra.
- González García, M<sup>a</sup>. F. (2008). *El mapa conceptual y el diagrama UVE. Recursos para la enseñanza Superior en el siglo XXI* (2<sup>a</sup> Edición). Madrid: Narcea.
- Goñi, J. (2005). *El Espacio Europeo de Educación Superior, un reto para la universidad*. Madrid: Octaedro.
- Mora, J. G. (2004). La necesidad del cambio educativo para la sociedad del conocimiento. *Revista Iberoamericana de Educación*, 35, 13-37.
- Schau, C. (2003). Students' attitudes: the "other" important outcome in statistics education. *Paper presented at the Joint Statistics Meetings*. San Francisco.

# CÓMO MEDIR LA METACOGNICIÓN EN ESTUDIANTES UNIVERSITARIOS

Aída Sandoval y Lisette Franchi  
Universidad del Zulia, Venezuela

Las Ciencias de la Educación avalan los instrumentos de medición como procedimientos sistemáticos que permiten medir aspectos de la condición humana tales como los procesos cognitivos. La metacognición es la capacidad que posee un individuo para conocer, regular y evaluar su propio conocimiento. El análisis de este constructo generó un instrumento que permite medirlo, considerando las modalidades más importantes para la resolución de problemas: metaatención, metacomprensión, metamemoria y metalenguaje; sustentado teóricamente en los aportes de Flavell entre otros. El instrumento diseñado, *Inventario de Metacognición*, es una herramienta que revela el conocimiento consciente que poseen los estudiantes acerca de la naturaleza de su propia cognición, además de la forma en que la controla y evalúa. Es un instrumento formal (Ruiz, 2002), que consta de 62 reactivos tipo Likert, agrupados en 4 sub-escalas, una para cada modalidad y al cual se le realizó un estudio técnico en relación al análisis de los ítemes, estimación de confiabilidad y estudio de la validez. Se realizó una prueba piloto a 134 estudiantes para analizar los reactivos y su grado de discriminación. Se concluyó que el instrumento es válido y confiable para determinar la metacognición en estudiantes universitarios.

Desde hace algunos años se ha generado una nueva conceptualización del aprendizaje como un proceso de construcción de conocimientos y no como una mera aplicación de los mismos. Este enfoque resulta particularmente adecuado para el caso de los aprendizajes relacionados con la formación del ingeniero donde el estudio de la matemática exige manejar la información, el planteamiento de conjeturas y una serie de acciones cognitivas que debe poner en práctica para construir sus propios conocimientos en esta área.

Una profesión con alto grado de abstracción como lo es la ingeniería requiere la formación de un profesional que demuestre capacidad para utilizar modelos matemáticos acordes con sus necesidades creando sus propias estrategias para la resolución de problemas.

Se han adelantado reformas curriculares en las universidades venezolanas, pero aún se debe dar mayor importancia a la enseñanza de habilidades metacognitivas y de estrategias de apoyo que permitan mejorar la calidad y el control del procesamiento de la información en cada una de las asignaturas que cursa el estudiante universitario.

Es importante enseñar a los estudiantes a “aprender a aprender”, a tener la capacidad de razonar sobre la forma en que aprenden, a manejar los procesos básicos y el conocimiento metacognitivo que poseen -qué saben y cómo procesan este conocimiento-.

En el caso particular de los estudiantes de ingeniería, son múltiples los esfuerzos que se hacen para ayudarlos a desarrollar herramientas de estudio efectivas; sin embargo, fracasan con mucha frecuencia en algunas asignaturas básicas y del área profesional.

Dada la relevancia que supone conocer las fortalezas y debilidades que presentan los estudiantes en sus habilidades metacognitivas, se analizaron dos instrumentos para determinarlas, éstos son: la Prueba LASSI de Weinstein, Zimmerman y Palmer (1988), que aunque no fue diseñada para medir únicamente metacognición sino habilidades de estudio, involucra en sus áreas de interés procesos

relacionados con el desarrollo de habilidades metacognitivas; y el Inventario de Sternberg (1977) que determina habilidades metacognitivas a través de ocho áreas, las cuales conducen a una medida general acerca de la metacognición de los estudiantes.

De este análisis se desprendió que sólo con estos instrumentos no era posible determinar la información necesaria para describir la actividad metacognitiva de los estudiantes de ingeniería, considerando por separado aspectos muy vinculados a la resolución de problemas tales como: la atención, la memoria, el lenguaje y la comprensión.

En consecuencia, se consideró pertinente diseñar un instrumento para medir específicamente las habilidades metacognitivas en estudiantes de ingeniería adaptado a la realidad educativa en la que se desarrolló esta investigación, considerando para ello las modalidades de metaatención, metalenguaje, metamemoria y metacompreensión.

Elaborar un instrumento para evaluar una situación particular que abarca el campo de la psicología y la educación como la metacognición, es una tarea que supone realizar un estudio técnico utilizando los pasos para su diseño y construcción, aplicándolo en sus diferentes versiones hasta lograr un instrumento confiable y válido que mida un aspecto de la conducta humana desde el punto de vista cognitivo. El instrumento así obtenido además del valor teórico que aporta, tiene una utilidad metodológica.

Se utilizaron las fases propuestas por Ruiz (2002) para la elaboración de instrumentos de medición, expresadas en la figura 1, tomando en cuenta la valiosa experiencia de este autor en la investigación educativa como profesor de diseño y construcción de instrumentos de investigación en la Maestría de Educación de la Universidad Pedagógica Experimental Libertador (UPEL, Venezuela).



Figura 1. Fases para la Construcción de Instrumentos de Medición (Ruiz, 2002)

Fuente: Sandoval y Franchi (2009)

### *PRIMERA FASE. Propósito del instrumento*

El instrumento diseñado, *Inventario de Metacognición*, tiene el propósito de medir la metacognición de estudiantes universitarios a través de las modalidades: metaatención, metamemoria, metacompreensión y metalenguaje. Asimismo, permite determinar separadamente cada una de éstas modalidades metacognitivas.

## SEGUNDA FASE. Decidir sobre el tipo de instrumento

Se decidió diseñar una Escala tipo Likert con 6 alternativas de respuesta que van desde completamente en desacuerdo (1) hasta totalmente de acuerdo (6).

## TERCERA FASE. Conceptualizar el constructo

De acuerdo con este autor, se debe iniciar el diseño de un instrumento con la conceptualización del constructo. Para esto se hizo una revisión cuidadosa de la literatura especializada, adoptando entre varias definiciones la de *John Flavell* establecida en su *Teoría del Monitoreo Cognitivo* donde relaciona la metacognición con la resolución de problemas.

### Conceptualización de la Metacognición

La metacognición “es el conocimiento de los propios procesos cognoscitivos, de los resultados de esos procesos y de cualquier aspecto que se relacione con ellos, es decir, el aprendizaje de las propiedades relevantes de la información”. (Flavell, 1979).

### Modalidades de la Metacognición

Existen tantas modalidades de metacognición como procesos cognoscitivos pues, según Carrasco (1997), la metacognición abarca el conocimiento, el control y la regulación de tales procesos. Para efectos de este estudio se tomaron en cuenta las modalidades: metaatención, metalenguaje, metacomprensión y metamemoria, como se refleja en la figura 2.

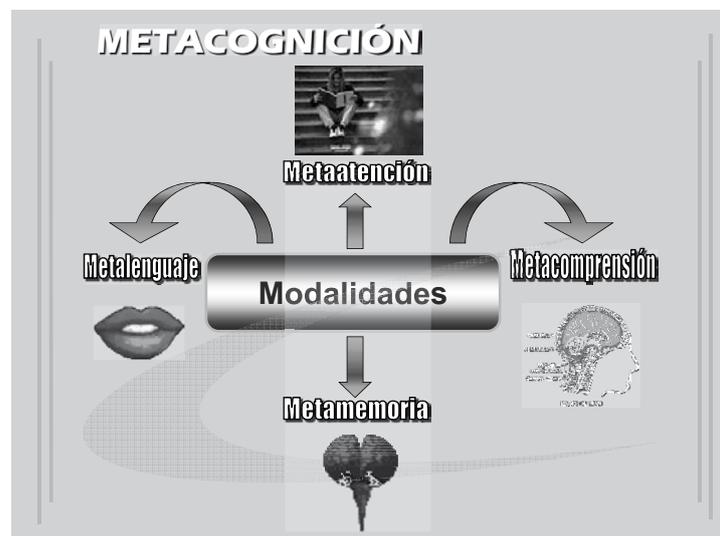


Figura 2. Modalidades de la Metacognición

Fuente: Sandoval y Franchi (2009)

La metaatención abarca el conocimiento de los procesos involucrados en la acción de atender: a qué se debe atender, cuáles operaciones mentales se deben realizar para atender, cómo hacer para evitar las distracciones internas o externas que interrumpen el acto de atender, cómo tener una atención más selectiva y menos dispersa.

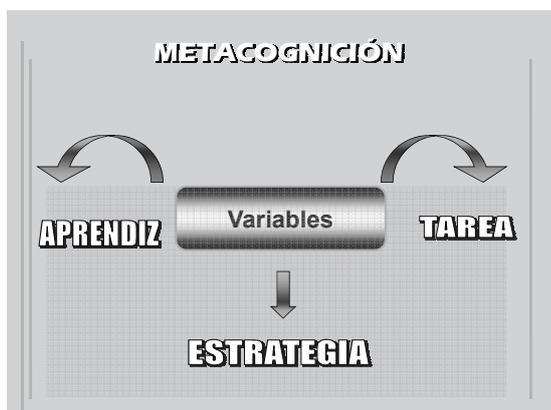
Metalingüaje es un lenguaje de segundo orden cuya única función es describir una lengua. Tomando en cuenta que el lenguaje es constitutivo de la acción cognitiva y que a través de éste el individuo piensa y se expresa, la habilidad metacognitiva que tiene por objeto el lenguaje incorpora las actividades de reflexión y control sobre todos los componentes lingüísticos.

La metacompreensión está relacionada con el conocimiento que puede tener el estudiante de su propio yo, de la tarea que se le asigna y de las estrategias para realizar con éxito esa tarea. Es quizás el aspecto más importante del aprendizaje, por tanto los alumnos con un alto nivel de comprensión son aquellos que conocen que están comprendiendo cuando de hecho lo está haciendo, o aquellos que se dan cuenta que no están comprendiendo cuando realmente no lo están.

El conocimiento sobre los procesos de la memoria y sus mecanismos autorreguladores, se denomina metamemoria. Hace referencia a nuestro conocimiento y conciencia acerca de la memoria y de todo aquello relevante para el registro, almacenamiento y recuperación de la información.

### *Variables de la Metacognición*

La tarea, las estrategias y el aprendiz constituyen las variables más importantes de la metacognición. Las variables del individuo, las tareas y las estrategias interactúan cuando los alumnos realizan actividades metacognoscitivas. Los estudiantes consideran la clase y las dimensiones del material por aprender (tarea), los diferentes métodos posibles para alcanzar ese aprendizaje (estrategia) y su habilidad para usar las distintas estrategias (aprendiz).



*Figura 3. Variables de la Metacognición*

Fuente: Sandoval y Franchi (2009)

### *Aprendiz*

Se toma en cuenta el conocimiento que posee, sus creencias, su nivel de habilidades para solucionar problemas y las motivaciones que animan el proceso.

### *Tarea*

Está relacionada no sólo con la naturaleza de los estímulos que ingresan a la memoria, sino también con el carácter de las demandas de la evocación y de la recuperación de la información. Si la tarea es muy fácil no se observará el uso de las estrategias metacognitivas, parte de la conciencia metacognitiva es el conocimiento de la relativa dificultad de aprender y recuperar de la memoria diversas informaciones. El darse cuenta de que una tarea es imposible de realizar se puede considerar como una manifestación metacognitiva producto de evaluación de lo que se sabe y no se sabe.

## Estrategia

Incluye el conocimiento disponible de la variedad de operaciones que un individuo puede voluntariamente llevar a cabo para ayudar a su sistema de memoria (de almacenamiento y evocación).

Es importante distinguir entre estrategias cognitivas y estrategias metacognitivas. Las cognitivas están diseñadas para alcanzar un objetivo cognitivo, mientras que las metacognitivas permiten elegir la estrategia cognitiva adecuada y controlar si se está alcanzando el objetivo.

### CUARTA FASE. Operacionalizar el constructo

Consiste en elaborar un cuadro que contiene el propósito del instrumento, la definición del constructo, sus dimensiones, los indicadores de cada dimensión y los ítemes.

En la figura 4 se muestra parte del cuadro de operacionalización del constructo en estudio.

DISEÑO DEL INSTRUMENTO ESCALA TIPO LIKERT		
4. Operacionalizar el constructo		
VARIABLE	Modalidad	Dimensiones
M E T A C O G N I C I Ó N	Metaatención	MA1: Conocimiento sobre la atención
		MA2: Control de los factores que impiden la atención
		MA3: Evaluación de la propia atención
	Metacomprensión	MC1: Conocimiento sobre la comprensión
		MC2: Control de los factores que impiden la comprensión
		MC3: Evaluación de la propia comprensión
	Metamemoria	MM1: Conocimiento sobre la memoria
		MM2: Control de los factores que impiden la memorización
		MM3: Evaluación de la propia memoria
	Metalenguaje	ML1: Consciencia del lenguaje a nivel sintáctico y semántico
		ML2: Control de los factores que impiden el uso correcto del lenguaje en forma oral o escrita
		ML3: Evaluación el uso propio del lenguaje

Figura 4. Modalidades y dimensiones de la variable  
Fuente: Sandoval y Franchi (2009)

### QUINTA FASE. Integrar el instrumento

En esta fase se realiza la validez de contenido del instrumento. A tal efecto, se seleccionaron 4 jueces o expertos para juzgar de manera independiente la bondad de los ítemes del instrumento, su relevancia, congruencia y claridad en la redacción, entre otros. La primera versión del instrumento, quedó estructurada por 87 ítemes distribuidos como aparece en la figura 5.



Figura 5. Distribución de ítems por modalidad de la primera versión

Fuente: Sandoval y Franchi (2009)

### *SEXTA FASE. Realizar la prueba piloto*

Se realizó una prueba piloto que consistió en aplicar el instrumento diseñado (Inventario de metacognición) y dos instrumentos estandarizados a los 164 participantes en esta investigación, con el objeto de poder analizar los ítems del inventario de metacognición, determinar su validez de constructo y su confiabilidad.

### *SÉPTIMA FASE. Estudio Técnico*

De acuerdo con Ruiz (2002) el estudio técnico consiste en realizar el análisis de ítems, la estimación de la confiabilidad, el estudio de la validez del instrumento, la estandarización y la normalización.

#### *Análisis de ítems*

Para realizar el análisis de los ítems se aplicó la prueba t de *Student* para grupos extremos, lo cual permitió determinar el poder discriminatorio de los ítems. Este análisis dio origen a la segunda versión del instrumento que quedó conformado por 62 ítems.

#### *Confiabilidad*

El coeficiente alfa de Cronbach obtenido fue de  $r=0,901$ , apuntando a una alta confiabilidad.

#### *Validación del Instrumento*

Se realizó una de las 4 formas propuestas por Thorndike (1989) y Helmstadter (1964), citados por Ruiz (2002): la validez convergente que consistió en correlacionar el constructo bajo estudio con otras variables en las que teóricamente era esperable una correlación positiva. Se correlacionaron los resultados del Inventario de Metacognición con los de la Pruebas LASSI y los del Inventario

de Sternberg. Dichas correlaciones fueron obtenidas a través de dos métodos: el método Momento Producto de Pearson y el de Spearman Brown. Ver tabla 1.

*Tabla 1. Correlaciones de la Variable Metacognición con las variables de otras pruebas*

La variable metacognición presentó correlaciones altas y significativas	
Método momento producto de Pearson	Método de Spearman Brown
r=0.850 con la variable habilidades metacognitivas (Sternberg)	r=0.848 con la variable habilidades megtacognitivas (Sternberg)
r=0.931 con la habilidades para el estudio (LASSI)	r=0.938 con la habilidades para el estudio (LASSI)

Fuente: Sandoval y Franchi (2009)

La versión final del instrumento elaborado se denominó Inventario de Metacognición. Presenta una estructura novedosa basada en 4 sub-escalas, correspondientes a las modalidades de la metacognición que más se relacionan con el proceso de resolución de problemas: metaatención, metamemoria, metacomprensión y metalenguaje. Es de aplicación sencilla y su aplicación tiene una duración de 30 minutos aproximadamente. Se evalúa de acuerdo con el baremo que aparece en la tabla 2.

*Tabla 2. Baremo*

Nivel	Rango
Muy bajo	$1.0 < \text{metacognición} \leq 2.0$
Bajo	$2.0 < \text{metacognición} \leq 3.0$
Moderado	$3.0 < \text{metacognición} \leq 4.0$
Alto	$4.0 < \text{metacognición} \leq 5.0$
Muy alto	$5.0 < \text{metacognición} \leq 6.0$

Fuente: Sandoval y Franchi (2009)

## Referencias

- Carrasco, J. (1997). *Hacia una enseñanza eficaz*. Madrid: Ediciones Rialp.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34, 906-911.
- Ruiz, C. (2002). *Instrumentos de Investigación Educativa. Procedimientos para su diseño y validación*. Barquisimeto: UPEL.
- Stenberg, R. (1986). *Intelligence Applied*. San Diego: Harcourt.
- Weinstein, C., Zimmerman, S. y Palmer, D. (1988). *Assesing learning strategies: the design and development of the Lassi*. En: C. Weinstein, E. Goetz y P. Alexander (Eds.). *Learning and study strategies*. New York: Academic Press.

# EL CONSTRUCTIVISMO, UN NUEVO SIGNIFICANTE PARA SIGNIFICADOS CON HISTORIA

Marta A. Tenutto

Facultad Latinoamericana de Ciencias Sociales

La formación docente en Argentina se constituyó a lo largo de un proceso que demarca hitos cuyas marcas perduran. En ellos se reconoce la convivencia de significados, en ocasiones contradictorios. Se presenta una investigación de diseño cualitativo concretada en un estudio de casos donde se indagó particularmente los significados atribuidos al “constructivismo” por los profesores de profesorado en la provincia de Buenos Aires, en el marco de la transformación educativa de 1990. También se indagó sobre el conductismo en tanto constituyó un par que actuó en tensión con el anterior. Los instrumentos de recolección de datos usados fueron: observación de clases, entrevistas y encuestas a docentes. Finalmente se hallaron recurrencias en la asignación de significados que se concretaron en nudos discursivos semánticos y funcionales. Se concluyó que se procuró instalar al constructivismo como parte del discurso pedagógico, ubicándolo en el polo del “deber ser”, haciéndolo jugar, como opuesto al conductismo que era sinónimo de la escuela existente y que se desea cambiar. Muchos profesores adoptaron ese discurso aunque manifestaban, en la mayoría de los casos, que carecían de elementos para poder llevarlo al aula.

El objetivo de la investigación fue: Indagar los significados otorgados al constructivismo por los profesores de los profesorados de formación docente, a la luz de los procesos de reforma educativa en Latinoamérica y particularmente en Argentina de la década de 1990. Fue entonces cuando, en el marco de la reforma del estado, la reforma educativa<sup>1</sup> fue incluida en la agenda de gobierno. Los dispositivos (Braslavsky y Gvirtz, 2001) que permitieron regular, promover y dar forma a los nuevos sistemas educativos fueron: fueron las leyes de educación marco (Ley Federal de Educación N° 24.195, Ley de la provincia de Buenos Aires N° 11.612), los contenidos básicos comunes, las reformas didácticas en la cotidianeidad escolar, los nuevos modos de gobierno del sistema educativo y de las instituciones escolares, las propuestas de capacitación y profesionalización docente y la construcción de sistemas nacionales de evaluación.

Esta investigación se concentró en los profesores de formación docente porque ellos juegan un papel fundamental a la hora de implementar propuestas de reforma, y además, no sólo significan lo que es transmitido por el lenguaje oficial sino también difunden la significación otorgada en la futura población docente. En ese contexto el concepto de “constructivismo” fue ubicado como significante privilegiado del lenguaje oficial de las reformas latinoamericanas de los 90, a pesar de que el este significante había sido introducido en Argentina varias décadas atrás.

En la formación docente en Argentina el Normalismo, la Terciarización y la Reforma conforman sus hitos constitutivos, marcas instituyentes que dejaron sus huellas en el presente. La instalación del significante constructivismo atravesó los dos primeros estratos de la formación docente. En el primer estrato, con la incorporación de las primeras obras de Piaget, en el segundo estrato considerando a la teoría a los fines de aplicarla, sin mediación, a la escuela. En este trabajo se profundizará sobre el tercer estrato de la formación docente.

---

[1] La reforma es talen tanto constituyó una “alteración fundamental de la política educativas nacionales que pueden afectar al gobierno y administración del sistema educativo y escolar, y a su estructura y financiación, al currículum –contenidos, metodología y evaluación- al profesorado –formación, selección o evaluación – y a la evaluación del sistema educativo” (Viñao; 2001: 26).

Bajo el término “constructivismo” se evoca una diversidad de teorías lo que impide considerarlo como una propuesta monolítica. A pesar de ello este término continúa siendo usado en singular. Al tratar de hallar aquello común a todas las propuestas, se apela usualmente a aspectos vinculados a la relación sujeto-objeto como “un ensayo por superar el dualismo entre el objeto y el sujeto de conocimiento.” (Castorina; 1996:15).

### Método

Se empleó un diseño de investigación cualitativo concretado en estudio de caso porque permite captar particularmente la naturaleza de los significados, que surgen de la interacción simbólica entre las personas (Ruiz Olabuénaga, 1999). Se relevaron las significaciones que los profesores de una institución de formación docente le asignan al significante “Constructivismo”. Se trata de una muestra no probabilística, elegida intencionalmente. El muestreo opinático (Ruiz Olabuénaga, 1999) permitió seleccionar a los informantes que componen la muestra con un criterio estratégico ya que se buscaba la profundización y riqueza en la información en lugar de pretender que las conclusiones sean generalizables. Se emplearon tres instrumentos de recolección de datos. Se realizaron entrevistas cara a cara semi-estructuradas a profesores de la institución que se hallaban a cargo de aquellas asignaturas en las que se aborda al constructivismo como contenido. Se administró una encuesta semi-estructurada a los profesores de los tres turnos, de la institución seleccionada. Se optó por una observación no participante, con registro de tipo etnográfico, donde relevó fundamentalmente los significados que surgen en las prácticas del aula. Se observaron clases donde se abordaba como contenido “conductismo” así como “constructivismo”. Se decidió incluir el contenido conductismo en la observación (además de constructivismo) así como la administración de encuestas a profesores que tienen a quienes tienen otras formaciones académicas para procurar evitar el sesgo que implicaría analizar solamente aquello que se quiere estudiar.

### Conclusión

Los profesores, en tanto sujetos, al tratar de expresar lo que piensan, creen o suponen no son totalmente libres. Se les imponen restricciones políticas, institucionales y discursivas que funcionan como filtros en el momento de optar por el uso de uno u otro término, de una u otra acepción, las que condicionan su elección. En tanto sujetos, se hallan inmersos en un mundo atravesado por el lenguaje, marcado por el equívoco y la polisemia. Las expresiones de estos hablantes son producto de sus subjetividades y de su historia pero también de los significantes que provee su comunidad. De esta forma construyen significados que tienen una relación arbitraria, pero motivada, con el significante, en espacios de interacción social (Wittgenstein, 1988), en relaciones de intercambio.

Al procurar relevar los significados<sup>[2]</sup> atribuidos al Constructivismo surge en el decir de los profesores una heterogeneidad de supuestos, creencias, tradiciones personales, institucionales, socio-políticas con las que cada uno fue conformando aquello que transmite. Los significados no resultan neutros ni asépticos, portan esos hilos con los que cada uno fue tejiendo el discurso, reconfigurando y resignificando los mismos. Los profesores construyen los significados en la interacción con los otros y es allí donde el significado surge. Los otros pueden hallarse presentes físicamente, como sucede en el aula, o mediatizados a través de la palabra escrita. En la clase, los profesores al procurar dar cuenta de lo que los autores han dicho, emplean un discurso indirecto narrativizado donde se hacen cargo de las palabras dichas por ellos y las integran en su propio discurso.

---

[2] Diversas investigaciones han permitido pensar que el significado se construye en interacción con el contexto especialmente aquellas que provienen de la Sociolingüística, las cuales han empleado la etnometodología: Mehan (1986), Cazden (1991) en EEUU, Delamont (1984), Stubbs (1987), Sinclair y Couthard (1975) en el Reino Unido.

La institución en la que se ha realizado esta indagación, tiene como misión formar docentes de todos los niveles. Al hacerlo asegura, en sus prácticas, la perpetuación de la escuela moderna en tanto produce sujetos sensibles a determinadas representaciones (Donald, 1995) y propicia un discurso sostenido en el principio de autoridad basado en el docente, en generalizaciones homogenizantes que anulan las diferencias y clasificaciones conformadas por una lógica dicotómica, donde para aludir a un término hay que buscar tomar otro que se le contraponga. Formar, dar forma, con-formar al estilo de las metáforas educativas. La presentación del conductismo en simple oposición al constructivismo, a asociado imágenes de la escuela autoritaria pero necesaria y eficiente, responde a la configuración de la escuela normalizadora y moralizante que marca el camino a recorrer para llegar a ser lo que se debe. El tercer estrato de la formación docente se ha entramado con los estratos anteriores, especialmente con el de la terciarización tecnicista, aportando los significados en cuya clave se leen los significantes que procuran instalarse en el discurso pedagógico. En el marco de la presión reformadora de la década de 1990, los profesores procuraron apropiarse de los nuevos saberes y se aproximaron a las nuevas propuestas reduciendo lo nuevo a lo conocido, despliegue táctico con lo que respondieron a la estrategia reformadora.<sup>[3]</sup> Ésta se dirigió a los profesores proponiendo la modificación de las prácticas a través de –entre otras cuestiones– la incorporación de nuevas nominaciones y conocimientos, pero los profesores conformaron significaciones propias a las otorgadas por las propuestas orgánicas y estratégicas de reforma. La táctica les permitió producir significaciones discrepantes con respecto a aquéllas propuestas, donde tuvieron un fuerte peso las situaciones personales, institucionales y profesionales. La dinámica de la reforma y su recepción puede ser analizada bajo la óptica de estrategia y táctica en función de sus posibilidades comunicativas. La reforma fue propiciada desde una política documentalista y estuvo sostenida en la escritura. Los módulos de capacitación docente de la provincia de Buenos Aires procuraron instalar al constructivismo como parte del discurso pedagógico, ubicándolo en el polo del “deber ser”, haciéndolo jugar, de este modo, como opuesto al conductismo, casi como sinónimo de la escuela existente y que se desea cambiar. Los profesores respondieron, entonces, con los recursos que disponían. Provistos de determinadas concepciones del mundo –aportadas por las experiencias personales y profesionales– y atravesados por las lógicas institucionales, los docentes significaron y significan las teorías de acuerdo a sus creencias, vivencias, representaciones, olvidando o reprimiendo lo que no se adecua con los patrones esperados. Es en este juego, de táctica y estrategia, en esta negociación de significaciones donde construyeron significados idiosincrásicos, en los que intervienen sus historias personales como su formación y experiencia. De este modo las tácticas simplificadoras desplegadas tuvieron una fuerte tendencia al reduccionismo binario. En esta “aparente incapacidad de constituirse uno mismo sin excluir al otro y la aparente incapacidad de excluir al otro, sin desvalorizarlo y, finalmente, sin odiarlo” (Castoriadis, 1985) se puede reconocer cómo se constituye o se instala un significante que opera en el polo de la confrontación con otro. De este modo, en el par conductismo – constructivismo, éste último es significado desde un lugar de oposición y como teoría superadora del Conductismo. Así, del lado del conductismo se halla lo estructurado, previsto, impuesto, automático, el estímulo, la respuesta, el control, el autoritarismo, la memoria, la rigidez, lo reforzamiento, lo repetitivo, la rapidez, la eficiencia, la centralidad del docente, las planificaciones rígidas y con evaluaciones objetivas. Del lado del constructivismo quedan la acción, la actividad, el diálogo, la diversidad, la asimilación y acomodación, la adaptación, el aprendizaje significativo, el aprendizaje por descubrimiento, la construcción, el equilibrio, el desequilibrio, la reestructuración, los saberes previos, la producción, los niños trabajando, el pasaje de un estado de desorganización a uno más estructurado, de una situación poco clara a una más clara, más elevada, más compleja, más ordenada, el juego, el grupo y el proyecto. El constructivismo emerge, entonces, como una propuesta que tiene ciertas limitaciones ya que, de acuerdo a las características grupales y a los objetivos que se plantee el docente deberá apelar a una propuesta conductista. Los autores asociados con conductismo son quienes conformaron la teoría y en menor medida quienes las llevaron a las aulas.

---

[3] La estrategia es la ciencia de los movimientos guerreros fuera del campo de visión de enemigo (Von Bulow citado por de Certeau, 1996), en tanto la táctica constituye el interior del campo, como lugar del poder y de la voluntad propios (de Certeau, 1996).

El constructivismo es colocado en el polo instituyente, y es llevado a transitar por el camino iniciado por las viejas dicotomías educativas. En estas oposiciones se articulan series rivales de creencias, percepciones y proposiciones las cuales se fueron conformando en el tiempo para luego olvidar el proceso que permitió construirlas. Pero a pesar de la fuerza de lo instituido, algunos profesores trataron de ir más allá de las dicotomías. De este modo, este tercer estrato no constituye una frontera que delimita un antes y un después de la reforma, sino que –por el contrario– conforma un piso donde los significados se yuxtaponen y superponen a nominaciones anteriores.

Este trabajo ha procurado indagar significados, tratando de evitar juicios de valor y de señalar a los docentes como resistentes al cambio, como sucede en numerosas ocasiones. Tampoco se ha buscado “detrás” de lo que dicen el verdadero sentido de lo dicho, ya que en el discurso de los profesores se halla aquello que creen. “No se busca en modo alguno, por debajo de lo manifiesto la garrulería casi silenciosa de otro discurso” (Foucault, 1997: 45).

Sabemos que la institución escolar aporta las formas que constituyen lo que se piensa y se hace, pero debemos reflexionar acerca de si es posible, en ellas, apartarse de la interpretación central y lineal con que se aborda los hechos.

### Referencias

- Braslavsky, C. y Gvirtz, S. (2001). *Educación Comparada. Política Educativa en Iberoamérica*. Madrid, Organización de los estados Iberoamericanos (OEI).
- Caruso, M. y Fairstein, G. (1997). Las puertas del cielo. Hipótesis acerca de la recepción de la psicogénesis y el constructivismo de raíz piagetiana en el campo pedagógico argentino. En A. Puiggrós, *Dictadura y utopías en la historia de la educación argentina (1955-1983)*. Buenos Aires: Galena.
- Castoriadis, C. (1985). *Reflexiones en torno al racismo*, Memorias del Coloquio “Inconsciente y cambio social”, de la *Association pour la Recherche et l’Intervention Psychologiques*, Francia.
- Castorina, J. A. (1996). El debate Piaget-Vigotsky, la búsqueda de un criterio para su evaluación. En A. Castorina, et al. *Piaget-Vigotsky: contribuciones para replantear el debate*. Paidós: Buenos Aires.
- De Certeau, M. (1996). *La invención de lo cotidiano I. Artes de hacer*. México: Universidad Iberoamericana.
- Donald, J., Faros del futuro: enseñanza, sujeción y subjetivación. En J. Larrosa (1995). *Escuela, poder y subjetivación*. Madrid: Ediciones de La Piqueta.
- Foucault, M. (1993). Lo que digo y lo que dicen que digo. En D. Lecourt et al. *Disparen contra Foucault*. Buenos Aires: Ediciones El Cielo por asalto.
- Foucault, M. (1997). *La arqueología del saber*. Buenos Aires: Editorial Siglo XXI.
- Ruiz Olabuénaga, J.I. (1999). *Metodología de la investigación cualitativa* (pp. 102-117). Segunda edición. Bilbao: Universidad de Deusto.

Viñao, A. (2001). *¿Fracasan las Reformas Educativas? La Respuesta de un Historiador*. Sociedad Brasileña de Historia da Educao. SBHE. Campinas – Sao Paulo.

Wittgenstein, L. (1988). *Investigaciones Filosóficas*. Barcelona: Editorial Crítica.

# DIFERENCIAS DE GÉNERO EN LA AUTOPERCEPCIÓN DE LA VOLUNTAD DE TRABAJO

Horacio F. Attorresi, Facundo J. P. Abal, Gabriela S. Lozzia,  
María Ester Aguerri y María Silvia Galibert  
Universidad de Buenos Aires, Argentina

Voluntad de Trabajo (VT) es un rasgo de la personalidad que describe la tendencia de un individuo a abocarse a sus obligaciones realizándolas con automotivación, responsabilidad y precisión. El objetivo de este trabajo es indagar las diferencias de género respecto de la VT autopercebida y comparar los resultados con estudios internacionales basados en constructos semejantes a dicho rasgo. Participaron 247 individuos (53% mujeres) residentes en la Ciudad de Buenos Aires y 870 estudiantes de Psicología (81.4% mujeres). Se administraron la Escala VT (Alfa=.83) y la escala Perseverancia del Big Five Questionnaire (Alfa=.71). Acorde a lo esperado según la revisión bibliográfica, no se hallaron diferencias significativas intergénero en la autopercepción de VT en la muestra de población general. El tamaño del efecto  $d$  de Cohen de la diferencia para VT en esta muestra resultó de 0.12 (a favor de las mujeres). Sin embargo, en la muestra de estudiantes de Psicología las mujeres reportaron un puntaje promedio significativamente más alto que los varones, alcanzando un valor  $d=0.55$ . Estos resultados fueron consistentes con los obtenidos a partir de la escala Perseverancia. Se concluye que la autopercepción diferencial de la VT según género es una característica particular de la población de estudiantes de Psicología.

La Voluntad de Trabajo (VT) es un rasgo de la personalidad que caracteriza la disposición de un individuo para dedicarse con responsabilidad y automotivación al cumplimiento de sus obligaciones aun cuando éstas pudieran desagradarle (Lozzia, Abal, Aguerri, Galibert & Attorresi, 2007). Investigaciones previas efectuadas en el marco de estudios de validación de una escala que mide esta variable mostraron que las estudiantes universitarias de Psicología tendían a puntuar significativamente más alto que sus pares varones (Galibert, Abal, Lozzia, Aguerri & Attorresi, 2006).

Sin embargo, estos resultados no son consistentes con los hallazgos obtenidos de estudios de género internacionales con constructos semejantes a VT. Diversas investigaciones realizadas en el marco de la Teoría de los Cinco Grandes de la personalidad no reportaron diferencias significativas entre varones y mujeres al medir la dimensión Escrupulosidad (Conscientiousness) ni en población general (Caprara, Barbaranelli & Borgogni, 1993; Costa, Terracciano & McCrae, 2001) ni en población universitaria (De Miguel, 2005).

Feingold (1994) realizó uno de los estudios de metaanálisis más extensos sobre diferencias intergénero en dimensiones de la personalidad. Este autor combinó datos extraídos de los inventarios de 16-PF, el EPQ-R y NEO-PI-R y obtuvo una medida del rasgo Responsabilidad que muestra características similares a VT. En esta investigación se concluyó que las mujeres tienden a percibirse con un nivel de responsabilidad más elevado que los varones. No obstante, la magnitud de esta diferencia fue reducida, ya que el tamaño del efecto estimado mediante la  $d$  de Cohen resultó de sólo 0.1.

El objetivo del presente trabajo es explorar las diferencias en la Voluntad de Trabajo percibida según el género del individuo. Este estudio se basa en la comparación de los perfiles obtenidos en una muestra de estudiantes universitarios y una muestra de población general.

## Método

### *Participantes*

*Muestra 1.* Colaboraron 870 alumnos de Estadística de la Licenciatura de Psicología (UBA). El 81.4% fueron mujeres. La edad media fue de 20.7 años (DE = 2.4 años) y la mediana fue 20 años.

*Muestra 2.* Participaron 247 adultos jóvenes de ambos sexos (53% mujeres) residentes en Ciudad de Buenos Aires y alrededores. El promedio de edad fue de 23.2 años (DE= 3.6 años).

### *Instrumentos*

*Escala de Voluntad de Trabajo.* Consta de 9 ítems con 7 opciones de respuesta tipo Likert (de *nunca a siempre*). La escala cuenta con evidencias suficientes de validez factorial y de criterio (Abal, Lozzia, Aguerrri, Galibert & Attorresi, 2007). El Alfa de Cronbach obtenido en la presente investigación resultó de .83.

*Escala Perseverancia del Big Five Questionnaire* (Caprara et al., 1993, adaptación de Bermúdez, 1995). Esta escala mide aspectos como la persistencia con que se efectúan las tareas y el cumplimiento de las promesas. Se compone de 12 ítems con 5 opciones de respuesta (de *completamente falso para mí a completamente verdadero para mí*). El Alfa Cronbach resultó de .71.

### *Procedimiento*

Se llevó a cabo un diseño no-experimental transeccional con fines descriptivos (Hernández-Sampieri, Fernández-Collado & Baptista, 2006). Para sendas muestras se realizó un muestreo no-aleatorio por accesibilidad. Con el propósito de controlar el efecto de la edad sobre las variables dependientes de este estudio se acotó el rango de variación entre 18 y 30 años para ambas muestras.

## Resultados

Considerando el grupo de estudiantes de Psicología, las mujeres obtuvieron un puntaje promedio significativamente más alto que los varones tanto en VT ( $t(224.1) = 5.9, p < .001$ ) como en Perseverancia ( $t(867) = 5, p < .001$ ). Pero estas diferencias intergénero no se observaron en la muestra de población general. Como se puede observar en la Figura 1, el tamaño del efecto  $d$  de Cohen de la diferencia entre los géneros en VT resultó pequeño para los adultos jóvenes de la población general ( $d = 0.12$ ) y medio para los estudiantes de Psicología ( $d = 0.55$ ). Los  $d$  obtenidos para las diferencias intergénero en la escala de Perseverancia resultaron de magnitud semejante. Esto es, mientras que el valor de  $d$  fue considerable para los estudiantes ( $d = 0.43$ ), en el caso de la población general resultó muy bajo y a favor de los varones ( $d = -0.08$ ).

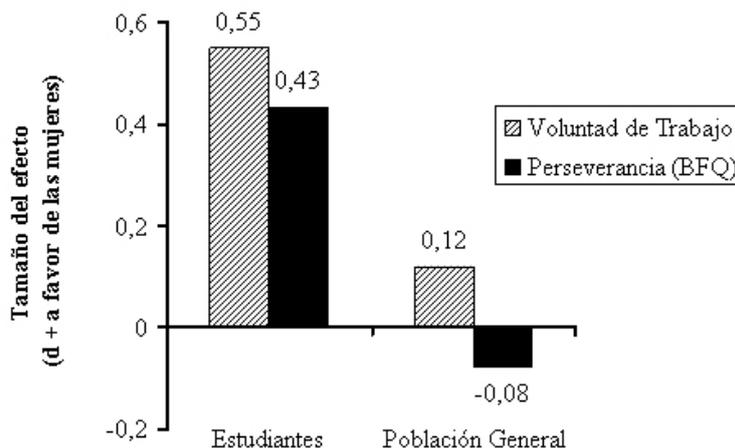


Figura 1. Tamaños del efecto de la diferencia intergénero en VT y Perseverancia

Al contrastar los puntajes promedio en VT de las mujeres estudiantes con las mujeres de población general no se encontró una diferencia significativa. Sin embargo, sí se halló diferencia entre los varones, en tanto que los estudiantes obtuvieron puntuaciones más bajas que los varones de población general ( $t(276) = 3.3, p < .001$ ). Como se observa en la Figura 2, la comparación de los promedios obtenidos por cada grupo permitió identificar que la VT autopercebida por los estudiantes de Psicología varones se distancia del resto de los grupos conformados por la interacción género x muestra.

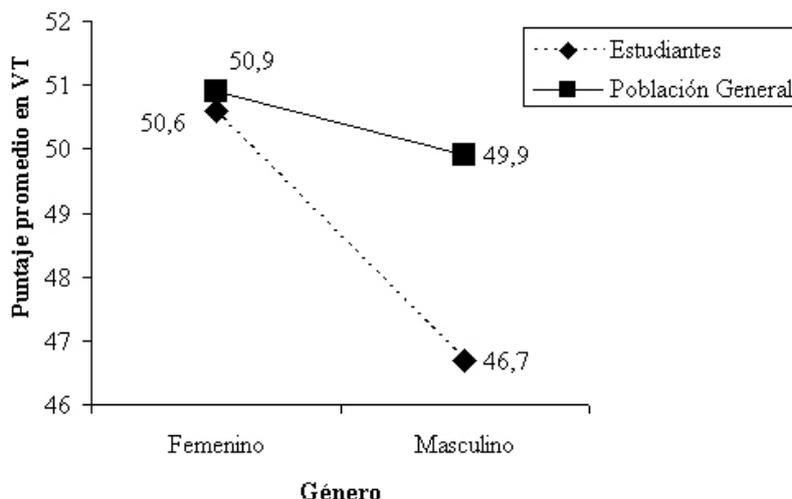


Figura 2. Medias en VT obtenidas en la interacción Género x Muestra

### Discusión

La ausencia de diferencias significativas en la percepción la VT de varones y mujeres de población general resulta acorde con los hallazgos reportados en la literatura (Caprara et al., 1993; Costa et al., 2001; de Miguel, 2005; Feingold, 1994). Este resultado no sólo es útil desde la perspectiva de los estudios de género sino que también aporta nuevas evidencias de validez de constructo para la Escala de VT.

Aunque no se corresponden con los estudios internacionales, las diferencias de género encontradas entre los estudiantes de Psicología son consistentes con los resultados obtenidos en investigaciones locales realizadas previamente (Abal et al., 2007; Galibert et al., 2006). El hecho de que se hayan encontrado similares resultados con la escala Perseverancia permite comprobar

que la diferencia no es generada artificialmente por el instrumento utilizado. En consecuencia, estos hallazgos permiten pensar que se trata de una característica particular de la población de estudiantes de Psicología.

No obstante, contrario a lo supuesto con anterioridad (Galibert et al., 2006), las diferencias intergénero de los estudiantes no parecen explicarse por una autopercepción más positiva de las mujeres. La comparación con los datos de población general mostró que son los alumnos varones quienes muestran una valoración más negativa de su VT. Futuros estudios ahondarán en esta visión negativa de los alumnos varones sobre su VT. Sería importante determinar si se trata sólo de una característica propia de los estudiantes de Psicología o podría extenderse a otros universitarios masculinos.

Otra conclusión importante que se desprende de los resultados es la distorsión que se podría generar en los puntajes de VT al adoptar a muestras de universitarios como grupo normativo. Esta es una práctica errónea aunque habitual en nuestra disciplina no recomendable para la Escala VT.

Agradecimientos. Este trabajo se llevó a cabo con las subvenciones UBACyT P043 y AGENCIA PICT 20909, de la Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires y de la Agencia Nacional de Promoción Científica y Tecnológica, respectivamente.

#### Referencias

- Abal, F., Lozzia, G., Aguerri, M. E., Galibert, M. S. & Attorresi, H. (2007). Evidencias acerca de la validez de constructo de la escala Voluntad de Trabajo. *Anuario de Investigaciones*, XV, 283-289.
- Bermúdez, J. (1995). *Manual del Cuestionario "Big Five" (BFQ)*. Madrid: TEA
- Caprara, G. V., Barbaranelli, C. & Borgogni, L. (1993). Big Five Questionnaire (BFQ). Manuale. Florencia: Organizzazioni Speciali.
- Costa, P.T., Terracciano, A. & McCrae, R.R. (2001). Gender Differences in Personality Traits across Cultures. Robust and surprising findings. *Journal of Personality and Social Psychology*, 81, 322-331.
- de Miguel, A. (2005). Diferencias de edad y género en el NEO-PI-R en dos muestras con distinto nivel académico. *International Journal of Psychology and Psychological Therapy*, 1, 13-31.
- Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological bulletin*, 116, 429-456.
- Galibert, M. S., Abal, F., Lozzia, G., Aguerri, M. E. & Attorresi, H. (2006). Voluntad de Trabajo en estudiantes de Psicología. *Anuario de Investigaciones XIV*. Tomo I, 269-275.
- Hernández-Sampieri, R., Fernández-Collado, C. & Baptista L., P. (2006). *Metodología de la investigación*. Cuarta edición. México: McGraw Hill.
- Lozzia, G., Abal, F., Aguerri, M. E., Galibert, M. S. & Attorresi, H. (2007). Delimitación del constructo Voluntad de Trabajo. *Summa Psicológica UST*, 4, 137-148.

# USO DEL ANÁLISIS ENVOLVENTE DE DATOS PARA LA EVALUACIÓN DE LOS PROGRAMAS DE FOMENTO DE LA I+D+I

Rosario C. Barea<sup>1</sup>, Esther Cabrera<sup>2</sup>, José Luis Pino<sup>1</sup> y Francisco M. Solís<sup>2</sup>

<sup>1</sup>Universidad de Sevilla

<sup>2</sup>Consejería de Innovación, Ciencia y Empresa

El Análisis Envolvente de Datos (DEA), es un método no paramétrico, basado en la programación lineal, para medir la productividad y la eficiencia relativa de unidades de organizaciones como empresas, universidades, hospitales, etc., que utilizan múltiples recursos para producir múltiples productos. El objetivo fundamental del DEA es hallar una frontera de eficiencia, formada por aquellas combinaciones de recursos que optimizan la cantidad de productos obtenidos minimizando los costos de producción, y a partir de esta frontera, evaluar la eficiencia relativa de las combinaciones de recursos que no pertenezca a la misma. En este trabajo, se analiza la eficiencia técnica de los grupos de investigación del área científico-técnica de las Tecnologías de la Información y Comunicación de Andalucía.

## *El Análisis Envolvente de Datos*

### *La Técnica DEA*

Es una técnica de programación matemática que permite la construcción de una superficie envolvente, frontera eficiente o función de producción empírica a partir de los datos disponibles del conjunto de Unidades de Toma de Decisión (DMUs) objeto de estudio, de forma que las DMUs que determinan la envolvente son denominadas eficientes y aquellas que no permanecen sobre la misma son consideradas ineficientes, DEA permite la evaluación de la eficiencia relativa de cada una de las DMUs.

### *Caracterización de los Modelos DEA*

Los modelos DEA pueden ser clasificados, básicamente, en función de:

- El tipo de medida de eficiencia que proporcionan: modelos radiales y no radiales.
- La orientación del modelo: Input orientado, Output orientado o Input-Output orientado.
- La tipología de los rendimientos a escala que caracterizan la tecnología de producción, entendida ésta como la forma (procedimientos técnicos) en que los factores productivos (Inputs) son combinados para obtener un conjunto de productos (Outputs), de tal forma que esa combinación de factores puede caracterizarse por la existencia de rendimientos a escala: constantes o variables a escala.

### *Orientación del modelo*

Siguiendo a Charnes, Cooper y Rhodes (1981), la eficiencia puede ser caracterizada con relación a dos orientaciones básicas, pudiendo hacer referencia a dos tipos de modelos:

1. Input Orientado: Dado un nivel de Outputs, se busca la máxima reducción proporcional en el vector de Inputs.
2. Output Orientado: Dado el nivel de Inputs, se busca el máximo incremento proporcional de los Outputs.

*Tipología de los rendimientos a escala*

Los rendimientos a escala, indican los incrementos de la producción que son resultado del incremento de todos los factores de producción en el mismo porcentaje, pueden ser constantes, crecientes o decrecientes:

1. Rendimientos constantes a escala:

$$\text{Si } f(cX) = c^t f(X) \text{ con } t = 1$$

2. Rendimientos crecientes a escala:

$$\text{Si } f(cX) = c^t f(X) \text{ con } t > 1$$

3. Rendimientos decrecientes a escala:

$$\text{Si } f(cX) = c^t f(X) \text{ con } t < 1$$

donde  $f(X)$  es la tecnología de producción,  $X$  es un vector de Inputs y  $c$  un escalar.

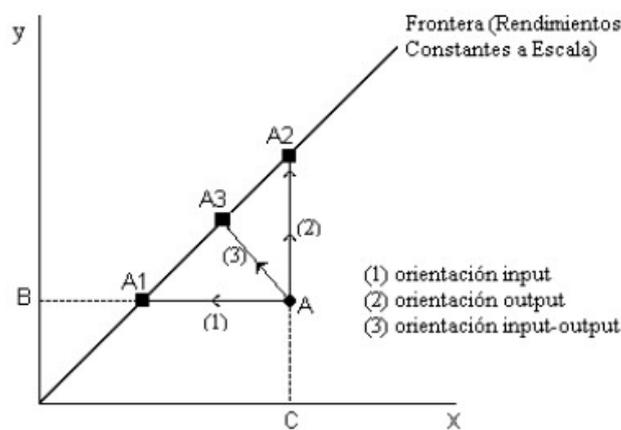


Figura 1. Tipos de orientaciones

*Clasificación de los modelos DEA*

Modelos DEA-CCR

El modelo DEA-CCR (Charnes, A., Cooper, W. W. y Rhodes, E. (1979)) proporciona medidas de eficiencias radiales, Input u Output orientadas y supone convexidad, fuerte eliminación gratuita de Inputs y Outputs y rendimientos constantes a escala.

Modelo DEA-CCR Input Orientado en Forma Envolvente

$$\begin{aligned} & \text{Min}_{\theta, \lambda} \quad z_0 = \theta \\ & \text{sa:} \\ & \quad Y\lambda \geq y_0 \\ & \quad \theta x_0 \geq X\lambda \\ & \quad \lambda \geq 0 \end{aligned}$$

Donde  $\lambda$  es el vector ( $n \times 1$ ) de pesos o intensidades, así  $\lambda_j$  es la intensidad de la DMU  $j$ , y  $\theta$  denota la puntuación de eficiencia (técnica) de la DMU evaluada. La eficiencia quedaría caracterizada de la siguiente forma:

Si  $\theta^* = 1$  y todas sus holguras son cero, la DMU será eficiente.

Si  $\theta < 1$  ó  $\theta^* = 1$  y alguna de sus holguras son distinta de cero, la DMU será ineficiente.

#### Modelo DEA-CCR Output Orientado en Forma Envolvente

$$\begin{aligned} & \text{Max}_{\varphi, \lambda} \quad z_0 = \varphi \\ & \text{sa:} \\ & \quad \varphi y_0 \geq Y\lambda \\ & \quad X\lambda \geq x_0 \\ & \quad \lambda \geq 0 \end{aligned}$$

La eficiencia técnica output de la DMU evaluada será  $\frac{1}{\varphi^*}$ .

#### Modelos DEA-BCC

El modelo DEA-BCC (Banker, R.D., Charnes, A. y Cooper, W.W. (1984)) no es más que una extensión del modelo DEA-CCR, mientras que el DEA-CCR considera rendimientos constantes a escala, el DEA-BCC permite que la tipología de rendimiento a escala sea variable.

#### Modelo DEA-BCC Input Orientado en Forma Envolvente

$$\begin{aligned} & \text{Min}_{\theta, \lambda} \quad z_0 = \theta \\ & \text{sa:} \\ & \quad Y\lambda \geq y_0 \\ & \quad \theta x_0 \geq X\lambda \\ & \quad \bar{1}\lambda \geq 1 \\ & \quad \lambda \geq 0 \end{aligned}$$

#### Modelo DEA-BCC Output Orientado en Forma Envolvente

$$\begin{aligned} & \text{Max}_{\varphi, \lambda} \quad z_0 = \varphi \\ & \text{sa:} \\ & \quad \varphi y_0 \geq Y\lambda \\ & \quad X\lambda \geq x_0 \\ & \quad \bar{1}\lambda \geq 1 \\ & \quad \lambda \geq 0 \end{aligned}$$

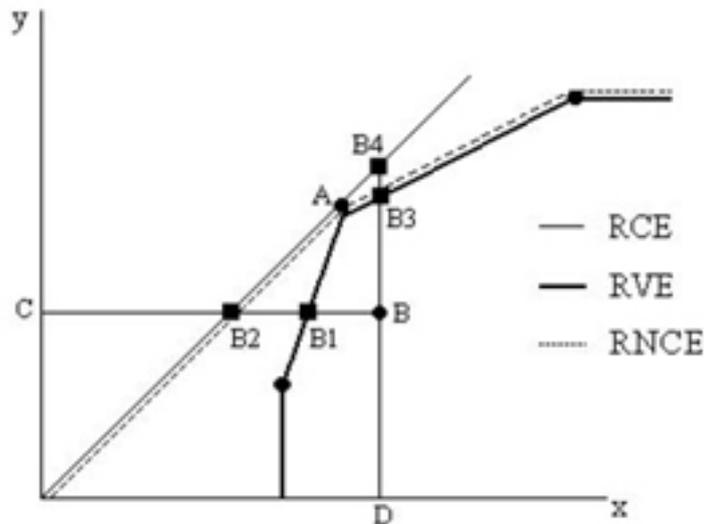


Figura 2. Tipología de rendimientos a escala

## Aplicación

### Objetivos

Analizar la eficiencia técnica de los grupos de investigación del área científico-técnica de las Tecnologías de la Información y Comunicación.

### Unidades Seleccionadas

La selección de las DMUs que van a formar parte de una investigación DEA debe estar íntimamente relacionado con las características del estudio que se va a desarrollar, para lo cual hay que tener en cuenta el objeto de la investigación. Además se requiere que las DMUs seleccionadas cumplan un conjunto de condiciones necesarias para el buen funcionamiento de esta, concretamente Torrico (2000) habla de tres condiciones:

1. Las DMUs deben ser lo suficientemente homogéneas para que se puedan comparar pero suficientemente heterogéneas para que se pueda extraer alguna información de esa comparación.
2. Deben tener responsabilidad en la gestión de los recursos para la producción, lo que implica capacidad para gestionar los recursos de que disponen.
3. Deben constituir un grupo suficientemente amplio para no limitar la capacidad discriminadora del análisis, condición que se encuentra relacionada con la dimensión del modelo.

Una vez analizadas las condiciones necesarias para una buena selección de unidades, en esta investigación las unidades seleccionadas corresponderán con cada uno de los grupos de investigación del área científico-técnica de la Tecnologías de la Información y Telecomunicaciones (TIC), en total 86 grupos de investigación.

### *Dimensión del modelo*

Para seleccionar tanto las unidades como las variables del análisis, nos encontramos en disposición de clarificar la dimensión del modelo para determinar si alcanza una dimensión suficiente o no.

La determinación de la suficiencia de dimensión se puede hacer a través de algunas reglas existentes para ello, como son:

- Golany y Roll (1989):  
 $n \geq 2 \times (p+q)$
- Charnes et al. (1990):  
 $n \geq 3 \times (p+q)$
- Murias (2004):  
 $n \geq p \times q$

En esta investigación, el análisis es referido a 86 unidades (grupos de investigación) y 8 variables (3 inputs y 5 outputs), por lo tanto, se cumplen todas las reglas sugeridas en relación a la dimensión suficiente del modelo.

- Golany y Roll (1989):  
 $86 \geq 2 \times (3+5)$
- Charnes et al. (1990):  
 $86 \geq 3 \times (3+5)$
- Murias (2004):  
 $86 \geq 3 \times 5$

### *Selección de las variables*

La selección de indicadores tanto de entradas como de salidas (inputs/outputs) en el sector público y referente a actividades relacionadas con los servicios, implica no solo establecer la relación recursos-productos, sino también determinar indicadores capaces de representar de forma adecuada la actividad realizada (Pina y Torres, 1995, Murias, 2004).

Como primera tarea para seleccionar las variables de la investigación estaría la creación de un listado, lo más exhaustivo posible, que recoja todas las variables que puedan influir en la actividad de las unidades de decisión, ya sea por el lado de las entradas o de las salidas (Golany y Roll (1989)).

Para la correcta selección de las variables, se considera indispensable contar con el juicio de expertos en la actividad, muy especialmente cuando el análisis de la eficiencia pretende realizarse desde fuera de la organización objeto de estudio (Golany y Roll, 1989; Bowlin, 1998).

Para la selección de las variables y del papel que desempeñan, es preciso disponer de un marco teórico de los objetivos de los programas de fomento de la investigación científica, el desarrollo y la innovación.

En nuestro caso asumimos que las políticas que desarrolla la Junta de Andalucía tienen como objetivo básico impulsar la capacidad de los grupos de I+D+i andaluces en la captación de recursos dentro

de los programas nacionales e internacionales, de ahí que se considere Inputs las subvenciones autonómicas y Outputs los recursos económicos obtenidos de fuentes ajenas a la Comunidad Autónoma.

Dada la heterogeneidad de los contratos con empresas y la dificultad de disponer de datos para algunos grupos se ha optado por incluir los fondos procedentes de instituciones públicas que tiene un carácter competitivo.

La última de las tareas que debe realizarse antes de proceder a la elección definitiva de las variables, tiene que ver con la recomendación realizada por Golany y Roll (1989) y Pedraja-Chaparro et al. (1999), referente a la realización previa de un análisis de correlaciones entre las variables.

*Tabla 1. Relación de Inputs y Outputs seleccionados*

<b>INPUTS</b>
Resto de Personal de I+D
Nº de Doctores Activos
Subvenciones recibidas por los grupos de investigación
<b>OUTPUTS</b>
Ingresos por proyectos de financiación pública
Publicación en revistas ISI
Nº de tesis total
Total otros proyectos
Nº de contratos total

### *Rendimientos de Escala*

Llegados a este punto, se requiere definir si se optará por la utilización de un modelo DEA-CCR, o un modelo DEA-BCC.

El presente estudio seguirá la hipótesis de rendimientos a escala constante (DEA-CCR), para los análisis de eficiencia que se lleven a cabo.

### *Orientación del modelo*

Centrando la atención en los inputs seleccionados, los grupos no poseen pleno poder para contratar al personal que deseen cuando lo necesitan. En este sentido la orientación output sería la más recomendable pues unos resultados que indiquen la reducción de las entradas no serían lógicos, al no poder modificar unos recursos que no se controlan.

Se estudiará la eficiencia de los grupos de investigación por la vía de las salidas (outputs). Esto es mediante la maximización de los resultados.

### *Análisis de los resultados*

Una vez ejecutado el modelo DEA, se obtienen una serie de resultados que deben ser analizados. En este sentido el análisis comportará un estudio de los límites, características y posibles soluciones del problema que surge ante las diferencias de eficiencia que aparecen entre los grupos de investigación dentro del área científico-técnica de las TIC.

Como muestra la figura 3, en la distribución de los resultados se refleja la existencia de 19 unidades 100% eficientes y 31 unidades cuyo grado de eficiencia varía entre el 60% y el 99 %. No obstante, existen 36 grupos de investigación muy ineficientes, no llegando ni casi a un 60 %.

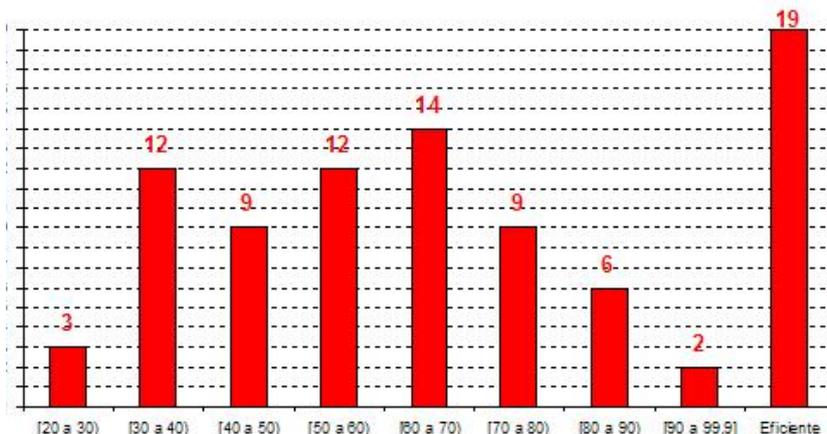


Figura 3. Distribución de los resultados

Según la técnica DEA, las unidades eficientes han alcanzado el máximo desempeño para el conjunto de unidades comparadas, en cambio las unidades menos eficientes se encuentran ante la necesidad de llegar a un desempeño eficiente y por lo tanto deben alcanzar la frontera de eficiencia para ponerse al mismo nivel de eficiencia que las mejores unidades. Para ello necesitan llevar a cabo una serie de mejoras, deben modificar su gestión. En este caso, la orientación output implica que las mejoras que deben realizar los grupos de investigación se encaminen hacia la mejora de la productividad, producir más utilizando los mismos recursos disponibles.

El grado de eficiencia que obtiene como resultado una unidad no eficiente se encuentra en función de las unidades que muestren mejor desempeño y además dispongan de unos recursos similares para la obtención de resultados. En este sentido el DEA crea grupos de referencia, cada unidad no eficiente es comparada con un grupo de unidades con semejantes recursos pero con un desempeño eficiente. Puede darse el caso de que una unidad 100% eficiente forme parte del grupo de referencia de más de una unidad ineficiente.

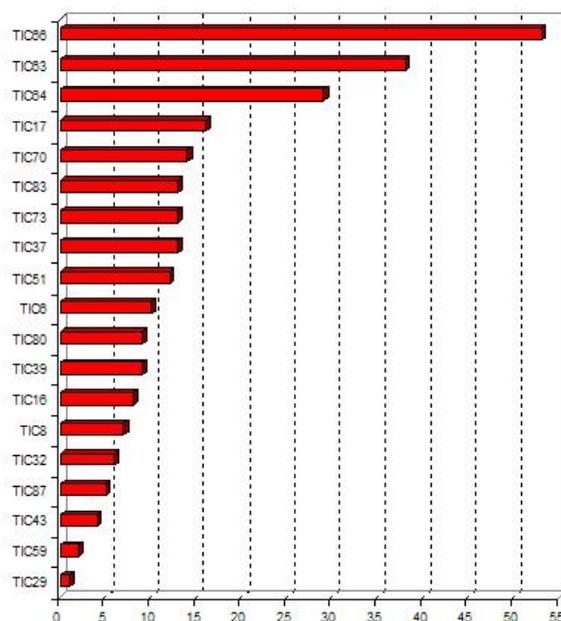


Figura 4. Frecuencia de grupos de referencia

Se observa que el grupo de investigación TIC66 sería el líder global.

El análisis de eficiencia no debe acabar en ese punto, sino que necesita ir un poco más allá y enseñar qué deberían hacer las unidades menos eficientes para alcanzar la frontera de eficiencia y así llegar a ser unidades 100% eficientes, es decir, mostrar cuáles son las mejoras potenciales que deben desarrollarse entre los grupos de investigación menos eficientes.

De forma ilustrativa se mostrará la mejora potencial de la unidad TIC45.

Lo primero que se debe de obtener es el conjunto de referencia de TIC45 (Tabla 2).

*Tabla 2. Conjunto de referencia*

Unidad	Valor de $\lambda$
TIC39	0,78904
TIC66	0,542262
TIC70	0,013008
TIC83	0,249909

Esta tabla significa que, para que la unidad que estamos evaluando llegue a ser eficiente, se debe guiar de las unidades TIC39, TIC66, TIC70 y TIC83.

Una vez que se tiene el conjunto de referencia el siguiente paso es determinar los valores objetivos, es decir, los valores que las variables deben tomar para llegar a ser eficientes, y por consiguiente el plan de mejora (Tabla 3).

*Tabla 3. Valores observados, objetivos y mejora potencial de la unidad evaluada*

Input/ Output	Valor Observado	Valor Objetivo	Mejora Potencial
x1	0	0	0
x2	5	5	0
x3	5,109	5,109	0
y1	559,607	581,242	21,635
y2	3	10	7
y3	1	1	0
y4	2	5	3
y5	17	18	1

### Conclusión

Los resultados muestran como la mayoría de los grupos de investigación presentan niveles de eficiencias medios, existiendo un total de 19 grupos de investigación 100% eficientes. Se ha obtenido cual ha sido el grupo de investigación que lidera, y por último, y de forma ilustrativa, se ha creado el plan de producción que debería seguir un grupo de investigación ineficiente para llegar a ser eficiente, guiándose de las DMUs eficientes.

## Referencias

- Banker, R.D., Charnes, A. y Cooper, W.W. (1984): "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis". *Management Science* 30(9): 1078-1092.
- Bowlin, W. F. (1998), "Measuring performance: an introduction to data envelopment analysis", *Journal of Cost Analysis, Autumn 98*, pp. 3-24.
- Charnes, A.; Cooper, W.W. y Rhodes, E. (1981): "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through". *Management Science*, 27(6): 668-697.
- Golany, R. y Roll, Y. (1989). "An application procedure for DEA", *Omega, International Journal of Management Science*, 17, 3, pp. 237 250.
- Murias Fernández, M.P. (2004). "Metodología de aplicación del análisis envolvente de datos; evaluación de la eficiencia técnica en la universidad de Santiago de Compostela", Ph. D. disertation, Santiago de Compostela University.
- Nunamaker, T.R. (1985). "Using Data Envelopment analysis to measure the Efficiency of Nos-profit Organizations: a Critical Evaluation", *Managerial an Decision Economics*, 6,1.
- Pedraja Chaparro, F., Salinas Jimenez, J. y Smith, P. (1999). "On the quality of the data envelopment analysis model", *Journal of Operacional Research Society*, 50, pp. 634-644.
- Pina, V. y Torres, L. (1995). "Evaluación del rendimientos de los departamentos de contabilidad de las universidades españolas", *Hacienda Pública Española*, 135, pp. 183-190.
- Torrico, A. (2000). *Técnicas cuantitativas para un análisis macroeconómico de la eficiencia y la financiación dentro de un sistema público de educación superior. Una aplicación para la toma de decisiones en la Universidad de Málaga*. Tesis Doctoral. Universidad de Málaga.

# APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES AL ESTUDIO DE LA CONDUCTA DE RATONES SALVAJES Y CARENTES DEL RECEPTOR LPA<sub>1</sub> EN EL LABERINTO DE HOYOS

Estela Castilla<sup>1</sup>, Jorge Sánchez López<sup>1</sup>, Carolina Hoyo<sup>2</sup>, Emma Zambrana<sup>1</sup>, Elisa Matas<sup>2</sup>, Jerold Chun<sup>3</sup>, Fernando Rodríguez de Fonseca<sup>2</sup>, Carmen Pedraza<sup>1</sup>, Guillermo Estivill<sup>2</sup> y Luís J Santín<sup>1</sup>

<sup>1</sup>Universidad de Málaga

<sup>2</sup>Unidad de Investigación, Fundación IMABIS, Hospital Regional Universitario Carlos Haya

<sup>3</sup>Department of Molecular Biology, Helen L. Dorris Child and Adolescent Neuropsychiatric Disorder Institute, The Scripps Research Institute, USA

En el estudio de la conducta animal, la interacción entre actividad, emoción y cognición dificulta la interpretación de resultados. Este trabajo investiga la relación entre estas variables en animales modificados genéticamente, estudiando cómo las alteraciones exploratorias y emocionales afectan a la memoria espacial en un laberinto de hoyos. Se aplicó un Análisis de Componentes Principales (ACP) a las variables obtenidas durante la habituación al laberinto y en la última sesión de un entrenamiento en memoria espacial, por separado en ratones salvajes y carentes del receptor LPA<sub>1</sub>. El ACP durante la habituación reveló 3 factores en cada genotipo (relacionados con actividad, ansiedad y arousal), confirmando que las medidas de actividad y ansiedad son independientes. En la tarea de memoria espacial, el ACP reveló 4 factores en el genotipo salvaje (actividad/memoria de trabajo, latencia, ansiedad y memoria de referencia) y 3 factores en el genotipo mutado (actividad, memoria/rearing y ansiedad), señalando que el incremento de ansiedad en los animales carentes de receptor LPA<sub>1</sub> no deteriora su rendimiento cognitivo. Nuestros datos indican que el ACP es una herramienta útil para investigar las funciones psicológicas en animales modificados genéticamente, permitiendo precisar el tipo de déficit inducido por la modificación génica.

La posible interacción entre la actividad, la emoción y la cognición es un hecho que dificulta el estudio de la conducta animal. Por ejemplo, la permanencia en la periferia del laberinto (lugar protegido por las paredes) es una medida de ansiedad en roedores (Crawley, 1999). Sin embargo, esta conducta puede verse afectada por la motivación del roedor para explorar o por posibles alteraciones motoras, del mismo modo que la ansiedad también puede inhibir o estimular la exploración (Strekalova, Spanagel, Dolgov y Bartsch, 2005). Por otra parte, es frecuente el uso de medidas espaciotemporales (p.ej. el tiempo en alcanzar una plataforma oculta) para evaluar la memoria en roedores, siendo probable que la ejecución locomotora o la propia exploración influyan en la ejecución de las tareas cognitivas (Kalueff y Murphy, 2007). Así mismo, las pruebas de memoria suelen implicar elementos aversivos, como la exposición a espacios abiertos o la natación forzada, lo que facilita que la ansiedad del animal pueda afectar a su rendimiento (Ohl, Roedel, Blinder y Holsboer, 2003).

Por otro lado, recientemente hemos demostrado que la ausencia del receptor del ácido lisofosfatídico LPA<sub>1</sub>, en animales modificados genéticamente, conlleva alteraciones en parámetros de plasticidad cerebral, como la neurogénesis hipocampal adulta (Matas-Rico et al., 2008), junto con déficit de memoria espacial, hipolocomoción y respuestas emocionales anómalas (Santín et al., 2009). Sin embargo, aún se desconoce la relación de estos déficit en este ratón transgénico, de modo que podrían ser independientes unos de otros o de forma contraria observarse algún grado de interdependencia entre ellos. El objetivo de este trabajo es estudiar la relación entre la actividad, la ansiedad y la memoria espacial en ratones normales y carentes del receptor LPA<sub>1</sub> en la prueba del laberinto de hoyos. Para ello empleamos el Análisis de Componentes Principales (ACP), una aproximación multivariante

que permite agrupar las variables en las dimensiones independientes (factores) que subyacen a la conducta.

## Método

### *Animales y procedimiento experimental*

Se utilizaron 37 ratones machos adultos, 23 de genotipo normal (salvajes) y 14 mutantes carentes del receptor LPA<sub>1</sub> (*knockout*), estabulados a una temperatura de 20±2 °C con un ciclo de 12 h de luz/oscuridad. La colonia de ratones mutantes empleada en este estudio está descrita en Estivill-Torrús et al. (2008). Durante los 4 días anteriores a la experimentación, los animales fueron manipulados manualmente (5 minutos al día) y sometidos a una dieta restrictiva que se mantuvo a lo largo del experimento. La evaluación conductual tuvo lugar durante 6 días en el laberinto de hoyos, un campo abierto (40 x 40 cm) con 16 hoyos equidistantes en los que podía depositarse un reforzador (0.02 gr de comida). Los 2 primeros días se dedicaron a la habituación de los animales al laberinto, con 1 sesión al día (1 ensayo de 5 minutos de duración) con todos los hoyos reforzados. En los 4 días posteriores se reforzó un patrón fijo de 4 hoyos y se llevó a cabo un entrenamiento en memoria espacial, con 4 ensayos al día distribuidos en 2 sesiones de 2 ensayos consecutivos cada una (intervalo entre sesiones de 2 horas). El ensayo finalizaba cuando el animal encontraba todos los reforzadores.

Se evaluaron las 2 sesiones de habituación (que representan, respectivamente, el comportamiento ante un contexto nuevo y ante un contexto familiar) y la última sesión de entrenamiento en memoria espacial (que evalúa el rendimiento cognitivo alcanzado). Para cada sesión, se registró la permanencia del animal en la periferia del laberinto (porcentaje de tiempo en los 6.5 cm interiores a las paredes), la locomoción (mm recorridos por segundo), el *head dipping* (exploración de hoyos) y el *rearing* (exploración vertical con apoyo en las patas traseras). Adicionalmente, en las sesiones de habituación se registraron el *grooming* (conductas de aseo), la valoración de riesgo (estiramiento del cuerpo con vuelta a la postura original, sin locomoción) y la defecación (heces depositadas en el laberinto). En la fase de entrenamiento en memoria espacial, además de registrarse la duración de cada ensayo, se calculó un índice de memoria a largo plazo (nº visitas y revisitas a hoyos reforzados / nº de visitas y revisitas a hoyos reforzados y no reforzados) y un índice de memoria de trabajo (nº visitas a hoyos reforzados / nº de visitas y revisitas a hoyos reforzados) para estudiar ambas funciones cognitivas. La locomoción y la permanencia en la periferia se evaluaron mediante el sistema de registro informatizado Ethovision XT (Noldus. The Netherlands). El resto de conductas se evaluaron observacionalmente. El procedimiento se llevó a cabo de acuerdo a la normativa legal vigente sobre animales de experimentación y transgénicos (Ley 32/2007, Real Decreto 1205/2005 y Ley 9/2003).

### *Análisis estadístico*

El ACP se llevó a cabo por separado para cada genotipo en cada una de las sesiones evaluadas. Anteriormente al análisis, los datos de los 2 ensayos de la sesión de entrenamiento se promediaron en una única medida. Sobre la solución del ACP se realizó una rotación ortogonal (varimax), que asegura que los factores obtenidos no correlacionan entre sí y por tanto reflejan procesos independientes. Los factores se eligieron hasta explicar al menos un 75% de la varianza total y se nombraron en función de las variables con mayor carga en el factor. Las cargas menores de 0.5 no se consideraron. El test de esfericidad de Bartlett resultó significativo en todos los casos ( $p < 0.05$ ), y la medida de adecuación muestral de Kaiser-Meyer-Olkin arrojó valores entre 0.5 y 0.6. Los análisis se realizaron con el software estadístico SPSS versión 13.0.

## Resultados

Tabla 1. Resultados del ACP en las sesiones de habituación

Primera sesión de habituación (contexto nuevo)						
	<i>Ratones salvajes</i>			<i>Ratones knockout de LPA<sub>1</sub></i>		
	Factor 1 Actividad	Factor 2 Ansiedad	Factor 3 Valorac. Riesgo	Factor 1 Actividad	Factor 2 Arousal	Factor 3 Ansiedad
Periferia		0.848				0.772
Locomoción	0.831			0.943		
Head dipping	0.755					-0.747
Rearing	0.924			0.858		
Grooming	0.653				0.591	
Valorac. Riesgo			0.962		0.819	
Defecación		0.799			-0.821	
<i>% Varianza</i>	<i>43.44</i>	<i>23.06</i>	<i>12.06</i>	<i>34.90</i>	<i>24.26</i>	<i>17.03</i>
Segunda sesión de habituación (contexto familiar)						
	<i>Ratones salvajes</i>			<i>Ratones knockout de LPA<sub>1</sub></i>		
	Factor 1 Ansiedad	Factor 2 Actividad	Factor 3 Groom., Head dip.	Factor 1 Ansiedad	Factor 2 Actividad	Factor 3 Grooming
Periferia	0.836			0.658		
Locomoción		0.757			0.787	
Head dipping			-0.718	-0.717		
Rearing		0.846			0.803	
Grooming			0.904			0.972
Valorac. Riesgo	0.810			0.857		
Defecación	-	-	-	0.604	0.712	
<i>% Varianza</i>	<i>30.92</i>	<i>26.75</i>	<i>21.11</i>	<i>39.12</i>	<i>23.13</i>	<i>15.90</i>

Se representan las cargas > 0.5. Valorac. Riesgo: valoración de riesgo; Groom.: grooming; Head dip.: head dipping; los ratones salvajes no defecaron en la segunda sesión.

Tabla 2. Resultados del ACP en la última sesión de aprendizaje espacial

Última sesión del entrenamiento en memoria espacial							
<i>Ratones salvajes</i>				<i>Ratones knockout de LPA<sub>1</sub></i>			
	Factor 1 Actividad, MDT	Factor 2 Latencia	Factor 3 Ansiedad	Factor 4 MLP	Factor 1 Actividad	Factor 2 Memoria, Rearing	Factor 3 Ansiedad
MLP				-0.991		0.739	
MDT	-0.949					0.839	
Periferia			0.965				0.966
Locomoción	0.776				0.964		
Duración		-0.942			-0.674	-0.639	
Head dipping		0.850			0.929		
Rearing	0.561		0.564			-0.774	
% Varianza	28.30	27.47	18.82	15.72	44.39	26.47	12.29

Se representan las cargas > 0.5. MLP: índice de memoria a largo plazo; MDT: índice de memoria de trabajo.

En la primera sesión de habituación, las soluciones factoriales explicaron el 78.56 % (salvajes) y el 76.19 % (*knockout*) de la varianza total (Tabla 1). En la segunda sesión de habituación, el ACP explicó un 78.79 % (salvajes) y un 78.15 % (*knockouts*) de la varianza (Tabla 1). En la sesión de entrenamiento en memoria espacial, la varianza explicada por los factores fue un 90.31 % (salvajes) y un 83.15 % (*knockouts*) (Tabla 2).

## Discusión

Este trabajo investiga las dimensiones que subyacen a la conducta de ratones salvajes y *knockout* para el receptor LPA<sub>1</sub> en el laberinto de hoyos, estudiando la relación entre la actividad, la ansiedad y la memoria. En las 2 sesiones de habituación al laberinto, el ACP confirmó componentes separados para la actividad y la ansiedad, lo cual concuerda con estudios previos (Fernandes, González, Wilson y File, 1999; Ohl et al., 2003). En este sentido, la asociación entre la locomoción y el *rearing* es frecuente y representa una dimensión de actividad general y exploración del entorno (e.j. Fernandes et al., 1999). En relación a la permanencia en la periferia, es importante destacar que no está relacionada con la locomoción. Por tanto, es independiente de la capacidad motora y constituye una buena medida del estado emocional (ansiedad) en los animales de este estudio (Crawley, 1999). La defecación presenta problemas como medida de ansiedad al no estar presente en los ratones salvajes durante todas las sesiones, mientras que el *grooming* y la valoración de riesgo han saturado en factores de difícil interpretación. En cuanto al *head dipping*, se considera una conducta de exploración dirigida a un objeto y su relación con la ansiedad y con la actividad es ambigua (revisado en Lister, 1990). Finalmente, es relevante comentar el cambio en el peso relativo de los componentes que se observó entre una sesión de habituación y la siguiente. Esto sugiere que cada sesión ha provocado una respuesta distinta y, por tanto, que la familiaridad del contexto afecta a la conducta (Holmes y Rodgers, 1998).

Tras el estudio de la fase de habituación, se investigó la relación de la actividad y la emoción con la cognición. El ACP en la sesión de entrenamiento en la tarea de memoria espacial reveló que la permanencia en la periferia y la locomoción eran independientes de los índices de memoria espacial a largo plazo y de trabajo en los ratones carentes del receptor LPA<sub>1</sub>. Previamente, nuestro grupo ha descrito

déficit de memoria espacial en estos animales (Santín et al., 2009). No obstante, se desconocía si estos déficits eran puramente cognitivos o se debían a la presencia de otras alteraciones no cognitivas. Los resultados de este trabajo demuestran que la ansiedad, la actividad y la memoria espacial representan dominios independientes en los ratones carentes del receptor LPA<sub>1</sub> y que su rendimiento cognitivo no se puede explicar por la presencia de alteraciones exploratorias ni emocionales. Nuestros resultados indican que el uso de técnicas de análisis multivariante, como el ACP, son de gran utilidad para estudiar la conducta de animales transgénicos, ya que permite conocer la dependencia o independencia entre diferentes dimensiones conductuales.

Agradecimientos. MEC SEJ2007-61187, I3SNS Programme, FIS 02/1643, FIS PI07/0629, Red CIEN (G03/06), CTS065, CTS433, Programa FPU (AP2006-02582; AP2007-03719), MH51699 y MH01723.

### Referencias

- Crawley, J. (1999). Behavioral phenotyping of transgenic and knockout mice: experimental design and evaluation of general health, sensory function, motor abilities and specific behavioural tests. *Brain Research*, 835, 18-23.
- Estivill-Torrus, G., Llebreg-Zayas, P., Matas-Rico, E., Santin, L., Pedraza, C., De Diego, I., Del Arco, I., Fernandez-Llebreg, P., Chun, J y De Fonseca, F.R. (2008). Absence of LPA<sub>1</sub> signaling results in defective cortical development. *Cerebral Cortex*, 18, 938-950.
- Fernandes, C., González, M.I., Wilson, C.A. y File, S.E. (1999). Factor analysis shows that female rat behaviour is characterized primarily by activity, male rats are driven by sex and anxiety. *Pharmacology, Biochemistry and Behavior*, 64, 731-738.
- Harrison, S.M., Reavill, C., Brown, G., Brown, J.T., Cluderay, J.E., Crook, B., Davies, C.H., Dawson, L.A., Grau, E., Heidbreder, C., Hemmati, P., Hervieu, G., Howarth, A., Hughes, Z.A., Hunter, A.J., Latcham, J., Pickering, S., Pugh, P., Rogers, D.C., Shilliam, C.S. y Maycox, P.R. (2003). LPA<sub>1</sub> receptor-deficient mice have phenotypic changes observed in psychiatric disease. *Molecular and Cellular Neuroscience*, 24, 1170-1179.
- Holmes, A. y Rodgers, L.J. (1998). Responses of Swiss-Webster mice to repeated plus-maze experience: Further evidence for a qualitative shift in emotional state? *Pharmacology, biochemistry and behaviour*, 60, 473-488.
- Kalueff, A.V., Fox, M.A., Gallagher, P.S. y Murphy, D.L. (2007). Hypolocomotion, anxiety and serotonin syndrome-like behavior contribute to the complex phenotype of serotonin transporter knockout mice. *Genes, Brain & Behavior*, 6, 389-400.
- Lister, R.J. (1990). Ethologically-Based animal models of anxiety disorders. *Pharmacology & Therapeutics*, 46, 321-340.
- Matas-Rico, E., Garcia-Diaz, B., Llebreg-Zayas, P., Lopez-Barroso, D., Santin, L., Pedraza, C., Smith-Fernandez, A., Fernandez-Llebreg, P., Tellez, T., Redondo, M., Chun, J., De Fonseca, F.R. y Estivill-Torrus, G. (2008). Deletion of lysophosphatidic acid receptor LPA<sub>1</sub> reduces neurogenesis in the mouse dentate gyrus. *Molecular and Cellular Neuroscience*, 39, 342-355.

- Ohl, F., Roedel, A., Blinder, E. y Holsboer, F. (2003). Impact of high and low anxiety on cognitive performance in a modified hole board test in C57BL/6 and DBA/2 mice. *European Journal of Neuroscience*, 17, 128-136.
- Santin, L.J., Bilbao, A., Pedraza, C., Matas-Rico, E., Lopez-Barroso, D., Castilla-Ortega, E., Sanchez-Lopez, J., Riquelme, R., Varela-Nieto, I., de la Villa, P., Suardiaz, M., Chun, J., De Fonseca, F.R. y Estivill-Torres, G. (2009). Behavioral phenotype of  $maLPA_1$ -null mice: increased anxiety-like behavior and spatial memory deficits. *Genes, Brain & Behavior*, 8, 772-784.
- Strekalova, T., Spanagel, R., Dolgov, O. y Bartsch, D. (2005). Stress-induced hyperlocomotion as a confounding factor in anxiety and depression models in mice. *Behavioural Pharmacology*, 16, 171-80.

# INFLUENCIA DE LOS FACTORES DEL AUTOCONCEPTO SOBRE CANTIDAD Y CALIDAD DE LA ACTIVIDAD FÍSICA EN JÓVENES ESPAÑOLES

Armando Cocca, Jesús Viciano, Francisco Salinas y Kamal Mohamed  
Universidad de Granada

La actividad física (AF) es un parámetro fundamental para la determinación de un estilo activo o sedentario. Muchos autores se han dedicado a su estudio, definiendo modelos como *The Lifespan model of Physical Activity and Health* y el *Young Physical Activity Promotion Model*, que concretan pautas para el incremento de la AF a través de factores físicos y psicológicos relacionados con ésta. Entre estos últimos el autoconcepto es el más estudiado por los investigadores. El objetivo de este estudio fue la evaluación de la influencia del autoconcepto sobre la AF en jóvenes españoles. Para ello se cogieron 3695 sujetos entre 8 y 23 años, estudiando en centros escolares de diferentes etapas educativas de Granada capital. Se les suministró un cuestionario creado por los autores a partir de otros anteriormente validados, tras obtener autorización de los directores de cada centro encuestado. Los resultados mostraron un decremento significativo en la AF entre estudiantes de Primaria y de Universidad, al igual que para el autoconcepto. La correlación entre estas dos variables fue muy significativa. Por tanto, el autoconcepto tiene una influencia muy fuerte sobre los parámetros de la AF, y en futuro debemos considerar esta variable para el incremento de su práctica.

La salud es un tema de preocupación actual que se manifiesta continuamente en nuestra sociedad (Naval, 2003). Los datos de estudios recientes confirman que existe una tendencia negativa con respecto a la salud pública e individual a nivel internacional, así como los análisis estadísticos que asociaciones a carácter médico llevan a cabo anualmente corroboran esta situación puesto que se indica un incremento de la morbilidad y de la mortalidad en edades inferiores a la expectativa de vida (Rosamond et al., 2008).

Es necesario por tanto intervenir con el fin de invertir la actual tendencia, enfocando la atención de la investigación científica, de la intervención política y en general, de la sociedad, hacia la estructuración de un estilo de vida saludable. Tener hábitos correctos y vivir sano influye positivamente en la duración y en la calidad de la vida (Peel et al., 2007). Además, mantener un estilo de vida saludable en el periodo del desarrollo psicofísico es determinante para el correcto crecimiento de los jóvenes (Stock et al., 2007). Se ha demostrado que mantener buenos hábitos desde edades tempranas es un método eficaz de prevención de afecciones como osteoporosis, cáncer o diabetes (Akesson et al., 2007; Dontas y Yiannakopoulos, 2007).

Las mejoras en salud pueden ser más significativas si se interviene desde edades tempranas, ya que en este caso sería posible aplicar una prevención primaria, suprimiendo las causas de futuros padecimientos (Driskell et al., 2007). De hecho, es más fácil erradicar prácticas diarias incorrectas cuando todavía no se han convertido en hábitos.

Existen muchos factores relacionados con la salud que influyen en la estructuración de un modelo de vida sano. Entre las variables más estudiadas, la AF se considera fundamental (Vicente-Rodríguez et al., 2007). Ésta interviene en varios aspectos fisiológicos mejorándolos, además de reforzar aspectos psicológicos ligados al estado saludable (Lakka y Laaksonen, 2007). El papel positivo de la AF en el desarrollo de la salud ha sido además demostrado en numerosas investigaciones médicas, fisiológicas, psicológicas, deportivas y educativas (Corbin, Pangrazi y Frank, 2000).

Algunos investigadores han creado modelos que estudian la relación entre AF, variables psicológicas y salud. Uno de ellos es el *Lifespan model of Physical activity and Health*, desarrollado por Welk et al. (2006). Este modelo otorga a la AF un impacto muy importante en el estado físico sano del adulto (HRF, Health Related Fitness), resaltando cómo el nivel de AF depende a su vez del HRF que se obtenga en la niñez y en la adolescencia. Paralelamente, se resaltan los efectos de las habilidades motoras y de la participación en las actividades deportivas sobre las pautas de la actividad en la edad adulta. Las habilidades motoras se consideran de hecho factores predictivos de la forma cardiorespiratoria y de la AF organizada en los jóvenes. Sin embargo, la consecuencia de la falta de dichos factores determina el aburrimiento y el abandono de la práctica. Por esto, Welk et al. (2006) concluyen que es necesario medir las variables psicológicas, consideradas mediadoras de la AF y del estilo de vida.

*The Young Physical Activity Promotion Model* (Welk, 1999) es otro modelo desarrollado con el fin de analizar la relación entre factores psicológicos, factores físicos y salud. Se señalan tres factores de influencia:

1. Factores de predisposición: son de tipo psicológico, basados en las teorías sociocognitivas, y predisponen los jóvenes a ser activos. Con respecto a esto, se ha encontrado una relación causal entre percepción de competencia y AF. De hecho, aunque un niño atribuya mucha importancia a la AF, si no se considera competente a nivel motor, no la practicará.
2. Factores que permiten hacer AF: se refieren a la oferta de AF (cantidad y tipo) y a las instalaciones deportivas y su accesibilidad, sin embargo no son agentes determinantes. El nivel de forma física y las competencias motrices son considerados de los agentes más influyentes en este sentido, ya que incrementan las habilidades deportivas y la motivación hacia la práctica. Esta es la razón por la cual variables psicológicas como el autoconcepto y la percepción de la competencia se relacionan con diferentes niveles de AF.
3. Factores que refuerzan la AF: se refieren a variables psicológicas y sociales como los otros significativos, que con sus comportamientos influyen sobre la percepción de la habilidad, la motivación y el autoconcepto.

Por todo esto, es necesario estudiar en cada etapa de la vida las variables psicológicas más influyentes, como la percepción de competencia, el autoconcepto físico, la expectativa, la motivación y la actitud hacia la AF o el refuerzo de los otros significativos, con el fin de observar cuánto y cómo influyen en las tasas de AF y por tanto en el nivel de salud y el establecimiento de un estilo de vida sano.

Entre las variables psicológicas mencionadas, el autoconcepto se considera un importante indicador del bienestar psicológico y un mediador de la conducta. Esto es un constructo que se puede definir como las percepciones que tiene el individuo sobre sí mismo, que se forman a través de la experiencia y de la interpretación de su entorno. Un autoconcepto positivo está en la base del buen funcionamiento personal, social y profesional. Esto puede influir sobre el rendimiento académico, los logros sociales y los hábitos saludables. El autoconcepto se construye a lo largo de toda la edad del desarrollo e incluso después de la adolescencia.

Algunos investigadores se han centrado en la relación entre deporte y algunas dimensiones específicas del autoconcepto que se definen 'autoconcepto físico' (Marsh, 2001). Varias investigaciones muestran que los sujetos que practican deporte poseen mayores puntuaciones en lo que se refiere a dicho componente (Marsh, 1997, 2001; Moreno, 1997). Al mismo tiempo, una alta percepción de competencia incrementa la motivación hacia la AF.

Asimismo, es posible que incrementando el nivel y la cantidad de la AF, los sujetos obtengan beneficios en el autoconcepto, al igual que un mayor autoconcepto podría incitarlos a una mayor práctica deportiva. Estudios sobre el autoconcepto en sujetos que compiten en distintas modalidades deportivas confirman que una práctica constante corresponde, en general, a valores medios-altos en todas las dimensiones del autoconcepto físico (Lozano et al., 2007).

Por lo tanto, el objetivo de este estudio fue evaluar la situación de la salud en jóvenes estudiantes españoles a través del análisis de los parámetros de la AF y del autoconcepto físico; además, se quiso estudiar la relación existente entre estas dos variables.

## Método

### *Participantes*

Los sujetos seleccionados asistían a diferentes centros educativos del área urbana de Granada capital durante el año académico 2007/08. Se utilizaron los datos facilitados por la Consejería de Educación para conocer la población total de estudiantes de Primaria, Secundaria y Bachillerato, así como los informes estadísticos de la Universidad de Granada para el caso de los estudiantes universitarios.

A partir de estos datos, conformamos la muestra con una técnica de muestreo aleatoria por conglomerados, estratificada y con afijación proporcional. Se consideró un error del 5% para los estudiantes de Educación Primaria, Secundaria y Bachillerato, mientras se asumió el 2,5% de error para los universitarios. Finalmente, se aplicó la corrección para población finita y se obtuvo la fracción de muestreo, que se situó en el 2.1% y en el 2.79%. En la tabla 1 se expone la muestra final del estudio y sus características.

*Tabla 1. Muestra final del estudio y su análisis descriptivo*

ETAPA	CICLO	GÉN	NÚM	EDAD	PESO (kg)	ALTURA (m)
Prim	2º	V	215	9.05 ± 0.819	34.79 ± 7.47	1.36 ± 0.13
	2º	M	210	9.00 ± 0.821	33.54 ± 6.12	1.39 ± 0.34
	3º	V	274	11.15 ± 0.77	43.64 ± 10.48	1.49 ± 0.10
	3º	M	281	11.24 ± 0.831	42.03 ± 8.75	1.51 ± 0.09
Sec	1º	V	186	13.36 ± 0.994	54.55 ± 10.23	1.65 ± 0.09
	1º	M	160	13.18 ± 0.865	50.26 ± 7.27	1.60 ± 0.06
	2º	V	276	15.08 ± 0.842	65.19 ± 13.45	1.72 ± 0.08
	2º	M	216	14.98 ± 1.15	55.05 ± 8.25	1.65 ± 0.06
Bach	1º-2º	V	243	16.95 ± 0.982	69.89 ± 10.03	1.76 ± 0.07
	1º-2º	M	266	17.03 ± 1.203	57.57 ± 8.18	1.66 ± 0.06
Univ	1º	V	437	21.33 ± 2.308	75.75 ± 11.38	1.77 ± 0.07
	1º	M	452	20.88 ± 2.32	59.79 ± 8.58	1.66 ± 0.06
	2º	V	209	24.17 ± 2.53	76.24 ± 9.74	1.78 ± 0.07
	2º	M	259	24.22 ± 3.942	60.40 ± 8.68	1.66 ± 0.06
TOTAL		TOT	3685	16.50 ± 5.28	57.82 ± 16.08	1.64 ± 0.15

El diseño aplicado se define como cuasiexperimental. La investigación tuvo las características de un estudio sociológico seccional trasversal, estudiando diferentes variables en la muestra seleccionada. Sólo tuvo lugar una sesión de medición y no se aplicó ninguna intervención para manipular las variables, definidas como interdependientes. Las variables estudiadas fueron la actividad física moderada/intensa, y el autoconcepto físico.

### *Instrumento*

La herramienta que utilizamos para la recogida de datos fue el cuestionario, creado por los mismos investigadores a partir de otros cuestionarios ya validados internacionalmente y usados en diferentes estudios previos.

Asimismo, el cuestionario final se compuso de ítems procedentes del IPAQ (Booth, 2000), que recoge datos relativos a la cantidad y calidad de la AF practicada, indicando el tiempo de práctica así como su intensidad; y de ítems procedentes del Physical Self-concept Questionnaire (Marsh, 1994) para la valoración del autoconcepto físico.

### *Procedimiento*

El primer paso fue solicitar la autorización a los directores de cada centro involucrado en la investigación, y sucesivamente a los profesores cuyas clases se vieran directamente afectadas por la cumplimentación del cuestionario. Obtenida la autorización, se entregaron los cuestionarios a los sujetos previstos, que los cumplimentaron con la supervisión de su profesor y de un encuestador, que solventaba cualquier tipo de duda o problema.

Los datos colectados se introdujeron en una hoja de trabajo del software SPSS 14.0 para proceder con el análisis estadístico.

### *Análisis de datos*

Todas las variables se adecuaron a escalas ordinales, permitiéndose de esta forma el análisis comparativo y el estudio de correlaciones.

Para interpretar los resultados obtenidos, en primer lugar y para comprobar la validez de los datos, se efectuaron pruebas robustas de igualdad de las medias mediante test de Welch y de Brown-Forsythe en todas las variables de estudio. A continuación, se aplicaron varios análisis estadísticos:

- Estadística descriptiva: para el estudio de las frecuencias, porcentajes, desviaciones típicas y medias.
- Estadística inferencial a través de ANOVA. Mediante pruebas *post hoc* para comparaciones múltiples (Test de Tamhane), se buscaron diferencias significativas en los valores obtenidos en cada estrato, según las secciones establecidas en la muestra. El test de Tamhane se utiliza cuando, como en el caso de este estudio, no se cumple la homogeneidad de la varianza entre las variables estudiadas. Para las variables psicológicas se utilizó la prueba no paramétrica de Kruskal-Wallis, ya que sus valores numéricos fueron el resultado de la codificación de datos cualitativos.
- Estadística correlacional: a través de la correlación de Pearson para variables paramétricas; usando la correlación de Spearman para las no paramétricas, en relación a las variables psicológicas.

### Resultados

Se halló una diferencia significativa entre estudiantes de Primaria y de Universidad ( $p = .001$ ) en lo que se refiere al nivel de práctica de AF moderada/intensa.

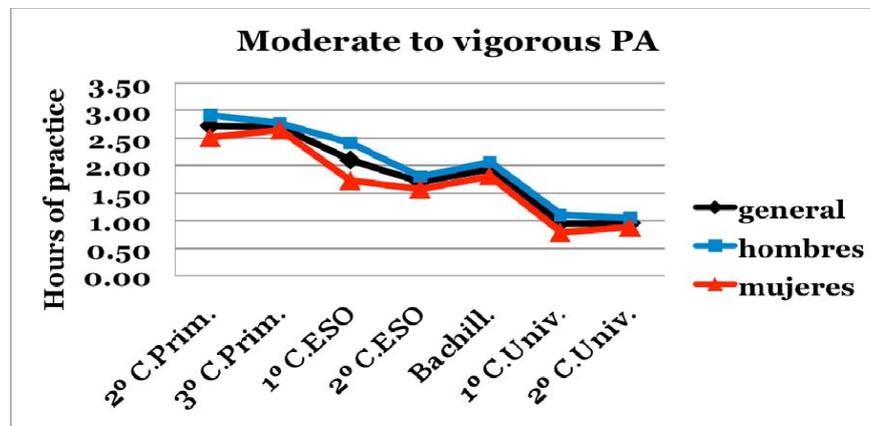


Figura 1. Tendencia de la práctica de AF a lo largo de las etapas estudiadas

Además, resultados significativos se encontraron también entre Primaria y Secundaria ( $p = .001$ ) y entre Bachillerato y Universidad ( $p = .001$ ). Si por un lado en Primaria los sujetos alcanzan casi las 3 horas diarias de AF, los universitarios no practican ni una hora (55'). No se hallaron diferencias significativas entre estudiantes de Secundaria y de Bachillerato.

En lo que al autoconcepto se refiere, también se halló una tendencia negativa, encontrándose diferencias significativas entre estudiantes de Primaria y Secundaria ( $p = .001$ ), de Secundaria y Bachillerato ( $p = .001$ ) y de Bachillerato y Universidad ( $p = .001$ ). Como es visible en la figura 2, también para el autoconcepto los valores se reducen a lo largo de las etapas educativas: en Primaria son muy altos (3.49) pero bajan hasta valores intermedios (2.52 en Universitarios).

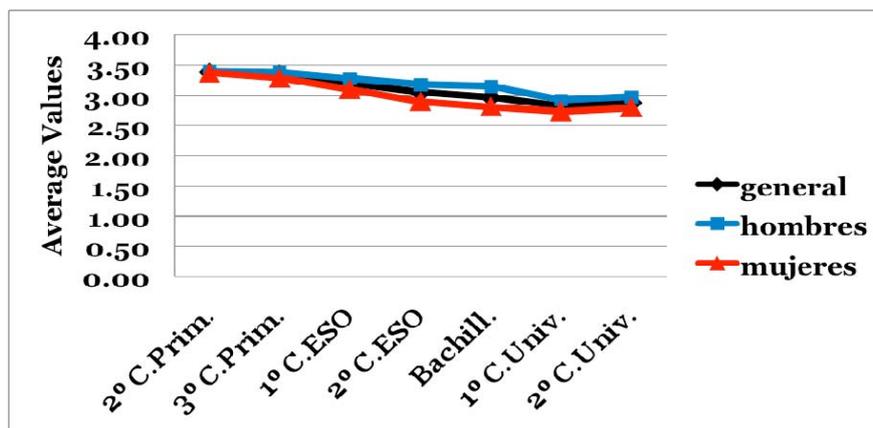


Figura 2. Autoconcepto y su tendencia en las etapas estudiadas

Además, se halló una relación muy fuerte entre estas dos variables, positiva y significativa a nivel estadístico ( $p = .001$ ; coeficiente de correlación  $\rho = .358$ ).

## Discusión

La AF realizada disminuye de forma muy significativa de Primaria a Universidad. Sin embargo si consideramos las más recientes recomendaciones sobre la práctica física para la salud (Canadian Society for Exercise Physiology, 2008), el nivel de los jóvenes de Granada es bueno y suficiente para mantenerse sano, incluso para la Universidad donde se hallaron los resultados más bajos.

De todos modos, permanece clara la tendencia negativa que asume la práctica física a lo largo de los años, disminuyendo significativamente. Es necesario encontrar los puntos cruciales de esta reducción e intervenir para contrarrestar esta tendencia. En los estudiantes de Granada estos puntos críticos parecen situarse entre primer y segundo ciclo de Secundaria, y al entrar en la Universidad.

La correlación hallada entre autoconcepto y AF nos indica que la primera puede influir positivamente sobre la segunda, por lo tanto podemos pensar de incrementarla de cara a mejorar los valores de práctica deportiva de los sujetos.

Para poder intervenir de forma eficaz, en futuro deberíamos plantearnos añadir otras variables a estudios como los otros significativos. Es indispensable estudiar como estos influyen sobre el comportamiento del joven, la importancia que asumen los profesores, los padres, los amigos o la pareja a lo largo de los años. El contexto de vida también es fundamental, porque de eso depende por ejemplo la facilidad de acceso a instalaciones deportivas. La motivación hacia la AF es otra variable psicológica muy importante y relacionada con el entorno socioafectivo. Para obtener un cuadro completo del nivel de salud, sería importante añadir otros aspectos que no pertenecen al área psicológica, como la nutrición.

Por tanto, es recomendable incrementar las variables de estudio, sean éstas otros importantes factores psicológicos o relativos a otros aspectos de la vida.

También sería interesante extender la muestra del presente estudio, abarcando en esto el primer ciclo de Primaria y el tercero de Universidad, para obtener un cuadro general más amplio y detallado. Además, es básico ampliar la investigación encontrando herramientas adecuadas para que sea posible realizar el estudio también sobre los trabajadores. De hecho el tema de la salud en el trabajo es muy actual, y además los trabajadores representan una fracción importante de la población joven y adulta.

## Referencias

- Akesson, A., Weismayer, C., Newby, P.K., y Wolk, A. (2007). Combined effect of low-risk dietary and lifestyle behaviors in primary prevention of myocardial infarction in women. *Archives of internal medicine*, 167(19), 2122-2127.
- Booth, M.L. (2000). Assessment of Physical Activity: An International Perspective. *Research Quarterly for Exercise and Sport*, 71(2), 114-20.
- Canadian Society for Exercise Physiology. *Canada's Physical Activity Guide to healthy active living*. 2008.
- Corbin, C., Pangrazi, R. y Frank, D. (2000). Definitions: Health, Fitness and Physical Activity. *President's Council on Physical Fitness and Sports Research Digest*, 3.

- Dontas, I.A., y Yiannakopoulos, C.K. (2007). Risk factors and prevention of osteoporosis-related fractures. *Journal of Musculoskeletal & neuronal interactions*, 7(3), 268-272.
- Driskell., M.M., Dymont, S., Mauriello, S., Castle, P. y Sherman, K. (2007). Relationship among multiple behaviors for childhood and adolescent obesity prevention. Extraído el 12 de Noviembre de 2007 desde <http://www.sciencedirect.com>
- Lakka, T.A., y Laaksonen, D.E. (2007). Physical activity in prevention and treatment of the metabolic syndrome. *Applied Physiology, Nutrition and Metabolism*, 32, 76-88.
- Lozano, L., Cocca, A., Salinas, F., Miranda, M.T. y Viciano, J. (2007). El autconcepto de los nadadores frente a otras modalidades deportivas. En Arellano (Eds.), *Swimming Science I* (203-208). Granada: Editorial Universidad de Granada.
- Marsh, H.W. (2001). A multidimensional physical self-concept: a construct validity approach to theory, measurement and research. *Paper presented at 10th World Congress of Sport Psychology*, May, Greece.
- Marsh, H.W. (1997). The measurement of Physical Self-Concept: A construct validation approach. En K.R. Fox (Ed.), *The physical self. From motivation to well-being* (27-58). Champaign: Human Kinetics.
- Marsh, H.W., Richards, G.E., Johnson, S., Roche, L. y Tremayne, P. (1994). Physical Self-Description Questionnaire: Psychometric Properties and a Multitrait-Multimethod Analysis of Relations to Existing Instruments. *Journal of Sport and Exercise Psychology*, 16 (3), 270-305.
- Moreno, Y. (1997). Propiedades psicométricas del Perfil de Autopercepción Física (PSPP). Tesis de Licenciatura. Valencia: Universitat de València.
- Naval, C. (2003). Orígenes recientes y temas clave en Educación para la ciudadanía democrática actual. *Revista de Educación*, número extraordinario (2003), 169-189.
- Peel, N.M., Bartlett, H.P., y McClure, R.J. (2007). Healthy Aging as an intervention to minimize injury from falls among older people. *Annals of the New York Academy of Sciences*, 1114, 162-169.
- Rosamond, W., Flegal, K., Furie, K., Go, A., Greenlund, K., Haase, N., Hailpem, S., Ho, M., Howard, V., Kissela, B., Kittner, S., Lloyd-Jones, D., McDermott, M., Meigs, J., Moy, C., Nichol, G., O'Donnell, C., Roger, V., Sorlie, P., Steinberg, J., Thom, T., Wilson, M., y Hong, Y. (2008). Heart Disease and Stroke Statistics—2008 Update. A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*.
- Stock, S., Miranda, C., Evans, S., Plessis, S., Ridley, J., Yeh, S., y Chanoine, J.P. (2007). Healthy Buddies: a novel, peer-led health promotion program for the prevention of obesity and eating disorders in children in elementary school. *Pediatrics*, 120(4), e1059-1068.
- Vicente-Rodriguez, G., Libersa, C., Mesana, M.I., Béghin, L., Iliescu, C., Moreno Aznar, L.A., Dallongeville, J., y Gottrand, F. (2007). Healthy lifestyle by nutrition in adolescence (HELENA). A new EU funded project. *Thérapie*, 62(3), 259-270.

- Welk, G. J., Eisenmann, J. C. & Dollman, J. (2006). Health-related physical activity in children and adolescents: a bio-behavioral perspective, in David Kirk, Doune Macdonald and Mary O'Sullivan (2006). *The Handbook of Physical Education*. 665-684. Sage Pub. London.
- Welk, G.J. (1999). The Youth Physical Activity Promotion Model: A Conceptual Bridge Between Theory and Practice. *Quest*, 51 (1), 5-23.

# PREVALENCIA DE LOS TRASTORNOS INTERNALIZANTES Y EXTERNALIZANTES EN UNA POBLACIÓN INFANTIL

Antonio Cortés<sup>1,2</sup>, Jesús Miranda<sup>2</sup> y María V. Trianes<sup>2</sup>

<sup>1</sup>AT. Gabinete Psicopedagógico. Unidad de Desarrollo Infantil y Atención Temprana. Fuengirola, Málaga

<sup>2</sup>Universidad de Málaga

Se describe la prevalencia de los principales problemas psicopatológicos (trastornos internalizantes y externalizantes) en una población infantil. Partiendo de una población accesible del municipio de Fuengirola de 2431 sujetos, se seleccionó una muestra de 423 sujetos de dicha población escolarizada, con edades comprendidas entre los 5 y 8 años, de los cuales 195 sujetos eran niños y 228 niñas. Para el estudio de prevalencia se utilizaron las puntuaciones obtenidas del *C-Teacher's Report Form* y el *Teacher's Report Form* (Achenbach y Rescorla 2000, 2001). Atendiendo a los rangos de riesgo y clínico los resultados muestran valores de prevalencia de los trastornos externalizantes sensiblemente más altos, en chicas frente a chicos, destacando la diferencia en el TDAH subtipo Hiperactivo-Impulsivo. A excepción de los Trastornos afectivos, los trastornos internalizantes presentaron prevalencias levemente más elevadas en los chicos. El género se muestra como una variable relevante para la realización de programas de prevención y minimización de los efectos de los trastornos que inciden de manera directa en el desarrollo normal tanto de chicos como de chicas.

Se ha llevado a cabo un estudio de prevalencia de carácter descriptivo para determinar la posible presencia de psicopatología desde el punto de vista evolutivo en la etapa infantil. El conocimiento de dichos datos se muestran de interés para contribuir a mejorar la intervención preventiva en el ámbito escolar (Ezpeleta, 2005; Angold y Costello, 2005). Diversos estudios (APA, 1994, 2000; Barkley, 2006; Kazdin, 1995; Méndez, Olivares y Bermejo, 2002) han determinado distintos intervalos de prevalencia para los Trastornos de inicio en la infancia, la niñez y la adolescencia. Según dichos estudios, para los Problemas de Ansiedad entre el 1% y 4%; para los Problemas Afectivos entre el 0.5% y el 2,5%. En los Trastornos Generalizados del Desarrollo, 15 casos de cada 10.000 nacimientos; el Trastorno por Déficit de Atención con Hiperactividad (TDAH) entre un 3% y un 7%; el Trastorno Disocial entre el 4% y el 10%; en último lugar, el Trastorno Oposicionista Desafiante (TOD) entre un 2% y un 16%.

La clasificación diagnóstica DSM-IV permite usar una organización temática para clasificar los trastornos anteriormente mencionados, pudiéndose determinar dos áreas generales de preocupación. La primera está relacionada con la demostración o la exteriorización de los problemas del comportamiento y englobaría a la población infantil con mayor riesgo de fracaso académico, problemas ajuste social, comportamientos desadaptativos, etc. Estos trastornos externalizantes están representados en el DSM-IV por el Trastorno por Déficit de Atención con Hiperactividad, el Trastorno Oposicionista Desafiante y el Trastorno de Conducta Antisocial (APA, 1994, 2000; House, A.E., 2003; Hudziak, J., Copeland, W., Stanger, C., y Wadsworth, M., 2004; Kazdin, 1995).

Una segunda área está relacionada con aquellos problemas de difícil exteriorización relacionados con síntomas emocionales y que al igual que los problemas externalizados provocan un desarrollo deficitario del niño. Los trastornos internalizantes quedan reflejados en el DSM-IV por los problemas afectivos (Distimia, Depresión Mayor); problemas de ansiedad (Ansiedad Generalizada, por Separación, Fobias Específicas) y del Desarrollo (House, A.E., 2003; Rescorla, L.A., 2005)

Otro aspecto a tener en cuenta es el género ya que además de hacer referencia a una diferenciación en el comportamiento social, también implica unas connotaciones cognitivas, afectivas y sociales (Ezpeleta, 2005). Según la investigación una de las posibles causas de la diferenciación por sexo en psicopatología, tiene su inicio en las diferencias genéticas, responsable de las diferencias biológicas. También tiene importancia la respuesta ante factores de riesgo diferentes y la gravedad para cada sexo. Así como en las diferencias observadas en cuanto a los mecanismos de riesgo que están implicados en la aparición de un trastorno. Aparece una predominancia en niños trastornos de inicio temprano que implican un deterioro del neurodesarrollo como el Trastorno por Déficit de Atención con Hiperactividad o los relacionados con los Trastornos de Conducta, incluidos en la clasificación de trastornos externalizantes. Mientras que en las niñas predominan los trastornos emocionales o internalizantes, como los Trastornos de Ansiedad o Depresión (Ezpeleta, 2005; Rutter, M., Caspi, A., y Moffitt, T.E., 2003).

### *Objetivo del estudio*

Con base en un planteamiento descriptivo, el objetivo es determinar la prevalencia por género de psicopatología, a nivel internalizante y externalizante, en sujetos de edad escolar residentes en la localidad de Fuengirola (Málaga).

## Método

### *Participantes*

Para la selección de los participantes parte de una población accesible del municipio de Fuengirola (Málaga) de 2431 sujetos escolarizados, con edades comprendidas entre los 5 y los 8 años. De los cuales se seleccionó una muestra incidental de 423 sujetos, compuesta por 195 chicos y 228 chicas, con una media de edad de 5,9 años y una desviación típica de 0,92 años.

### *Instrumentos*

Se han empleado cuestionarios de espectro amplio pertenecientes al *Achenbach System of Empirically Based Assessment -ASEBA* (Achenbach y Rescorla, 2000; 2001). Estos cuestionarios permiten valorar las conductas manifiestas de los niños, desde el punto de vista del profesor. Esta prueba admite realizar dos niveles de análisis, uno para determinar problemáticas que implican la posible presencia de problemas de carácter temporal. Un segundo análisis permite corresponder los resultados obtenidos con criterios diagnósticos DSM-IV (APA, 1994).

La presente investigación utiliza el segundo nivel de análisis para determinar la presencia de trastornos tanto internalizantes como externalizantes, con base en los resultados obtenidos mediante las versiones en castellano del *Caregiver-Teacher Report Form (C-TRF 1 ½ -5)* y el *Teacher's Report Form (TRF 6-18)*.

Los criterios diagnósticos se han agrupado en Trastornos Internalizantes y Externalizantes. Los Trastornos Internalizantes recogen: Problemas Afectivos (Distimia, Depresión Mayor); Problemas de Ansiedad (Ansiedad Generalizada, por Separación, Fobias Específicas). Los Trastornos Externalizantes recogen: Trastorno por Déficit de Atención con Hiperactividad (TDAH) (Subtipos Combinado, Inatento e Hiperactivo-Impulsivo); Trastorno Negativista Desafiante (TOD); y en último lugar, la conducta Antisocial (Trastorno Disocial, tipo de inicio infantil).

Las puntuaciones de ambos se han dispuesto en tres rangos: Rango Normal (< 50% - 93%); Rango de Riesgo (93% - 98%); y Rango Clínico (98% - 100%).

### Procedimiento

Para la obtención de los datos, se solicitó la autorización de la Delegación de Educación de Málaga dependiente de la Consejería de Educación y Ciencia de la Junta de Andalucía. Posteriormente se acudió a los Centros Educativos de Educación Infantil y Primaria, donde fue aprobada la propuesta de colaboración en sus correspondientes consejos escolares. Las familias participantes firmaron las correspondientes autorizaciones sobre la participación en el estudio. En último lugar, se repartieron los dossiers con los cuestionarios para el profesorado de los alumnos que aceptaron participar, dándoles las instrucciones verbales y por escrito para su cumplimentación. Así mismo se les informó del carácter confidencial de los datos y que estos serían utilizados únicamente con fines estadísticos para la investigación.

### Resultados

Los resultados referentes a Trastornos Externalizantes se presentan en la tabla I. Se aprecian valores de prevalencia sensiblemente más altos, en chicas frente a chicos, en todos los subtipos de TDAH, Trastorno Oposicionista-desafiante y Trastorno de Conducta Antisocial, destacando la diferencia en el TDAH subtipo Hiperactivo-Impulsivo (.002-.035) en niños frente a un intervalo de (.025-.083) en niñas.

Tabla 1

<b>Prevalencia de Trastornos Externalizantes.</b>				
	<b>Chicos</b>		<b>Chicas</b>	
	<b>% Prevalencia</b>	<b>(Intervalos de confianza 95%)</b>	<b>% Prevalencia</b>	<b>(Intervalos de confianza 95%)</b>
<b>TDAH.</b>				
<b>Subtipo Combinado</b>				
Rango Normal	.442	(.394-.490)	.477	(.429-.526)
Rango de Riesgo	.021	(.009-.040)	.042	(.025-.066)
Rango Clínico	.007	(.001-.020)	.009	(.002-.024)
<b>TDAH.</b>				
<b>Subtipo Inatento</b>				
Rango Normal	.402	(.340-.466)	.524	(.459-.588)
Rango de Riesgo	.016	(.004-.041)	.020	(.006-.046)
Rango Clínico	.008	(.000-.029)	.028	(.011-.057)
<b>TDAH. Subtipo Hiperactivo-Impulsivo</b>				
Rango Normal	.402	(.340-.466)	.512	(.447-.576)
Rango de Riesgo	.012	(.002-.035)	.048	(.025-.083)
Rango Clínico	.012	(.002-.035)	.012	(.002-.035)
<b>Trastorno Oposicionista Desafiante</b>				
Rango Normal	.444	(.396-.493)	.513	(.464-.561)
Rango de Riesgo	.014	(.005-.030)	.018	(.008-.036)
Rango Clínico	.002	(.000-.013)	.007	(.001-.020)
<b>Trastorno de Conducta Antisocial</b>				
Rango Normal	.402	(.340-.466)	.528	(.464-.592)
Rango de Riesgo	.008	(.000-.029)	.032	(.014-.063)
Rango Clínico	.016	(.004-.041)	.012	(.002-.035)

En la tabla II se recogen los valores de prevalencia encontrados en trastornos internalizantes. Atendiendo a los rangos de riesgo, los valores son levemente más altos en chicos en cuanto a Problemas de Ansiedad y Problemas de Desarrollo, mientras que ocurre lo contrario en cuanto a Trastornos Afectivos, siendo algo más altos los valores de prevalencia en rango de riesgo y rango clínico en chicas.

Tabla 2

	Chicos		Chicas	
	% Prevalencia	(Intervalos de confianza 95%)	% Prevalencia	(Intervalos de confianza 95%)
<b>Problemas Afectivos</b>				
Rango Normal	.425	(.377-.474)	.482	(.433-.531)
Rango de Riesgo	.071	(.039-.117)	.078	(.047-.121)
Rango Clínico	.002	(.000-.013)	.014	(.052-.030)
<b>Problemas de Ansiedad</b>				
Rango Normal	.427	(.380-.476)	.520	(.471-.568)
Rango de Riesgo	.018	(.008-.036)	.007	(.001-.020)
Rango Clínico	.014	(.005-.030)	.011	(.003-.027)
<b>Problemas de Desarrollo</b>				
Rango Normal	.497	(.421-.573)	.480	(.404-.556)
Rango de Riesgo	.011	(.001-.040)	.005	(.000-.031)
Rango Clínico	.000	(.000-.020)	.005	(.000-.031)

Como resultado complementario, los tres trastornos de mayor prevalencia en rango de riesgo o en rango clínico se presentan en los gráficos recogidos en la figura I. Representan las medias de edad encontradas en los rangos normal, de riesgo y clínico en chicos y chicas. Destaca la diferencia de edad entre chicos y chicas en el caso del TDAH de Subtipo Combinado.

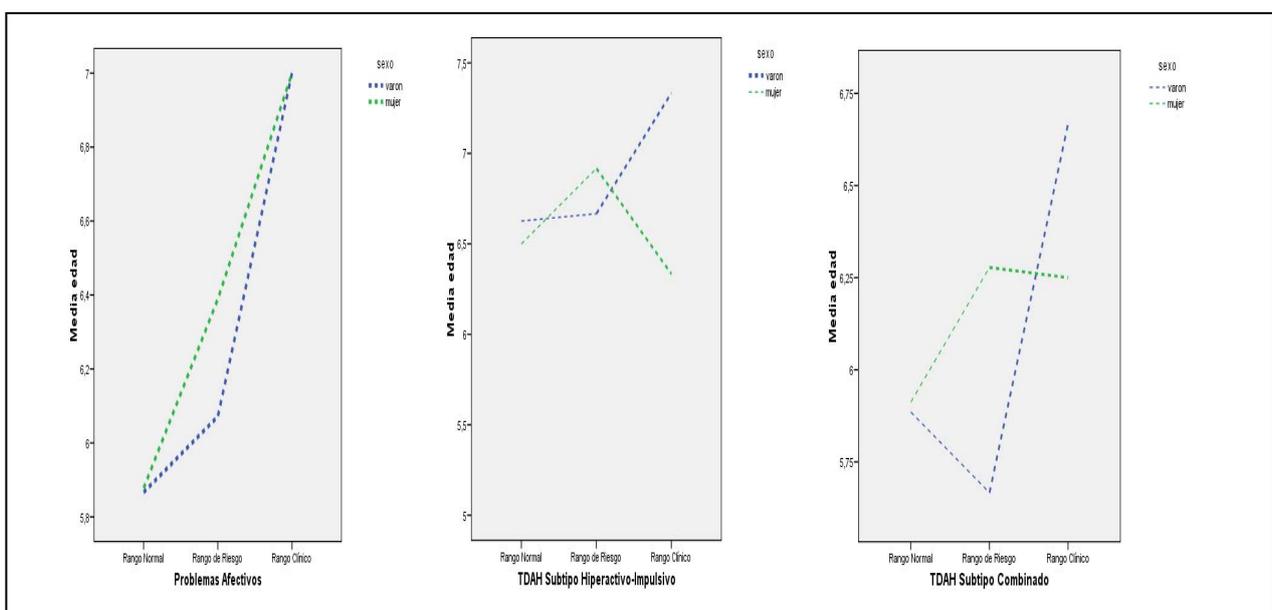


Figura 1

## Discusión

El propósito de este estudio preliminar de carácter descriptivo, ha sido identificar la prevalencia de los trastornos internalizantes y externalizantes en una población infantil. Como aspecto importante de los resultados destaca el género como variable relevante (Hudziak, J., Copeland, W., Stanger, C., y Wadsworth, M., 2004; Kazdin, 1995; Rutter, M., Caspi, A., y Moffitt, T.E., 2003). Aunque en la mayoría de las revisiones relacionan la presencia de trastornos internalizantes con chicas y la presencia de trastornos externalizantes con chicos, los resultados muestran como existe una prevalencia sensiblemente más alta de los trastornos externalizantes en chicas que en chicos. Mientras que los chicos presentan mayor prevalencia en trastornos internalizantes relacionados con la ansiedad y los trastornos afectivos. En el caso del Trastorno por Déficit de Atención con hiperactividad normalmente se ha considerado que las niñas presentan un predominio de la inatención sobre la hiperactividad/impulsividad, asociada a menores tasas de conductas externalizadoras (Gaub, M., y Carlson, CL., 1997).

Sin embargo los resultados manifiestan una diferencia en el rango de riesgo en cuanto al TDAH de subtipo Hiperactivo-Impulsivo en las niñas frente al intervalo presentado por los niños. En el caso de las chicas el diagnóstico del TDAH se produce a una edad mayor que el de los chicos, coincidiendo las complicaciones con la de los chicos, siendo imprescindible una detección temprana para poder minimizar los riesgos posteriores (Correas, J., 2009).

Con base en los resultados obtenidos, se puede considerar el género como una variable relevante para determinar cuales serán los efectos de los trastornos tanto en chicos como en chicas. Cuestión que permitiría un mejor desarrollo de programas de prevención en problemas psicológicos que interfieren de manera directa en el desarrollo normal de dicha población.

Una mayor profundización sobre el papel modulador de las variables género y la edad permitirá completar este estudio preliminar, analizando su repercusión sobre la prevalencia en los rangos de riesgo y clínico.

## Referencias

- Achenbach, T.M., y Rescorla, L.A. (2000). *Manual for de ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.
- Achenbach, T.M., y Rescorla, L.A. (2001). *Manual for de ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4.<sup>a</sup> edición) (DSM-IV). Washington, DC: Autor.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4.<sup>a</sup> edición Texto revisado) (DSM-IV-TR). Washington, DC: Autor.
- Angold, A. y Costello, E.J. (2005). Epidemiología del desarrollo: la naturaleza del riesgo en los trastornos psiquiátricos. En Ezpeleta (dir.), *Factores de riesgo en psicopatología del desarrollo*. Barcelona: Masson.
- Barkley, R.A. (2006). *Attention-Deficit Hyperactivity Disorder. A handbook for diagnosis and treatment*. New York: The Guilford Press.

- Correas, J. (2009). Género y trastorno por déficit de atención e hiperactividad. En Quintero, F.J., Correas, J. y Quintero, F.J. *Trastorno por déficit de atención e hiperactividad (TDAH) a lo largo de la vida*. 3ª ed. Barcelona: Elsevier-Masson.
- Ezpeleta, L. (dir.) (2005). *Factores de riesgo en psicopatología del desarrollo*. Barcelona: Masson.
- Ezpeleta, L. (2005). Prevención en psicopatología del desarrollo. En Ezpeleta (dir.), *Factores de riesgo en psicopatología del desarrollo*. Barcelona: Masson.
- Gaub, M. y Carlson C.L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal American Academy of Child and Adolescent Psychiatry*, 36, 1036-46.
- House, A.E. (2003). *DSM-IV. El diagnóstico en la edad escolar*. Madrid: Alianza Editorial.
- Hudziak, J., Copeland, W., Stanger, C. y Wadsworth, M. (2004). Screening for DSM –VI externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry*, 45 (7), 1299-1307.
- Kazdin, A.E. (1995). *Conduct disorder in childhood and adolescence* (2.ª edición). Newbury Park, CA: Sage.
- Méndez, F.J.; Olivares, J. y Bermejo, R.M. (2002). Características clínicas y tratamiento de los miedos, fobias y ansiedades específicas. En V.E. Caballo y M.A. Simón (dir.), *Manual de Psicología Clínica Infantil y del Adolescente. Trastornos generales*. Madrid: Editorial Pirámide.
- Rescorla, L.A. (2005). Assessment of young children using the Achenbach System of Empirically Based Assessment (ASEBA). *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 226-237.
- Rutter, M., Caspi, A. y Moffitt, T.E. (2003). Using sex differences in psychopathology to study casual mechanisms: Unifying issues and research strategies. *Journal of Child Psychology and Psychiatry*, 44, 1092-1115.

# CONSIDERACIONES METODOLÓGICAS ACERCA DE LA DETECCIÓN DE ASOCIACIONES MEDIANTE LA INDUCCIÓN AL CONSUMO DE SAL

Alfredo Espinet, José M. Caramés, Rafael Alonso, Ana González Galán,  
Arian Quero y Trinidad Vaquero  
Universidad de Málaga

La inducción al consumo de sal se ha utilizado recientemente como una herramienta para detectar la existencia de asociaciones excitatorias e inhibitorias entre estímulos compuestos de sabores. Una serie de experimentos llevados a cabo en nuestro laboratorio sugiere la conveniencia de realizar siempre una prueba final de consumo de sal tras las pruebas principales relacionadas con el diseño experimental.

Una cantidad importante de las investigaciones llevadas a cabo durante las dos últimas décadas en el campo del aprendizaje animal se ha dirigido a estudiar procesos en los que intervienen las asociaciones que se establecen entre dos o más estímulos neutros cuando éstos se presentan juntos formando un estímulo compuesto. Los estímulos neutros producen, en sus primeras presentaciones, una respuesta de orientación que indica que están siendo procesados por los animales. Este procesamiento tiene como resultado principal el establecimiento de relaciones espacio-temporales entre los estímulos presentados, de forma que los animales aprenden cuáles van juntos y cuáles van separados. La primera demostración de esta capacidad de los animales para relacionar estímulos inicialmente irrelevantes la proporcionó Brodgen (1939) quien con su procedimiento de precondicionamiento sensorial mostró que, de manera análoga a lo que ocurre cuando emparejamos un estímulo condicionado con un estímulo incondicionado, si dos estímulos neutros se presentan emparejados se establece entre ellos una asociación excitatoria.

Las asociaciones excitatorias establecidas entre estímulos neutros constituyen una pieza clave en la articulación de las recientes teorías asociativas del aprendizaje perceptivo (p. e., McLaren y Mackintosh, 2000; Hall, 2003). El “aprendizaje perceptivo” se produce por mera exposición a dos estímulos muy parecidos y consiste en la adquisición de la capacidad de discriminarlos. Gran parte de la investigación sobre aprendizaje perceptivo se ha llevado a cabo empleando como estímulos dos compuestos de sabores, AX y BX, en los que A y B son sabores diferenciales (por ejemplo, dulce y salado) a los que se añade un tercer sabor común X (por ejemplo, ácido) en una proporción tal que los dos compuestos AX y BX resultan difíciles de discriminar. Durante la década de los 90 se llevaron a cabo numerosos experimentos en los que se utilizó el condicionamiento de aversión al sabor como forma de medir la capacidad discriminativa de los animales. En estos experimentos durante una primera fase de preexposición se presentan dos sabores compuestos AX y BX en ensayos alternos (AX, BX, AX, BX, ..., AX, BX) con un largo intervalo entre ensayos. Posteriormente se condiciona aversivamente el compuesto AX emparejando su consumo con el malestar provocado por una inyección de cloruro de litio y, finalmente, en una fase de prueba se mide el consumo de BX. Consumos elevados de BX indican que los animales discriminan bien entre AX y BX. Por el contrario, consumos pequeños de BX indican que los animales no son capaces de diferenciar AX de BX. Los efectos de la preexposición alterna se comparan con los efectos de una preexposición por bloques en la que todos los ensayos con AX preceden a los ensayos con BX o viceversa (AX, AX, AX, ..., BX, BX, BX). De esta forma se evitan diferencias en inhibición latente que afecten al posterior condicionamiento de AX, ya que tanto

en el grupo de preexposición alterna como en el de preexposición por bloques los estímulos AX y BX se presentan el mismo número de veces. El resultado frecuentemente obtenido es que la preexposición alterna facilita la discriminación posterior entre AX y BX mucho más que la preexposición por bloques (Symonds y Hall, 1995). A partir de este resultado una buena parte de las investigaciones en aprendizaje perceptivo han tenido como objetivo desvelar los procesos asociativos subyacentes en cada una de estas dos formas de preexposición. El experimento que se presenta a continuación tenía por objetivo conocer parte de estos procesos, pero nuestra intención al presentarlo aquí es llamar la atención sobre algunos aspectos metodológicos que conviene tener en cuenta para llevar a cabo una adecuada interpretación de los datos.

Es un hecho comprobado (Rescorla y Cunningham, 1978) que entre los dos estímulos que forman un compuesto de sabores se establecen asociaciones excitatorias ( $A \leftrightarrow X$  o bien  $B \leftrightarrow X$ ). Se ha sugerido (Hall, 2003) que estas asociaciones permanecen activas a lo largo de toda la preexposición alterna mientras que en la preexposición por bloques la asociación  $A \leftrightarrow X$  formada a lo largo del primer bloque de ensayos (presentaciones de AX) se extingue durante el segundo bloque de ensayos (presentaciones de BX) ya que en este segundo bloque X se presenta sin ir acompañado por A. Sin embargo, algunos resultados obtenidos en nuestro laboratorio llevan a pensar que cuando la preexposición por bloques es larga (más de doce ensayos con cada compuesto) es posible que no se produzca la extinción de la asociación  $A \leftrightarrow X$ . Esto puede ponerse a prueba mediante un diseño de preconditionamiento sensorial, condicionando X tras la preexposición y comprobando si la aversión adquirida por X se transfiere con la misma o con diferente intensidad hacia A que hacia B. Sin embargo, este procedimiento tiene algunas desventajas. Por ejemplo, durante la fase de condicionamiento X debe presentarse separado de A. Esto puede debilitar la asociación previa establecida entre A y X. Por otra parte, esta no es una medida directa, sino indirecta de la asociación entre A y X. Durante la fase de prueba la aversión suscitada por el sabor A dependerá de la fuerza de la asociación  $A \rightarrow X$  establecida durante la preexposición, pero también dependerá de la fuerza de la asociación  $X \rightarrow$  Malestar adquirida durante el condicionamiento. En consecuencia, con este procedimiento podemos acumular error por diferencias entre grupos en el nivel de condicionamiento X-Malestar o por generalización hacia cualquier otro sabor, de la aversión adquirida por X.

Un procedimiento más directo para evaluar la permanencia de las asociaciones intracompuesto lo proporciona la técnica de inducción al consumo de sal. Fudim (1978) descubrió que las ratas a las que se ha provocado un descenso de los niveles de sodio (p. e., mediante una inyección de furosemida) consumen mucho de cualquier sabor previamente asociado a la sal. Esto permite prescindir del condicionamiento evitando los anteriores problemas y proporciona una medida más directa de la asociación, cuando se utilizan sabores y uno de los estímulos es sal.

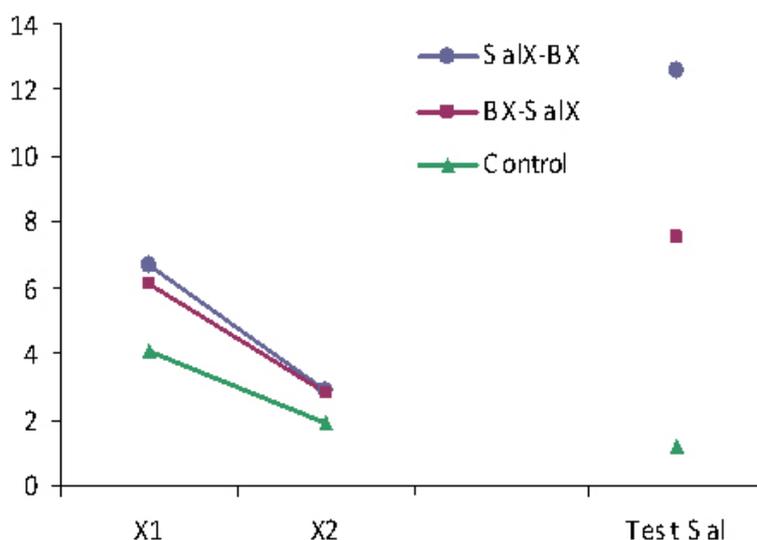
Utilizando esta técnica llevamos a cabo el experimento cuyo diseño se presenta a continuación:

GRUPO	PREEXPOSICIÓN	INYECCIÓN F	PRUEBA
SalX-BX	SalX... BX	Sí	X ?
BX-SalX	BX... SalX	Sí	X ?
Control	SalX...BX / BX...SalX	No	X ?

Todos los grupos recibieron preexposición por bloques a los compuestos SalX y BX. En el grupo SalX-BX las presentaciones del compuesto SalX ocurrieron durante el primer bloque de ensayos, por lo

que la asociación  $\text{Sal} \leftrightarrow X$  podría extinguirse a lo largo del segundo bloque de ensayos (presentaciones de BX). Por el contrario, en el grupo BX-SalX la asociación  $\text{Sal} \leftrightarrow X$  debería estar firmemente establecida al final de la preexposición dado que las presentaciones del compuesto SalX ocurrieron durante el segundo bloque de ensayos. En el grupo de control la mitad de los sujetos recibieron SalX en el primer bloque de ensayos y la otra mitad en el segundo bloque. Tras la preexposición se inyectó furosemida a los grupos SalX-BX y BX-SalX y cuando los animales experimentaban una imperiosa necesidad de sal se les ofreció el sabor X en dos pruebas separadas por un intervalo de 30 minutos.

Los resultados de estas pruebas se muestran en la siguiente figura. Como puede apreciarse los grupos SalX-BX y BX-SalX consumieron cantidades similares de X, el sabor asociado con la sal, en los dos ensayos de prueba, mientras que el grupo Control que no estaba bajo los efectos de la furosemida consumió cantidades menores de X. Los análisis estadísticos revelaron que los grupos SalX-BX y BX-SalX no se diferenciaron significativamente entre sí en sus consumos del sabor X. El grupo Control, cuyos sujetos no estaban bajo los efectos de la necesidad de sal, consumió cantidades significativamente menores de X que los otros dos grupos.



Por tanto, aparentemente, la asociación  $\text{Sal} \leftrightarrow X$  se mantuvo con la misma fuerza en el grupo BX-SalX que en el grupo SalX-BX, es decir, no se extinguió. Sin embargo, una prueba posterior de consumo de sal realizada 30 minutos después del segundo ensayo de prueba con X reveló que el grupo BX-SalX experimentaba una necesidad de sal significativamente menor que el grupo SalX-BX, por lo que este dato invalida la conclusión establecida anteriormente. Si los sujetos del grupo BX-SalX hubieran experimentado la misma necesidad de sal que los sujetos del grupo SalX-BX, ¿no habrían bebido una mayor cantidad de X en los ensayos de prueba?

Estas consideraciones son importantes porque en muchos diseños de este tipo no se realiza una prueba final de consumo de sal tal y como la presentada en este experimento. Dado que estos experimentos se llevan a cabo con un N pequeño no resulta difícil que errores cometidos al inyectar a un par de sujetos pasen inadvertidos y nos lleven a conclusiones erróneas como la que establecimos inicialmente. Por otra parte, como puede apreciarse en la figura, a pesar de que los sujetos del grupo SalX-BX habían consumido abundante cantidad de líquido en los dos ensayos con el sabor X, ingirieron una enorme cantidad de sal en la prueba final en la que se les presentó esta sustancia. Así pues, los resultados de la prueba final de consumo de sal no parecen verse afectados por la ingesta previa de otros líquidos. Por el contrario, esta prueba resulta un indicador muy fiable de la necesidad de sal provocada por el tratamiento. En consecuencia sugerimos que una valoración adecuada de lo que ocurre en las

pruebas en las que se emplea el procedimiento de inducción al consumo de sal para valorar la fuerza de las asociaciones entre sabores debería incluir necesariamente una comprobación de que la inyección ha provocado efectos similares en todos los grupos, es decir, realizar una prueba final de consumo de sal como la presentada en este experimento.

Agradecimientos. Esta investigación ha sido financiada por el Ministerio de Educación y Ciencia (Proyecto SEJ7-67409-C0202).

### Referencias

- Brodgen, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, 25, 323-332.
- Fudim, O. K. (1978). Sensory preconditioning of flavors with a formalin-induced sodium need. *Journal of Experimental Psychology: Animal Behavior Processes*, 4, 276-285.
- Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *Quarterly Journal of Experimental Psychology*, 56B, 43-55.
- McLaren, I. P. L. y Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning and Behavior*, 28, 211-246.
- Rescorla, R. A. y Cunningham, C. L. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes*, 4, 267-275.
- Symonds, M. y Hall, G. (1995). Perceptual learning in flavor aversion conditioning: Roles of stimulus comparison and latent inhibition of common stimulus elements. *Learning and Motivation*, 26, 203-219.

# EL PROCESAMIENTO DE HOMÓGRAFOS INTERLINGUALES: DATOS DE UNA SIMULACIÓN COMPUTACIONAL

Alejandro J. Estudillo y Javier García Orza  
Universidad de Málaga

Los homógrafos interlinguales (HI) son aquellas palabras que teniendo la misma ortografía tienen significados diferentes en dos idiomas (e.g., once). Estos estímulos han servido para investigar si los bilingües al leer una palabra en su lengua no natal activan candidatos de su lengua natal. Usando tareas de decisión léxica algunos estudios han mostrado que los bilingües presentan un retraso en el reconocimiento de los HI, lo que sugiere la existencia de un léxico común para ambos idiomas. El modelo BIA asume estos principios al considerar que el reconocimiento de un HI por parte de un bilingüe conlleva la activación en paralelo de las dos posibles lecturas del homógrafo (una por cada idioma). El presente trabajo pretende simular computacionalmente la conducta de los bilingües ante los HI en el marco del modelo BIA, para ello se presenta una red neuronal artificial multicapa, que funciona con la regla de retropropagación del error. Se utilizó una tasa de aprendizaje de 0.8 y un total de 10000 ensayos de aprendizaje. Los resultados muestran que una red de tipo conexionista es capaz de simular el nivel de activación de un HI en la segunda lengua.

Según la RAE un homógrafo hace referencia a una palabra que, teniendo distinta significación que otra, se escribe de igual manera que ella, ej.: haya. Este fenómeno es muy común dentro de una misma lengua.

Un caso lingüístico más que curioso podemos observarlo en la canción *angel eye* de Sting. Es muy probable que el cantante no supiese que en holandés la palabra *angel* significa *sting*. Este ejemplo es una muestra clara de lo que se conoce como homógrafo interlingual (HI). En español con respecto al inglés tenemos algunos ejemplos que si bien no son tan claros, sí pueden ayudarnos a entender este fenómeno. (Ej.: once).

Los investigadores han hecho uso de los HI para investigar el léxico de los bilingües. Siguiendo a Clark (2002), el léxico o lexicón mental es como un “diccionario” que tenemos en nuestra cabeza, es decir, es el lugar dónde se encuentran todas las palabras que conocemos. En este sentido cabe preguntarse si los bilingües al leer una palabra en su lengua no natal, activan candidatos de su lengua natal. De hecho si un HI es reconocido más rápido o más lento que palabras con significación en una sola lengua, esto tiene que implicar algún tipo de influencia interlingüística entre las dos lecturas (lengua natal y segunda lengua) del homógrafo. Esto anterior solo puede ocurrir si todas las palabras independientemente del idioma se encuentran en el mismo léxico. Esta posición es la que se conoce como visión no selectiva del acceso léxico (Brysbaert, Van Dyck, y Van de Poel, 1999).

El modelo BIA asume estos principios. Según este modelo el reconocimiento por parte de un bilingüe de un HI conlleva la activación en paralelo de las dos posibles lecturas del homógrafo (una por cada idioma). Por ejemplo la palabra *angel* tiene una representación ortográfica conectada a las características fonológicas y semánticas, tanto en inglés como en holandés. Como consecuencia el reconocimiento implica una competición entre las dos posibles lecturas. Este proceso está modulado por la frecuencia del homógrafo en las distintas lenguas. Por lo tanto, cuando un homógrafo tiene una frecuencia más alta en la primera lengua que en la segunda, se produce un proceso extra de competición que resulta en la inhibición de la palabra en la segunda lengua. Existe tanto evidencia conductual (Dijkstra y van Heuven, 1999) como con potenciales evocados (Kerkhofs, Dijkstra, Chwilla, de Bruijn, 2006) que apoyan este modelo.

Con el presente trabajo pretendemos simular mediante una red conexionista el nivel de activación que presenta un concepto homógrafo interlingual en la lengua no natal.

### Simulación

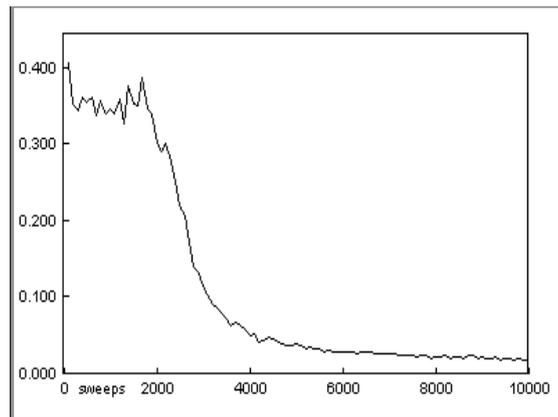
La simulación se llevó a cabo mediante el software Tlearn (Plunkett y Elman, 1997). Hemos utilizado una red neuronal multicapa, con doble capa de unidades ocultas, ya que existía un solapamiento importante entre los vectores (Hinton, 1990). Cada una de estas capas estaba formada por cinco unidades. Por su parte, la capa de unidades de entrada estaba formada por 12 neuronas, dos de las cuales no recibieron ningún entrenamiento hasta la segunda fase, además de que eran reservadas para comprobar el aprendizaje de la red en la fase test. La capa de salida tan solo estaba formada por una neurona. En cada ensayo tan solo se entrenaban dos neuronas. Los valores 1 y 0.3 representaban respectivamente, alta y baja frecuencia. Por otro lado, los valores de salida 0.5, 1 y 0.3 representaban el grado de activación del concepto en la lengua no nata. En este sentido, cuando la red se enfrenta a un homógrafo con una frecuencia alta en la lengua natal y una frecuencia baja en la segunda lengua, el concepto debería estar más activado en la lengua natal, por lo que existirá un proceso extra de competición, por dicha razón la activación de dicho concepto en la segunda lengua será baja. Por el contrario, cuando el concepto presente una baja frecuencia en la lengua natal y una alta frecuencia en la segunda lengua, la activación de este concepto en esta lengua será mayor. Por otro lado, cuando la frecuencia del concepto sea alta en ambas lenguas, el nivel de activación del concepto en la segunda lengua será medio. Por último cuando la frecuencia sea baja en ambas lenguas, el nivel de activación del concepto en la segunda lengua será bajo. Un resumen de esto anterior se puede observar en la tabla 1. Como regla de aprendizaje utilizamos la regla delta, con una tasa de aprendizaje de 0.7 y dos fases de entrenamiento, con 10000 ensayos de aprendizaje cada una.

*Tabla 1. Patrón de vectores*

<b>Frec. L1</b>	<b>Frec L2</b>	<b>TUTOR</b>
1	1	0.5
0.3	1	1
1	0.3	0.3
0.3	0.3	0.5

### *Resultados Fase I*

Los resultados de esta primera fase de entrenamiento se pueden observar en el gráfico 1. Esta gráfica representa el error. Como se puede observar en dicho gráfico, el aprendizaje es óptimo a partir del ensayo 6000 aproximadamente.



*Gráfico 1. Tasa de error entrenamiento 1*

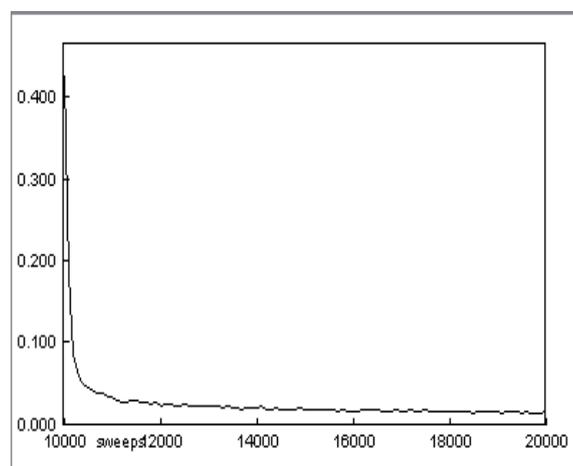
### *Simulación de la fase II*

En esta fase utilizamos exactamente los mismos parámetros que en la primera fase. Tan solo entrenamos a las dos neuronas últimas, las cuales no habían sido entrenadas en la primera fase, con dos vectores aleatorios (ej.: 0 0 0 0 0 0 0 0 0 0 1 1 y 0 0 0 0 0 0 0 0 0 0 .3 .3), y congelamos los pesos resultantes del aprendizaje anterior, con el fin de comprobar si el aprendizaje previo es capaz de resolver el nuevo problema.

### *Resultados fase II*

Los resultados de esta fase de entrenamiento se pueden observar en el gráfico 2. Como se puede observar, la tasa de error es prácticamente inexistente hacia el ensayo 1000.

Por último, en la fase test, presentamos los dos vectores que no habían sido presentados en la fase anterior.



*Gráfico 2. Tasa de error segundo entrenamiento*

### *Resultados fase test*

El nivel de activación de las neuronas de salida ante los vectores 0 0 0 0 0 0 0 0 0 0 .3 1 y 0 0 0 0 0 0 0 0 0 0 .3 .3 fue de 0.93 y 0.45, respectivamente. Estos resultados muestran que nuestra red es capaz de simular el nivel de activación de un homógrafo interlingual en la segunda lengua.

### Conclusión

Kerkhofs y colaboradores (2006) han demostrado que los efectos inhibitorios de la segunda lengua serían reducidos introduciendo una relación semántica *prime-target*, debido a que el *prime* aumenta la activación de la palabra *target*. Los mayores efectos serían obtenidos para homógrafos que tienen una baja frecuencia en la lengua natal y una frecuencia alta en la segunda lengua. Según estos autores, el *prime* ayudaría al sujeto a “contextualizarse” dentro de una lengua determinada. Una nueva simulación debería de tener en cuenta tales resultados.

### Referencias

- Brysbaert, M., Van Dyck, G. & Van de Poel, M. (1999). Visual Word Recognition in Bilinguals: Evidence From Masked Phonological Priming. *Journal of Experimental Psychology-Human Perception and Performance*, 25(1), 137-148.
- Dijkstra, T. & van Heuven, W. J. B. (1998). The BIA-model and bilingual word recognition. In J. Grainger, & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Erlbaum.
- Dijkstra, T. y Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175-197.
- Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, 46, 47-76.
- Kerkhofs, R., Dijkstra, A., Chwilla, D. y De Bruijn, E. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and ERP effects. *Brain Research*, 1068, 170-183.
- Plunkett, K. & Elman, J. L. (1997). *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. Cambridge, MA: MIT Press.

# PREDICCIÓN DE LOS RESULTADOS EN PISTA PARA PIRAGÜISTAS DE ALTA COMPETICIÓN

José Luis García Soidán y Pilar García Soidán  
Universidad de Vigo

Las variables medidas en el laboratorio, conjuntamente con las pruebas realizadas en pista, proporcionan indicadores sobre la condición física de los deportistas. En este sentido, sería de gran interés modelizar la relación existente entre ellas, lo cual permitiría predecir los objetivos al alcance de los deportistas en cada momento de la temporada. Lo anterior resume el objetivo de este trabajo, enfocado al piragüismo español de alta competición. Para llevarlo a cabo se contó con la colaboración de diversos integrantes de la selección junior nacional de piragüismo, a los cuales se les realizaron pruebas de kayak-ergómetro y otros análisis de laboratorio, conjuntamente con tests en pista de 1000 m. Con los datos resultantes fue posible establecer y validar un modelo de regresión lineal, cuya aplicación permite predecir los resultados que podrían alcanzar los deportistas en las pruebas en pista, a partir de las variables medidas en el laboratorio más correlacionadas.

Los entrenadores habitualmente recurren a la realización de tests de campo específicos, adaptados a los distintos deportes y a sus diferentes modalidades, para comprobar la condición física de los deportistas en cada etapa y así poder evaluar el desarrollo de su programa anual de preparación. Los resultados de estos tests se suelen complementar con datos obtenidos en el laboratorio, donde se realizan pruebas de esfuerzo específicas, que proporcionan indicadores del estado de forma del deportista, además de otros parámetros relacionados con aspectos biomecánicos y fisiológicos. A partir de esta información, sería de gran ayuda poder modelizar la relación existente entre los datos obtenidos en los tests de pista y los de laboratorio, que permitiese inferir el resultado que previsiblemente alcanzarían los deportistas en competición. Teniendo presente que España es un país con gran tradición en las distintas especialidades que conforman el piragüismo de competición, se enfocó el presente trabajo hacia este deporte.

Para el piragüismo, los tests en pista se realizan en diferentes momentos de la temporada, sobre las distancias contempladas en estas pruebas en competición. Por otra parte, desde hace varios años se utiliza el kayak-ergómetro como medio de entrenamiento e investigación del trabajo realizado por el deportista y existen diversos estudios que avalan su uso con este propósito (Van Someren & Dumbar, 1996).

En los últimos años, distintos autores han encontrado relaciones estadísticamente significativas entre los tests de pista y los realizados en el laboratorio (Mitchell & Swaine, 1998; Palmer, Phillips & Van Someren, 2000). Atendiendo a las necesidades tanto aeróbicas como anaeróbicas del piragüismo, se han establecido distintos parámetros fisiológicos, como indicadores del rendimiento del piragüista: la capacidad aeróbica, la concentración de lactato en sangre o la frecuencia cardíaca, entre otros. Otros trabajos han analizado las variables cinemáticas relacionadas con la distribución del esfuerzo en 500 y 1000m. (Issurin, 1998; Bishop, Bonetti & Dawson, 2002), así como los ritmos de palada comparados entre kayak-ergómetro y tests de pista.

El presente trabajo se enfocó al estudio del piragüismo de alta competición, con objeto de establecer un modelo que permitiese predecir los resultados que podrían alcanzar los deportistas en las pruebas en pista, a partir de las variables medidas en el laboratorio más correlacionadas. Para llevarlo a cabo se contó con la colaboración de diversos integrantes de la selección junior nacional de piragüismo, a los cuales se les realizaron pruebas de kayak-ergómetro y otros análisis de laboratorio,

en dos períodos de tiempo. Además, con un margen de 15 días de diferencia con respecto a las pruebas anteriores, se llevaron a cabo tests en pista de 1000 m. El procedimiento seguido y los resultados obtenidos se describen en las secciones siguientes.

## Método

### Participantes

Integran la muestra 6 piragüistas de la selección nacional junior de canoa, que en el momento de la investigación realizaban sus entrenamientos en el Centro Gallego de Tecnificación Deportiva de Pontevedra y figuraban entre los mejores de Europa en su categoría.

### *Diseño del estudio*

El piragüismo es un deporte que necesita de una gran capacidad tanto anaeróbica como aeróbica que viene definida por las distintas distancias de competición. Para cuantificar los parámetros asociados, se llevaron a cabo los siguientes tests en dos momentos diferentes de la temporada (marzo y julio), con un margen de 15 días entre ellos: pruebas de kayak-ergómetro y de laboratorio y tests en pista.

Previamente a la realización de las pruebas, se informó de las mismas a los deportistas y a sus padres o tutores legales, solicitando su consentimiento para proceder con el estudio. Durante las 48 horas inmediatamente anteriores, los participantes no deberían ejecutar ningún esfuerzo físico ni psíquico intenso y la última ingesta debería realizarse con 3 horas de antelación como mínimo. De este modo, en la hora fijada para cada deportista, se llevó a cabo un reconocimiento médico de aptitud deportiva, con especial atención a los aparatos cardiovascular y locomotor. Además, se efectuó un test maximal con kayak-ergómetro (modelo Dansprint, DK ®), con una duración de 240 segundos. En cuanto a los tests en pista, éstos se realizaron en pista homologada de 1000 metros. Para el tratamiento de la información obtenida, se diseñó una aplicación en formato SPSS para la grabación de datos, que luego fueron depurados y analizados utilizando el mismo programa estadístico SPSS.

### *Procedimiento*

Con la información recabada se trató de establecer la correlación existente entre los parámetros observados en el laboratorio y los resultados obtenidos en pista, de manera que fuese posible determinar la eficacia del sistema de entrenamiento y, lo que es más importante, prever el tiempo de ejecución de la misma al alcance del deportista. Para ello, se tuvo presente que, para cada participante en el estudio y para cada una de las variables medidas, se disponía de dos valoraciones correspondientes a los dos momentos de la temporada seleccionados,  $X_1$  y  $X_2$ . No obstante, a la hora de plantear el estudio, se utilizó la tasa de variación absoluta para cada deportista y para cada variable,  $Y=X_2-X_1$ , que resultó significativa para la construcción de un modelo de regresión lineal.

## Resultados

Tras un primer análisis exploratorio de las diferentes variables medidas en el laboratorio, se decidió considerar en el estudio las que presentaban mayor correlación: el peso (P), la potencia promedio desarrollada (Wmed), la frecuencia cardiaca máxima (FCmax), lactato máximo (LACTmax) y  $VO_2$  promedio ( $VO_{2med}$ ). En la tabla 1 se resumen los valores obtenidos para las correlaciones entre cada

una de las variables anteriores y el tiempo conseguido en pista, deduciéndose que la mayor correlación corresponde a  $VO_2$ med, seguida por la frecuencia cardiaca máxima y el peso absoluto.

*Tabla 1. Valores del coeficiente de correlación lineal*

Variable	Correlación con el tiempo
$VO_2$ med	0,951
FCmax	0,610
P	0,364
LACTmax	0,169
Wmed	-0,089

Seguidamente se procedió a aplicar un análisis de regresión lineal múltiple, tomando como variable dependiente el tiempo alcanzado en pista y las restantes como independientes, incorporando estas últimas en el modelo de regresión mediante el método de introducción por pasos hacia delante. Al aplicar este procedimiento, las variables independientes que resultaron significativas fueron  $VO_2$ med y Wmed y, según se desprende de la tabla 2, entre ambas variables independientes explican un 97,7% de la varianza del modelo y reducen el error típico en la estimación a 0,37846.

*Tabla 2. Medidas de la bondad del ajuste del modelo de regresión lineal en cada etapa*

Etapa	R <sup>2</sup>	R <sup>2</sup> corregido	Error típico	Durbin-Watson
1	0,904	0,880	0,86880	
2	0,986	0,977	0,37846	2,483

Por otra parte, se puede afirmar que existe una relación lineal significativa entre las variables incluidas en el modelo y la variable dependiente, al nivel de significación del 5%, como se desprende de la tabla 3.

*Tabla 3. Tabla ANOVA del modelo de regresión lineal final*

Modelo	Suma de cuadrados	Grados de libertad	F	Significación
Regresión	31,124	2	108,647	0,002
Residual	0,430	3		
Total	31,553	5		

A partir de la información que figura en la tabla 4, sobre los coeficientes del modelo final, se puede aceptar que los parámetros que intervienen en el modelo de regresión son no nulos, ya que el nivel de significación empírico de cada test no supera el 5%.

Tabla 4. Coeficientes del modelo de regresión lineal final

Variable	Coef. no estand.		Coef. estand.	t	Sig.
	B	Error típico	Beta		
(Constante)	- 0,757	0,222		-3,417	0,042
VO <sub>2</sub> med	0,009	0,001	1,066	14,682	0,001
Wmed	0,037	0,009	0,309	4,252	0,024

Además, los valores estimados para los coeficientes no estandarizados, conducen a la siguiente ecuación de regresión, donde TV representa la tasa de variación:

$$TVTiempo = -0,757 + 0,009 \cdot TVVO_2med + 0,037 \cdot TVWmed$$

Cabe destacar también que, aunque *a priori* la variable Wmed parecía no estar muy relacionado con el tiempo alcanzado en pista, sí que aporta una influencia importante al controlar las restantes variables independientes. De hecho, tal y como se deduce de la tabla 5, la correlación total entre Wmed y el tiempo es relativamente escasa (-0,089), pero cambia significativamente la situación al eliminar de ambas el efecto de las restantes variables independientes, modificándose el valor del coeficiente de correlación parcial hasta 0,926.

Tabla 5. Correlaciones de orden cero y parciales en el modelo de regresión lineal final

Variable	Correlaciones	
	Orden cero	Parcial
VO <sub>2</sub> med	0,951	0,993
Wmed	-0,089	0,926

Podemos concluir que no hay colinealidad significativa en el conjunto de datos, a la vista de los resultados de la tabla 6, ya que los índices de condición toman valores pequeños. Estos índices de condición se obtienen como la raíz cuadrada del autovalor más grande y cada uno de los restantes, donde los autovalores proporcionan información sobre el número de dimensiones subyacentes en el conjunto de variables independientes utilizadas.

Tabla 6. Diagnósticos de colinealidad

Dimensión	Autovalor	Índice de condición
1	2,238	1,000
2	0,469	2,185
3	0,293	2,762

Para finalizar el análisis, se procedió a verificar la validez del modelo. La condición de independencia de los residuos se puede aceptar a partir del test de Durbin-Watson, que proporciona un valor comprendido entre 1,5 y 2,5 (tabla 3).

Por otra parte, la estimación de los valores residuales del modelo, reflejados en el gráfico siguiente, permite concluir las hipótesis de media cero y homocedasticidad.

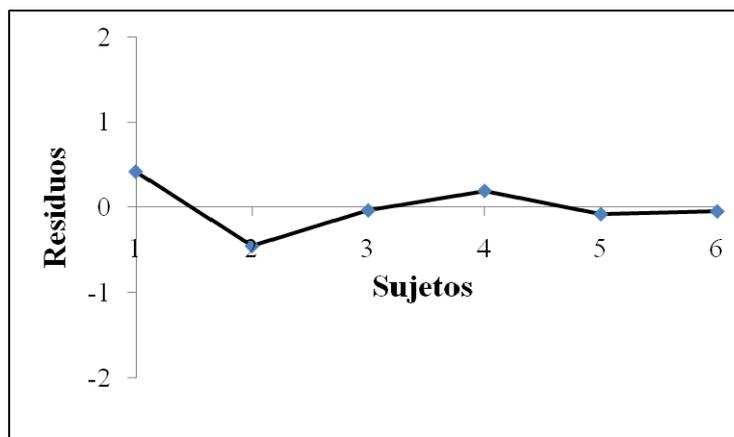


Figura 1. Gráfico de dispersión de los valores residuales del modelo

Finalmente, la normalidad de los residuos se puede aceptar a partir del test de Kolmogorov-Smirnov, cuya significación supera ampliamente el valor 0,05, según se refleja en la tabla 7.

Tabla 7. Prueba de Kolmogorov-Smirnov para los residuos

Parámetros normales		Diferencias más extremas			Test de K-S	Sig. asint.
Media	Desv típica	Absoluta	Positiva	Negativa		
0,000	0,293	0,227	0,214	-0,227	0,556	0,916

### Conclusión

Los resultados obtenidos en el presente trabajo proporcionan una perspectiva interesante de estudio, al abrir la posibilidad de establecer una ecuación de ajuste entre los tiempos obtenidos en las pruebas de pista por los deportistas y los valores medidos previamente en laboratorio, tomando como base las variables más correlacionadas. Ciertamente, el tamaño reducido de la muestra impone la necesidad de contrastar la significatividad de la regresión establecida sobre un conjunto más amplio de datos y de hacer un seguimiento de su tendencia y evolución a lo largo del tiempo, para determinar patrones de comportamiento de las mismas u otros factores que tengan relevancia. Sin embargo es un punto de partida novedoso, que aporta luz sobre un mecanismo que contribuye a determinar el mejor momento de forma de un deportista, a partir de información objetiva obtenida en las pruebas de laboratorio.

Particularmente en este estudio se ha puesto de manifiesto que dos variables clave para determinar el rendimiento en este deporte son el  $VO_2$  promedio y la potencia promedio desarrollada, frente a un conjunto más amplio de factores de observación, que globalmente explican casi el 98% de la tasa de variación de los tiempos. Además, la posibilidad de establecer un ajuste válido, en este caso lineal, permite cuantificar los grados de relación entre las variables significativas, así como prever las marcas futuras de los deportistas, por lo que puede ser de gran ayuda para medir la eficacia de los programas de entrenamiento y cuantificar el estado de forma de los piragüistas de élite en cada momento de la temporada.

Agradecimientos. El presente trabajo ha sido desarrollado en el marco del Proyecto 17/UPB10/07 “Correlación y transferencia entre los resultados de laboratorio obtenidos con kayakergómetro y los tests de entrenamiento en pista, realizados en piragüistas de élite”, financiado por el Consejo Superior de Deportes-Ministerio de Educación y Ciencia.

## Referencias

- Bishop, D., Bonetti, D. & Dawson, B. (2002). The influence of pacing strategy on VO<sub>2</sub> max and supramaximal kayak performance. *Medicine And Science In Sports And Exercise* 34, 1041-1047.
- Cuesta, G., Polo, J.M. & Padilla, S. (1991). Correlación entre la marca deportiva obtenida en test de campo y parámetros fisiológicos obtenidos en laboratorio, en piragüistas adolescentes. *Apunts* 18, 131-142.
- Issurin, V. (1998). Analysis of the race strategy of world-class kayakers. En V. Issurin (Ed.), *Science & practice of canoe/kayak high-performance training: selected articles in memory of junior world champion Nevo Eitan* (pp. 27-39). Tel-Aviv: Elite Sport Department of Israel.
- Mitchell, R. & Swaine, D. (1998). A biomechanical analysis of the Olympic-style flatwater kayak stroke. *Medicine & Science in Sports & Exercise*, 12, 183-188.
- Palmer, G.S., Phillips, G.R. & Van Someren, K.A. (2000). Comparison of physiological responses to open water kayaking and kayak ergometry. *International Journal of Sport Medicine*, 21, 200-204.
- Van Someren, K.A. & Dumbar, G.M. (1996). An investigation into the use of a kayak ergometer for the determination of blood lactate profiles in international kayakers. *Journal of Sport Sciences*, 14, 102.
- Van Someren, K.A. & Oliver, J.E. (2002). The Efficacy of ergometry determined heart rates for flatwater kayak training. *International Journal of Sport Medicine*, 23, 28-32.
- Zamparo, P., Capelli, C. & Guerrini, G. (1999). Energetics of kayaking at submaximal and maximal speeds. *European Journal of Applied Physiology*, 80, 542-548.

# DISCREPANCIAS ENTRE LOS CONOCIMIENTOS AUTO- PERCIBIDOS Y LOS CONOCIMIENTOS REALES EN SEXUALIDAD

Elena García-Vega<sup>1</sup>, Elena Menéndez-Robledo<sup>2</sup>, Paula Fernández García<sup>1</sup> y  
Alejandra Menéndez<sup>3</sup>

<sup>1</sup>Universidad de Oviedo

<sup>2</sup>Psicóloga y enfermera de los Servicios de Salud del Principado de Asturias

<sup>3</sup>Socióloga del Principado de Asturias

La juventud actual es la generación con mayor acceso a la información sobre sexualidad, sin embargo el número de embarazos no deseados e infecciones de transmisión sexual (ITS) está aumentando. El objetivo de este estudio es conocer las diferencias entre los conocimientos que los adolescentes tienen sobre sexualidad y los que ellos creen tener. La muestra estuvo constituida por 815 adolescentes (52.8% chicas, 47.2% chicos), entre los 13 y los 19 años, de tres centros públicos de enseñanza de España. Los resultados revelaron que los adolescentes creyeron estar mejor informados de lo que en realidad estaban. Las chicas demostraron más conocimientos que los chicos, sobre todo en materia de anticoncepción e ITS. Dieron mucho crédito a la información obtenida de sus padres. No ven a los profesores como fuentes de adquisición de conocimientos en sexualidad.

En España, en 2008 se produjeron un 13,48% de interrupciones voluntarias del embarazo (IVE) en menores de 20 años, situándose Asturias en el puesto decimoprimer en número total de IVEs. (MSC. Plan Estadístico Nacional sobre Interrupciones Voluntarias del embarazo, 2008).

En la actualidad, los jóvenes de entre 15 y 24 años representan el 45% de todas las nuevas infecciones por el VIH en adultos. Según estos datos, los jóvenes deben ser objeto de acciones preventivas, dado que la UNESCO estima que 40 millones de adolescentes se infectarán hasta el año 2030 (ONUSIDA 2007, 2008. Informe sobre la epidemia mundial de sida, Ginebra).

Según la encuesta Daphne sobre anticoncepción, llevada a cabo en nuestro país en 2007, solo un 42,4% de las menores de 19 años declararon usar el preservativo en sus relaciones (Equipo Daphne, 2007).

A la luz de todos estos datos vertidos a modo de radiografía de la situación de la salud sexual de nuestros jóvenes, cabe plantearse que algo no está funcionando a nivel de prevención y promoción en salud sexual y reproductiva.

Organismos como la OMS considera a la población adolescente como una población de riesgo en materia de sexualidad, atribuyéndole prioridad a la hora de poner en marcha intervenciones a nivel preventivo y educativo (Palenzuela, 2006).

Si queremos conseguir que nuestros/as jóvenes adopten métodos de protección eficaces, deberemos trabajar sobre determinadas variables que pueden estar influyendo negativamente, como puedan ser una escasa información procedente de fuentes no fiables o una falsa percepción de la información con la que cuentan (Oliva, Serra y Vallejo, 1997; Santín 2003).

La información que nuestros jóvenes reciben de sus iguales en materia de sexualidad tiene más

influencia en ellos que aquella impartida desde la escuela. Esta información no formal que reciben contribuye a perpetuar valores y creencias erróneas (Santín, 2003; Testa, 2007), y les lleva a no asimilar bien toda aquella información sobre conductas preventivas frente a ITS y SIDA (Palenzuela, 2006).

El objetivo del presente estudio es el de evaluar la discrepancia existente entre los conocimientos reales y los auto-percibidos por los jóvenes en sexualidad, considerando que esta puede ser una variable que esté influyendo en que los jóvenes adopten conductas sexuales de riesgo.

## Método

### *Participantes*

La muestra estuvo constituida por 815 estudiantes de 2º de la ESO a 2º de bachiller, de tres institutos de enseñanza secundaria de Oviedo (Asturias), con edades comprendidas entre los 13 y los 19 años ( $M=15,65$  y  $DT=1,42$ ). El 52,8% fueron mujeres y el 47,2% hombres.

La elección de la muestra se llevó a cabo mediante un muestreo estratégico dado que debieron quedar fuera del estudio aquellos institutos privados o concertados, por no estar obligados a llevar a cabo programas de educación sexual regulados por el Ministerio de Educación. Del universo de 8 institutos, se seleccionaron al azar un instituto del centro de la ciudad y dos de barrios periféricos, uno al norte y otro al sur.

De un universo muestral de 12.258 jóvenes con edades comprendidas entre 13 y 19 años en la ciudad de Oviedo (Fuente: INE. Padrón 01-01-08), fueron estudiados un 7% de esta franja de edad, lo que supone, a un nivel de confianza del 95%, asumir un margen de error del 3%.

### *Instrumentos*

Se elaboró un cuestionario *ad hoc* para la investigación. Se establecieron como variables *sociobiográficas* la edad, el sexo, la nacionalidad familiar, el número de hermanos y su lugar ocupado, las creencias religiosas y políticas, y la orientación sexual, por estar documentada la influencia de las mismas en la conducta sexual (Ballester y Gil, 2006). En cuanto a la variable *conocimientos*, se trató de establecer las fuentes de obtención de información así como los mitos que en materia de sexualidad siguen estando vigentes en nuestros días (López-Villaverde, 2004; Santín, Torrico, López y Revilla, 2003).

El análisis de la variable *conducta sexual* se centró en establecer las prácticas sexuales llevadas a cabo por los jóvenes, su primera relación sexual, las medidas de protección usadas, la frecuencia de sus relaciones, ... (Torres, Walker, Gutierrez y Bertozzi, 2006; Santín et al. 2003). La *conducta sexual de riesgo* fue evaluada analizando las situaciones en las que no se usó el preservativo y el número de prácticas sexuales de riesgo (sexo oral, anal y vaginal) llevadas a cabo por los jóvenes encuestados.

### *Procedimiento*

La recogida de datos fue colectiva mediante la aplicación de auto-informes. Para cubrir el cuestionario se colocó a los alumnos en una situación tipo examen, intentando conseguir una cierta intimidad y así, la máxima colaboración y sinceridad posible.

Se les informó del carácter anónimo y voluntario del cuestionario. Se aclararon las dudas surgidas, procurando no influir en la respuesta de los sujetos. Los encuestadores estuvieron presentes durante la administración del cuestionario para proporcionar ayuda.

Diseño: El diseño se corresponde con un estudio descriptivo mediante auto-informes con diseño transversal (Montero y León, 2002).

Las técnicas de análisis de datos aplicadas fueron los análisis descriptivos de las variables y análisis no paramétricos, empleando como paquete estadístico el SPSS en su versión 15.0.

## Resultados

Los adolescentes encuestados refirieron estar bien o muy bien informados en materia de sexualidad en un 75.9%, encontrándose diferencias estadísticamente significativas entre el grado de información auto-percibido y la edad ( $F= 5.450$ ,  $p<0.000$ ), así se observó que el grupo de 15 a 16 años eran los que se consideraba mejor informado, frente a un 17.9% de los de 13 a 14 años.

*Tabla 1. Conocimientos auto-percibidos*

CONOCIMIENTOS AUTO-PERCIBIDOS					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Poco o mal informado/a	196	24,0	24,1	24,1
	Bien o muy bien informado/a	616	75,6	75,9	100,0
	Total	812	99,6	100,0	
Perdidos	Sistema	3	,4		
Total		815	100,0		

No se encontraron diferencias estadísticamente significativas correlacionando sexo con grado de información declarado.

Un 72.8% del alumnado dijo haber adquirido sus conocimientos sobre sexualidad de los/as amigos/as, un 55.7% de libros y/o televisión, un 38.9% de los padres y un 29.6% de la experiencia.

De los que adquirieron conocimientos de sus padres, un 81.1% se consideró bien o muy bien informado, un 73.8% de los que habían obtenido información de sus amigos y un 87% si los conocimientos los habían conseguido mediante la propia experiencia.

Se confirmó que a más edad, más obtuvieron sus conocimientos mediante la experiencia ( $F= 12.831$ ,  $p<0.000$ ).

Solo se hallaron diferencias estadísticamente significativas al correlacionar sexo con fuentes de obtención de conocimientos para padres (chicas=45.1%, chicos=31.9%;  $F=15.322$ ,  $p<0.000$ ), profesionales sanitarios (chicas=19.4%, chicos=11%;  $F= 11.293$ ,  $p< 0.01$ ), libros/TV... (chicas=49.3%, chicos=62.9%;  $F= 15.513$ ,  $p<0.000$ ) y experiencia (chicas=25%, chicos=34.7%;  $F= 9.303$ ,  $p<0.01$ ).

Del alumnado que declaró estar bien o muy bien informado, un 44.2% demostró tener unos

conocimientos deficientes ( $F=36.201$ ,  $p<0.000$ ).

Tabla 2. Conocimientos demostrados

CONOCIMIENTOS AUTOPERCIBIDOS		CONOCIMIENTOS REALES		Total
		Conocimientos deficientes	Conocimientos aceptables o buenos	
FININFOR	Poco o mal informado/a	70,9%	29,1%	24,1%
	Bien o muy bien informado/a	44,2%	55,8%	75,9%
Total		100%	100%	100%

Un 40.4% de la muestra total demostraron tener unos conocimientos entre aceptables y buenos. Analizando la temática de estos conocimientos, solo un 26.2% demostraron conocimientos aceptables o buenos en métodos anticonceptivos/ITS, frente a un 71.6% en fisiología/sexualidad.

Tabla 3. Conocimientos demostrados, por materias

CONOCIMIENTOS DEMOSTRADOS, POR MATERIAS		Conoc. total	Conoc. Anticoncep.	Conoc. Fisiolog.
Válidos	Conocimientos deficientes	59.6%	73.8%	28.4%
	Conocimientos aceptables o buenos	40.4%	26.2%	71,6%

Se encontró que, de los que habían recibido información de profesionales sanitarios, un 58.7% demostró tener un nivel de conocimientos aceptable o bueno ( $F=4.245$ ,  $p<0.05$ ) siendo este porcentaje del 64.6% en aquellos que habían obtenido la información de la experiencia ( $F=25.854$ ,  $p<0.000$ ).

Se encontraron diferencias estadísticamente significativas, en función del sexo, para los conocimientos ( $F=9.876$ ,  $p<0.01$ ), de tal manera que un 58.1% de las chicas obtuvieron una puntuación aceptable o buena, frente a un 41.9% de los chicos. Analizando por separado los mitos relacionados con anticoncepción/ITS y con fisiología/sexualidad, encontramos que solo existen diferencias estadísticamente significativas en los primeros ( $F=11.150$ ,  $p<0.000$ ), donde las chicas demostraron en mayor porcentaje tener más conocimientos; no hallándose tal diferencia con los mitos sobre fisiología/sexualidad.

Tabla 4. Conocimientos demostrados, por material y por sexo

CONOCIMIENTOS DEMOSTRADOS, POR MATERIA Y POR SEXO	CONOCIMIENTOS DEFICIENTES		CONOCIMIENTOS ACEPTABLES O BUENOS	
	Chicas	Chicos	Chicas	Chicos
MITOS GENERALES	46%	54%	58.1%	41.9%
MITOS EN ANTICONCEPCIÓN/ITS	48.7%	51.3%	63.5%	36.5%
MITOS EN FISIOLOGÍA/ SEXUALIDAD	50.3%	49.7%	55.2%	44.8%

$F=9.876$ ,  $p<0.01$

## Discusión

Pese a que la mayoría de los alumnos encuestados se declararon bien o muy bien informados, casi la mitad de estos demostraron tener unos conocimientos deficientes.

Detrás de esto puede estar el hecho de que la información que obtuvieron en materia de sexualidad fue principalmente de amigos y medios de comunicación, todos ellos fuentes no formales de adquisición de conocimientos, a los que atribuyen gran veracidad, y que parecen estar contribuyendo a mantener las creencias erróneas sobre sexualidad.

Los que más conocimientos demostraron habían obtenido la información de su propia experiencia y de los profesionales sanitarios. Algo a tener en cuenta es que los profesores no son vistos por el alumnado como una fuente de adquisición de conocimientos.

De otra parte, la buena información demostrada se basaba en conocimientos sobre fisiología/sexualidad, más que en conocimientos sobre anticoncepción/ITS (Testa, 2007; López-Villaverde, 2004).

Una variable que debería tenerse en cuenta en los programas formativos, es el desequilibrio que, en materia de información, tienen chicos y chicas, debido a que la información que estos reciben es emitida con un sesgo sexista. Así nos encontramos, por ejemplo, con que son las chicas las que más conocimientos tienen sobre métodos anticonceptivos e infecciones de transmisión sexual, cuando la anticoncepción y la prevención debe ser cosa de dos. Campañas como la del uso del ‘doble método’ están encaminadas a implicar a los chicos en la prevención, en su caso, de las infecciones de transmisión sexual, mientras que la chica asume la responsabilidad de evitar el embarazo no deseado. Por eso, nuestros esfuerzos deben dirigirse a educar a los adolescentes en la responsabilidad compartida, teniendo que hacer especial hincapié, en el caso de los chicos, en la formación en afectividad, en la educación en el compromiso y en la responsabilidad.

Otro punto sobre el que incidir es el hecho de que los jóvenes declararon concederle gran crédito a la información recibida por parte de sus padres (Rodrigo, Márquez, García, Mendoza, Rubio, Martínez y Martín, 2004; Rose, Koo, Bhaskar, Anderson, White and Jenkins, 2005; Ogle, Glasier and Riley, 2008), lo que nos debe llevar a plantearnos estrategias donde los padres jueguen un papel principal en la formación de sus hijos para tratar de restar importancia a las fuentes de información a las que los jóvenes están recurriendo, los amigos y los medios de comunicación.

El presente estudio viene a aportar información valiosa de cara a tener en cuenta en la elaboración de programas formativos en materia de sexualidad, al mostrarnos a los padres como una posible fuente de transmisión de conocimientos, al revelarnos que los profesores no son vistos como formadores en esta materia, que existen diferencias formativas en cuestión de sexo, y al hecho de que tenemos que ajustar la percepción de los jóvenes con la realidad, ya que esta discrepancia puede estar detrás de las conductas sexuales de riesgo emitidas por los jóvenes (Bradley-Stevenson, 2007; López-Villaverde, 2004).

## Referencias

- Ballester, R. y Gil, M.D. (2006). La sexualidad en niños de 9 a 14 años. *Psicothema*, 18(1), 25-30.
- Bradley-Stevenson, C. (2007). Adolescent sexual health. *Paediatrics and child health*, 17/12, 474-479.
- Equipo Daphne. Encuesta sobre anticoncepción en España, 2007. [www.equipodaphne.es](http://www.equipodaphne.es)
- Fuente: INE. Padrón 01-01-08.
- López-Villaverde, P. (2004). Influencia de los mitos y falacias en las prácticas sexuales de la población adolescente madrileña: proyecto de encuesta. *Revista Profesional Española de Terapia Cognitivo-Conductual*, 2, 87-93.
- Montero, I. y León, O.G. (2007). Guía para nombrar los estudios de investigación en Psicología. *Internacional Journal of Clinical and Health Psychology*. 7(3), 847-862.
- MSC. Plan Estadístico Nacional sobre Interrupciones Voluntarias del embarazo, 2008.
- Ogle, S., Glasier, A. and Riley, S.C. (2008). Communication between parents and their children about sexual health. *Contraception*, 77, 283-288.
- Oliva, A., Serra, L. y Vallejo, R. (1997). Patrones de comportamiento sexual y contraceptivo en la adolescencia. *Infancia y Aprendizaje*, 77, 19-34.
- ONUSIDA, 2007-2008. *Informe sobre la epidemia mundial de sida*, Ginebra.
- Palenzuela, A. (2006). Intereses, conducta sexual y comportamientos de riesgo para la salud sexual de escolares adolescentes participantes en un programa de educación sexual. *Análisis y modificación de conducta*, 32 (144), 451-496.
- Rodrigo, M.J., Márquez, M.L., García, M., Mendoza, R., Rubio, A., Martínez, A. y Martín, J.C. (2004). Relaciones padres-hijos y estilos de vida en la adolescencia. *Psicothema*, 16(2), 203-210.
- Rose, A., Koo, H.P., Bhaskar, B., Anderson, K., White, G. and Jenkins, R. (2005). The influence of primary caregivers on the sexual behavior of early adolescents. *Journal of Adolescent Health*, 37, 135-144.
- Santín, C., Torrico, E., López, M.J. y Revilla, C. (2003). Conocimiento y utilización de los métodos anticonceptivos y su relación con la prevención de enfermedades de transmisión sexual en jóvenes. *Anales de psicología*, 19(1), 81-90.
- Testa, A. (2007). Sexual health knowledge, attitudes and behaviours among an ethnically diverse sample of young people in the UK. *Health Education Journal*, 66, 1, 68-81.
- Torres, P., Walker, D.M., Gutierrez, J.P. y Bertozzi, S.M. (2006). Estrategias novedosas de prevención de embarazo e ITS/VIH/SIDA entre adolescentes escolarizados mexicanos. *Salud Pública de México*, 48(4), 308-316.

# EVOLUCIÓN DE LA EDAD DE INICIO EN EL CONSUMO DE DROGAS Y LA DEMORA EN EL TRATAMIENTO

Francisco Herrero<sup>1</sup>, Pilar Jara<sup>1</sup>, Marta Salvador<sup>1</sup> y Francisco Verdú<sup>2</sup>

<sup>1</sup>Universidad 'Jaume I' de Castellón

<sup>2</sup>Generalitat Valenciana

El presente trabajo expone los tiempos de demora que se producen desde el inicio de consumo de drogas hasta la petición de ayuda para su tratamiento. Los resultados confirman diferencias significativas entre las provincias Alicante, Castellón y Valencia, tanto en los momentos de inicio de consumo como en la demora hasta pedir ayuda en un Centro de Tratamiento. También, aparecen diferencias significativas en cuanto a género.

El análisis de varianza univariado indica diferencias significativas ( $F= 205,8$ , sig: .000) en inicio de consumo de la droga principal entre las tres provincias, siendo en Castellón el inicio más temprano seguido de Valencia y Alicante, aunque su tamaño del efecto es bajo ( $\eta^2=.008$ ). En cuanto a la influencia del género en el inicio de consumo de la droga principal, existen diferencias significativas entre ambos ( $t = -11,264$ , sig = .000), siendo ellos quienes se inician antes. También encontramos diferencias significativas en la demora en la solicitud de tratamiento entre sexos ( $t = 6,629$ , sig = .000); ellas, inician antes el tratamiento, además también aparecen diferencias significativas ( $F = 219.9$ , sig = .000) entre las tres provincias como muestran las pruebas de comparaciones múltiples, pero el tamaño de efecto importante ( $\eta^2=.004$ ), no es importante.

## *Muestra*

La muestra general consta de 110253 sujetos admitidos a tratamiento en la Comunidad Valenciana entre enero de 1998 y diciembre de 2007 en 37 Unidades de Conductas Adictivas (U.C.A.'s) y otros centros notificadores. Alicante cuenta con 41176 (37.6%) casos, Castellón 11455 (10.4%) y en Valencia 57622 (52.3%). Del total 87.346 (79.2%) son hombres y 22.907 (20.8%) son mujeres. La admisión se realizó de acuerdo con las siguientes demandas: el 41.78% en desintoxicación ambulatoria, el 15.95% en psicoterapia, el 12.28% en metadona, el 5.62% en centro de día, el 1.70% en desintoxicación hospitalaria, el 0.90% en Comunidad terapéutica (UDR), el 0.87% en medicación de urgencia, el 14,40% no especifica la demanda y el 6.49% especifican otras.

En líneas generales, la droga principal que motiva la admisión a tratamiento es mayoritariamente la ingesta de alcohol (33.5%), heroína (23.3%), cocaína (22.8%), tabaco (7.9%), ludopatía (4.40%) y Cannabis (3.16%). Si bien, los porcentajes se modifican considerados por provincias.

## *Análisis*

Para llevar a efecto los objetivos planteados se consideró pertinente el uso de análisis descriptivos y gráficos, los análisis de varianza (univariados), para comprobar las diferencias entre las tres provincias, y la prueba T para medidas independientes con objeto de conocer las diferencias debidas al género.

### Resultados

#### Alicante

En la figura 1 podemos comprobar la distribución de la edad de inicio de las diferentes drogas principales en la ciudad de Alicante. La figura 2 muestra el histograma comparativo del inicio en el consumo y el tiempo de demora que transcurre hasta que los integrantes de la muestra piden ayuda para iniciar su tratamiento.

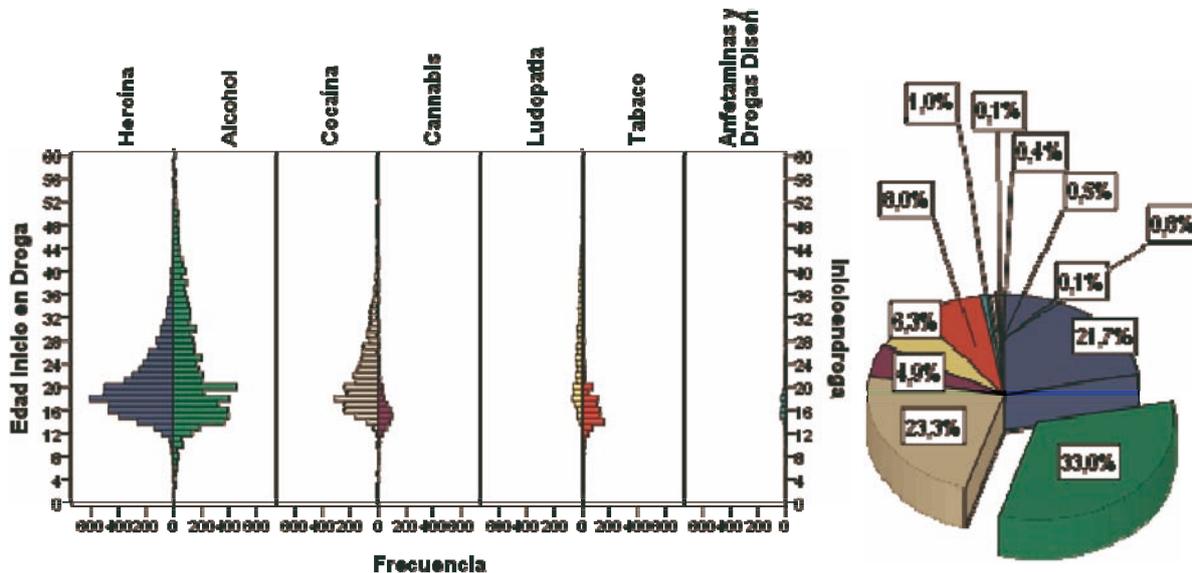


Figura 1. Distribución de la edad de inicio de consumo (Alicante)

El cuadro 1 recoge información sobre los contrastes por género de inicio de consumo y demora en tratamiento en Alicante.

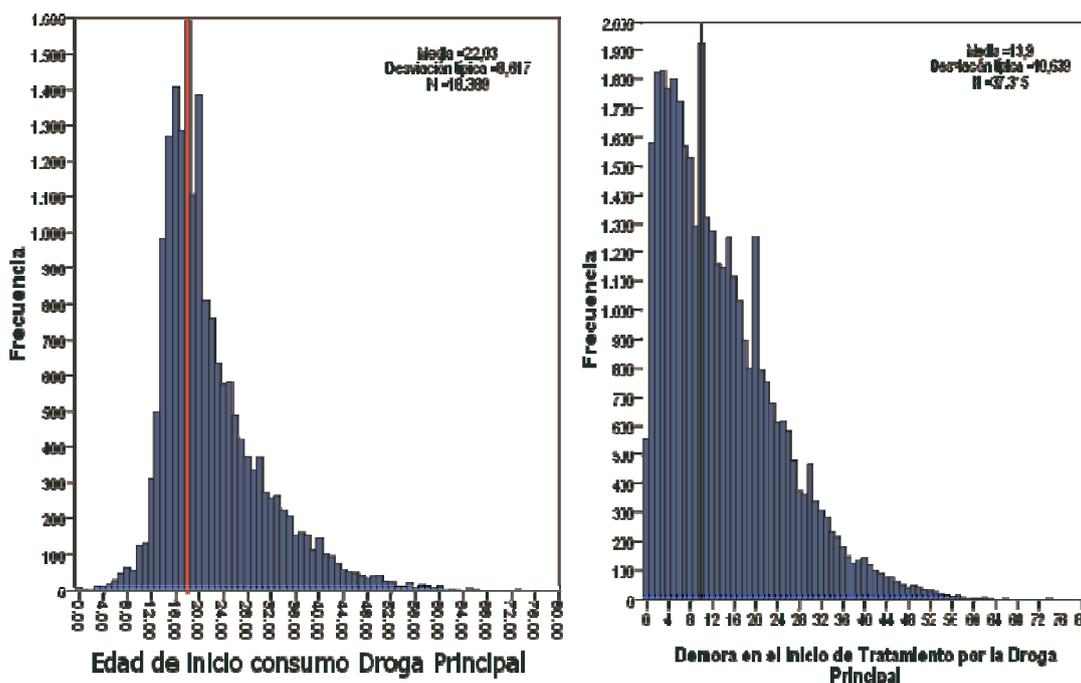


Figura 2. Edad de inicio consumo y demora en inicio de tratamiento (Alicante)

Cuadro 1. Medias y contrastes, por género, en inicio de consumo y demora en tratamiento (Alicante)

ALICANTE				
	SEXO	N	$\bar{X}$	Sx
Edad en que se inicia el consumo de la droga	Hombre	14847	21,82	8,381
	Mujer	3532	22,97	9,441
diferencia año de inicio de consumo y año de tratamiento	Hombre	29768	13,93	10,707
	Mujer	7547	13,77	10,363

ALICANTE Prueba de muestras independientes <sup>a</sup>										
		Prueba de Levene		Prueba T para la igualdad de medias					95% I.C.	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	E. típ. de la dif.	Inferior	Superior
Edad inicio consumo de la droga	Varianzas iguales	120,847	,000	-7,176	18377	,000	-1,155	,161	-1,470	-,839
	No varianzas iguales			-6,670	4937,71	,000	-1,155	,173	-1,494	-,815
diferencia año inicio consumo y año tratamiento	Varianzas iguales	3,140	,076	1,144	37313	,252	,157	,137	-,112	,426
	No varianzas iguales			1,167	11961,7	,243	,157	,134	-,107	,420

Castellón

En la figura 3 podemos comprobar la distribución de la edad de inicio de las diferentes drogas principales en la ciudad de Castellón. La figura 4 muestra el histograma comparativo del inicio en el consumo y el tiempo de demora que transcurre hasta que los integrantes de la muestra piden ayuda para iniciar su tratamiento

El cuadro 2 recoge información sobre los contrastes por género de inicio de consumo y demora en tratamiento en Castellón.

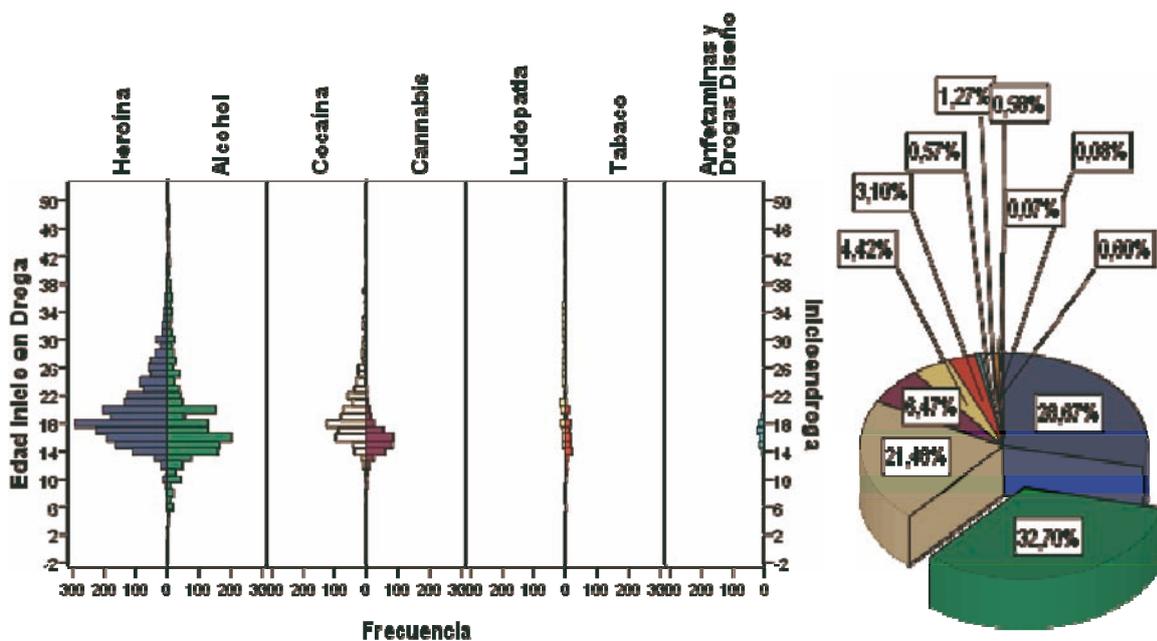


Figura 3. Distribución de la edad de inicio de consumo (Castellón)

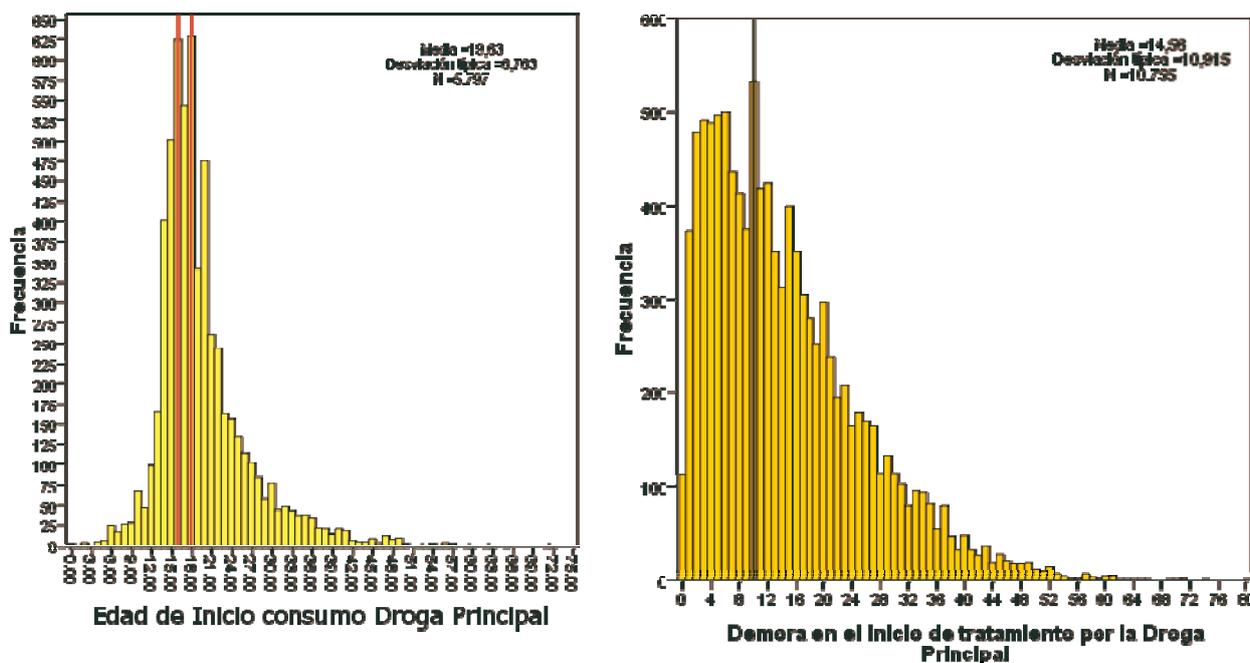


Figura 4. Edad de inicio consumo y demora en inicio de tratamiento (Castellón)

Cuadro 2. Medias y contrastes, por género, en inicio de consumo y demora en tratamiento (Castellón)

CASTELLÓN				
	SEXO	N	$\bar{X}$	Sx
Edad en que se inicia el consumo de la droga	Hombre	4872	19,53	6,670
	Mujer	924	20,23	7,186
diferencia año de inicio de consumo y año de tratamiento	Hombre	8946	14,81	11,020
	Mujer	1809	13,30	10,291

CASTELLÓN Prueba de muestras independientes <sup>a</sup>										
		Prueba de Levene		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	E. típ. de la dif.	95% I.C.	
									Inferior	Superior
Edad inicio consumo de la droga	Varianzas iguales	11,386	,001	2,883	5794	,004	-,699	,242	-1,174	-,224
	No varianzas iguales			2,741	1242,9	,006	-,699	,255	-1,199	-,199
diferencia año inicio consumo y año tratamiento	Varianzas iguales	9,654	,002	5,404	10753	,000	1,519	,281	,968	2,070
	No varianzas iguales			5,655	2714,2	,000	1,519	,269	,992	2,045

Valencia

En la figura 5 podemos comprobar la distribución de la edad de inicio de las diferentes drogas principales en la ciudad de Castellón. La figura 6 muestra el histograma comparativo del inicio en el consumo y el tiempo de demora que transcurre hasta que los integrantes de la muestra piden ayuda para iniciar su tratamiento

El cuadro 3 recoge información sobre los contrastes por género de inicio de consumo y demora en tratamiento en Valencia.

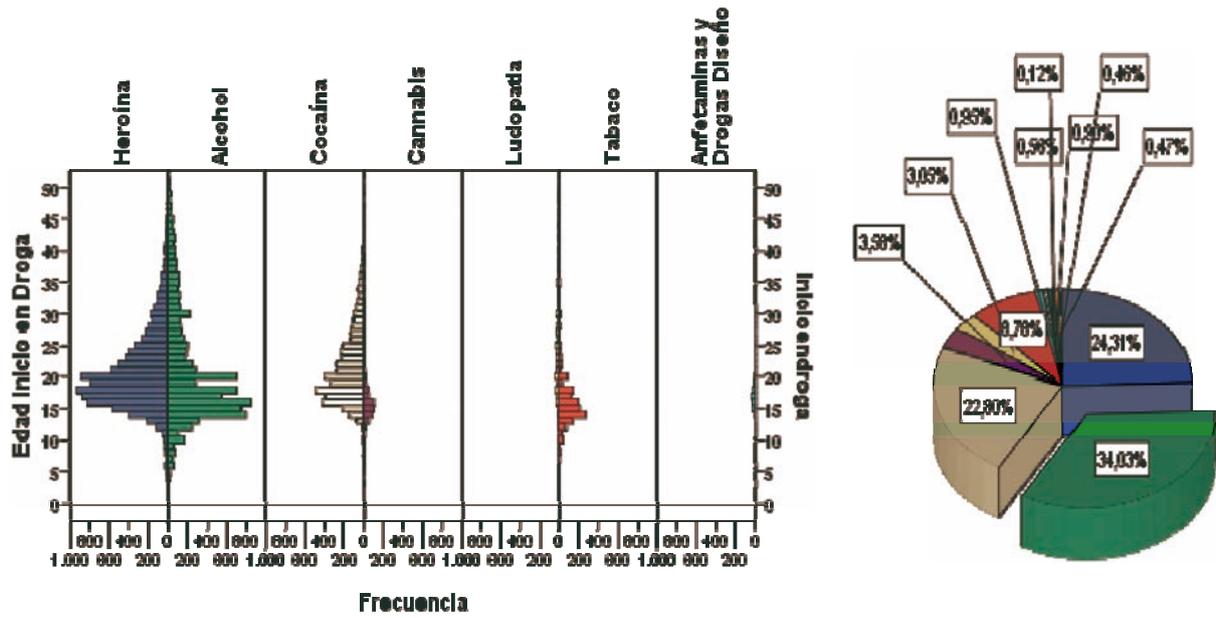


Figura 5. Distribución de la edad de inicio de consumo (Valencia)

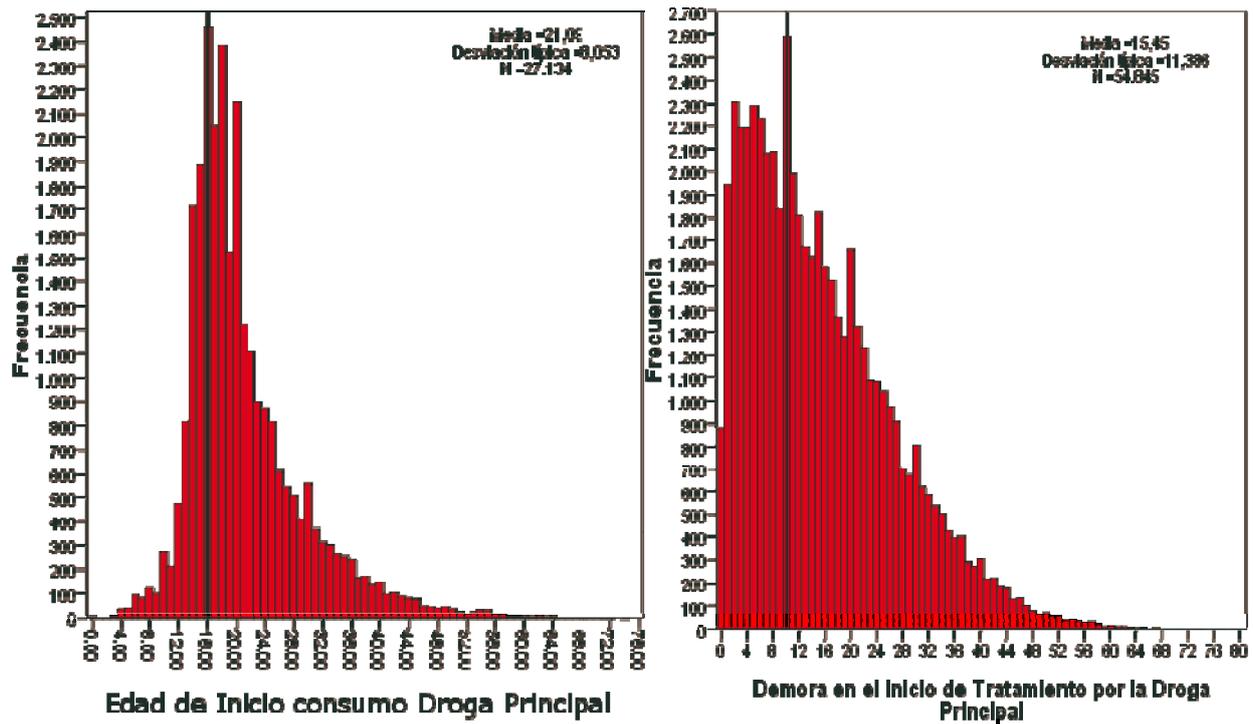


Figura 6. Edad de inicio consumo y demora en inicio de tratamiento (Valencia)

Cuadro 3. Medias y contrastes, por género, en inicio de consumo y demora en tratamiento (Valencia)

VALENCIA				
	SEXO	N	$\bar{X}$	Sx
Edad en que se inicia el consumo de la droga	Hombre	21560	20,87	7,825
	Mujer	5573	21,95	8,823
diferencia año de inicio de consumo y año de tratamiento	Hombre	42769	15,61	11,521
	Mujer	12076	14,86	10,874

VALENCIA Prueba de muestras independientes <sup>a</sup>										
		Prueba de Levene		Prueba T para la igualdad de medias					95% I.C.	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	E. típ. de la dif.	Inferior	Superior
Edad inicio consumo de la droga	Varianzas iguales	153,600	,000	-8,925	27131	,000	-1,078	,121	-1,315	-,842
	No varianzas iguales			-8,317	7982,70	,000	-1,078	,130	-1,332	-,824
diferencia año inicio consumo y año tratamiento	Varianzas iguales	11,216	,001	6,408	54843	,000	,752	,117	,522	,981
	No varianzas iguales			6,618	20364,6	,000	,752	,114	,529	,974

### Conclusión

Se puede observar que se produce un inicio temprano en el consumo de drogas, generalmente antes de los 14 años, tanto en las legales como en drogas ilegales. Es por ello que el “target” o la población diana de las campañas de prevención deben fijarse en un rango de edad inferior a los 14 años.

Además, se observa que la mayor demora en solicitud de tratamiento, hasta 4 años en el mejor de los casos, se produce en el Tabaco, dado que quizá sigue habiendo una gran aceptación social del consumo de dicha sustancia, llegando incluso a no considerarla como una droga.

Con respecto a las diferencias entre provincias en el inicio de consumo, encontramos que en Castellón la edad de inicio es menor que en Valencia y por último en Alicante, observando que los hombres inician antes su consumo que las mujeres. Así mismo, la demora en inicio de tratamiento es menor en Alicante, seguida de Castellón y por último Valencia, siendo las mujeres las que se demoran menos en la demanda de inicio de tratamiento.

### Referencias

- Bickel, W. K. & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction (Abingdon, England)*, 96(1), 73-86.
- Bickel, W. K., Johnson, M. W., Loewenstein, G., Read, D. & Baumeister, R. (2003). *Delay discounting: A fundamental behavioral process of drug dependence Time and decision: Economic and psychological perspectives on intertemporal choice*. (pp. 419-440). New York, NY US: Russell Sage Foundation.
- PNSD (2006). *Indicador Admisión a Tratamiento*. from <http://www.pnsd.msc.es /Categoria2/observa/pdf/AdmisionesTratamiento.pdf>.
- Rapp, R. C., Xu, J., Carr, C. A., Lane, D. T., Wang, J. & Carlson, R. (2006). Treatment barriers identified by substance abusers assessed at a centralized intake unit. *Journal of Substance Abuse Treatment*, 30, 227 - 235.

# ANSIEDAD Y OBJETIVIDAD EN LAS INVESTIGACIONES DEL COMPORTAMIENTO

Samuel Hurtado  
Universidad Central de Venezuela

A la ansiedad que siempre se activa en el científico –especialmente en el científico del comportamiento- es necesario que se le dé una solución. No orientada hacia una terapia de la personalidad, sino hacia una episteme de la ciencia. Esta orientación no es suficiente plantearla como cuestión descriptiva en cuanto invención histórica de conceptos y teorías con sus desarrollos, sino como un problema lógico, tocante a la cuestión operatoria de la objetividad, sin la cual la realidad científica luce cuesta arriba. Junto a la naturaleza impasible, se encuentra la observación pro-activa con la que el científico “enuncia” los datos. Antes que los conceptos, la observación comporta una epistemología aplicada, que, desencadenada la ansiedad, pretende solucionar como problema incorporándola a la construcción de los datos, merced a dos operaciones fundamentales: 1) elaboración de la referencia específica de la investigación (variable intermedia o constructo); 2) incorporación de la ansiedad, para aprovecharla con miras a la objetividad, en el juego recíproco de las inter-subjetividades, cuya distorsión incoada definirá la operación de los datos. Originados éstos en la psique del observador con su cultura y circunstancias, su “realidad” se cierra en la contra-observación del observado, punto estructural que sienta la objetividad del dato.

“Cuanto más las tendencias neuróticas del observador tienden a distorsionar la realidad, más pánico sentirá ante los hechos que recoja y más tendencia tendrá a recurrir a la lógica formal o a las teorías rígidas, con el fin de quitar a estos hechos su fuerza traumatizante, escondiéndolos bajo un caparazón de conceptos fáciles” (Bastide, 18).

Si el comportamiento vivo existe por sí mismo como natural o normal, el inducido no puede por menos que necesitar como acompañante al comportamiento vivo. De ninguna manera puede suprimir a éste porque haría imposible su propia existencia inducida. Lo importante aquí es que el comportamiento inducido permite la entrada a la observación y la coloca en posición de ser utilizada como dato científico. Para nuestro tema planteado, la observación no es técnicamente pura, está revestida siempre de elementos tanto culturales como de personalidad. Es decir, es técnicamente impura, porque está interferida por significaciones localizadas y por deseos y sentimientos individualizados. Uno de estos sentimientos es la ansiedad. La observación está cargada de ansiedad, y los medios de observación son una expresión anxiógena. A la elaboración de los datos se incorpora esta perturbación subjetiva, que marca siempre a su inflexión de relatividad, y a su explicación, aunque completa, de particularidad.

## *Olvido, artulugios y percepción*

La ansiedad no es sentida como un estado que se padece; es vivida en lo íntimo, incorporándose a nuestro ser moral. Como puede ser estimulante o paralizadora, modificar nuestras defensas activas o señalar un desarreglo profundo, esta forma atenuada de angustia suele pasar, como la etnocultura, inadvertida en la relación de observador y observado. Como un duende, expulsado de la conciencia, vuelve a ingresar a ella actuando ahora impunemente, al producir distorsiones graves en la observación.

Las dificultades en la construcción de la epistemología metodológica en la investigación del comportamiento, provienen del olvido de la reciprocidad interactiva que acontece en la observación

del investigador y la contra-observación del individuo investigado que forma parte de la información. Es preciso plantear la lógica del “lugar” en que uno y otro pretenden construir las demarcaciones que normalmente tienen carácter de distinta índole e intencionalidad. Se puede decir que un mismo género de observación surge como propósito del común encuentro cognitivo. Pero el género observacional no es suficiente si no se definen las especies de relaciones de observación a determinar por investigador e investigado desde su propio poder autónomo asimétrico.

Decenas de artilugios acuden al campo de la observación. Muchos se miden como resistencias con el fin de hacer pasar inadvertidos aspectos o rasgos de uno con respecto al otro. Otras veces puede acontecer una agonía de muerte por la competencia en decidir quién tiene la última palabra en estas situaciones diádicas. El observado dispone sólo de la reciedumbre de la experiencia directa, y con enunciar “y esto percibo” quiere cerrar la situación; un cierre a la fuerza al dejar de escuchar las pertinencias de los enunciados del observador.

El observador sabe que sus enunciados sobrepasan la experiencia directa y contienen la flexibilidad de una segunda argumentación más compleja que pretende alcanzar el cierre con una pertinencia de mayor calibre teórico. Se trata de la lógica reflexiva mediante la cual los enunciados proyectan las conciencias del observador y del observado. Es un comportamiento por la conciencia lo que le convierte en un enunciado de segundo orden. Implica un enunciado completo y significa una información como dato de observación producido. “La mejor prueba de que la conciencia perceptiva es una realidad irreductible en la investigación de la conducta es precisamente el hecho de que incluso el electrón podría decirse que contraobserva al físico, si no se trata la conciencia como un componente *sine qua non* de la observación y la contraobservación en la ciencia de la conducta” (Devereux, 1989, 335).

La conciencia como enunciado de conducta proyecta una información que impulsa la génesis psicoterapéutica. La transferencia sola es insuficiente; representa un cortocircuito en la comunicación proyectiva de los enunciados en la situación diádica. He aquí el problema de la lucha a muerte en la competencia por la decisión de la última palabra en el transcurso de los intercambios de decir los enunciados con más pertinencia. Únicamente se vence si se deponen las armas antagonistas a favor de la co-participación en la difusión de la conciencia perceptiva del observador y el observado.

### *Internalización del objeto y lucha por la pertinencia enunciativa*

Lo auténtico de la co-participación no consiste en que el observador se reserva su inconsciente, logrando entrar impunemente en el inconsciente del observado. La conciencia perceptiva para acceder a la conciencia del otro tiene que arrancar de su autopercepción y no puede descabalgarse de ella. Sólo cuando el observador coloca reflexivamente su inconsciente de cara al observado, comienza genuinamente a “moverse” su inconsciente. No se produce la comunicación diádica mediante un contacto directo. El observador no puede hablar en términos del observado. La referencia de su habla son sus propios términos o experiencia personal. Aún cuando cree que actúa dentro del observado, lo que hace es actuar sobre sí mismo, y esta especie de actuación le permite objetivar la información interior que le ofrece el observado.

La lucha por la decisión de los enunciados con más pertinencia señala que el desafío de la demarcación de los “lugares” del observador y del observado es móvil y que su desplazamiento, recuerda Devereux (1989, 335), no es continuo sino discontinuo, pero que al mismo tiempo se regenera constantemente buscando un lugar apropiado para las necesidades operacionales. La curva de Jordan puede señalar que parte del espacio que antes quedaba fuera del observador, ahora quede dentro, y viceversa. Indicará el avance o retroceso en la cualidad de la observación y contra-observación.

Cuando parecía que la construcción del dato estaba acabada, algún imprevisto o añadido obliga a comenzar de nuevo la observación.

Los desplazamientos de la curva de Jordan empiezan, como una centrífuga, a regresarse para terminar en el observador. La orientación esperada de la observación se encuentra siempre en el observador. Este la crea y puede recrearla sin fin. No es que el sujeto observado sea mera ocasión del juego del observador. Éste no podría dar un paso si no tuviera la referencia de las actividades y reacciones del observado. La ciencia etnológica también llega a esta conclusión, pues el descubrimiento del *Otro* como depositario de valor social para el *Yo* identifica comprensivamente el objeto de la etnología moderna, siendo casi el único aporte de la cultura occidental al acervo societario del “homo sapiens” (Levi-Strauss, 2000).

El observado, representante del otro, tiene la facultad fenomenológicamente de cerrar genuinamente la configuración de la observación en su trasunto de contra-observación. (Aunque, como hemos dicho, el observador tiene la facultad de la pertinencia explicativa de los enunciados últimos). De este modo la conciencia del investigador es creadora de su propia existencia, y como existente resulta ser la principal fuente de los estímulos desencadenantes de su propio comportamiento, normalmente sometido a perturbaciones ante la presencia del observado.

La observación de los fenómenos comienza en las reacciones del observador, que se coloca “adentro” de la curva de Jordan demarcada por la atención que demanda el fenómeno. El observador, al interiorizar dicha atención, hace que el fenómeno ingrese dentro de su subjetividad. Así la luna se encuentra definitivamente “dentro” del astrónomo, y el “lugar”, en que la radica, limita con sus intereses anxiógenos. Los aparatos de la observación pertenecen también al “adentro” del investigador, aunque considerados de otro modo pertenecen al objeto observado. El pedal es prolongación del pie, y de otro modo pertenece al objeto bicicleta. Es un juicio interesado de acuerdo al momento en que se quiera pensar el lugar del deslinde “hacia adentro”: lo percibo como prolongación del sujeto; o “hacia fuera”: lo percibo como parte del objeto. Se diseña la curva de acuerdo a las necesidades operacionales.

No se trata de los cambios en los límites del yo, como si éste existiera con límites. El yo es un límite en sí mismo. Tampoco se trata de los límites del aparato. Éste no responde de un modo restrictivo a la naturaleza del fenómeno, como si el aparato produjera el fenómeno. Cualquier amoldadura del yo que trate de excluir la percepción de lo real, en la que el observador no tenga conciencia de sus construcciones teóricas, destruye el fenómeno que quiere observar. La interferencia (=observación) deviene en negativa. Una psicologización genuina con respecto al cambio de los límites del observador puede corresponder perfectamente con la necesidad lógica del procedimiento de investigación epistemológica del comportamiento.

El punto oportuno de la observación puede determinarse convencionalmente según un criterio técnico. Pero éste no constituye el marco de observación, sino el criterio teórico construido por el observador. Un psicólogo considerará que el punto debe encontrarse antes de la teoría metodológica. Un psicoanalista pensará que está en el comienzo del terreno de las hipótesis. Hay una negativa a tomar en cuenta el terreno firme de la teoría y ha determinar metodológicamente la posición del objeto: si está en movimiento o está estático, porque es imposible observar al mismo tiempo y con similar precisión las dos posiciones. Siempre existe la posible imprecisión que genera la incertidumbre en el sistema, y la posibilidad de algún error. Pero ello no obedece a un obstáculo técnico, sino a un atributo esencial de la realidad siempre en movimiento inacabado.

### *Alteración ansiógena y cierre de configuración objetiva*

No es de extrañar que el deslinde esté ubicado en el lugar de la perturbación, zona simbiótica del principio de complementariedad tratado con certidumbre por la práctica pluridisciplinaria entre el psicoanálisis y la antropología cultural, dando origen a la sub-disciplina de etnopsiquiatría o etnopsicoanálisis. Dicho punto/instante coincide innumerables veces con un acontecimiento interior profundo del investigador. La consecuencia es que tal punto/instante termina estando en la perturbación desencadenada del investigador. Aquí se señala que la perturbación/ansiedad es un punto muy significativo para orientar un diseño metodológico en la investigación del comportamiento.

En el análisis del comportamiento se necesita intensificar la perturbación como dato básico. “Para él (psicoanalista ideal) no es un subproducto, indeseable aunque inevitable, sino más bien la meta de sus actividades de recolección de datos psicoanalíticos... como debiera ser también la de la labor de investigación del científico de la conducta. El psicoanalista ideal canaliza deliberadamente los estímulos que dimanan *directamente* hacia su propio inconsciente y, en grado menor, también a su preconsciente. Además, usa de aparato de percepción -o receptor- aquella porción de su psique que la mayoría de los demás científicos de la conducta tratan de esconder y emparejar y empezará por elaborar esos estímulos por medio de su actividad mental de proceso primario. Deja llegar -y llegar adentro- a su paciente. Permite la creación de una perturbación dentro de sí y después la estudia más atentamente aún que los dichos del paciente. Entiende a su paciente psicoanalíticamente sólo hasta donde entiende los trastornos que su paciente causa dentro de él. Dice ‘y esto percibo’ sólo en relación con las reverberaciones que halla ‘en sí mismo’”(Devereux, 1989, 360-361).

El psicoanalista observa al paciente, o el etnógrafo al grupo etnocultural, con el fin de que las emociones se interioricen en su psique como investigador. Su subjetividad, perturbada, es lo único genuino que va a examinar. No tiene otro aparato para el análisis. Los resultados del análisis es lo que puede mostrar cuando dice: esto es lo que percibo de ti o lo que he recolectado en el grupo étnico. Sólo una esencial subjetivación permite objetivar lo que queda de “allá afuera”. El fenómeno seguirá su curso objetivo merced al esfuerzo subjetivo del observador. El científico debe contar con sus inquietudes afectivas que emergen de su mar de fondo subjetivo. La función de la subjetividad es clave en la elaboración de la objetividad: “Muchos científicos de la conducta tratan ‘lo subjetivo’ como una causa de error sistemático, mientras que el psicoanalista lo trata como su principal fuente de información, simplemente porque su análisis didáctico lo capacitó para tolerar tales informaciones subjetivas”(Devereux, 1989, 361).

Aunque el psicoanálisis es presentado como el modelo de tratamiento epistemológico de la perturbación de anormalidad (Devereux, 1989, 360), la antropología cultural también puede generar similar tratamiento del desorden etnocultural.

Nosotros venimos trabajando los desórdenes étnicos relativos al negativismo social con dicho tratamiento (1999). Este esfuerzo nos ha colocado en situaciones de ansiedad con respecto a la crítica cultural nacional. Hemos tratado de remontarlas incorporando nuestra ansiedad positivamente. No es fácil enfrentar al propio colectivo nacional y sobrevivir en la comunidad científica reacia a entender nuestros resultados. A la comunidad científico-social le es cómodo trabajar con un interpretacionismo subjetivista conducente a repetir textos ajenos, origen de sus grandes ansiedades científicas.

Los alumnos han sido y son el campo normal del drenaje de nuestras ansiedades. Frente a las ansiedades negativas de la comunidad científica nacional, ellos nos ayudan a transformar las nuestras en positivas o creadoras, abiertas a las nuevas percepciones e informaciones. Así informando a una

periodista sobre el entierro del jefe de malandros<sup>[1]</sup> de un barrio marginal de Caracas, la “inspección” del rito nos generó una ansiedad, al pensar en el negativismo social inscrito, en dicho caso, en un desorden etnotípico (Devereux, 1973). Mi crítica cultural iba a desentonar de la “crítica con alabanza” producida por el colectivo étnico nacional venezolano.

### Referencias

- Bastide, R. (1973). Prefacio. En G. Devereux, *Ensayos de Etnopsiquiatría General* (pp. 9-19). Barcelona: Seix Barral.
- Devereux, G. (1973). *Ensayos de Etnopsiquiatría General*, Barcelona: Seix Barral.
- Devereux, G. (1989). *De la Ansiedad al Método en las Ciencias del Comportamiento*. México: Siglo XXI.
- Hurtado, S. (1999). *La Sociedad tomada por la Familia*. Caracas: La Biblioteca, Universidad Central de Venezuela.
- Levi-Strauss, C. (2000). *Raza y Cultura*. Madrid: Cátedra.

---

[1] Un malandro: es el designado en Venezuela como el malandrín, hasta puede llegar a ser un asesino. Incursiona en los barrios de Caracas y también en la ciudad urbanizada en busca de bienes indebidos. Frente a esta figura, el colectivo venezolano opera con una envidia de admiración, según un doble código etnopsiquiátrico.

# CONDUCTA ALIMENTARIA COMO FACTOR DE RIESGO EN LOS TRASTORNOS DEL COMPORTAMIENTO ALIMENTARIO

Juan Manuel Llopis y M. Victoria Acosta  
Universidad de Almería

Este trabajo tiene como objetivo principal conocer las conductas alimentarias de los adolescentes de la Comunidad Andaluza, centrándose en la conducta dietética y en determinar si se presentan diferencias en función del sexo. Sabiendo que en nuestro entorno no abundan los estudios de este tipo, pensamos que sería conveniente conocer la posible aparición de estos factores de riesgo entre nuestros adolescentes. Las hipótesis de partida eran que se darían diferencias por edades y sexo con respecto a la práctica de la conducta alimentaria restrictiva. Para realizar el estudio se parte de una población de 2000 sujetos, se seleccionó una muestra aleatoria de 530 sujetos de los cuales el 56% eran chicas (n=297) y el 43% eran chicos (n=233). El rango de edades oscilaba entre 12 y 18 años, siendo la media de edad 14'75 años y la desviación típica 1'66. Los resultados obtenidos apuntan a la idea de que a menor edad de los sujetos, mayor porcentaje de conducta alimentaria restrictiva presentan. Este efecto se produce tanto para los hombres como para las mujeres. Estudiando exclusivamente la prevalencia por sexos se encuentra, tal como sucede en muestras estudiadas en otros lugares, que entre el sexo femenino es mucho más elevada.

**LA CONDUCTA DIETÉTICA.** Hay que considerar que la preocupación por alcanzar un bajo peso corporal es un hecho notorio en nuestra sociedad, en contraposición a otras culturas como la árabe o la china donde el ideal de belleza se centra en un cuerpo relleno como principal síntoma de salud, prosperidad y capacidad de reproducción.

Como producto de la cultura actual, la preferencia por la delgadez y el rechazo a la obesidad se han vuelto muy populares, eso ha generado que un número cada vez mayor de personas adopten y practiquen toda clase de dietas y planes alimentarios con la finalidad de intentar conseguir la estilización de su figura y poder imitar a los modelos que de forma continuada se exponen a través de los medios de difusión social. Esta manera de actuar, producto de la cultura moderna, está acarreado cada vez más un mayor número de alteraciones y trastornos alimentarios.

Otra observación, claramente visible la encontramos en que muchas personas intentan bajar de peso utilizando los procedimientos que constantemente difunde el mercado de la dieta, la autoadministración va a ser la actuación primordial de dichos sujetos.

El término “dieta” hay que entenderlo como un programa alimentario que incluye cualquier tipo de alimento y en cualquier cantidad. Con base a esto se puede decir que existen diferentes tipos de dietas:

- La hipocalórica, es la más utilizada entre la población en general, es aquella baja en calorías y utilizada fundamentalmente para bajar de peso. Este tipo de dietas debe de conservar todas las características de una dieta correcta, es decir, ser completas, suficientes en todo excepto en energía y equilibradas. Hay que insistir en que una dieta cuanto más hipoenergética es más difícil será lograr que sea suficiente en todo tipo de nutrientes. Lo ideal estará determinado por la reducción proporcional de las tres fuentes de energía: glúcidos, lípidos y aminoácidos o también la reducción de alguno de estos componentes. Con bastante frecuencia las dietas suelen causar deficiencias de vitaminas y nutrimentos inorgánicos debido al desequilibrio que presentan en sus componentes.

- La dieta blanda, utilizada para el control de determinadas enfermedades.
- La hipercalórica, esta es rica en grasa y carbohidratos y se suele utilizar para que ganen peso aquellas personas que, por circunstancias diversas, se encuentran en estado de emaciación o bajo peso.

De las dietas mencionadas con anterioridad, las más utilizadas entre los adolescentes y jóvenes, son aquellas bajas en calorías ya que son la mejor alternativa para adelgazar. No obstante la palabra dieta puede significar desde el consumo de alimentos bajos en calorías hasta la omisión de alguna comida como puede ser el desayuno, el almuerzo o la cena. Hay que hacer notar que la palabra “hacer dieta” es una expresión subjetiva, pues lo que para una persona es “dieta” para otra puede ser una ingesta normal.

Para algunos autores la palabra dieta es un concepto negativo, significa un programa de no comer, no comer muchas calorías, no comer comidas que engorden, no comer las comidas favoritas etc. (Gordon, 1993, pág. 4). Esto se puede interpretar como una forma de autocastigo que sería el extremo opuesto del comer en exceso y visto como una conducta auto-indulgente. Este planteamiento de dos extremos es lo que desemboca en el síndrome de la “dieta-atasque yo-yo”, es decir, de la privación a la autoindulgencia o el comer en exceso. La solución a este péndulo se plantea como el comer de manera equilibrada, o lo que es lo mismo sin privación ni comilonas.

Según Hill, Weaver y Blundell (1990) la popularidad de las dietas se debe a diferentes razones: al poder de los estereotipos sociales sobre delgadez y obesidad. Con frecuencia, la delgadez es asociada a belleza, salud y autocontrol, en contraste con la obesidad que es percibida como un estado indeseable y que está provocado por la autoindulgencia, glotonería y pereza.

El seguimiento de dietas hipocalóricas y la práctica de conductas alimentarias anómalas, las cuales son consideradas, cada vez más, como comportamientos normales. La práctica de estos comportamientos representa una variante de la automedicación. En muy pocas ocasiones las personas acuden al médico o/y nutriólogo a solicitar una dieta, con bastante frecuencia es tomada de una revista de “cotilleo” o es recomendada por alguna amiga o algún familiar.

Existe consenso, en la literatura científica dedicada a esta temática, en que uno de los factores de riesgo que influye notoriamente en las conductas alimentarias anómalas, es la práctica y seguimiento de dietas reductoras de peso. Igualmente, los investigadores han encontrado, que dichas dietas llevan a la deshinibición en el comer o a la sobrealimentación y que esa actitud, precede al desarrollo de desórdenes del comer, sin importar la edad del dietante. (Herman y Polivy, 1980, Striegel-Moore, Silberstein y Rodin, 1986).

De acuerdo con Patton, Johnson-Sabine, Wood, Mann y Wakeling (1990) el riesgo relativo de los dietantes de convertirse en casos clínicos a los 15 años, es ocho veces mayor que el de los no dietantes. Resulta cada vez más evidente que los niños están adquiriendo la neurosis de los adultos, la obsesión por un peso y forma corporal tendente hacia una delgadez difícil de alcanzar.

En varios estudios se confirma que la edad de algunos dietantes se sitúa entre los 10 y 12 años, (Hill, Weaver y Blundell, 1990; Moore, 1988; Nassar, Hodges y Ollendick, 1992). Incluso estos autores afirman que los niveles totales de restricción en las niñas de 12 años son comparables con las de las personas adultas. Entre estos, están los realizados en México (Gómez Peresmitré y Ávila, 1998) que confirman dichos resultados y que se describen a continuación.

Gómez Peresmitré y Ávila (1998) encontraron en una muestra de 487 escolares mexicanos, de ambos sexos y con una media de edad de 10 años y ocho meses, que el 25% de ellos había realizado dieta tanto para bajar como para subir de peso.

El seguimiento de dieta en las preadolescentes está relacionado con la insatisfacción, con la figura corporal o con la preocupación por la misma. Los mismos investigadores han encontrado entre las adolescentes, prácticas alimentarias más extremas, tales como las prácticas de evitación: ayuno, saltarse comidas, uso de inhibidores del hambre, ejercicio excesivo, e igualmente prácticas de expulsión como vómito autoinducido, uso de laxantes y de diuréticos.

Una evidencia notoria es que, hacer dieta reductora de peso durante la niñez es una cuestión extremadamente problemática, es necesaria una ingesta calórica adecuada para el desarrollo saludable, por el contrario una reducción sustancial del peso puede resultar muy arriesgada para la mayoría de los niños (Bonjour, Thientz, Buchs, Slosman y Rizzoli, 1991, Kriepe y Forbes, 1990).

Queda claro que muchas personas intentan bajar de peso, pero numerosos estudios han demostrado que, en un plazo de 5 años, la mayoría de ellas vuelven a subir. De hecho, las dietas contribuyen más al problema que a la solución. El metabolismo baja durante la dieta para compensar la falta de comida, de ahí que el cuerpo queme las calorías más lentamente. El cuerpo aprende a hacer más con menos por lo que mantener el peso después de una dieta será cada vez más difícil.

Se afirma que cuando se tiene una dieta inadecuada se puede manifestar en determinados síntomas clínicos tales como hipotiroidismo y cuyas manifestaciones más comunes son: agotamiento físico, intolerancia al frío, resequedad, etc. y todo ello aunque los resultados de las pruebas resulten normales.

Es necesaria para ello una educación tanto familiar como escolar acerca de la salud y de los riesgos que conlleva un desequilibrio alimentario.

Se debe de enseñar que la imagen corporal deseable es aquella que representa a un cuerpo sano y no la de un cuerpo delgado. Una buena educación alimentaria, implica el desarrollo de unos hábitos saludables, que por supuesto, van a optimizar la calidad de vida.

## Método

### *Participantes*

De una muestra de adolescentes, de la Comunidad Autónoma Andaluza, se extrajeron 530 sujetos igualados en edad y en nivel académico, de los cuales 297 eran mujeres y 233 hombres, con una media de edad de 15,47 años. La edad mínima fue de 12 años y la máxima de 18 años.

*Tabla 1. Tamaño de la muestra total, medias y desviación típica de la variable edad por sexo*

Grupo	N	Media edad	D.T. edad
Hombres	233	15,29	1,73
Mujeres	297	15,61	1,78
Total	530	15,47	1,76

### *Instrumentos*

Para la obtención de los datos se aplicó la Escala de Factores de Riesgo Asociados con Trastornos de la Alimentación (EFRATA) de Gómez Peresmitré, validada en población española. El Cuestionario ha arrojado, en estudios previos, un coeficiente de clasificación correcta de 87.27% (Análisis discriminante) y coeficientes de consistencia interna de .80 a .85.

### *Procedimiento*

La aplicación del cuestionario fue llevada a cabo por 2 investigadores experimentados. El cuestionario se aplicó en Institutos de Enseñanza Secundaria de Andalucía, previa autorización oficial. La aplicación fue colectiva y voluntaria, se realizó en las aulas de los centros.

### *Análisis de datos*

Los datos fueron analizados con el programa SPSS 17.0. Una vez introducidos los datos, se calcularon las diferencias del Índice de Masa Corporal (IMC) por edad mediante un ANOVA. Luego se calcularon las diferencias del IMC por sexo mediante una prueba t de medias independientes.

## Resultados

Una vez realizados los cálculos, se obtuvieron los siguientes resultados:

*Tabla 2. ANOVA de IMC x Edad*

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupo	373,626	6	62,271	7,359	,000
Intra-grupos	4425,327	523	8,461		
Total	4798,953	529			

*Tabla 3. Medias de IMC x Sexo*

	SEXO	N	Media	Desviación típ.	Error típ. de la media
IMC	mujer	297	20,62	2,833	0,164
	hombre	233	21,51	3,162	0,207

## Conclusión

Los resultados confirman que, en cuanto al IMC, éste es menor cuanto menor es la edad de los sujetos ( $p < ,001$ ).

De la misma forma, se demuestra que el IMC de las mujeres es menor que el IMC de los hombres ( $t = -3,353$ ;  $p < ,001$ ).

Al analizar los ítems de la escala, referidos a preocupación por el peso y la comida, los resultados nos muestran que, cuando se comparan por sexo, son las mujeres las que presentan mayor preocupación por la comida que los hombres ( $p < ,000$ ), se confirma así, lo que la literatura internacional y nacional frecuentemente han informado, es decir: que son las mujeres las que en comparación con sus contrapartes los hombres de la misma edad y nivel de estudios las que desean, independientemente de su peso real, ser más delgadas o tener menor peso, igualmente muestran mayor preocupación por el peso y la comida; también son las que siguen en mayor proporción dietas restrictivas.

Está claro que en las mujeres se presentan mayores índices de riesgo alimentario que en los hombres. Toro (1999), indica que la influencia de las imágenes o reproducciones corporales que aparecen en los medios de comunicación social, afecta a la población en general, pero especialmente a las mujeres jóvenes.

En Canadá ya desde 1985, apareció que el 70% de todas las mujeres adultas canadienses, (con sobrepeso, peso normal, delgadas) deseaban reducir el mismo. El 36% de las mujeres con peso normal creían que tenían sobrepeso (Millar, 1991) y se informaba que niñas de 5 años restringían sus alimentos por temor a ser gordas (Feldman, Feldman y Goodman, 1986).

### Referencias

- Bonjour J.P., Thientz G., Buchs B., Slosman D. y Rizzoli, R. (1991). Critical years and stages of puberty for spinal and femoral bone mass accumulation during adolescence. *Journal of Clinical Endocrinology & Metabolism*, 73, 555-563.
- Feldman, W., Feldman, E. y Goodman, J. T. (1986). Culture versus biology: children's attitudes towards thinness and fatness. *Pediatrics*, 81, 190-194.
- Gómez Peresmitré, G. y Ávila, E. (1998). Conducta Alimentaria y Obesidad. *Revista Iberoamericana*, 6 (2), 10-22.
- Gómez Peresmitré, G. (1993). Detección de anomalías de la conducta alimentaria en estudiantes universitarios: obesidad, bulimia y anorexia nervosa. *Revista Mexicana de Psicología*, 10 (1), 17-27.
- Gordon, T. (1993). P.E.T. *Padres eficaz y técnicamente preparados*. México: Diana.
- Herman, C. P. y Polivy, J. (1980). *Restrained eating*. In A. J. Stunkard (Ed.), *Obesity*. Philadelphia: Saunders.
- Hill, A. J. Weaver, C. y Blundell, J.E. (1990). Dieting concerns of 10-year-old girls and their mothers. *British Journal of clinical Psychology*, 29, 346-348.
- Kriepe, R.E. and Forbes, G.B. (1990). Osteoporosis: A new morbidity for dieting female adolescence? *Pediatrics*. 86, 478-80.
- Millar, W. (1991). A trend to a healthier lifestyle. *Health Reports*, 3, 363-370.
- Moore, D.C. (1988). Body image and eating behaviour in adolescent girls. *American Journal of Child Disorders*, 142, 1114-1118.

- Nassar, C., Hodges, P. & Ollendiek, T. (1992). Self-concept, eating attitudes and dietary patterns in young adolescent girls, *School counselor*, 39(5), 338-343.
- Patton, G.C., Johnson-Sabine, E., Wood, K., Mann, A.H. y Wakeling, A. (1990). Abnormal eating attitudes in London schoolgirls- a prospective epidemiological study: Outcome at twelvemonth follow-up. *Psychological Medicine*, 20, 383-394.
- Striegel-Moore, R. H., Silberstein, L. R., y Rodin, J. (1986). Toward an understanding of risk factors for bulimia. *American Psychologist*, 41, 246-263.
- Sánchez-Carracedo, D., Raich, R. M., Figueras, M., Torrás, J. y Mora, M. (1999). Adaptación preliminar del cuestionario de alimentación de Stunkard y Messick ( Thee factor eating questionnaire, TFEQ) con una muestra española universitaria. *Psicología Conductual* 7 (3), 393-416.
- Toro, J. (1999). *El cuerpo como delito*. Ariel. Barcelona.

# ESTUDIO EMPÍRICO DEL USO EXCESIVO DE LAS TECNOLOGÍAS DE ENTRETENIMIENTO DE LOS ADOLESCENTES DE BARCELONA

Olatz López Fernández, M. Luisa Honrubia y Montserrat Freixa  
Universidad de Barcelona

Esta investigación se centra en el “uso excesivo” de las tecnologías de entretenimiento, en concreto de la videoconsola, el Internet y el teléfono móvil por parte de los adolescentes. Esta última década, el estudio del uso excesivo, problemático o, incluso adictivo, en este sector de la población respecto a esta forma de entretenimiento ha dado lugar a una serie de publicaciones científicas que indican cierta alerta de carácter psico-social y clínica a considerar en nuestro país, que es el cuarto a nivel europeo en consumo de este tipo de entretenimientos (así como el sexto a nivel mundial). El propósito del estudio empírico es determinar la existencia de uso excesivo y uso problemático de estas tecnologías en el caso de los adolescentes de la ciudad de Barcelona. En concreto, primero se revisa el estado de la cuestión para delimitar el concepto “uso excesivo” o “uso problemático” en la adolescencia y, segundo, se observa si puede estar presente en las tres tecnologías y en qué medida, mediante los datos obtenidos a través de un cuestionario elaborado *ad hoc*. Finalmente, se expondrán los resultados hallados, se compararán con los obtenidos en un estudio piloto previamente realizado y, finalmente, se contrastarán con estudios similares.

En la última década los entretenimientos tecnológicos han ido consolidándose como una de las principales formas de ocio de la juventud de los países industrializados. En nuestro contexto, los adolescentes barceloneses han crecido con estas tecnologías que empiezan a ser asequibles a nuestro mercado a partir del año 2000, por el que en la actualidad parece que forman parte de su comportamiento habitual fuera del ámbito escolar y lo que, ya es una realidad constatada, es que se han ido consolidando desde el pasado año (según la Asociación Española de Distribuidores y Editoras de Software y el Observatorio de la Producción Audiovisual) como uno de los principales juegos consumidos por este sector de la población, englobando tanto lo que hace referencia a consolas como videojuegos y siendo nuestro país en este momento, el cuarto de Europa más consumidor de esta tipología de entretenimiento (tendencia a que va en aumento: el 2006 hubo un incremento del 12% respecto al 2005, llegando a los 1000 millones de euros en las ventas en España).

A la vez que evoluciona el uso de los juegos mediante tecnologías, la alerta social de adicción comportamental hacia este tipo de tiempo libre ha aparecido tanto a nivel periodístico como científico. En el primer caso, en diferentes noticias recogidas a través de la principal prensa catalana (El Periódico y La Vanguardia) se ha observado que desde 2006 han comenzado a aparecer noticias, tanto a nivel local como internacional, sobre el uso excesivo de los videojuegos y las consecuencias negativas que empiezan a producirse en algunos adolescentes. En el segundo caso, a pesar de que la literatura del tema de la adicción a las tecnologías de entretenimiento adolescente es relativamente reciente (Griffiths & Hunt, 1995; Young, 1996) se demuestra que el problema de la dependencia existe y que se deriva de un uso patológico de estas formas de entretenimiento, a pesar de que al mismo tiempo se reconoce que aún faltan estudios empíricos que determinen la magnitud de este nuevo trastorno psicológico.

En referencia a los tipos de patrones de uso del juego, la literatura de nuestro contexto se ha basado en la terminología del estudio del juego con apuestas y la ludopatía para determinar tres principales tipologías de juego sin apuestas: el patrón de uso social (o habitual o controlado), que haría referencia al comportamiento de juego ocasional o regular con tecnologías, haciendo un uso

razonable del tiempo y utilizándolas como forma de entretenimiento puntual que no afectaría a ningún área de la vida cotidiana; el patrón de uso de riesgo (o problemático), que se centraría en la existencia de cierta dependencia hacia la tecnología y los videojuegos que comportaría un tiempo de dedicación considerable con una frecuencia prácticamente diaria, que afectaría en la dedicación a otras actividades cotidianas del adolescente (deberes, tareas del hogar, etc.) y/o relaciones con el entorno, demostrando dificultad por el auto-control; el patrón de uso adictivo (o patológico), que mostraría una dependencia emocional ningún este comportamiento y una pérdida de control que interferiría con su vida cotidiana, en otras palabras, esta última tipología sería la más grave y correspondería básicamente al no poder dejar de jugar con videojuegos y estar continuamente conectado a la tecnología, lo que implicaría aislarse para entretenerse el máximo tiempo posible con el videojuego y, aparte de el ámbito social (relación con amigos, familiares, etc.) y los hábitos básicos de la vida cotidiana (comida, dormir, etc.).

En cuanto al diagnóstico de dicho síndrome, es preciso aclarar que, en parte, debido a la novedad y, por lo tanto, de los pocos estudios empíricos, de carácter epidemiológico, que se han realizado hasta el momento presente no constituye un trastorno psiquiátrico propio del manual Diagnóstico y Estadístico de Desórdenes Mentales (DSM-IV) (APA, 1994).

## Método

### *Participantes*

La procedencia de la muestra ha sido de centros de secundaria pertenecientes a diferentes distritos de la ciudad de Barcelona. En el estudio piloto con el cuestionario *ad hoc* se recogieron datos de un total de 116 adolescentes, entre 11 y 18 años ( $M=15$ ;  $DT=2,103$ ), donde el 43,1% eran hombres y el 56,9% eran mujeres. La participación en el estudio dependió de los sujetos que accedieran a colaborar, o sea, que el muestreo fue no probabilístico de tipo accidental.

En este punto, hay que señalar también que previamente se los pidió su consentimiento de forma escrita y/u oral, donde se los garantiza la confidencialidad y el anonimato de las respuestas, que los datos obtenidos serían sólo utilizadas para la presente investigación universitaria y que fuesen honestos en todo cuanto comunicasen. Al finalizar, se les agradeció su participación. Para participar en el estudio no era un requisito estar familiarizado con las tecnologías de entretenimiento, pues se trataba de hacer una exploración sobre su uso en la adolescencia, por el que sólo se limitó el intervalo de edad.

### *Instrumentos*

Para este estudio piloto de carácter cuantitativo, se diseñó un cuestionario *ad hoc*, con los siguientes bloques de contenido: (1) variables socio-demográficas: género, edad, lugar de residencia, el número de miembros del núcleo familiar, variables indirectos del nivel socio-económico de los padres –si habían acabado los estudios, si trabajaban y en qué profesión, el curso escolar, si tenía ningún hábito potencialmente adictivo y si tenía ningún hobby; (2) videojuegos de consolas; (3) videojuegos de ordenador y otros entretenimientos de Internet: las mismas variables que a la consolas; (4) teléfono móvil.

### *Procedimiento*

Se solicitó permiso a los directores de los centros, padres y/o estudiantes. Se realizó la administración del cuestionario durante una hora de tutorías. Durante ésta el investigador principal estaba en el aula con

los estudiantes para facilitar cualquier ayuda (resolución de dudas). El cuestionario era auto-aplicado y se realizó individualmente en silencio, para garantizar la validez de la información recogida.

Posteriormente, se realizó un análisis estadístico a nivel descriptivo de la muestra en relación a los tres entretenimientos tecnológicos principales: videojuegos de consuela, juegos de ordenador y webs y, del teléfono móvil. Se realizó también pruebas de carácter inferencial para determinar la relación que las variables de las tecnologías podían tener con el género y la edad, variables socio-demográficas principales que eran equilibradas en la muestra, así se obtuvieron los primeros resultados preliminares con SPSS (v. 14).

## Resultados

En resumen, se procede a enumerar los resultados principales del estudio:

1. La mayoría de adolescentes son propietarios de las tecnologías de entretenimiento (por orden de mayor a menor): de ordenadores con Internet (90%), teléfonos móviles (85%) y videoconsolas (80%).
2. La edad de inicio promedio de las tecnologías de entretenimiento es (por orden de menor a mayor): por las videoconsolas hacia los 8 años, por los ordenadores hacia los 10 años y por el teléfono móvil hacia los 11 años. A pesar de que esta tendencia está cambiante en función de los cohortes de edad, pues los adolescentes más jóvenes (entre 11 y 14 años) comienzan a jugar y a hacer uso de estas tres tecnologías antes que los más grandes (entre 15 y 18 años).
3. El aspecto más positivo de estos entretenimientos es: por la videoconsola, la diversión y en algún caso el aprendizaje, en cambio por el ordenador también la diversión pero seguida de la comunicación y una minoría mencionan el aprendizaje o el buscar de información. En el caso del teléfono móvil destacaba la comunicación como elemento más beneficioso.
4. El aspecto más negativo de estos entretenimientos es: en el caso de la videoconsola y el ordenador, que enganchan y provocan malestar. En cambio, en el caso del teléfono móvil, lo que producía más preocupación era el elevado coste económico de su uso.
5. Las tecnologías las usan con frecuencia (por orden de mayor a menor) con uno promedio de: unas 4 veces por semana, en el caso de los ordenadores y, en cambio, unas 3 en las videoconsolas con sesiones entre 1 y 2 horas por ambas tecnologías.
6. El comportamiento con el juego depende de la tecnología y de las diferencias individuales: con las videoconsolas, dependiendo del momento de juego en general se juega tanto solo como acompañado, en cambio con el ordenador el juego es más bien solitario y desde casa.

## Conclusión

Este estudio es uno de los pocos realizados para determinar los patrones de uso de estas tecnologías por parte de los adolescentes. En este momento, se ha perfeccionado el instrumento y aplicado de forma masiva en la ciudad de Barcelona, donde nos va a permitir obtener datos de mayor relevancia respecto a este uso y posible uso excesivo, o incluso, problemático de la videoconsola, Internet y Teléfono Móvil.

## Referencias

- APA (American Psychiatric Association) (1994). *Diagnostic and statistical manual of the mental disorders*. 4th edition (DSM-IV). Washington, DC: APA.
- Griffiths, M.D. & Hunt, N. (1995). Computer game playing in adolescence. Prevalence and demographic indicators. *Journal of Community and Applied Social Psychology*, 5, 189–193.
- Young, K.S. (1998). Internet addiction: the emergence of a new clinical disorder. *Cyberpsychology and Behaviour*, 1, 237-244.

# A DEFINITION OF POTENTIAL ENTREPRENEUR FROM A PROBABILISTIC POINT OF VIEW

Jorge López Puga<sup>1</sup>, Juan García García<sup>1</sup>, Carlos J. Cano<sup>1</sup>, Ana B. Gea<sup>2</sup> y Leticia de la Fuente<sup>1</sup>

<sup>1</sup>Universidad de Almería

<sup>2</sup>Fundación Mediterránea-Universidad de Almería-Empresa

Entrepreneurship “is a role that individuals undertake to create organizations” and entrepreneurial activity has been related to organizational leadership. We propose defining entrepreneur in a probabilistic way. In our view, a person can be classified as an entrepreneur depending on the probability s/he shows a set of traits. To test this hypothesis we have concentrated on the definition of potential entrepreneurs: undergraduate students who think they might be entrepreneurs in the future but have not yet owned and managed a business. We asked a sample of undergraduate students (n = 1,111; age average = 23.37, SD = 4.28, range = 17-56; male = 35%, female = 64.6%) to fill a questionnaire on attitudes towards entrepreneurship. The test had ten dimensions. We built a Naïve Bayes Net Classifier and a convergent Bayesian network in order to assess the influence of the test dimensions on the attitude towards organizations creation. Our results show that the convergent model is able to predict more than the 87% of the entrepreneurial tendency. We conclude that our probabilistic model is highly efficient predicting entrepreneurship. As a result, our framework considering entrepreneurship as a matter of probabilistic nature has been reinforced.

The fact that entrepreneurship plays an important role in the productive system has been emphasized in several contexts (Corman, Lussier, and Nolan, 1996). Entrepreneurs have a double effect on economy. On the one hand, entrepreneur people have the power to regulate employment, introduce innovation or make economy more dynamic and those changes can be measured from a microeconomic point of view. On the other hand, from a macroeconomic point of view, they enrich the business web of countries.

The entrepreneur tries to make a profit from a creative view of the world while a manager earns a living from a non innovative activity. More specifically, the entrepreneur is a person or group who tries to exploit a business opportunity (McKenzie, Ugbah, and Smothers, 2007). Samuelson (1970) noted that entrepreneurs are characterized by a vision, originality, courage and tendency to introduce instead of inventing things. From Gartner’s (1989) point of view, an entrepreneur “is a role that individuals undertake to create organizations”. Secondly, entrepreneurial activity has been related to organizational leadership (Antonakis and Autio, 2006; Bjerke and Hultman, 2003). Thus, the entrepreneur is a kind of leader who assumes the creation of organizations. However, the definition of entrepreneur is elusive and Rogoff and Lee (1996) noted that entrepreneurship has confused researchers in social sciences the way subatomic particles have puzzled physicists.

We propose defining entrepreneur in a probabilistic way instead of describing it as an all-or-nothing phenomenon. In our view, a person can be classified as an entrepreneur depending on the probability s/he shows a set of traits. To test this hypothesis we have concentrated on the definition of potential entrepreneurs suggested by Huefner, Hunt and Robinson (1996): undergraduate students who think they might be entrepreneurs in the future but have not yet owned and managed a business. We used Bayesian networks as analytic tools to model entrepreneurship. More specifically, we built two types of models (Naïve Bayes Classifier and a convergent Bayesian network) to assess the predictive power of the dimensions of a scale on attitudes towards business creation. Our results show that the convergent Bayes net is far more predictive than the Naïve Classifier. Overall, convergent model was

able to predict above 87% of the tendency to entrepreneurship. As a result, our theoretical framework of considering entrepreneurship as a matter of probabilistic nature has been reinforced.

## Method

### *Participants*

We asked a sample of 1,111 (male: 35% and female: 64.6%) undergraduate students from University of Almería to fill a questionnaire on attitudes towards entrepreneurship. The participants' age ranged from 17 to 56 and the averaged age was 23.37 (*SD*: 4.28). All the degrees studied at the Universidad de Almería were tested and they were classified into three clusters (a) Human and Low Sciences, b) Technical Sciences and c) Business Sciences) to carry out a no-probabilistic stratified sampling procedure.

### *Materials*

We used a test (ACEMP) about attitudes towards business creation we have developed in previous researches (i. e., Cano, García, and Gea, 2003). The test had 29 multiple-choice items with a Likert response scale with four options. There were 13 items in a negative sense so these items were recoded before getting the final score on attitude to entrepreneurship. After inverting the negative items the scale sense is positive (ranging from 29 to 116), indicating a higher value a more positive attitude towards business creation. The test had ten dimensions: negotiation, perseverance, independence, creativity, risk taking, internal locus of control, competitiveness, risk tolerance, self-confidence and self-organization.

### *Procedure*

The scale towards business creation was into a booklet containing other scales aimed to collect information under a wide research project program. The booklet was provided to the students, previous consent of the professor, before or after sessions of compulsory subjects in their classrooms. The test was self-administrated in groups and participants neither receive any reward nor payment for filling in the questionnaire but a few words of thanks were given to them.

### *Data Analysis*

We built a Naïve Bayes Classifier (it is also called Simple Bayes Classifier and divergent Bayesian network) and a convergent Bayesian network in order to assess the influence of the test dimensions on the attitude towards organizations creation. We used Netica 4.08 (*Norsys Software Corp.*) to build the models and the parameters were estimated using the maximum likelihood procedure corrected with Laplace's rule. The variable of convergence or divergence, depending on the model, was the answer to the question *Do you wish to set up your own business?* The answer to this question only took two possible values, *Yes* or *Not*.

### Results

On average, the models predict the desirability of setting up a business with an 80.26% of accuracy. However, as can be seen on Figure 1, the convergent Bayesian network produces a better rate of correct classifications.

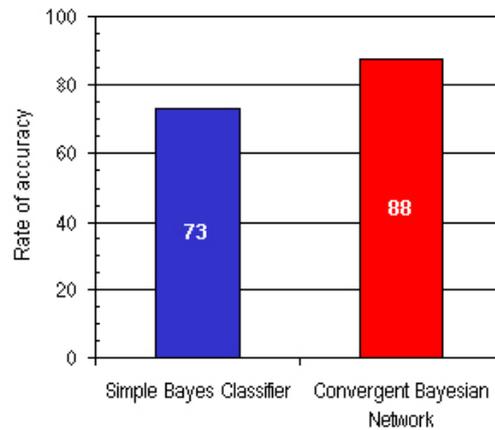


Figure 1. Comparison between models

These differences remain when we use statistics for the goodness of fit. In Table 1 you can see the logarithmic loss, quadratic loss and the spherical payoff for each model. As can be seen, the convergent model obtains better values in all of these parameters.

Table 1. Goodness of fit for the models

Model	Logarithmic Loss	Quadratic Loss	Spherical payoff
Simple Bayes Classifier	0.5696	0.3816	0.7864
Convergent Bayesian Network	0.4332	0.2598	0.8632

As regards to the fit of the test dimensions considered independently, we can see on Table 2 that the dimensions of negotiation and perseverance got the best values whereas the traits of confidence and organization reach the worst values.

Table 2. Goodness of fit for each node

Node	% of hits	Deviation rate compared to class node	
		Naïve	Convergent
Negotiation	87.71	14.72	0.19
Perseverance	80.94	7.95	-6.58
Independence	60.82	-12.17	-26.7
Creativity	60.02	-12.97	-27.5
Risk taking	57.08	-15.91	-30.44
Internal locus of control	56.46	-16.53	-31.06
Competitiveness	55.30	-17.69	-32.22
Risk tolerance	53.61	-19.38	-33.91
Confidence	52.09	-20.9	-35.43
Organization	51.74	-21.25	-35.78
Average		-11.413	-25.943

## Discussion

We have modelled propensity to entrepreneurial activity with Bayesian networks and found that our model is highly efficient. Comparatively, the convergent Bayesian network has reached a better level of accuracy and fit. However, in terms of the average of goodness of fit, taking into account the deviation accuracy in the features variables, the simple Bayesian classifier generates better values (Table 2). In terms of the dimensions studied, we have noted that negotiation and perseverance are the best for predicting entrepreneurship but confidence and organization are the worst. In general, our framework considering entrepreneurship as a matter of probabilistic nature has been reinforced.

The results from this research are significant in two ways, methodologically and theoretically. On the one hand, we have reinforced the idea of using Bayesian networks as analytic tools and we have shown it is useful in the context of researching the key features of entrepreneurship (i. e., García, López, Cano, Gea y De la Fuente, 2006; López, 2009). Secondly, our results can be considered as new evidence on the validity of our scale to measure attitudes towards business creation (i. e., Cano et. al., 2003; García, Cano y Gea, 2005). On the other hand, from a theoretical point of view, our results could be useful to characterise the profile of potential entrepreneur (Huefner et. al., 1996). And that could be useful to guide local, regional and national policies regarding entrepreneur activity promotion or to design training and optimization programs in order to improve entrepreneur's abilities.

## References

- Antonakis, J. and Autio, E. (2006). Entrepreneurship and leadership. In J. B. Baum, M. Frese, R. Baron (Eds), *The Psychology of Entrepreneurship* (pp. 189-207). Mahwah, NJ: Laurence Erlbaum.
- Bjerke, B. and Hultman C.M. (2003). A dynamic perspective on entrepreneurship, leadership and management as a proper mix for growth. *International Journal of Innovation and Learning*, 1, 72-93.
- Cano, C. J., García, J. and Gea, A. B. (2003). *Actitudes emprendedoras y creación de empresas en los estudiantes universitarios*. Almería: Servicio de Publicaciones de la Universidad de Almería / Consejo Social de la Universidad de Almería.
- Corman, J., Lussier, R. and Nolan, K. G. (1996). Factors that encourage entrepreneurial start-ups and existing firm expansion: a longitudinal study comparing recession and expansion periods. *Academy of Entrepreneurship Journal*, 1, 43-55.
- García, J., Cano, C. J. y Gea, A. B. (2005). Actitudes emprendedoras en estudiantes universitarios y empresarios. Evidencias de validez de un instrumento. *Iberpsicología*, 10 (8), art. 12.
- García, J., López, J., Cano, C. J., Gea, A. B. y De la Fuente, E. I. (2006, Septiembre). *Aplicación de las redes bayesianas al modelado de las actitudes emprendedoras*. Comunicación presentada en el IV Congreso de Metodología de Encuestas. Pamplona.
- Gartner, W. B. (1988). "Who is an entrepreneur?" Is the wrong question. *American Journal of Small Business*, 12 (4), 11-32.
- Huefner, J. C., Hunt, H. K., and Robinson, P. B. (1996). A comparison of four scales predicting entrepreneurshp. *Academy of Entrepreneurship Journal*, 1, 56-80.

López, J. (2009). *Modelos predictivos en actitudes emprendedoras: análisis comparativo de las condiciones de ejecución de las redes bayesianas y la regresión logística*. Tesis doctoral no publicada, Facultad de Psicología, Universidad de Almería.

McKenzie, B., Ugbah, S. and Smothers, N. (2007). “Who is an entrepreneur” is still the wrong question? *Academy of Entrepreneurship Journal*, 13, 23-43.

Rogoff, E. G., and Lee, M. S. (1996). Does firm origin matter? An empirical examination of types of small business owners and entrepreneurs. *Academy of Entrepreneurship Journal*, 1, 1-17.

Samuelson, P. A. (1970). *Economics* (8<sup>a</sup> ed.). New York: McGraw-Hill.

# REVISIÓN DE INSTRUMENTOS DE MEDIDA DE CANTIDAD-CALIDAD DE ACTIVIDAD FÍSICA-SALUD EN EDUCACIÓN FÍSICA

Kamal Mohamed, Bertok Szabolcs, Luis Lozano, Armando Cocca, Francisco Salinas, Juan Carlos Martínez Álvarez y Jesús Viciano  
Universidad de Granada

El objetivo del presente estudio fue comprobar la eficacia de tres instrumentos para medir la implicación energética y el autoconcepto; la primera medida en número de pasos totales (NP), METs promedio (MP) y frecuencia cardíaca media (FCM), y la segunda en AC con 11 categorías; en clase de Educación Física entre chicos y chicas de Educación Primaria (EP) y Secundaria-Bachillerato (ES-B). Participaron 214 sujetos, 115 de EP (56 chicos y 59 chicas de  $10.93 \pm 2.07$  años) y 99 de ES-B (50 chicos y 49 chicas de  $15.46 \pm 3.54$  años), pertenecientes los de EP al C.E.PR. Nuestra Señora de las Nieves y los de ES-B al I.E.S. Albayzín de Granada. Los instrumentos de recogida de datos empleados fueron unos brazaletes (Armband de Sensewear, Bodymedia) con sensores, pulsómetros Polar RS y un cuestionario sobre autoconcepto. Para analizar los datos se aplicó una estadística inferencial comparativa por género entre EP y ES-B, obteniendo los datos en 10 sesiones de Educación Física. Los resultados mostraron por género entre etapas que existen diferencias altamente significativas ( $p < 0.001$ ) en la variable NP, muy significativas ( $p < 0.01$ ) en MP y significativas ( $p < 0.05$ ) en FCM y AC. Finalmente, llegamos a la conclusión de que los resultados obtenidos en las distintas variables medidas con estos instrumentos son bastante fiables, sin olvidar que existen ciertos problemas para determinar con la máxima exactitud dichos resultados.

Son diversos los estudios que al respecto de la actividad física (AF) y/o deportiva de los escolares o adolescentes se han realizado en los últimos años en España, en los que se pone de manifiesto una pobre práctica de AF por parte de los jóvenes, así como un nivel de condición física mejorable (Carratalá y García, 1999; Castillo y Balaguer, 2001; García Ferrando, 1997; Mendoza, 2000; Rodríguez, 2000; Ruiz, García y Hernández, 2001; Tercedor, 1998; Zabala, Lozano y Viciano, 2002). El descenso en esta práctica de AF se da fundamentalmente en la adolescencia, tal y como lo expresan los estudios de Mendoza (2000) o el de García Ferrando (1997), cuya tendencia no deja de aumentar. Son diversas las circunstancias que han ayudado a este descenso; entre ellas, la mayor sedentarización debido a causas tecnológicas y sociales (Castillo et al., 2001; García Ferrando, 1997; Mendoza, 2000). El sedentarismo es tanto más acusado cuanto mayor es la edad de los alumnos, especialmente en el caso de las chicas. La proporción de las que afirman que nunca hacen deporte fuera del horario escolar pasa del 29% a los once años al 48 % a los dieciocho años (Mendoza, 2000). Esta reducción del ejercicio físico ha provocado un descenso en el estado de la condición física de dicha población, con la consiguiente influencia sobre la salud y la calidad de vida, así como el desarrollo de enfermedades cardiovasculares, obesidad y osteoporosis (Guerra, Duarte y Mota, 2001; Piéron, 1993; Ribero, Guerra, Pinto, Oliveira, Duarte y Mota, 2003).

Actualmente se reconoce la importancia de la AF y el deporte como herramienta de prevención y rehabilitación contra las secuelas del sedentarismo propio de la sociedad tecnificada (Bauman, 2004), teniendo presente que los numerosos beneficios para la salud del ejercicio regular dependen del tipo, intensidad y volumen de actividad desarrollada por el individuo. Ante este preocupante panorama, la Educación Física (EF) se convierte en un área fundamental para promover la salud de los escolares (Sánchez, Lozano, Zabala, Gargallo, Zabala, M. y Viciano, 2004; Sánchez, Ramírez, Zabala, M. y Viciano, 2005; Zabala, 2004). Desde la EF se puede dotar a los alumnos de los conocimientos teórico-prácticos necesarios para mantener y mejorar su salud de manera efectiva y autónoma. Desde este área

podemos procurar a los alumnos un mayor conocimiento y concienciación de lo que les puede suponer una adecuada práctica de AF en relación a su salud, de tal forma que conozcan y apliquen de manera autónoma esos conocimientos para así realizar individualmente una práctica de AF de mayor calidad y enfocada de manera intencional a cumplir los propósitos que ellos mismos se marquen.

Los jóvenes deben comenzar cuanto antes a aprender que el ejercicio físico diario realizado de manera autónoma es un hábito muy importante para consolidar un estilo de vida saludable (Pangrazi, 1987). Y la EF, puede y debe cumplir esa función social, educativa y formativa (Álamo, Amador y Pintor, 2002; Zabala, 2004). Parece claro que existe una necesidad de educar a la población sobre los beneficios de ciertos cambios comportamentales en el estilo de vida seleccionado -dieta y actividad física- (Zabala, 2004; Zabala et al., 2002). Si los alumnos entienden los beneficios que les aporta la práctica de AF aeróbica sobre su salud, gozarán de una mayor autonomía e intencionalidad en su práctica físico-deportiva (Zabala, 2004). Las mejoras en la resistencia aeróbica que acompañan a las actividades en las clases de EF son el resultado de muchas adaptaciones al estímulo de esas actividades. En relación a las adaptaciones cardiovasculares destacamos: incremento de las dimensiones internas del ventrículo izquierdo y el grosor de la pared ventricular, aumento del volumen sistólico en reposo, mayor contractilidad ventricular izquierda, reducción de la frecuencia cardiaca en reposo y durante el ejercicio, etc. Al igual que el sistema cardiovascular, el sistema respiratorio experimenta adaptaciones como: la frecuencia respiratoria permanece estable en reposo, aumenta la diferencia arteriovenosa de oxígeno, etc. (Wilmore y Costill, 1998). El ejercicio aeróbico reduce el estrés mental y la depresión, mejora la función endocrina, y reduce las lipoproteínas de baja densidad (LDL), disminuyendo de ese modo el colesterol total (Virus y Smirnova, 1995). Uno de los trabajos más relevantes acerca de la AF aconsejada para que sea vinculada a beneficios saludables, se basan en la incidencia que ésta presenta respecto al desarrollo de la resistencia cardiorespiratoria (Tercedor, 2001). Se debe tener en consideración la intensidad, duración, frecuencia y tipo de ejercicio.

Por último, uno de los aspectos a tener muy en cuenta en las clases de EF es el tema de la motivación. Los alumnos manifiestan una tendencia natural hacia el movimiento que repercute en la motivación que experimentan al participar en actividades físico-deportivas. Sin embargo, conforme avanza la etapa de la infancia y comienza la adolescencia, se observa que existe un alto porcentaje de sujetos que abandonan la práctica físico-deportiva. Los estudios realizados en el campo de la Psicología aplicada al deporte abordan el análisis de las motivaciones que inducen a los sujetos a la participación deportiva, decantándose por aquellas motivaciones intrínsecas centradas en el aprendizaje de las tareas por ser las que aseguran una cierta continuidad deportiva y, como consecuencia, un estilo de vida más saludable.

En el estudio que hemos desarrollado se han utilizado algunos instrumentos de medida que se han aplicado en la EF con la finalidad de medir la cantidad y calidad de AF que realizan los escolares. El objetivo de este estudio fue comprobar la eficacia de tres instrumentos para medir la implicación energética y el autoconcepto; la primera medida en número de pasos totales (NP), METs promedio (MP) y frecuencia cardíaca media (FCM), y la segunda en AC con 11 categorías; en clase de EF entre chicos y chicas de Educación Primaria (EP) y Secundaria-Bachillerato (ES-B).

## Método

### *Sujetos*

En el presente estudio ha participado una muestra de 214 sujetos, de los cuales 115 era alumnado de EP (56 chicos y 59 chicas de  $10.93 \pm 2.07$  años) y 99 de ES-B (50 chicos y 49 chicas de  $15.46 \pm 3.54$  años). Los de EP pertenecientes al C.E.PR. Nuestra Señora de las Nieves y los de ES-B al I.E.S. Albayzín de Granada.

### *Diseño*

El diseño empleado en la investigación fue un diseño multigrupo mixto de cuatro variables dependientes (NP, MP, FCM y AC), dos grupos cuasi-experimentales (EP y ES-B) y dos variables independientes (chicos y chicas). Las tareas, actividades y juegos seleccionadas y ejecutadas en las sesiones eran representativas de un clase típica de EF. Como posible variable contaminadora se controló el tiempo utilizado en cada una de las sesiones con los diferentes grupos.

### *Aparatos y material*

Los instrumentos de recogida de datos empleados para medir las variables dependientes fueron unos brazaletes (Armband de Sensewear, Bodymedia) que poseen unos sensores, pulsómetros Polar RS y un cuestionario sobre AC (Marsh, Ascí y Tomas-Marco, 2002). Para analizar los datos se aplicó una estadística inferencial comparativa por género entre EP y ES-B. La recogida de datos se ha obtenido en 10 sesiones de EF con diferentes grupos de alumnado de EP y ES-B.

## Resultados

Los valores medios y desviación estándar de las cuatro variables dependientes por etapa y género fueron las siguientes: en EP el NP ( $2147,82 \pm 451,42$  en chicos y  $1905,23 \pm 344,08$  en chicas), MP ( $4,22 \pm 0,97$  en chicos y  $4,23 \pm 0,97$  en chicas), FCM ( $156,14 \pm 15,74$  en chicos y  $161,50 \pm 19,48$  en chicas), AC ( $4,74 \pm 1,19$  en chicos y  $4,31 \pm 1,10$  en chicas); en ES-B el NP ( $2617,59 \pm 600,30$  en chicos y  $2485,71 \pm 422,65$  en chicas), MP ( $5,51 \pm 1,30$  en chicos y  $5,80 \pm 1,18$  en chicas), FCM ( $156,08 \pm 13,47$  en chicos y  $159,79 \pm 15,84$  en chicas), AC ( $4,32 \pm 1,05$  en chicos y  $3,77 \pm 1,12$  en chicas).

*Tabla 1. Resultados de las variables dependientes (NP, MP, FCM y AC) por etapa (EP y ES-B) y género (♂ y ♀)*

Etapa	Género	NP	MP	FCM	AC
EP	♂	2147,82±451,42	4,22±0,97	156,14±15,74	4,74±1,19
	♀	1905,23±344,08	4,23±0,97	161,50±19,48	4,31±1,10
ESB	♂	2617,59±600,30	5,51±1,30	156,08±13,47	4,32±1,05
	♀	2485,71±422,65	5,80±1,18	159,79±15,84	3,77±1,12

Los resultados mostraron por género entre etapas que existen diferencias altamente significativas ( $p < 0.001$ ) en la variable NP, muy significativas ( $p < 0.01$ ) en MP y significativas ( $p < 0.05$ ) en FCM y AC.

## Discusión

El objetivo principal de nuestro estudio fue comprobar la eficacia de tres instrumentos para medir la implicación energética y el autoconcepto en clase de EF entre chicos y chicas de EP y ES-B. En relación a estos instrumentos son diversos los estudios que al respecto de la AF de los escolares y adolescentes se han llevado a cabo en los últimos años tanto a nivel internacional (Balaguer, 1998; Byrne y Gavin, 1996; Duncan, Al-Nakeeb, Woodfield y Lyons, 2007; Tudor-Locke, Pangrazi, Corbin et al., 2004; Webber, Cresanta, Voors y Berenson, 1983; Wei, Kampert, Barlow et al., 1999) como por grupos de investigación de nuestro país (Mohamed, Cocca, Bertok, Martínez, Viciano, 2009; Moleiro y Cid, 2008; Núñez Pérez, González-Pienda, Rodríguez, González-Pumariaga, Montero, Álvarez Pérez y Torres 2008; Pastor, Baleguer y García-Merita, 2008). Muchas veces la eficacia de estos instrumentos no han permitido los resultados exactos ni homogéneos que quisiéramos (Welk, Eisenmann y Dollman, 2006). A pesar de ello, en las diferentes investigaciones llevadas a cabo en nuestro ámbito se ha llegado a la conclusión de que los resultados obtenidos en las distintas variables medidas con estos instrumentos son bastante fiables.

## Conclusión

Actualmente se reconoce la importancia que tienen los instrumentos de medida en EF, pero es evidente que existen ciertos problemas a la hora de aplicarlos en las numerosas investigaciones llevadas a cabo en nuestra área. Entre esos problemas, y atendiendo a los instrumentos que hemos utilizado en nuestro estudio, cabe destacar las posibles interferencias que se producen entre los pulsómetros, el retraso de los brazaletes para detectar la temperatura corporal y efectuar el correspondiente registro de datos, la formulación de excesivas y repetidas preguntas en el cuestionario. En definitiva, son problemas que hasta ahora se encuentra la ciencia para determinar con exactitud los resultados obtenidos en las investigaciones (Welk, Eisenmann y Dollman, 2006).

## Referencias

- Álamo, J.M., Amador, F. & Pintor, P. (2002). Función social del deporte escolar. El entrenador del deporte escolar. *Lecturas: revista digital*, 45.
- Balaguer, I. (1998). *Self-concept, physical activity and health among adolescents*. Conferencia presentada al *24th International Congress of Applied Psychology*, San Francisco.
- Bauman, A.E. (2004). Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. *J Sci Med Sport*, 7(1), 6-19.
- Byrne, B.M. & Gavin, D.A.W. (1996). The Shavelson model revisited: Testing for the structure of academic self-concept across pre-, early, and late adolescents. *Journal of Educational Psychology*, 88(2), 215-228.
- Carratalá, V. & García, A. (1999). Diferencias en las motivaciones y actividades de ocio y tiempo libre entre adolescentes deportistas y no deportistas. *Apunts*, 12, 13-22.
- Castillo, I. & Balaguer, I. (2001). Dimensiones de los motivos de práctica deportiva de los adolescentes valencianos escolarizados. *Apunts, Educación Física y Deportes*, 63, 22-29.

- Duncan, M.J., Al-Nakeeb, Y., Woodfield, L. & Lyons, M. (2007). Pedometer determined physical activity levels in primary school children. *Preventive Medicine*, 44, 416-420.
- García Ferrando, M. (1997). *Los españoles y el deporte, 1980-1995. Un estudio sociológico sobre comportamientos, actitudes y valores*. CSD/Tirant lo Blanch. Valencia.
- Guerra, S.; Duarte J. & Mota, J. (2001). Physical activity and cardiovascular disease risk factors in schoolchildren. *European Physical Education Review*, 7(3), 269-281.
- Marsh, H.W., Ascí, F.H. & Tomas-Marco, I. (2002). Multitrait-multimethod analyses of two physical self-concept instruments: A cross-cultural perspective. *Journal of Sports and Exercise Psychology*, 24, 99-119.
- Mendoza, R. (2000). Diferencias de género en los estilos de vida de los adolescentes españoles: implicaciones para la promoción de la salud y para el fomento de la actividad físico-deportiva. En *Educación física y salud. Actas del 2º congreso internacional de educación física*. FETE-UGT. Jerez de la Frontera. Cádiz.
- Mohamed, K., Cocca, A., Bertok, Sz., Martínez, F.S. & Ramírez, J.V. (2009). Gender Differences in Relation to the Energy Involvement in Physical Education Classes with Students of Primary and Secondary Education. En: M. González-Gross, D. Cañada, J. Valtueña, U. Albers y P.J. Benito (eds.) *Physical Activity and Health Education in European Schools* (pp. 197-199). Madrid: Universidad Politécnica.
- Moleiro, M.Á. & Cid, F.V. (2008). El papel de la autoeficacia en el entrenamiento para controlar la frecuencia cardíaca durante pruebas de esfuerzo. *Psicothema*, 16(1), 50-57.
- Núñez Pérez, J.C., González-Pienda, J.A. Rodríguez, M.G., González-Pumariega, S., Montero, C.R., Álvarez Pérez, L. & Torres, M<sup>a</sup>.C.G. (2008). Estrategias de aprendizaje, autocenpeto y rendimiento académico. *Psicothema*, 10(1), 97-109.
- Pangrazi, R.P. (1987). Health related fitness for young children. En: Stuart Biddle (ed.) *Foundations of health related fitness in Physical Education*. The Ling Publishing house, (pp. 103-109).
- Pastor, Y., Baleguer, I. & García-Merita, M. (2008). Relaciones entre el autoconcepto y el estilo de vida saludable en la adolescencia media: un modelo exploratorio. *Psicothema*, 18(1), 18-24.
- Piéron, M. (1993). Analyser l'enseignement pour miewx enseigner. *Revue E.P.S.* París.
- Ribero, J., Guerra, S., Pinto, A., Oliveira, J., Duarte J. & Mota, J. (2003). Overweighth and obesity in children and adolescents: relationship with blood pressure, and physical activity. *Annals Human Biology*, 30(2), 203-213.
- Rodríguez Allen, A. (2000). *Adolescencia y deporte*. Nobel. Oviedo.
- Ruiz, F., García, M.E. y Hernández, A.I. (2001). Comportamientos de actividades físico-deportivas de tiempo libre del alumnado almeriense de enseñanza secundaria post-obligatoria. *Motricidad*, 7, 113-143.

- Sánchez, C., Lozano, L., Zabala, J., Gargallo, E., Zabala, M. & Viciano, J. (2004). Pautas para el control nutricional en alumnos de ESO desde el área de Educación Física. *I Congreso Nacional de Educación Física "Investigación y Experimentación innovadora en el ámbito de la Actividad Física"*. Jerez.
- Sánchez, C., Ramírez, J., Zabala, M. & Viciano, J. (2005). Investigaciones en el ámbito de la salud desde el área de Educación Física. *II Jornadas Científicas de Ciencias de la Salud*. Granada.
- Tercedor, P. (1998). *Estudio sobre la relación entre actividad física habitual y condición física-salud en una población escolar de 10 años de edad*. Tesis Doctoral. Universidad de Granada.
- Tercedor, P. (2001). *Actividad física, condición física y salud*. Wanceulen, Sevilla.
- Tudor-Locke, C., Pangrazi, R.P., Corbin, C.B. et al. (2004). BMI-referenced standars for recommended pedometer-determined steps/day in children. *Preventive Medicine*, 38, 857-864.
- Viru, A. & Smirnova, T. (1995) Health promotion and exercise training. *Sports Medicine*, 19, 123-136.
- Webber, L.S., Cresanta, J.L., Voors, A.W. & Berenson, G.S. (1983). Tracking of cardiovascular disease risk factor variables in school-age children. *Journal of Chronic Diseases*, 36, 647-660.
- Wei, M., Kampert, J.B., Barlow, C.E. et al. (1999). Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. *Journal of the American Medical Association*, 282, 1547-1553.
- Welk, G.J., Eisenmann J.C. & Dollman, J. (2006). Health-related physical activity in children and adolescents: a bio-behavioral perspective. En: D. Kirk, M. O'Sullivan and D. Macdonald (eds.). *The Handbook of physical Education* (pp. 665-684). London: Sage.
- Wilmore, J.H. & Costill, D.L. (1998). *Fisiología del esfuerzo y del deporte*. Paidotribo. Barcelona.
- Zabala, M. (2004). *Influencia de un programa de intervención basado en el biofeedback de la frecuencia cardíaca sobre la percepción de la intensidad de esfuerzo en alumnos de Educación Secundaria Obligatoria*. Tesis Doctoral. Universidad de Granada.
- Zabala, M., Lozano, L. & Viciano, J. (2002). La práctica deportiva extralectiva y extraescolar y su relación con la EF lectiva desde la perspectiva de profesores y alumnos de ESO. En *Educación Física, ocio y recreación. Actas del Tercer Congreso Internacional de Educación Física*. FETE - UGT. Jerez de la Frontera (Cádiz), 807-820.

# DISEÑOS GENÉTICAMENTE INFORMATIVOS: EL REGISTRO DE GEMELOS DE MURCIA

Juan R. Ordoñana<sup>1</sup>, Irene Rebollo-Mesa<sup>2</sup>, Francisca González-Javier<sup>1</sup>, Eduvigis Carrillo<sup>1</sup>, Juan Francisco Sánchez-Romera<sup>1</sup>, Lucía Colodro<sup>1</sup> y José M. Martínez-Selva<sup>1</sup>

<sup>1</sup>Universidad de Murcia

<sup>2</sup>Section of General Psychiatry. Institute of Psychiatry. King's College. Londres, GB

Los diseños genéticamente informativos y, en particular, los estudios de gemelos, constituyen la metodología más utilizada para analizar la contribución relativa de factores genéticos y ambientales a la variabilidad inter-individual. Básicamente, consisten en comparar la mayor o menor similitud, con respecto a un carácter seleccionado, entre gemelos monocigóticos (MZ) y gemelos dicigóticos (DZ). Una muestra de estas características permite aplicar modelos biométricos para cuantificar la contribución relativa de influencias genéticas y ambientales sobre las diferencias individuales. El ajuste de estos modelos permite la separación de la varianza observada en sus componentes genéticos [Factores genéticos aditivos (A)] y ambientales [Compartidos por los miembros de una familia (C); e individuales (E)]. Los gemelos MZ comparten la totalidad de A y de C, mientras que los DZ comparten la totalidad de C pero solo la mitad de A. El Registro de Gemelos de Murcia (RGM) es un registro poblacional, que permite llevar a cabo estudios de estas características. Al finalizar el primer contacto, el RGM incluye a 862 mujeres gemelas adultas (MZ 46% y DZ 54%). Dispone de datos sobre cigosidad, antropometría, y comportamientos relacionados con la salud y en breve contará con un biobanco de ADN ligado al proyecto.

Los diseños genéticamente informativos y, en particular, los estudios de gemelos constituyen un valioso recurso para la investigación sobre las causas de variación fenotípica en humanos. Estos, constituyen la metodología más utilizada para separar las causas genéticas de las ambientales en el parecido familiar (Neale y Cardon, 1992; Plomin, De Fries, McClearn y McGuffin, 2008). Básicamente, consisten en comparar la mayor o menor similitud, con respecto a un carácter seleccionado, entre gemelos que provienen de un único cigoto escindido o monocigóticos (MZ) y gemelos que provienen de dos cigotos diferentes o dicigóticos (DZ). Dado que los primeros son genéticamente iguales mientras que los segundos comparten sólo, de media, la mitad de su genoma, su comparación proporciona una valiosa información para estudiar la influencia genética en diversos fenotipos. De esta forma se puede analizar la importancia relativa de factores hereditarios y ambientales en el desarrollo de caracteres complejos y, en particular, fenotipos conductuales.

Una muestra de estas características permite aplicar modelos biométricos para cuantificar la contribución relativa de influencias genéticas y ambientales sobre las diferencias individuales. El ajuste de estos modelos permite la separación de la varianza observada en sus componentes genéticos [Factores genéticos aditivos (A)] y ambientales [Factores ambientales compartidos por los miembros de una familia (C); y factores ambientales individuales (E)]. Los gemelos MZ comparten la totalidad de A y de C, mientras que los DZ comparten la totalidad de C pero solo la mitad de A. (Figura 1).

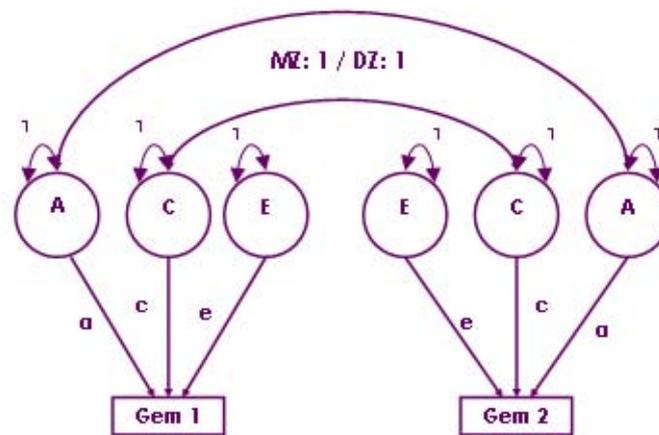


Figura 1: Diagrama básico para datos de gemelos. Las variables latentes se representan con círculos y las observadas con rectángulos. A = Varianza genética aditiva; C = Varianza ambiental compartida; E = Varianza ambiental no compartida

La investigación en gemelos concita la colaboración entre diversas disciplinas (psicología, biología, medicina, estadística,...) y métodos de investigación (Martin, Boomsma y Machin , 1997), dadas las evidentes ventajas que presentan los diseños con este tipo de muestras como eliminación de variables de confusión, facilidad de emparejamiento o control de factores ambientales compartidos. Así, los registros de gemelos han sido utilizados por la comunidad científica para llevar a cabo distintos tipos de estudios, permitiendo tanto diseños transversales como longitudinales. Entre estos podemos mencionar: 1.- Estudios biométricos clásicos de gemelos y de familias de gemelos, que intentan cuantificar el papel de factores genéticos y ambientales en la variación de caracteres y susceptibilidad a trastornos; 2.- Estudios de caso-control en gemelos, utilizando parejas de gemelos discordantes para factores ambientales específicos, o discordantes en cuanto a diagnóstico o medidas de morbilidad; 3.- Estudios de intervención, en los que los gemelos de un par son asignados aleatoriamente a un determinado tratamiento de forma que se genera un emparejamiento natural por edad, sexo y susceptibilidad genética; o 4.- Estudios longitudinales, en los que se analiza la evolución en el tiempo de las variables de interés.

Adicionalmente, en los últimos años, los avances teóricos e instrumentales que han tenido lugar han llevado a que los modernos análisis de gemelos vayan más allá de la clásica estimación de la heredabilidad de un rasgo, para servir como recurso en estudios de epidemiología genética, el análisis de las interacciones genotipo-ambiente, o el estudio de las relaciones genotipo-fenotipo a partir del análisis de genes candidatos o de barridos genómicos (GWAS) (Boomsma, Bushjan y Peltonen, 2002; Bushjan, 2002; Johnson, Turkheimer, Gottesman y Bouchard, 2009).

Ineludiblemente, este tipo de estudios depende de la existencia y accesibilidad de muestras de sujetos con estas características. Estos registros constituyen una infraestructura de investigación básica y están presentes en casi todos los países científicamente avanzados. Dichos registros, presentan un especial interés cuando tienen una base poblacional y obviamente, ganan en calidad, potencial investigador y posibilidad de interconexión y participación con otros grupos a nivel nacional e internacional, cuando aúnan diferentes tipos de datos, incluyendo muestras biológicas (i.e., ADN) unidas a datos fenotípicos y de estilos de vida de los participantes (Ohm-y Derom, 2006).

*El Registro de Gemelos de Murcia*

A pesar de la importancia de este modelo de investigación y la larga tradición de mantenimiento de estas cohortes en otros países, la implantación de registros de estas características en nuestro entorno apenas se ha iniciado. El Registro de Gemelos de Murcia (RGM) (Ordoñana et al., 2006) es una iniciativa para el desarrollo de un registro de gemelos de base poblacional ([www.um.es/registrogemelos](http://www.um.es/registrogemelos)). Su principal objetivo se centra en el análisis de los factores genéticos y ambientales que están involucrados en el desarrollo de conductas de salud, percepciones de salud, calidad de vida en relación con la salud y utilización de servicios sanitarios. Adicionalmente, el RGM pretende establecer las bases para un estudio longitudinal a largo plazo sobre estos aspectos. El RGM es gestionado desde el Área de Psicobiología de la Universidad de Murcia y en su desarrollo participa activamente la Consejería de Sanidad de la Comunidad Autónoma de la Región de Murcia.

Método

La población de referencia del RGM comprende, por el momento, todos los pares de gemelos, residentes en la Región de Murcia, nacidos entre 1940 y 1966, ambos inclusive. La información sobre la muestra se obtiene de las bases de datos digitales disponibles en el sistema sanitario de la C.A. Inicialmente, el RGM se ha centrado exclusivamente en mujeres adultas. Durante 2007 se procedió a la detección, selección y realización de una entrevista telefónica a todas las posibles gemelas identificadas, sobre distintas variables relacionadas con la salud. Las gemelas contactadas son incluidas en el RGM si cumplen los dos criterios de inclusión: pares con ambos miembros vivos en el momento de inclusión en el registro, y no padecer trastornos o discapacidades que puedan limitar su participación activa y consciente. La colaboración con el RGM es voluntaria, no remunerada y cumple los requisitos legales sobre protección de datos personales.

Un total de 2387 casos fueron identificados inicialmente como posibles participantes. El 58.7% fueron excluidas por no ser gemelas. Al finalizar el primer contacto, 912 gemelas habían sido localizadas. 26 (2.9%) no cumplieron los criterios de inclusión y 24 (2.6%) no estaban interesadas en participar. Un total de 862 gemelas (94.5%) respondieron al cuestionario inicial, de las que 850 aceptaron participar en el RGM (MZ 46.2% y DZ 53.8%). Esto representa aproximadamente un 85% de la población total de gemelas residentes en la Región para esta franja de edad (Asumiendo una proporción similar de casos no-gemelares entre los no localizados). (Figura 2)

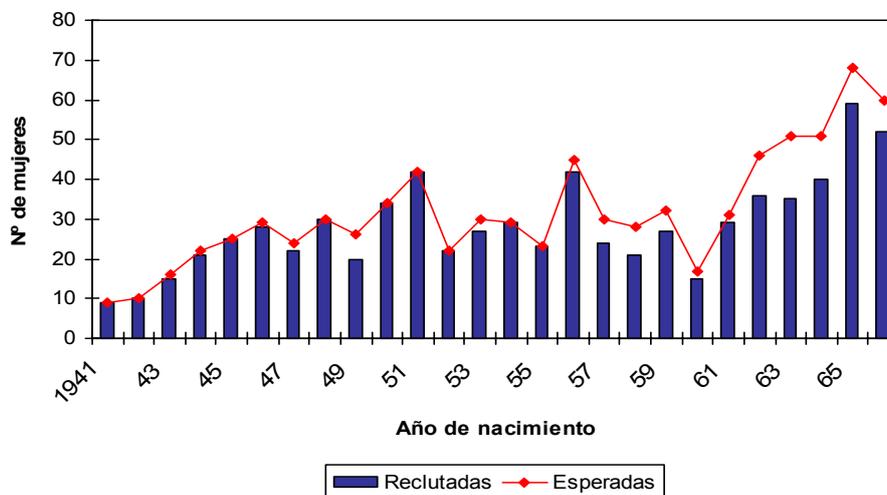


Figura 2: Distribución de casos esperados y reclutados por año de nacimiento

La cigosidad de las gemelas ha sido establecida inicialmente por medio de cuestionario. Este instrumento ha sido elaborado específicamente para este Registro, dado que actualmente no existe ninguno disponible en español, adaptando el de 12 ítems utilizado por el Registro de Gemelos de Holanda. Este método asume un error en la clasificación de, aproximadamente, un 5% (Willemsen, Posthuma, & Boomsma, 2005).

## Resultados

Si bien está abierto a otros temas de interés, los objetivos principales de investigación del RGM se centran en la aplicación de este tipo de diseños al análisis de los comportamientos relacionados con la salud. Los cambios en los patrones de enfermedad que se vienen produciendo en las sociedades occidentales durante las últimas décadas han provocado un incremento de la importancia de los factores conductuales y medioambientales en la etiología de los trastornos de salud con mayor incidencia. Así, la obesidad, el sedentarismo, la alimentación, el consumo de tabaco, alcohol y otras sustancias, la adherencia al tratamiento, o el desarrollo de conductas sexuales de riesgo, entre otras, constituyen actualmente áreas de intervención y, por supuesto, de investigación de gran relevancia.

En los últimos años se ha observado que muchas de estas variables, además de por factores ambientales de tipo psicosocial, están influidas por factores genéticos (Institute of Medicine, 2006; Bazzett, 2008; Plomin et al., 2008). Siguiendo esta vía de investigación, el RGM ha descrito recientemente una heredabilidad (i.e. proporción de la varianza fenotípica que puede ser explicada por las diferencias genéticas entre los individuos de una población) elevada o moderada en variables como Índice de Masa Corporal (Ordoñana et al., 2007; Rebollo-Mesa y Ordoñana, 2009), conducta de lactancia natural (Ordoñana, Rebollo-Mesa, Colodro, et al., 2008), salud percibida (Ordoñana, Rebollo-Mesa, Sánchez-Romera, et al., 2009), actividad física (Ordoñana, Sánchez-Romera, Pérez Valera, et al., 2009), o la conducta de participación en programas de detección precoz del cáncer de mama (Ordoñana, Rebollo-Mesa, Pérez-Riquelme et al., 2008).

En este último caso, nuestros datos han confirmado los resultados hallados por otros autores previamente (Treloar, McDonald y Martin, 1999), según los cuáles las diferencias individuales encontradas en el desarrollo de conductas de prevención del cáncer de mama son explicables, en parte, por factores genéticos que implicarían una predisposición hacia la aceptación de las recomendaciones de las autoridades sanitarias y la participación en este tipo de programas. Las estimaciones de heredabilidad en este estudio oscilaron entre un 50% para la realización de mamografía y un 37% para el auto-examen de mama. Por su parte, nuestros datos arrojan unas estimaciones similares, incorporando el efecto de la edad, que implica una variación de las estimaciones de heredabilidad que van desde un 41% para la muestra global (40-67 años) hasta un 56% en las menores de 50. Este efecto es ambiental y se produce, con toda probabilidad, por la incorporación de las mujeres a los programas de cribado mamográfico, llevados a cabo por las autoridades sanitarias, a partir de los 50 años (Ordoñana, Rebollo-Mesa, Pérez-Riquelme et al., 2008).

## Conclusión

El procedimiento utilizado ha permitido un desarrollo correcto del Registro y alcanzar tasas de localización y respuesta elevadas. El método es viable para analizar la contribución de factores genéticos y ambientales sobre los comportamientos de salud estudiados. Actualmente el RGM está realizando una segunda oleada de recogida de datos, incluyendo muestra de ADN de las participantes con objeto de crear un biobanco ligado al Registro. Paralelamente se está iniciando el proceso para incorporar gemelos varones y de distinto sexo, en los mismos grupos de edad contemplados hasta ahora.

El RGM está diseñado, principalmente, como recurso de investigación. Como tal está abierto a colaboración y sus bases de datos están disponibles para investigadores interesados en esta metodología. Cualquier propuesta interesada en utilizar la información del banco de datos del RGM o recoger datos específicos en la muestra de gemelos disponible, será revisada y analizada de forma individualizada, en función de los requerimientos científicos y éticos establecidos por la comisión científica del Registro.

Agradecimientos. El Registro de Gemelos de Murcia está financiado por la Fundación Séneca (08633/PHCS/08) y el Ministerio de Ciencia e Innovación (PSI2009-11560).

### Referencias

- Bazzett, T.J. (2008). *An introduction to behavior genetics*. Sunderland, MA: Sinauer Associates Pub.
- Boomsma, D., Busjahn, A. y Peltonen L. (2002). Classical twin studies and beyond. *Nature Reviews*, 3, 872-882.
- Busjahn, A. (2002). Twin registers across the globe: what's out there in 2002? *Twin Res*, 5: v-vi.
- Institute of Medicine (2006). *Genes, behavior, and the social environment*. Washington, D.C.: The National Academies Press.
- Johnson, W., Turheimer, E., Hottelmann, II. y Bouchard, T.J. (2009). Beyond heritability. Twin studies in behavioral research. *Current Directions in Psychological Science*, 18, 217-220.
- Martin, N., Boomsma, D. y Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics*, 17, 387-392.
- Neale, M.C. y Cardon, L.R. (1992). *Methodology for genetic studies of twins and families*. NATO.
- Ohm Kyvik, K. y Derom, C. (2006). Data collection on multiple births -- establishing twin registers and determining zygosity. *Early Human Development*, 82, 357-363.
- Ordoñana, J.R., Pérez-Riquelme, F., González-Javier, F., Carrillo, E., Gómez-Amor, J. y Martínez-Selva, J.M. (2006). An initiative in Spain for the study of women's health: the Murcia Twin Registry. *Twin Res Hum Genet*, 9, 865-867.
- Ordoñana, J.R., Rebollo-Mesa, I., González-Javier, F., Pérez-Riquelme, F., Martínez-Selva, J.M., Willemsen, G. et al. (2007). Heritability of body mass index: a comparison between the Netherlands and Spain. *Twin Res Hum Genet*, 10, 749-756.
- Ordoñana, J.R., Rebollo-Mesa, I., Colodro, L., Pérez-Riquelme, F., Sánchez-Romera, J.F. y Martínez-Selva, J.M. (2008). Heritability of Breastfeeding Behavior. *Behavioral Genetics Association Conference: Louisville (USA)*.
- Ordoñana, J.R., Rebollo-Mesa, I., Pérez-Riquelme, F., Sánchez-Romera, J.F., Colodro, L. y Martínez-Selva, J.M. (2008). Heritability of Early Breast Cancer Detection Behaviors in Female Twins. *10th World Congress on Psycho-Oncology: Madrid*.

- Ordoñana, J.R., Rebollo-Mesa, I., Sánchez-Romera, J.F., Pérez-Riquelme, F. y Boomsma, D.I. (2009). Heritability of Self-Rated Health: Cross Cultural Comparison between The Netherlands and Spain *21st Convention of the Association for Psychological Science* San Francisco (USA).
- Ordoñana, J.R., Sánchez-Romera, J.F., Pérez-Valera, S., Pérez-Riquelme, F. y Rebollo-Mesa, I. (2009). Moderating effects of physical exercise on genetic and environmental influences on body mass index of adult female twins. *Behavioral Genetics Association Conference: Minneapolis (USA)*.
- Plomin, R., De Fries, J., McClearn, G. y McGuffin, P. (2008). *Behavioral Genetics*. New York: Worth Publishers.
- Rebollo-Mesa, I. y Ordoñana, J.R. (2009). Childbirths moderate the genetic and environmental influences on BMI in adult twins. *Obesity, 17*, 1646-1647.
- Treloar, S.A., McDonald, C.A. y Martin, N.G. (1999). Genetics of early cancer detection behaviours in Australian female twins. *Twin Res, 2*, 33-42.
- Willemsen, G., Posthuma, D. y Boomsma, D.I. (2005). Environmental factors determine where the Dutch live: results from the Netherlands Twin Register. *Twin Research and Human Genetics, 8*, 312-317.

# MODELADO ESTADÍSTICO DE VARIABLES ASOCIADAS A LA ACCIDENTALIDAD EN MOTOCICLETA

Patricia Pérez-Fuster, María F. Rodrigo, María Luisa Ballestar y Jaime Sanmartín  
Universidad de Valencia

El propósito fundamental de este trabajo consiste en analizar las circunstancias que describen los accidentes de motocicleta en España. Los datos analizados son 66722 registros de accidentes ocurridos en España entre 2002 y 2006 en los que ha habido al menos una motocicleta implicada, y han sido recogidos mediante el cuestionario de accidentes con víctimas que actualmente posee la Dirección General de Tráfico (DGT). El análisis descriptivo de estos datos muestra que los accidentes ocurridos con una sola motocicleta, producidos por salidas de vía y en carretera, son los que tienen consecuencias más graves. Sobre este conjunto de accidentes se aplica un modelo de regresión *logit* múltiple siendo la variable de respuesta dicotómica la gravedad del accidente con dos categorías: mortal o grave *vs.* leve. Se incluyen en el modelo variables predictoras relacionadas con el motorista, la vía y el entorno. Los resultados indican que entre las variables predictoras más relevantes de la gravedad del accidente se encuentran las infracciones de velocidad y la no utilización de los accesorios de seguridad.

Según datos de la DGT, en los últimos años se ha producido un incremento notable de la accidentalidad en motocicleta, registrándose un aumento de la mortalidad del 72% en el año 2007 con respecto a 2003. En 2008 la cifra de siniestralidad de motocicletas en carretera se redujo en un 27,7% con respecto al año anterior. Sin embargo, en los datos recogidos hasta el segundo trimestre de 2009, la cifra de víctimas mortales en este tipo de accidente representa un 2% más, respecto del total de accidentes de tráfico, que en el segundo trimestre del año anterior.

En la década de los 90, Shankar, V. y Mannering, F. L. (1996) llevaron a cabo un estudio sobre la gravedad del accidente en motocicleta y concluyeron que factores como la no utilización del casco, unido al consumo de alcohol, aumentaban la probabilidad de siniestralidad. Más recientemente, Savolainen, P. y Mannering, F. (2007) estudiaron la gravedad de los accidentes de motocicleta con colisiones en función de la implicación de uno o varios vehículos. Entre otros resultados, obtuvieron que factores como la poca visibilidad o las colisiones contra objetos fijos jugaban un papel decisivo en la gravedad de las lesiones de los motoristas.

## *Objetivos y análisis*

El objetivo de este trabajo consiste en estudiar las variables que se relacionan con la gravedad de los accidentes de motocicleta en España. El análisis de datos llevado a cabo en este estudio se divide en dos fases:

En primer lugar, se realizan análisis descriptivos sobre una base de datos que cuenta con un total de 66722 accidentes, que han sido recogidos mediante el cuestionario de accidentes con víctimas que posee actualmente la DGT y ocurridos entre 2002 y 2006 en los que hay al menos una motocicleta implicada. Mediante la utilización de los criterios *tipo de accidente* (colisión, contra obstáculo, atropello, vuelco, salida de vía u otro), *tipo de vía* (carretera o zona urbana) y *número de vehículos implicados* (uno, dos o accidente múltiple), se obtienen resultados que permiten la selección de un subconjunto formado por aquellos accidentes de motocicleta que se producen con mayor frecuencia y que tienen consecuencias más graves.

Y, partiendo del subconjunto de datos obtenido mediante el análisis anterior, formado por un total de 4493 accidentes de motocicleta, se procede a analizar los factores relacionados con el motorista, la vía y su entorno para explicar y predecir la gravedad del accidente con este tipo de vehículo. Para ello se aplica un procedimiento estadístico “backward” tomando como modelo inicial el modelo de regresión *logit* múltiple que incluye los efectos principales de todas las variables explicativas, e incluyendo como variable de respuesta la gravedad del accidente con dos categorías: grave o mortal vs. leve, entendiéndose accidente grave o mortal aquel en el que al menos un motorista fallece o resulta herido de gravedad, y entendiéndose accidente leve aquel en el que al menos un motorista resulta herido leve.

### Resultados

Los análisis descriptivos realizados con un total de 66722 accidentes de motocicleta ocurridos en España entre 2002 y 2006 indican que los accidentes de motocicleta con consecuencias más graves son aquellos que ocurren por salidas de vía, en carreteras convencionales y en los que hay una sola motocicleta implicada.

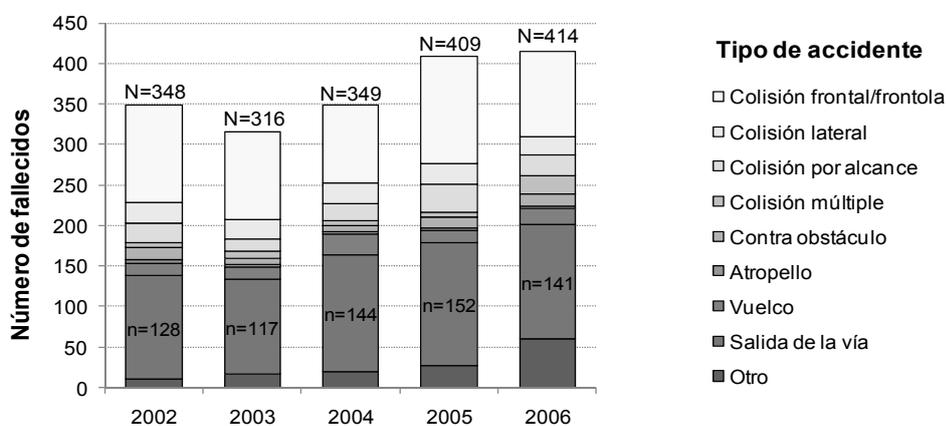


Figura 1. Fuente: DGT. Evolución del número de víctimas de accidente de motocicleta en función del tipo de accidente

Como se observa en la figura 1, el tipo de accidente con víctimas más frecuente entre las motocicletas a largo de los años estudiados es la salida de la vía, que se encuentra seguida de las colisiones frontales o frontolaterales, también muy frecuentes en este tipo de accidente.

Por otro lado, en zona urbana solamente un 5% de los accidentes con víctimas se producen a causa de salidas de vía mientras que en carretera, las salidas de vía se encuentran presentes en aproximadamente un 38% de los accidentes (Figura 2).

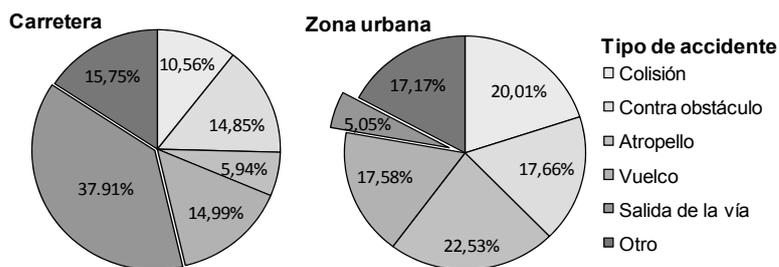


Figura 2. Fuente: DGT. Distribución porcentual de accidentes de motocicleta con víctimas en función del tipo de vía

Y, en cuanto al número de vehículos implicados en el accidente, puede observarse (Figura 3) que los accidentes de motocicleta con dos vehículos implicados son los más frecuentes, pero que sin embargo resultan más graves, es decir, hay un mayor número de fallecidos, cuando el accidente ocurre con solamente una motocicleta implicada.

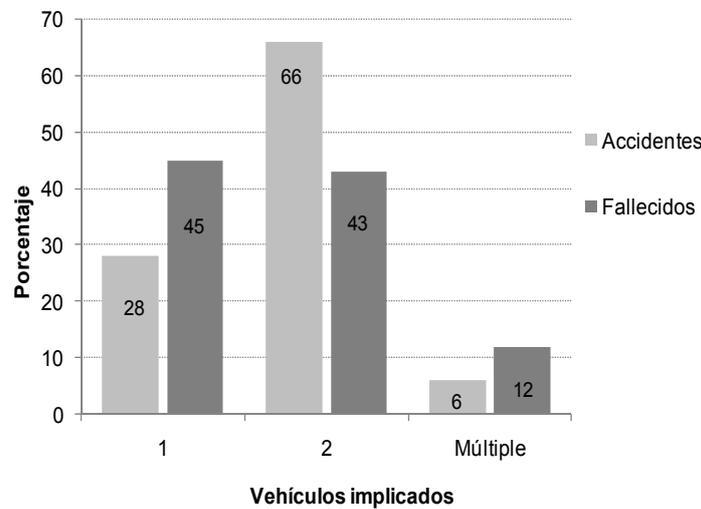


Figura 3. Fuente: DGT. Distribución porcentual de accidentes y de motoristas fallecidos en función del número de vehículos implicados en el accidente

Estos primeros resultados dan lugar a la selección de un subconjunto de datos que cuenta con un total de 4493 accidentes de motocicleta, ocurridos por salidas de vía, en carretera y con una sola motocicleta implicada, ocurridos entre 2002 y 2006.

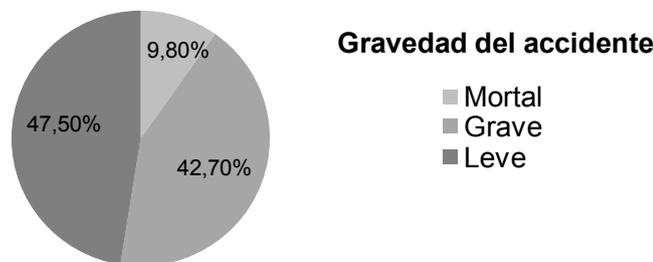


Figura 4. Fuente: DGT. Distribución porcentual de la gravedad del accidente en los accidentes de motocicleta ocurridos por salidas de vía, en carretera y con una sola motocicleta implicada

Un análisis descriptivo realizado sobre este subconjunto de datos (Figura 4) muestra que más de la mitad de los accidentes que cumplen con estas características tienen consecuencias graves o mortales, o dicho de otra manera, más de la mitad de los motoristas resultan fallecidos o gravemente heridos.

Con el objetivo de conocer cuáles son las variables que predicen la gravedad de estos accidentes, se aplica un modelo de regresión *logit* múltiple y los resultados del análisis se presentan a continuación.

Tabla 1. Estadísticos descriptivos de las variables explicativas incluidas en el modelo inicial

VARIABLES EXPLICATIVAS	DESCRIPCIÓN	PROPORCIÓN
<b>Edad del motorista</b>	Hasta 24 años (*Ref.)	11,60%
	Entre 25 y 32 años	40,80%
	Entre 33 y 40 años	24,10%
	41 años en adelante	23,50%
<b>Barrera de seguridad</b>	Sí	38,00%
	No (*Ref.)	62,00%
<b>Accesorios de seguridad</b>	Todos los ocupantes (*Ref.)	92,30%
	Uno de los ocupantes	0,70%
	Ninguno de los ocupantes	4,70%
<b>Características de la vía</b>	Recta (*Ref.)	16,90%
	Curva suave	20,10%
	Curva fuerte	56,20%
	Intersección	6,70%
<b>Infracción de velocidad</b>	Sí	64,90%
	No (*Ref.)	22,70%
<b>Condiciones de la superficie</b>	Seca y limpia	90,20%
	Condiciones adversas (*Ref.)	9,80%
<b>Luminosidad</b>	Pleno día (*Ref.)	81,40%
	Crepúsculo	3,80%
	Noche	14,80%
<b>Motivo de desplazamiento</b>	Trabajo (*Ref.)	8,80%
	Ocio/vacaciones	79,60%
<b>Día de la semana</b>	Entre semana (*Ref.)	37,50%
	Fin de semana	62,50%
<b>Arcén</b>	Inferior a 1,5 metros	87,80%
	Igual o superior a 1,5 metros (*Ref.)	12,20%
<b>Paneles direccionales</b>	Sí (*Ref.)	40,60%
	No	59,40%
<b>Captafaros</b>	Sí (*Ref.)	32,90%
	No	67,10%

\*Ref.: Categoría de referencia en el análisis de regresión.

En la tabla 1 pueden observarse todas las variables explicativas que han sido incluidas en el modelo de regresión inicial, con la descripción de sus correspondientes categorías e indicándose cuál de ellas se toma como categoría de referencia en el análisis. Puede apreciarse que se toman en cuenta un total de 12 variables que recogen información del motorista que se accidenta (edad, uso de accesorios de seguridad, infracción de velocidad, motivo de desplazamiento), de la vía por la cual circula el motorista (barrera de seguridad, características de la vía, arcén, paneles direccionales, captafaros) y del entorno que envuelve a éste en el momento del accidente (condiciones de la superficie, luminosidad, día de la semana).

Tabla 2. Coeficientes del modelo final seleccionado

VARIABLES	ODDS RATIO	P-VALUE	IC 95%	
			Inferior	Superior
<b>Barrera de seguridad</b>				
Sí	1,210	0,007	1,054	1,388
<b>Accesorios de seguridad</b>				
Uno de los ocupantes	2,424	0,042	1,033	5,690
Ninguno de los ocupantes	3,491	0,000	2,387	5,106
<b>Características de la vía</b>				
Curva suave	1,534	0,000	1,213	1,939
Curva fuerte	1,456	0,001	1,178	1,799
Intersección	0,920	0,610	0,667	1,269
<b>Infracción de velocidad</b>				
Sí	1,879	0,000	1,595	2,214
<b>Condiciones de la superficie</b>				
Seca y limpia	1,641	0,000	1,306	2,062
<b>Luminosidad</b>				
Crepúsculo	1,664	0,006	1,158	2,393
Noche	1,452	0,000	1,187	1,776
<b>Día de la semana</b>				
Fin de semana	1,257	0,001	1,094	1,444

Ajuste del modelo:  $\chi^2$  Hosmer-Lemeshow=8,897 Sig.=0,260;  $\chi^2$  Comparación con el modelo nulo=241,556 g.l.=11 Sig.=0,000.

El análisis de regresión indica que variables del motorista como la edad o el motivo de desplazamiento, y variables de la vía como el tamaño del arcén, la ausencia de paneles direccionales o de captafaros, no han resultado significativas en la predicción de la gravedad del accidente. No obstante, el modelo final seleccionado incluye siete variables predictoras cuyos *odds ratio*, significación e intervalos de confianza se presentan en la tabla 2. Los resultados se interpretan en las secciones que aparecen a continuación.

### *Barrera de seguridad*

Por barrera de seguridad se entiende cualquier protección presente en los bordes de las carreteras cuyo fin es evitar la salida del vehículo de la vía. El análisis indica que la *odds* de accidente grave o mortal se multiplica por 1,210 (IC95%=1,054-1,388) cuando hay barreras de seguridad. Por tanto, la presencia de quitamiedos en la carretera incrementa la probabilidad de que el motorista resulte fallecido o herido de gravedad tras el accidente.

### *Accesorios de seguridad*

El casco es el accesorio de seguridad obligatorio más común entre los motoristas aunque también pueden contemplarse en el análisis motocicletas que precisan cinturón de seguridad en vez de casco debido al diseño de su carrocería. Los resultados muestran que el valor de la *odds* de los accidentes graves o mortales se multiplica por 2,424 (IC95%=1,033-5,690), si prescinde uno de los ocupantes de los accesorios de seguridad, o por 3,491 (IC95%=2,387-5,106), si prescinden ambos, respecto a si hacen uso de los mismos todos los ocupantes de la motocicleta. Dicho de otro modo, si uno de los ocupantes no lleva puesto el accesorio de seguridad, aumenta la probabilidad de que el accidente resulte grave o mortal, y más consecuencias negativas tiene el accidente si ninguno de los dos ocupantes hace uso del mismo.

### *Características de la vía*

Curva suave, curva fuerte e intersección son las categorías que se consideran en esta variable para el análisis de regresión, tomando como categoría de referencia las rectas. Los datos obtenidos indican que circular por un tramo con curva fuerte incrementa en un 45,6% (OR=1,456; IC95%=1,178-1,799) la probabilidad de accidente grave o mortal y en un 53,4% (OR=1,534; IC=1,213-1,939) circular por tramo con curva suave, comparado a circular en tramo recto. La categoría intersección no resulta significativa en el análisis.

### *Infracción de velocidad*

Circular infringiendo los límites de velocidad permitidos, así como circular a una velocidad inadecuada debido a determinadas condiciones, son las infracciones que recoge esta variable. Los resultados obtenidos muestran que las infracciones de velocidad predicen significativamente los accidentes graves o mortales, multiplicándose el valor de la *odds* de los mismos por 1,879 (IC95%= 1,595-2,214) al comparar esta categoría con la ausencia de infracción de velocidad.

### *Condiciones de la superficie*

En la categoría condiciones adversas de la superficie se contempla superficie mojada, helada, nevada, barrosa, con gravilla suelta o con restos de aceite, entre otras. Por lo tanto, superficie seca y limpia recoge toda situación que no sea ninguna de las anteriores. El análisis muestra que el valor de *odds* de accidentes graves o mortales se multiplica por 1,641 (IC95%=1,306-2,062) al pasar de condiciones adversas a superficie seca y limpia. Dicho de otra manera, en condiciones de superficie óptimas los accidentes tienen consecuencias de mayor gravedad.

### *Luminosidad*

El día se divide en tres momentos según su nivel de luminosidad: día, noche y crepúsculo. Los datos indican que los accidentes ocurridos en el momento del crepúsculo (OD=1,664; IC95%=1,158-2,393) y durante la noche (OD=1,452; IC95%=1,187-1,776) tienen peores consecuencias que aquellos ocurridos durante el día, si bien el crepúsculo es el momento en el que los accidentes tienen consecuencias más graves.

### *Día de la semana*

Los accidentes ocurridos de lunes a viernes forman la categoría denominada *entre semana*, y los accidentes ocurridos en sábado o domingo se incluyen en la categoría *fin de semana*. Según los datos obtenidos, la probabilidad de accidente grave o mortal se incrementa en los desplazamientos producidos en fin de semana, multiplicándose su *odds* por 1,257 (IC95%= 1,094-1,444) al compararlo con los accidentes ocurridos entre semana.

## Discusión

Los resultados obtenidos se sitúan en la línea de la investigación llevada a cabo por Shankar, V. y Mannering, F.L. (1996) coincidiendo con el hallazgo de los efectos negativos que tiene la no utilización de los accesorios de seguridad sobre la gravedad del accidente.

Por otro lado, el análisis pone de manifiesto información relativa a las conductas de riesgo. Los motoristas se muestran más cautelosos bajo ciertas condiciones como puede ser conducir sobre pavimento mojado o con gravilla suelta, produciéndose en estas circunstancias lesiones de menor gravedad. Y contrariamente, en situaciones en las que no hay peligro de superficie adversa porque se encuentra en condiciones óptimas, las consecuencias de los accidentes son de mayor gravedad. El aumento de velocidad podría estar jugando en este tipo de situación un papel fundamental. Asimismo, los datos predicen accidentes más graves en tramos de curva suave que con curva fuerte lo que posiblemente implica una actitud favorable hacia el riesgo en condiciones o situaciones para las cuales se cree poseer suficiente destreza. Por todo ello, una extensión de este trabajo consistiría en llevar a cabo un estudio sobre percepción de riesgo de los motoristas. Además podrían incorporarse variables relacionadas con las características del vehículo, como puede ser su cilindrada, o con la conducta del motorista, como la velocidad a la que circulaba en el momento del accidente, ya que la base de datos que se ha manejado en este trabajo no recoge este tipo de información y sabemos, por el estudio llevado a cabo por Savolainen, P. y Mannering, F. (2007), que la gravedad de los accidentes aumenta a medida que se incrementa la velocidad de la motocicleta.

Finalmente, en este último estudio, los autores hallaron que las colisiones contra objetos fijos jugaban un papel decisivo en la gravedad de las lesiones de los motoristas y los resultados obtenidos evidencian este hecho en España. Las barreras de seguridad situadas estratégicamente en las carreteras con el objetivo de proteger y evitar salidas de vehículos tienen, por otra parte, consecuencias graves e incluso mortales sobre los motoristas. Un estudio más preciso sobre la interacción del motorista con estas barreras ofrecería una información clave para la prevención de lesiones en este tipo de conductor.

### Referencias

- Chen, C. (2009). Personality, safety attitudes and risky driving behaviors—Evidence from young Taiwanese motorcyclists. *Accident Analysis and Prevention*, 41(5), 963-968.
- Elliott, M. A., Baughan, C. J. & Sexton, B. F. (2007). Errors and violations in relation to motorcyclists' crash risk. *Accident Analysis & Prevention*, 39(3), 491-499.
- Haque, M. M., Chin, H. C. & Huang, H. (2009). Modeling fault among motorcyclists involved in crashes. *Accident Analysis & Prevention*, 41(2), 327-335.
- Ichikawa, M., Chadbunchachai, W. & Marui, E. (2003). Effect of the helmet act for motorcyclists in Thailand. *Accident Analysis & Prevention*, 35(2), 183-189.
- Lin, M. & Kraus, J. F. (2009). A review of risk factors and patterns of motorcycle injuries. *Accident Analysis and Prevention*, 41(4), 710-722.
- Majdzadeh, R., Khalagi, K., Naraghi, K., Motevalian, A. & Eshraghian, M. R. (2008). Determinants of traffic injuries in drivers and motorcyclists involved in an accident. *Accident Analysis and Prevention*, 40(1), 17-23.
- Mannering, F. L. & Grodsky, L. L. (1995). Statistical analysis of motorcyclists' perceived accident risk. *Accident Analysis & Prevention*, 27(1), 21-31.

- Natalier, K. (2001). Motorcyclists' interpretations of risk and hazard. *Journal of Sociology*, 37(1), 65-80.
- Oluwadiya, K. S., Kolawole, I. K., Adegbehingbe, O. O., Olasinde, A. A., Agodirin, O. & Uwaezuoke, S. C. (2009). Motorcycle crash characteristics in Nigeria: Implication for control. *Accident Analysis and Prevention*, 41(2), 294-298.
- Savolainen, P. & Mannering, F. (2007). Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Accident Analysis & Prevention*, 39(5), 955-963.
- Shankar, V. & Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research*, 7(3), 183-194.
- Wong, J., Chung, Y. & Huang, S. (2010). Determinants behind young motorcyclists' risky riding behavior. *Accident Analysis and Prevention*, 42(1), 275-281.
- Wong, T., Lee, J., Phoon, W. & Yiu, P. (1990). Driving experience and the risk of traffic accident among motorcyclists. *Social Science & Medicine*, 30(5), 639-640.
- Yannis, G., Golias, J. & Papadimitriou, E. (2005). Driver age and vehicle engine size effects on fault and severity in young motorcyclists accidents. *Accident Analysis & Prevention*, 37(2), 327-333.

# VARIABLES DISCRIMINANTES DEL FUNCIONAMIENTO EJECUTIVO EN NIÑOS Y ADOLESCENTES CON TRASTORNO POR DÉFICIT DE ATENCIÓN CON HIPERACTIVIDAD

Emilio Verche, Raquel Martín, Alejandra Machado, Pedro A. González-Pérez, Itziar Quintero, Jezabel Bravo y Sergio Hernández Sánchez  
Universidad de La Laguna

Desde la Neuropsicología se considera que el TDAH se asocia con déficit en las funciones ejecutivas (FE). Sin embargo, existen discrepancias sobre qué FE están alteradas en el TDAH. **Objetivo.** Identificar aquellas FE que maximizan las discrepancias entre TDAH y controles, relacionándolas con el ciclo vital. **Método.** Se realizaron dos análisis discriminantes, uno con dos grupos (TDAH y Controles) y otro con cuatro grupos (Niños-Controles, Niños-TDAH, Adolescentes-Controles y Adolescentes-TDAH). A los 65 sujetos se les administró un protocolo de evaluación neuropsicológica de las FE que mide *span* verbal y visual, fluidez fonética, planificación, flexibilidad e inhibición. **Resultados.** Primer análisis: la función discriminante fue significativa, describiendo que cuanto mayor es la puntuación en la función, mejor es la planificación, la flexibilidad, el *span* verbal y visual y la inhibición. El 78,5% de los casos fue clasificado correctamente a los grupos. Segundo análisis: dos de las tres funciones discriminantes fueron significativas, explicando un 96,8% de la varianza. La primera función está formada por fluidez fonética e inhibición y la segunda por *span* verbal, planificación y flexibilidad. Se clasifican correctamente el 75,4% de los casos. **Conclusiones.** Los resultados indican que las FE evaluadas discriminan correctamente entre sujetos controles y patológicos, niños y adolescentes.

El Trastorno por Déficit de Atención con Hiperactividad (TDAH) es la patología neuroconductual más común en la infancia (Bradley y Golden, 2001). El Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-IV-TR (American Psychiatric Association, 2003) define al TDAH atendiendo a la presencia de tres síntomas nucleares: Inatención, Hiperactividad e Impulsividad. El trastorno es más frecuente en varones con ratios de 2:1/9:1 en función del tipo y las circunstancias de la clínica. La prevalencia del TDAH se sitúa en torno al 3-7% de los niños en edad escolar (American Psychiatric Association, 2003). Las pautas oficiales que utilizan los profesionales sanitarios para el diagnóstico del TDAH se encuentran recogidas en los criterios diagnósticos del DSM-IV-TR y la CIE-10. Se requiere que el trastorno se dé en más de un ambiente y éste afecte a la vida diaria del sujeto. Por ello, el diagnóstico se hace a partir de escalas de comportamiento completadas por padres y profesores. Sin embargo, estas medidas no dejan de plantear problemas como la falta de correlación con tests objetivos y, por lo tanto, se hace necesario la consideración de otras medidas, además de las escalas, en el proceso diagnóstico (Quintero, Bravo, Martín, González-Pérez, Verche y Hernández, 2009). En la tabla 1 se recoge los criterios diagnósticos establecidos por el DSM-IV-TR para esta patología.

*Tabla 1. Criterios diagnósticos del DSM-IV-TR para el TDAH  
(American Psychiatric Association, 2003)*

<p>A. Debe cumplir 6 ó más de los síntomas siguientes para el déficit de atención y/o hiperactividad, síntomas que deben haber estado presentes durante más de 6 meses.</p> <p>Criterios de falta de atención (al menos 6):</p> <ol style="list-style-type: none"> <li>1. No pone atención a los detalles y comete errores frecuentes por descuido.</li> <li>2. Tiene dificultad para mantener la atención en las tareas y en los juegos.</li> <li>3. No parece escuchar lo que se le dice cuando se le habla directamente.</li> <li>4. No sigue las instrucciones, no termina las tareas en la escuela, no termina los “recados”, a pesar de entenderlos.</li> <li>5. Dificultades para organizar sus tareas y actividades.</li> <li>6. Evita o rechaza realizar tareas que le demanden esfuerzo.</li> <li>7. Pierde sus útiles o cosas necesarias para hacer sus actividades obligatorias (lapiceros, libros...).</li> <li>8. Se distrae fácilmente con estímulos irrelevantes.</li> <li>9. Olvidadizo en las actividades de la vida diaria.</li> </ol> <p>Criterios de hiperactividad-impulsividad (al menos 6):</p> <ol style="list-style-type: none"> <li>1. Molesta moviendo las manos y los pies mientras está sentado.</li> <li>2. Se levanta del puesto en la clase o en otras situaciones donde debe estar sentado.</li> <li>3. Corretea, trepa... en situaciones inadecuadas.</li> <li>4. Dificultad para relajarse o practicar juegos donde debe permanecer quieto.</li> <li>5. Está continuamente en marcha como si tuviera un motor por dentro.</li> <li>6. Habla excesivamente.</li> <li>7. Contesta o actúa antes de que se terminen de formular las preguntas.</li> <li>8. Tiene dificultad para esperar turnos en los juegos.</li> <li>9. Interrumpe las conversaciones o los juegos de los demás.</li> </ol>
B. Algunos síntomas presentes antes de los 7 años.
C. Síntomas presentes en 2 ó más lugares (escuela, casa...).
D. Clara evidencia de afectación social, académica u ocupacional.
E. Exclusión previa de otros trastornos del desarrollo que puedan estar justificando la sintomatología a estudio.

El DSM-IV-TR plantea la existencia de tres subtipos de TDAH. En el *TDAH tipo combinado* deben darse seis o más síntomas de inatención durante al menos seis meses y seis o más síntomas de hiperactividad-impulsividad durante al menos el mismo tiempo. En el *TDAH tipo inatento* deben darse seis o más síntomas de inatención durante al menos seis meses, pero menos de seis síntomas de hiperactividad-impulsividad. Por último, para el *TDAH tipo hiperactivo-impulsivo* deben cumplirse seis o más síntomas de hiperactividad-impulsividad durante al menos seis meses, pero menos de seis síntomas de inatención.

En cuanto a la evolución del trastorno, se observan diferencias sintomatológicas entre niños y adultos. En concreto, en los primeros años de escolarización, los síntomas del TDAH suelen ser más prominentes y según avanzamos en el desarrollo tienden a atenuarse, siendo menos frecuentes los signos de actividad motora excesiva en las últimas etapas de la infancia y comienzos de la adolescencia. Hetchman (2003) define tres subgrupos en los que evoluciona el TDAH en la adolescencia: sujetos con pocos problemas (10-20%), sujetos que continúan con el trastorno junto a problemas de carácter social, académico o emocional (70%) y sujetos con síntomas del TDAH y con conducta antisocial (25%). Existe controversia en cuanto a los datos, pero se estima que aproximadamente un 50% de los adolescentes con TDAH continuarán presentando el trastorno en la edad adulta (Silver, 2004).

La neurobiología del TDAH pone de manifiesto que estos síntomas son compatibles con un déficit en el funcionamiento de la corteza prefrontal, así como de las conexiones dopaminérgicas y noradrenérgicas que esta región mantiene con estructuras subcorticales. En población neurológicamente

normal, el conjunto de regiones cerebrales citadas, parece vincularse al funcionamiento ejecutivo. La consideración conjunta de todos estos datos, apoya la teoría de que los síntomas del TDAH son subsidiarios de un déficit primario en las funciones ejecutivas (e.g., Arnsten, Steere y Hunt, 1996; Pliszka, McCracken y Maas, 1996; Barkley, 1997; Goldberg, 2002; Willcut, Doyle, Nigg, Faraone y Pennington, 2005) pero, a pesar de la amplia literatura al respecto, no existe un consenso sobre qué dominios específicos están afectados en esta patología (Sergeant, Geurts y Oosterlaan, 2002; Doyle, 2006). Las funciones ejecutivas las podemos entender como: “*La capacidad para responder de manera adaptativa a nuevas situaciones y son la base de muchas habilidades cognitivas, emocionales y sociales*” (Lezak, Howieson y Loring, 2004, p. 611). Representan las capacidades cognitivas que se sitúan en el nivel supraordinado más elevado de la jerarquía cognitiva (Tranel, Anderson y Benton, 1994). En este constructo psicológico se incluyen, generalmente y no sin discusión, las siguientes funciones: atención sostenida y dirigida, memoria de trabajo, planificación, inhibición, flexibilidad cognitiva, fluidez verbal, fluidez de diseño y programación motora.

En un estudio de metaanálisis Sergeant et al. (2002) describen un déficit en inhibición en sujetos con TDAH, aunque también se objetiva en pacientes con Trastorno Oposicionista Desafiante y Trastorno de Conducta. Los resultados no son tan homogéneos en flexibilidad cognitiva medida con el Wisconsin Card Sorting Test ni en planificación entre sujetos controles y TDAH. En cambio, las tareas de fluidez diferencian entre TDAH y controles, especialmente ante consignas fonéticas. Por otra parte, Willcut et al (2005) encuentran déficits de moderada magnitud en inhibición de respuesta, flexibilidad, memoria de trabajo espacial, planificación y organización. Frazier, Demaree y Youngstrom (2004) señalan efectos similares en atención alternante, errores de omisión y de comisión, pero efectos pequeños en flexibilidad y organización, medidas con el Wisconsin Card Sorting Test y la Figura Compleja de Rey. No obstante, en estos dos últimos estudios también encuentran que los TDAH tienen un menor funcionamiento intelectual general, una menor velocidad de procesamiento y pobre vigilancia, lo que sugiere que los déficits en esta patología se extienden más allá del funcionamiento ejecutivo (Doyle, 2006). Las distintas investigaciones ponen de manifiesto que los déficits en las funciones ejecutivas no son suficientes ni necesarios para causar el TDAH (Willcutt et al., 2005). Sin embargo, los problemas en el funcionamiento ejecutivo constituyen una parte importante de la compleja etiología de este trastorno.

El objetivo de este estudio es identificar aquellas funciones ejecutivas que maximicen las discrepancias entre sujetos TDAH y sujetos controles sanos y estudiar si esos dominios del funcionamiento ejecutivo discriminan entre niños y adolescentes con y sin TDAH, dado el diferente patrón madurativo entre estos grupos de edad.

## Método

### *Participantes*

La muestra se compone de 65 sujetos, entre 7 y 17 años con una media de edad de 11,6 años y una desviación típica de 3,36. En la tabla 2 se detallan los datos sociodemográficos de la muestra. De éstos, 33 han sido diagnosticado de Trastorno por Déficit de Atención con Hiperactividad, subtipo combinado, según los criterios del Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-IV-TR (American Psychiatric Association, 2003) por parte de psiquiatras del Hospital Universitario de Canarias. Los restantes 32 sujetos constituyen la muestra de comparación, equiparados en edad y años de escolaridad. Los sujetos no contaban con otra patología neurológica y/o psiquiátrica ni trastornos del aprendizaje comórbidos y el cociente intelectual en ambos grupos es superior a 85 medido con la Escala de Inteligencia de Weschler para Niños 4ª Edición y la Escala de Inteligencia de Weschler para Adultos (Wechsler, 2005 y 1999). Los sujetos están escolarizados en centros públicos o concertados de

la provincia de Santa Cruz de Tenerife, acuden regularmente a clases y en el momento de la evaluación estaban sin tratamiento farmacológico.

Tabla 2. Descripción de la muestra

	CONTROLES (n=32)			TDAH (n=33)		
	N	A	Total	N	A	Total
Edad						
Media	9,5	15,6	11,7	9,7	15,5	11,4
(SD)	(1,9)	(0,8)	(3,4)	(2,9)	(0,9)	(3,4)
Rango	7-13	14-17	7-17	7-13	14-16	7-16
Género (V/M)	10/10	8/4	18/14	19/4	7/3	26/7

Niños; A=Adolescentes

### Aparatos

A todos los sujetos se les administró un protocolo de evaluación neuropsicológica de las funciones ejecutivas que mide memoria de trabajo verbal, memoria de trabajo visual, fluidez verbal, planificación, flexibilidad e inhibición cognitiva.

- *Memoria de trabajo verbal*: El test de memoria de trabajo de frases (Siegel y Ryan, 1989) consiste en oraciones a las que les falta la última palabra y el sujeto debe elicitarla. Después de escuchar dos frases, el sujeto tiene que recordar las dos palabras que faltaban en el mismo orden que las elicató. Se realizan tres ensayos por nivel y se va incrementando la dificultad con más palabras que elicitar hasta un máximo de cinco. Es una medida de *span*.
- *Memoria de trabajo visual*: el subtest de *span* espacial de la Wechsler Memory Scale-III (Wechsler, 1997) consiste en un tablero tridimensional en el que el sujeto tiene que reproducir la secuencia que realice el evaluador en orden directo e inverso. Para este estudio sólo se ha empleado el orden inverso que constituye una medida de memoria de trabajo visual.
- *Fluidez verbal*: Para la evaluación de la fluidez verbal se empleó el Controlled Oral Word Association Test. El sujeto tiene que producir la mayor cantidad de palabras ante una consigna fonética (letras F, A y S), con un minuto para cada letra. Se registra el número de palabras elicítadas, las intrusiones y las perseveraciones (Lezak et al, 2004).
- *Planificación*: La planificación se midió con la *Torre de Hanoi* (Simon, 1975). Esta prueba consiste en varias piezas de distinto tamaño que deben moverse desde una posición inicial hasta una posición final con la restricción de que sólo se debe coger una pieza cada vez y que una pieza no puede estar sobre otra de menor tamaño. Se realizan tres ensayos de práctica y seis de evaluación, uno de ellos a los 30 minutos. La puntuación viene dada por el número de movimientos, los errores y el tiempo empleado. Es una medida de planificación, aprendizaje procedimental y memoria procedimental (Lezak et al., 2004).
- *Flexibilidad*: La flexibilidad cognitiva se midió a través del Wisconsin Card Sorting Test (Berg, 1948). A los sujetos se les da 60 cartas en las que están impresas entre uno o cuatro símbolos (triángulo, estrella, cruz o círculo) en rojo, verde, amarillo o azul y tienen que colocarlas debajo de

las cartas estímulos de acuerdo a un principio que el sujeto debe deducir del patrón de respuestas del evaluador. Se registra el número de categorías completadas, las respuestas perseverativas y no perseverativas, los errores perseverativos y no perseverativos. Evalúa la habilidad de realizar cambios de estrategias y la identificación de categorías abstractas.

- *Inhibición cognitiva*: el test de Stroop (Stroop, 1935) consiste en tres láminas que el sujeto debe leer en 45 segundos. En la primera lámina (Palabra) tiene que leer palabras con el nombre de un color; en la segunda lámina (Color) se nombra el color en el que están unas marcas (“xxxx”), y en la tercera lámina (Palabra-Color), la de interferencia, tiene que decir el color en el que está escrita la palabra que, a su vez, denomina a un color distinto. Se recoge el total de palabras leídas en cada lámina y el número de errores. Constituye una medida de inhibición y de atención selectiva.

### *Procedimiento*

Las evaluaciones se realizaron en dos sesiones de dos horas cada una en días consecutivos en las dependencias de la Facultad de Psicología de la Universidad de La Laguna. Las mismas se llevaron a cabo por personal especializado y en un ambiente sin distractores. En la primera sesión se evaluó la inteligencia con el WISC-IV o WAIS-III, según la edad del sujeto. En la segunda sesión se administró el protocolo de evaluación neuropsicológica de las funciones ejecutivas, descrito anteriormente.

Para cada función ejecutiva evaluada se analizó un indicador, a saber, en memoria de trabajo verbal y visual el número total de ítems procesados, en fluidez verbal el número total de palabras producidas en los tres ensayos, en planificación el número total de movimientos, en flexibilidad los errores perseverativos y en inhibición cognitiva el número de ítems correctos en la lámina de interferencia. Por ello, las puntuaciones se tipificaron en valores  $z$  para el análisis discriminante.

Se llevaron a cabo dos análisis discriminantes. En el primer análisis, la variable dependiente fue el Grupo: control y TDAH y las variables discriminantes las que componen el protocolo de evaluación de funciones ejecutivas. En el segundo análisis, se dividieron a los sujetos en cuatro grupos: controles niños, controles adolescentes, TDAH niños y TDAH adolescentes. Todos los análisis se realizaron con el paquete estadístico SPSS 15 para Windows.

### Resultados

Comentaremos en primer lugar los resultados relativos a la comparación entre TDAH y controles y, posteriormente, dividiéndolos por grupo de edad.

En el primer análisis discriminante, la función es significativa ( $\lambda$  de Wilks=0,54;  $\chi^2_6=36,57$ ;  $p \leq 0,001$ ) y el 45,7% de la varianza es común a la función discriminante y los dos grupos (controles y TDAH). La prueba M de Box de igualdad de matrices de varianzas y covarianzas no es significativa (M de Box=27,73;  $F_{21,14567,31}=1,19$ ;  $p \geq 0,05$ ). La matriz de estructura (tabla 3) revela que cuanto mayor es la puntuación en la función, menos movimientos se realizan en la Torre de Hanoi, se cometen menos errores perseverativos y mayor es la puntuación en memoria de trabajo verbal y visual, en inhibición y fluidez verbal.

Tabla 3. Matriz de estructura

VARIABLE	CORRELACIÓN
Movimientos Torre de Hanoi	-0,62
Memoria Trabajo Verbal	0,56
Errores perseverativos	-0,49
Memoria Trabajo Visual	0,48
Inhibición	0,35
Fluidez verbal	0,25

El 78,5% de los casos se agruparon correctamente. Sólo 5 controles y 9 TDAH se clasifican erróneamente (tabla 4).

Tabla 4. Resultados de la clasificación entre el grupo Control y el grupo TDAH

	Grupo de pertenencia pronosticado		Total
	Control	TDAH	
Control	27 84,4%	5 15,6%	32 100%
TDAH	9 27,3%	24 72,7%	33 100%
Total	36 55,38%	29 44,62%	65 100%

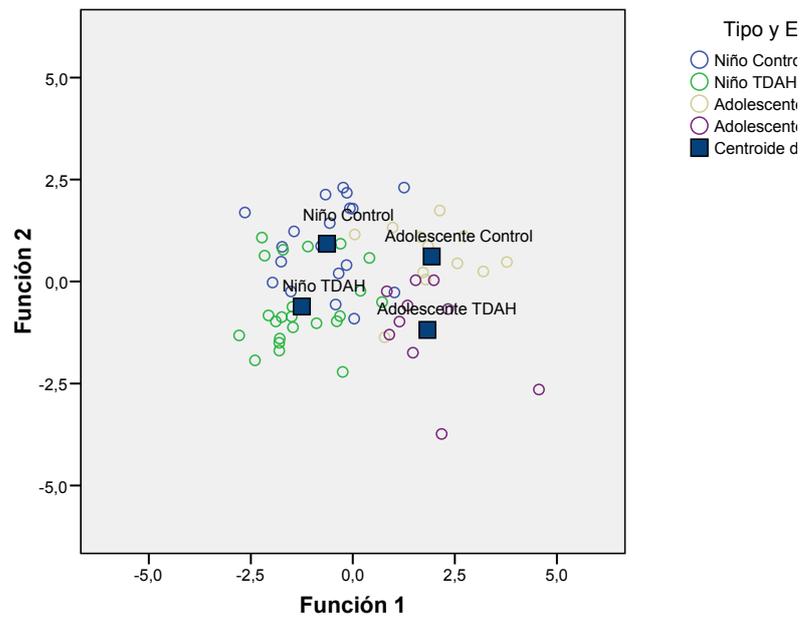
Cuando dividimos a los sujetos en función de su edad: niños controles (N-Control), niños TDAH (N-TDAH), adolescentes controles (A-Control) y adolescentes TDAH (A-TDAH) encontramos que dos de las tres funciones son significativas (función 1: lambda de Wilks=0,16;  $\chi^2_{18}=108,28$ ;  $p \leq 0,001$ ; función 2: lambda de Wilks=0,50;  $\chi^2_{10}=40,55$ ;  $p \leq 0,001$ ; función 3: lambda de Wilks=0,91;  $\chi^2_4=5,5$ ;  $p \geq 0,05$ ). La prueba M de Box de igualdad de matrices de varianzas y covarianzas no es significativa (M de Box=80,72;  $F_{63,4116,55}=1,01$ ;  $p \geq 0,05$ ). Un 68,69% de la variación es común entre la primera función y los grupos, un 44,89% entre la segunda función y los grupos y un 9% entre la tercera función y los grupos. La matriz de estructura rotada (tabla 5) nos indica que la primera función está formada por fluidez e inhibición, la segunda por memoria de trabajo verbal, planificación y flexibilidad y la tercera por memoria de trabajo visual.

Tabla 5. Matriz de estructura rotada de niños y adolescentes TDAH y controles\*

	Función 1	Función 2	Función 3
Fluidez verbal	0,81		
Inhibición	0,75		
Memoria de trabajo verbal		0,72	
Planificación		-0,62	
Flexibilidad		-0,43	
Memoria de trabajo visual			0,83

\*Se muestran sólo los coeficientes mayores de 0,3 en valor absoluto

En la gráfica 1 se muestran los sujetos representados en el espacio discriminante y los centroides de los cuatro grupos.



Gráfica 1. Sujetos de cada grupo en el espacio discriminante

En este análisis se predicen correctamente el 70% de los niños controles, el 69,6% de los niños TDAH, el 91,7% de los adolescentes controles y el 80% de los adolescentes TDAH. Globalmente, se clasifican correctamente el 75,4% de los casos (tabla 6).

Tabla 6. Resultados de la clasificación entre los grupos por edades

	Grupo de pertenencia pronosticado				Total
	N-C	N-TDAH	A-C	A-TDAH	
N-C	14 70,0%	4 20,0%	2 10,0%	0 0%	20 100%
N-TDAH	6 26,1%	16 69,6%	0 0%	1 4,3%	23 100%
A-C	0 0%	0 0%	11 91,7%	1 8,3%	12 100%
A-TDAH	0 0%	0 0%	2 20,0%	8 80,0%	10 100%
Total	20 30,77%	20 30,77%	15 23,08%	10 15,38%	65 100%

N-C: niños controles; N-TDAH: niños TDAH; A-C: adolescentes controles; A-TDAH: adolescentes TDAH

## Discusión

Este estudio pretende encontrar aquellas funciones ejecutivas que diferencien entre sujetos TDAH y sujetos sanos, dado la gran heterogeneidad de resultados que hay al respecto y las divergencias en el material de evaluación que se emplea. Además, debido a las diferencias madurativas en el desarrollo del funcionamiento ejecutivo en niños y adolescentes, se trata de identificar si esas funciones ejecutivas son capaces de discriminar entre niños y adolescentes TDAH y controles. El protocolo de evaluación empleado en esta investigación conformado por 6 pruebas que evalúan los dominios de memoria de trabajo verbal, memoria de trabajo visual, planificación, flexibilidad cognitiva, fluidez verbal e inhibición cognitiva constituye un buen discriminador entre sujetos con y sin patología, por encima de lo esperable al azar, pues clasifican correctamente al 78% de la muestra. Sólo 5 controles se clasifican como TDAH y 9 TDAH como controles, lo que puede ser debido a la inclusión en los dos grupos de niños y adolescentes, los cuales están en diferentes procesos madurativos del funcionamiento ejecutivo. Un sujeto que puntúe alto en la función (sujeto control) tiene mejor fluidez verbal, capacidad de inhibición, memoria de trabajo verbal y visual y comete menos errores perseverativos y de planificación. Estos resultados son compatibles con un déficit en el funcionamiento ejecutivo en TDAH (Barkley, 1997; Goldberg, 2002; Frazier et al., 2004; Willcut et al., 2005) como consecuencia de una disfunción de los circuitos dopaminérgicos y noradrenérgicos frontales (Himelstein, Schulz, Newcorn y Halperin, 2000).

En el análisis realizado diferenciando los grupos por edades, los resultados indican que se clasifica correctamente un 75,4% de los casos de los cuatro grupos, lo cual está en consonancia con el desarrollo ontogenético de las funciones ejecutivas. El rango de edad que se establece para el desarrollo de las funciones ejecutivas va aproximadamente de los 3 a los 12 años y tienen una especial relevancia para la maduración de las mismas el período comprendido entre los 6 y los 12 años, aunque diversos procesos ejecutivos no están totalmente establecidos hasta la adolescencia tardía o edad adulta temprana (Anderson, 2002). Las dos funciones discriminantes significativas indican que cuanto mayor es la fluidez verbal y la capacidad de inhibición, mayor es la puntuación en la primera función discriminante, y cuanto mayor es la puntuación en memoria de trabajo verbal y menor el número de errores perseverativos (mejor flexibilidad) y el total de movimientos en planificación, mayor es la puntuación en la segunda función discriminante.

Las valoraciones diagnósticas del TDAH basándose en escalas que cumplimentan padres y profesores (e.g., Conners, 1989) muestran un acuerdo medio entre informantes (Amador, Idiázabal, Sangorrín, Espadaler y Forns, 2002) lo cual hace necesario una valoración complementaria que confirme el diagnóstico. Es en este momento en el que la evaluación neuropsicológica de las funciones ejecutivas se conforma como una herramienta auxiliar con una sólida base teórica apoyada por los hallazgos neuroanatómicos que hablan de disfunción de los lóbulos frontales. El conjunto de pruebas empleadas constituye un protocolo con un buen poder discriminante del funcionamiento ejecutivo para servir a esta complementariedad diagnóstica. Sin embargo, no es suficiente esa capacidad discriminante, sino que el protocolo de evaluación debe tener en cuenta el desarrollo madurativo del sujeto, pues los niños y adolescentes con TDAH no comparten el mismo patrón de deterioro o no en la misma magnitud. En este sentido, la memoria de trabajo verbal y visual, la planificación, la flexibilidad cognitiva, la inhibición y la fluidez verbal medidas con las pruebas descritas tienen esa capacidad y constituyen dominios necesarios para una evaluación neuropsicológica del funcionamiento ejecutivo del TDAH en niños y adolescentes.

En definitiva, nuestros resultados ponen de manifiesto un déficit ejecutivo en los sujetos con TDAH y un funcionamiento diferencial según la edad. Asimismo, el protocolo de evaluación empleado se constituye como suficiente, debido al gran poder discriminante de las seis pruebas empleadas. Sin embargo, en futuros estudios se hace necesario la ampliación de la muestra, especialmente en los

adolescentes y la inclusión de otros grupos patológicos en los que se haya descrito problemas en el funcionamiento ejecutivo para su comparación con el TDAH.

### Referencias

- Amador, J. A., Idiázabal, M. A., Sangorrín, J., Espadaler, J. M. y Forn, M. (2002). Utilidad de las escalas de Conners para discriminar entre sujetos con y sin trastorno por déficit de atención con hiperactividad. *Psicothema*, 14 (2), 350-356.
- American Psychiatric Association (2003). *Manual Diagnóstico y Estadístico de los Trastornos Mentales* 4ª Edición Revisada (DSM-IV-TR). Barcelona: Masson.
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8 (2), 72-82.
- Arnsten A. F. T., Steere, J. C. y Hunt, R.D. (1996). The contribution of  $\alpha$ 2 noradrenergic mechanisms to prefrontal cortical cognitive functions: potential significance to attention-deficit hyperactivity disorder. *Archives of General Psychiatry*, 53, 448-455.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin*, 121, 65-94.
- Berg, E. A. (1948). A simple objective treatment for measuring flexibility in thinking. *Journal of General Psychology*, 39, 15-22.
- Bradley J. D. D. y Golden C. J. (2001). Biological contributions to the presentation and understanding of Attention-Deficit/Hyperactivity Disorder: a review. *Clinical Psychology Review*, 21 (6), 907-929.
- Conners, C.K. (1989). *Conners' Rating Scales*. Toronto: Multi-Health Systems.
- Doyle, A. E. (2006). Executive functions in Attention-Deficit/Hyperactivity Disorder. *Journal of Clinical Psychiatry*, 67 (suppl 8), 21-26.
- Frazier, T. W., Demaree, H. A. y Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, 18 (3), 543-555.
- Goldberg, E. (2002). *El Cerebro Ejecutivo*. Madrid: Ed. Crítica.
- Hechtman, L. (2003). Evolución del trastorno por déficit de atención con hiperactividad en subgrupos de adultos. En T. E. Brown (Ed.) *Trastornos por déficit de atención y comorbilidades en niños, adolescentes y adultos*. Barcelona: Masson.
- Himelstein, J., Schulz, K. P., Newcorn, J. H. y Halperin J. M. (2000). The neurobiology of attention deficit hyperactivity disorder. *Frontiers in Bioscience*, 5, 461-478.
- Lezak, M. D., Howieson, D. B. y Loring, D. W. (2004). *Neuropsychological Assessment*. Nueva York: Oxford University Press.

- Pliszka, S.R., McCracken, J.T. y Maas, J.W. (1996). Catecholamines in attention-deficit hyperactivity disorder: current perspectives. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 264-272.
- Quintero, I., Bravo, J., Martín, R., González-Pérez, P. A., Verche, E. y Hernández, S. (2009). Discrepancies between parents and teachers rating scales in the evaluation of Attention-Deficit/Hyperactivity Disorder: relationship with objective assessment of attention. *ADHD Attention Deficit and Hyperactivity Disorders*, 1 (1), 95.
- Sergeant, J. A., Geurts, H. y Oosterlaan, J. (2002). How specific is a deficit of executive functioning for Attention-Deficit/Hyperactivity Disorder? *Behavioural Brain Research*, 130, 3-28.
- Siegel, L. y Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60, 973-980.
- Silver, L. B. (2004). *Trastorno por déficit de atención con hiperactividad, guía clínica de diagnóstico y tratamiento para profesionales de la salud*. Barcelona: Psiquiatría Editores.
- Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, 7, 268-288.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Tranel, D., Anderson S. W. y Benton, A. L. (1994). Development of the concept of “executive function” and its relationship to the frontal lobes. En F. Boller y J. Grafman (Eds.), *Handbook of neuropsychology*, pp. 125–148. Amsterdam: Elsevier.
- Wechsler, D. (1997). *Wechsler Memory Scale Third Edition*. San Antonio: The Psychological Corporation.
- Wechsler, D. (1999). *Escala de Inteligencia de Wechsler para Adultos*. 3ª Edición (WAIS-III). Madrid: TEA Ediciones.
- Wechsler, D. (2005). *Escala de Inteligencia de Wechsler para Niños*. 4ª Edición (WISC-IV). Madrid: TEA Ediciones.
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V. y Pennington, B. F. (2005). Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biological Psychiatry*, 57, 1336-1346.

# MODELADO ESTADÍSTICO DE LA LATENCIA DEL P300 EN MEMORIA EXPLÍCITA E IMPLÍCITA

Agustín Wallace, Manuel Pelegrina, Álvaro Pelegrina y Rafael Marín  
Universidad de Málaga

En esta investigación mostramos la utilidad del enfoque del modelado estadístico para el estudio de procesos cognitivos en el contexto de la memoria de reconocimiento de ítems. Hemos utilizado el paradigma de la detección de señales y hemos registrado la latencia del P300. Para el análisis de los datos obtenidos utilizamos el modelado estadístico mediante el programa *R* con el cual hemos evaluado el ajuste de diferentes modelos. Esto nos permite inferir que hay diferencias entre ambos tipos de memoria, así como procesos diferenciados en los componentes de cada uno de los tipos de memoria analizados.

En este trabajo exponemos la utilidad del Modelado Estadístico (v.g.: Dobson, 1983; Everitt, y Hothorn, 2006; Lindsey, 1997; Grupo ModEst, 2000a y 2000b; McCullagh y Nelder, 1989) aplicado al ámbito de la neurociencia (Gross, 1999) mediante potenciales evocados y el paradigma de la teoría de detección de señales (TDS) (Tanner y Swets, 1954).

Para ello nos hemos centrado en la evaluación de la memoria explícita e implícita, manipulando los estímulos y las condiciones experimentales para contrastar las diferencias entre ambos tipos de memoria (Baddeley, 1999).

La tarea que nos planteamos consistió en el estudio del resultado obtenido en el rendimiento de un observador al valorar una señal. El procedimiento más sencillo consiste en establecer juicios de valor, basados en respuestas SI- NO. Desde la TDS partimos del supuesto de que el sujeto elige un valor como criterio de respuesta basado en la máxima verosimilitud. En nuestra investigación, los valores de sensación de la señal sobre los valores de sensación del ruido en el ámbito de la memoria de palabras, ver tabla 1 (Egan, 1958).

Tabla 1

Variable de fila (señal y ruido)	Variable de columna (respuestas)		
	Si ( $b_1$ )	No ( $b_2$ )	Total filas
Señales ( $a_1$ )	$b_1/a_1$ A	$b_2/a_1$ B	Total fila A+B
Ruidos ( $a_2$ )	$b_1/a_2$ C	$b_2/a_2$ D	Total fila C+D
Total columnas	Total columna A+C	Total columna B+D	Total N= A+B+C+D

Por otra parte, la aplicación del modelado estadístico nos permite obtener el modelo mejor ajustado de los datos, utilizando variables cualitativas y cuantitativas. Una de las ventajas del modelado es su capacidad para superar limitaciones importantes de los métodos de análisis más extendidos como la independencia, la homogeneidad y la distribución de los datos (Agresti, 2002). Pero también, la posibilidad de modelar los datos obtenidos, y no solo, la obtención de diferencias significativas o no entre las distintas condiciones experimentales.

Mediante el modelado estadístico tratamos de conocer la diferencia entre los valores empíricos y los valores ajustados (Agresti, 2002; Ato y López, 1996). Así pues, los datos obtenidos en nuestras investigaciones nos permiten realizar un “análisis adaptado a los datos”. De esta manera, superamos las limitaciones de las pruebas de hipótesis estadísticas (como hemos comentado antes) que solo se refieren al contraste; y no a la representación global de los datos empíricos. Este es un aspecto a tener muy en cuenta debido a la naturaleza de los datos obtenidos. En efecto, a menudo trabajamos por ejemplo con ruido que no se distribuye según los supuestos, condiciones de aplicación o requisitos de la estadística clásica. Este es el caso, entre otros muchos, de los datos registrados en potenciales evocados en particular y en neurociencia en general.

Los datos de este estudio proceden de una serie de estímulos sensoriales que proponemos al participante y que elicitán respuestas electrofisiológicas en el cerebro, provocando secuencias características de ondas en el trazado electroencefalográfico (EEG) (Hillyard y Kutas, 1983). Estas ondas son diferentes para cada modalidad sensorial y su variabilidad depende del tipo de estímulo.

En nuestra aplicación utilizamos los potenciales relacionados con eventos (ERP en inglés). Consideramos el tiempo que tarda en activarse un P300 como variable de respuesta ante los estímulos, la señal y el ruido durante diferentes sesiones de registro.

Mediante estas técnicas podemos aislar y registrar áreas del cerebro específicas. En la práctica, la forma de onda de la P300 se debe evocar utilizando un estímulo provisto por alguna de las modalidades sensoriales. El procedimiento típico es el paradigma *odd-ball*, en el que un estímulo diana se presenta entre estímulos de fondo más frecuentes. También puede usarse un estímulo distractor para asegurarse de que la respuesta se debe a la diana en lugar de a un cambio en el patrón de fondo o para manipular la atención. El paradigma *odd-ball* clásico ha sufrido muchas variaciones, pero al final la mayor parte de los protocolos utilizados para evocar la P300 implica alguna forma de hacer consciente una situación o una toma de decisiones. Es en este contexto en el que se introduce el paradigma de la TDS.

La investigación científica a menudo descansa en la medición de la P300 para examinar potenciales relativos a eventos, especialmente cuando se trata de la toma de decisiones, ya que existe una relación bastante directa entre la reducción de la incertidumbre del sujeto y la amplitud del P300.

Por otra parte el deterioro cognitivo a veces está relacionado con modificaciones en el P300, de manera que se puede utilizar como medida para detectar la eficiencia de la respuesta en la evaluación de diferentes procesos cognitivos. Así es que hay un amplio rango de usos del P300 que va, por ejemplo, desde el estudio de la depresión y la drogadicción, a los trastornos de ansiedad, o los procesos cognitivos, como la atención, percepción y memoria.

Las hipótesis que queremos contrastar aquí se refieren a la constatación de las diferencias entre memoria implícita y explícita, añadiendo a ello la obtención del modelo mejor ajustado, y determinando además en qué medida contribuyen a ello las diferentes condiciones experimentales y variables o categorías del experimento.

## Método

### *Participantes*

Realizamos un estudio piloto con un estudiante especialmente entrenado que participó durante 8 sesiones de 20 minutos cada sesión.

### *Material*

En primer lugar obtuvimos una muestra de 40 palabras a partir de su frecuencia de uso y clase. Las palabras fueron presentadas mediante el programa *E-prime* (Schneider, Eshman y Zucolotto, 2002). El registro se realizó con un equipo de potenciales evocados de 72 canales. Se utilizaron 11: 3 frontales (F3, FZ, F4), 3 centrales (C3, CZ, C4), 3 parietales (P3, PZ, P4), y 2 occipitales (O1, O2)

### *Procedimiento*

Se aplicó el paradigma de la TDS en memoria de reconocimiento de ítems. La mitad de las palabras, 20 para cada condición, fueron consideradas señal y otras 20 ruido. En la condición de memoria implícita el participante realizaba una tarea distractora al observar la primera lista de palabras. Esta tarea consistía en evaluar la simetría de dos números colocados encima y debajo de la palabra a recordar. Además el participante no era informado de que estaba ante una prueba de memoria. En la condición de memoria explícita aparecía la palabra a memorizar y el participante era informado de que debía someterse posteriormente a una prueba de memoria. La aparición de cada palabra estaba precedida por un punto negro en la pantalla durante 500 ms. El participante respondía mediante respuestas SI-NO en caso de que la palabra perteneciera o no a la lista original. Se realizaron 8 sesiones durante 20 minutos por sesión, registrando los datos del EEG en el programa *Recorder* (de Brainvision). La obtención del P300 se realizó mediante el programa *Analyzer* (de Brainvision). Ambos programas son comercializados para tareas de investigación por BRAIN PRODUCTS (BRAINPRODUCTS.COM )

## Resultados

De los resultados obtenidos en el registro de potenciales evocados se seleccionaron para el análisis los pertenecientes a los electrodos frontales (F3, FZ, F4). Para el ajuste de estos datos utilizamos el programa “*R*” (Crawley, 2007). El modelo mejor ajustado resultó ser un modelo lineal de interacción entre las variables: tipo de memoria, señal-ruido y sesión (ver cuadro 1)

Cuadro 1

```

> summary (M13)

Call:
lm(formula = p ~ 1 + m * sr * s)

Residuals:
    Min       1Q   Median       3Q      Max
-29.667 -11.167   1.000   7.333  51.333

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -167.67     129.06  -1.299  0.21230
m              274.67      81.62   3.365  0.00394 **
sr            252.00      81.62   3.087  0.00706 **
s             191.33      81.62   2.344  0.03231 *
m:sr         -147.67      51.62  -2.860  0.01133 *
m:s          -110.67      51.62  -2.144  0.04777 *
sr:s          -96.33      51.62  -1.866  0.08047 .
m:sr:s         53.00      32.65   1.623  0.12406
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.99 on 16 degrees of freedom
Multiple R-squared:  0.6173,    Adjusted R-squared:  0.4499
F-statistic: 3.687 on 7 and 16 DF,  p-value: 0.01453

> deviance (M13)
[1] 6396
> AIC (M13)
[1] 220.1580

```

## Discusión

En el cuadro 1 observamos la salida del programa R. El modelo que mejor hemos ajustado es un modelo lineal de interacción que incluye las variables m (tipo de memoria), la variable sr (señal-ruido), y la variable s (sesión), es decir el modelo m:sr:s. Este modelo incluye m que resulta de la comparación de los dos tipos de memoria y que es el resultados más significativo desde el punto de vista estadístico. Pero también observamos otros datos relevantes, por ejemplo la significación entre la señal y el ruido sr como dos elementos distintos del experimento. De las misma forma hubo diferencias estadísticamente significativas a lo largo de las diferentes sesiones s. También resultó estadísticamente significativa el ajuste de los modelos m:sr y m:s. Es decir, que los tipos de memoria, además de ser estadísticamente significativos entre sí, se comportan de forma diferente y varían de forma estadísticamente significativa en función de la señal-ruido y a lo largo de las 8 sesiones.

Considerando los resultados obtenidos tras el ajuste de los modelos, y teniendo en cuenta que estamos trabajando con un modelo exploratorio, podríamos decir en primer lugar que ambos tipos de memoria son diferentes, y en segundo lugar que los sujetos harían uso de un tipo u otro de memoria en función de los estímulos presentados y de las sesiones. Por lo que concluimos que la memoria, al ser la más significativa de todas las comparaciones incluye una variabilidad propia y sólo relativamente independiente de la señal-ruido y de las sesiones. En efecto, éstas últimas variables incorporan una variabilidad propia asociada a la memoria, pero con entidad propia y diferenciada, como vemos, ya que la significación estadística, a pesar de que existe, es menor a la comparación específica de tipos de memoria.

Todo ello aporta un modelo estructural general (saturado y articulado por partes), de acuerdo con unas variables fundamentales: memoria explícita e implícita; así como unos modelos parciales que permiten evaluar todos los efectos relevantes que se detectan en la respuesta de los participantes a partir de las variables estímulo y de las sesiones de la investigación, en la línea seguida por DeCarlo (1998) y Wallace, Pelegrina, Ruiz-Soler (2002) en el ámbito de la memoria, la detección de señales y el modelado estadístico. Además, en el presente trabajo hemos incorporado los potenciales evocados en el mismo contexto.

### Referencias

- Agresti, A. (2002). *Categorical data analysis* (2ª Ed.). New Jersey: Wiley-Interscience.
- Ato, M. y López, J.J. (1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis Psicología.
- Baddeley, A.D. (1982). *Psicología de la memoria*. Madrid: Editorial Debate.
- Crawley, M.J. (2007). *The R Book*. New York: John Wiley.
- DeCarlo, L.T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, 3, 186-205.
- Dobson, A.J. (1983). *Introduction to Statistical Modelling*. London: Chapman and Hall.
- Everitt, B.S. & Hothorn, T. (2006). *A Handbook of Statistical Analyses using R*. Boca Raton: Chapman and Hall.
- Gross, C.G. (1999). *Brain, Vision, Memory: Tales in the History of Neuroscience*. Cambridge, Mass.: The MIT Press.
- Grupo ModEst (2000a). *Análisis de Datos - Del Contraste de Hipótesis al Modelado Estadístico*. Tarrasa: CBS.
- Grupo ModEst (2000b). *Análisis de Datos - Modelo Lineal Generalizado*. Tarrasa: CBS.
- Lindsey, J.K. (1997). *Applying Generalized Linear Models*. New York: Springer-Verlag.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models* (2ª Ed.). London: Chapman and Hall.
- Wallace, A., Pelegrina, M. y Ruiz-Soler, M. (2002). Aplicación de los Modelos Lineales Generalizados en la Evaluación de Modelos sobre Memoria de Reconocimiento. *Metodología de las Ciencias del Comportamiento*. VE, 570-572.

## ÍNDICE DE AUTORES

Abad, F. J.	307	Castillo, M.	112, 136
Abal, F. J. P.	503	Chacón, J.	394
Acosta, M. V.	567	Chica, E.	143
Aguerri, M. E.	503	Chun, J.	516
Alarcón, R.	429, 436	Cocca, A.	469, 522
Algarañá, S.	290	Colodro, L.	588
Allueva, P.	245	Cortés, A.	530
Alonso, R.	536	Coscolluela, A.	423
Andrade, E.	130	Delgado, E.	107
Andrés, A.	297	Díaz-Loving, R.	136
Anguera, M. T.	224, 245	Dios, H.	417
Antivilo, A.	307	Escobar, M.	334, 340
Aranburu, M.	70	Espinet, A.	536
Arce, C.	130, 313	Estivill, G.	516
Arce, I.	130, 313	Estrada, L. I.	401
Arch, M.	176	Estudillo, A. J.	540
Ardá, T.	229	Extremera, N.	368
Aritzeta, A.	93, 101	Fàbregues, S.	374
Arnau, J.	17, 18, 24, 30, 56	Feixas, G.	183
Ato, M.	213	Fernández Berrocal, P.	101, 368, 380
Attorresi, H. F.	503	Fernández Borrero, M. A.	350
Ballestar, M. L.	594	Fernández García, P.	24, 45, 46, 51, 193, 199, 208, 213, 219, 550
Ballester, M. G.	483	Fernández-Baena, F. J.	334, 340
Balluerka, N.	93, 100, 101, 122, 123	Flores-Galaz, M. M.	136
Barea, R. C.	507	Franchi, L.	491
Bechger, T. M.	330	Francisco, C. De	130, 313
Benítez, I.	112, 117, 136, 319	Freixa, M.	169, 573
Benítez, S.	176	Frontera, M.	245
Blanca, M. J.	334, 340, 429, 436	Fuente, E. I. de la	76,83, 88
Blanco, A.	142	Fuente, L. de la	407, 577
Bono, R.	18, 24, 30, 56	Fuente, V. de la	394
Bravo, J.	602	Galibert, M. S.	503
Brutti, C.	290	Gálvez, P.	154
Cabrera, E.	507	García García, C.	344
Canales, J. J.	34	García García, J.	407, 577
Cano, C. J.	577	García Orza, J.	540
Cantos, S.	340	García Pérez, J. F.	204, 411
Cañadas, G. R.	76, 83, 88	García Soidán, J. L.	544
Caramés, J. M.	536	García Soidán, P.	417, 544
Carmona, J.	362	García-Vega, E.	550
Carrera, M. J.	183	Garrido, J.	130, 313
Carretero-Dios, H.	107	Garrido, M. R.	350
Carrillo, E.	588	Gea, A. B.	577
Casal, C. A.	229	Gómez-Benito, J.	117, 136, 297, 319
Castilla, E.	516	González Galán, A.	536
Castillo Fuentes, M. del	204, 411		

González-Javier, F.	588	Matas, A.	401
González-Pérez, P. A.	602	Matas, E.	516
Gordovil, A.	162	Mena, E.	441
Gorostiaga, A.	123	Menéndez-Robledo, E.	550
Gracia, E.	204, 411	Menéndez, A.	550
Guàrdia, J.	160, 162, 169, 176, 183	Miranda, J.	429, 436, 530
Guillén, M.	267	Miranda, M. T.	469
Gutiérrez, J.	357	Mohamed, K.	469, 522, 582
Hermosilla, D.	70	Molina, J. F.	262
Hernández Mendo, A.	143, 257	Morales, V.	142, 143, 149, 154
Hernández Sánchez, S.	602	Murgui, S.	204
Hernández-Lloreda, M.J.	394	Olea, J.	297
Herrero, F.	61, 556	Oliver, J. C.	447
Herrero, M.	245	Ordoñana, J. R.	588
Hidalgo, M. D.	117	Ortiz, M.	267
Honrubia, M. L.	423, 573	Padilla, J. L.	100, 112, 117, 122, 136, 319
Hoyo, C.	516	Padrón, H.	454
Hurtado, S.	562	Palacios, B.	387
Isasi, X.	123	Paré, M. H.	239
Jara, P.	61, 556	Pedraza, C.	516
Jarne, A.	176	Pelegrina, A.	612
Lila, M. S.	411	Pelegrina, M.	612
Livacic-Rojas, P.	46, 51, 193, 199, 208, 219	Pérez Fernández, T. A.	357
Llopis, J. M.	567	Pérez Moreno, P. J.	350
López Fernández, E.	253	Pérez-Fuster, P.	594
López Fernández, O.	224, 253, 262, 423, 573	Pérez-Gómez, L.	362
López Puga, J.	577	Peró, M.	169
López-Cuadrado, J.	357	Pino, J. L.	507
López-Montiel, D.	429, 436	Ponsoda, V.	344
López-Montiel, G.	429, 436	Quero, A.	536
Lozano, L.	582	Quintero, I.	602
Lozano, L. M.	76, 83, 88	Rebollo-Mesa, I.	588
Lozzia, G. S.	503	Rodrigo, M. F.	594
Machado, A.	602	Rodríguez de Fonseca, F.	516
Malapeira, J. M.	423	Rodríguez-Naveiras E.	273
Maldonado, E. F.	34	Romero, R. P.	149
Manolov, R.	30, 56	Rosel, J. F.	34, 61
Maria, L.	256	Ruiz Gallego-Largo, T.	394
Marín, R.	612	Ruiz Soler, M.	454, 460, 464
Maris, G.	330	Saavedra, F. J.	285
Márquez, Z.	257	Salazar, P.	447
Martín Salguero, J.	101	Saldaña, C.	297
Martín-Baranera, M.	319	Salinas, F.	469, 522, 582
Martín, M.	83	Salvador, M.	556
Martín, R.	602	Sánchez Garzón, V.	88
Martínez Álvarez, J.C.	582	Sánchez López, J.	516
Martínez-Arias, M. R.	394	Sánchez Prada, A.	387
Martínez-Selva, J. M.	588	Sánchez Ramos, J. L.	362
		Sánchez Santamaría, J.	475, 483
		Sánchez- Romera, J. F.	588

Sánchez-García, M.	368, 374
Sánchez-Iglesias, I.	380
Sandoval, A.	491
Sanmartín, J.	594
Santín L. J.	516
Serradilla, F.	394
Sierra, A.	344
Solanas, A.	30, 56
Solís, F. M.	507
Sueiro, M. J.	380
Szabolcs, B.	582
Tejero, P.	319
Tenutto, M. A.	290, 498
Tójar, J. C.	441
Torrado, J.	130,313
Torrente, P.	34
Torres, E.	68, 70
Trianes, M. V.	334, 340, 530
Tuero, E.	46, 51, 193, 199, 213, 219
Turbany, J.	169
Vadillo, J. A.	357
Valenzuela, B.	267
Vallejo, G.	18, 24, 46, 51, 93, 199, 208, 213, 219
Vaquero, T.	536
Vargas, C.	76
Vega, M. De	350
Verche, E.	602
Verdú, F.	556
Viader, M.	423
Viciana, J.	469, 522, 582
Villalba, E.	350
Wallace, A.	464, 612
Zambrana, E.	516

