



Artículo de investigación

Batería de Rendimiento Intelectual para niños escolares cubanos

Battery Intellectual Performance for Cuban children

Yaser Ramírez-Benítez^{1*}, Rosario Torres-Díaz² y Valeska Amor-Díaz²

1 Universidad de Cienfuegos “Carlos Rafael Rodríguez”. Cienfuegos, Cuba.

2 Centro de Neurociencias de Cuba (CNEURO). Cubanacán, La Habana, Cuba.

Resumen

Introducción. En Cuba, no existen instrumentos propios para la evaluación de la inteligencia. Para solucionar el problema, el Centro de Neurociencias de Cuba, y en especial, el Departamento de Neurociencias Educativas, diseñó una Batería de Rendimiento Intelectual (BRI) para niños escolares cubanos (conformado por Escala capacidad Intelectual, ECI y Analogías verbales, AV). **Objetivos:** (1) Completar y validar la BRI para la evaluación de Gf y Gc en niños cubanos de edad escolar, (2) **Método.** Determinar la contribución única de Gf y Gc, evaluada con la BRI, en el aprovechamiento académico en niños cubanos. Se aplicaron dos estudios. a) Validación de la ECI con una muestra de 370 niños entre 7 y 11 años (N1 – 270 para calcular punto de corte y N2 para validar los puntos de corte). b) Diseño de la AV con una muestra de 36 niños entre 10 y 11 años. **Resultados.** La ECI resulta una prueba apropiada para detectar al niño escolar con riesgo intelectual. El resultado se sustenta en la alta coincidencia (89%) con el WISC-R para clasificar niños con trastornos intelectuales. El diseño de la AV presenta apropiadas propiedades psicométricas, pues presenta significativas correlaciones con la sub-prueba de Vocabulario del WISC-R. **Conclusiones.** La BRI presenta apropiadas propiedades psicométricas para detectar el riesgo intelectual en niños cubanos.

Palabras clave: inteligencia fluida, inteligencia cristalizada, evaluación, aprendizaje

Abstract

Introduction. In Cuba not have own instruments for the assessment of intelligence. To resolve the problem, the Neuroscience Center of Cuba, and the Department of Educational Neuroscience, design a Battery of Intellectual Performance (BIP) for school children (Intellectual Capacity Scale, ICE and Verbal Analogy, VA). **Objectives:** (1) to complete and validate the BIP for the evaluation of Gf and Gc in Cuban children of school age, (2) **Method.** Determine the unique contribution of Gf and Gc, evaluated with the BIP, on academic achievement in Cuban children. Two studies were applied. a) Validation of the ICE with a sample of 370 children between 7 and 11 years (N1 - 270 to calculate the cut-off point and N2 to validate the cut-off points). b) Design of the VA with a sample of 36 children between the ages of 10 and 11 years. **Results.** The ICE is a proper test to detect the child with intellectual risk. The result is based on the high coincidence (89%) with the WISC-R to classify children with intellectual disorders. The design of the VA presents appropriate psychometric properties, because it presents significant correlations with the sub-vocabulary test of the WISC-R. **Conclusions.** The BIP presents appropriate psychometric properties to detect risk Cuban intellectual in school children.

Keywords: fluid intelligence, crystallized intelligence, evaluation, learning

Introducción

La evaluación de la inteligencia en la edad escolar tiene especial relevancia dada a su estrecha relación con el desempeño académico y el aprendizaje en general. El uso de instrumentos apropiados para su medición permite detectar problemas intelectuales leves en niños que cursan la enseñanza regular. Así, la atención temprana a estos niños contribuye a prevenir problemas de aprendizaje y deserción escolar futura.

Una de las teorías de la inteligencia más conocidas de todos los tiempos es la teoría del factor g de Spearman (1904); y plantea que la inteligencia puede ser explicada en función de un factor de capacidad mental general “g” y un gran número de factores específicos “s”. El factor g es una habilidad fundamental que interviene en todas las operaciones mentales, representa la energía mental y se moviliza en tareas no automatizadas. Las tareas cognitivas más cargadas de g son aquellas que exigen razonamiento deductivo o inductivo, visualización espacial, razonamiento cuantitativo, y razonamiento verbal; y demandan exigencias mínimas de conocimiento especializado.

Otra de las teorías de la inteligencia de más amplia aceptación dentro de la comunidad científica es la teoría bi-factorial de Cattell & Horn (Cattell, 1971, Horn, 1991). En este caso, se explica el rendimiento intelectual a partir de dos tipos de inteligencias: inteligencia fluida (Gf) e inteligencia cristalizada (Gc). Gf se define por el uso intencionado de diversas operaciones mentales en la resolución de problemas nuevos, incluye la formación de conceptos e inferencias, clasificación, generación y evaluación de hipótesis, identificación de relaciones, comprensión de implicaciones, extrapolación y transformación de información (McGrew, 2009). En cambio, Gc se refiere a la riqueza, amplitud y profundidad del conocimiento adquirido en una cultura dada (Cattell, 1963, 1971). Según Ackerman (1996), Gf y Gc son dos capacidades generales (inteligencia como proceso vs inteligencia como conocimiento), que están involucradas en el funcionamiento cognitivo.

Se plantea que Gf constituye la base de Gc, en tanto hace posible la adquisición de habilidades y conocimientos (Cattell, 1971). La relación entre Gf y Gc es explicada por Cattell a través de la teoría de la inversión. Esta teoría postula que la capacidad máxima de aprendizaje de un niño depende de los niveles de Gf, siempre y cuando las tareas demanden la comprensión

* Correspondencia: Yaser Ramírez Benítez, Carretera Central, Cuatro Caminos, Cienfuegos, Cuba. Tel. 43 549 608. Correo Electrónico: yramirezbenitez@gmail.com.

de relaciones complejas, tales como la lectura, la aritmética y el razonamiento abstracto (Cattell, 1971). De este modo, los niveles de Gf se “invierten” en el proceso de aprendizaje para adquirir habilidades y conocimientos (Gc), esencialmente en los inicios de la etapa escolar. En este sentido, la teoría establece que Gf es más importante que Gc en los primeros años escolares, pues garantiza la adquisición de conocimientos, en tanto, Gc muestra un impacto más tardío en el desarrollo (Horn, 1991).

Estudios sobre la relación entre estos dos tipos de inteligencia y el rendimiento académico han aportado evidencia empírica acerca de su importancia en el aprendizaje escolar general desde edades tempranas (Ritchie, Bate & Plomin, 2015; Stevenson et al, 2014; Pina et al, 2014; Van Bergen et al, 2013; Vukovic & Lesaux, 2013; Tong & Fu, 2013; Archibald et al, 2013). Por ejemplo, Van Bergen y colaboradores (2013), en un estudio longitudinal, evaluaron ambos tipos de inteligencias en un grupo de niños de 4 años. Posteriormente, midieron la habilidad lectora y matemática de estos niños a los 8 años. Sus resultados mostraron que Gc constituye un predictor significativo del desarrollo lector futuro de los niños; y que Gf lo es tanto del rendimiento lector como matemático. Otros autores, sin embargo, han encontrado evidencia de la relevancia de Gc también para el desarrollo de habilidades aritméticas (Pina, et al, 2014)

En la actualidad, las pruebas de más amplio uso a nivel mundial incluyen la evaluación de estas dos inteligencias. Algunos ejemplos son la Batería de Inteligencia Woodcock–Johnson III (Woodcock, McGrew, & Mather, 2001), la Escala de Inteligencia de Stanford–Binet (Roid, 2003), la Batería de Inteligencia de Kaufman para niños (Kaufman & Kaufman, 2004) y la Escala de Inteligencia de Wechsler para niños (Wechsler, 2003).

El uso de las pruebas de inteligencia, a pesar de su amplia y probada utilidad práctica, ha generado muchas críticas a lo largo de los años (Sternberg, et al, 2011). Desde el punto de vista teórico, se ha señalado que las tareas verbales están sesgadas, lo que afecta el resultado de grupos socialmente menos favorecidos por la educación (Gómez-Benito, 2010). Una solución a esta limitación ha sido realizar adaptaciones de las pruebas teniendo en cuenta las características propias del lenguaje y la cultura de los lugares donde van a ser usadas, pero estos estudios resultan sumamente costosos y muchos países pobres no pueden costearlos. En cambio, los tests no verbales no tienen este sesgo cultural pues no utilizan vocabulario ni miden conocimientos aprendidos; por lo que resultan más útiles para evaluar niños de zonas rurales apartadas y niños preescolares que no han completado la adquisición del código de la lecto-escritura (Flanagan y Kaufman, 2009).

Por otro lado, algunos problemas prácticos de la evaluación de la inteligencia están relacionados con el tiempo de aplicación de los tests, la complejidad del procedimiento y el pago de los derechos de autor por el uso de las pruebas. Para dar solución a estos problemas, algunos investigadores de habla hispana han diseñado pruebas de inteligencia, tanto para emitir criterios diagnósticos como pesquisar el riesgo intelectual en niños y adolescentes (Sánchez – Sánchez & Arriba, 2014; Sánchez, Santamaría & Abad, 2015; Riveros, Sepúlveda, Figueroa & Rosas, 2015)

En el caso de Cuba, no contamos con instrumentos propios para la evaluación de la inteligencia; y los que tenemos disponibles provienen de contextos culturales alejados del nuestro, lo que afecta la interpretación de los resultados obtenidos, como se ha comentado antes. Para dar solución a estos problemas, el Centro de Neurociencias de Cuba, y en especial, el Departamento de Neurociencias Educativas, se encuentra desarrollando un paquete tecnológico orientado a la evaluación temprana de los escolares y con ello a la protección del neurodesarrollo en estas edades. Así, se requiere el desarrollo de pruebas que permitan la evaluación válida y confiable de la capacidad intelectual de niños de edad escolar. El departamento desarrolló en el año 2011, la Batería de Rendimiento Intelectual compuesta por dos instrumentos que permiten detectar aquellos niños en riesgo de tener problemas intelectuales: la Escala de Capacidad Intelectual (ECI) orientada a detectar riesgo de problemas en la inteligencia fluida y la prueba Analogías Verbales (AV), en la inteligencia cristalizada. En general, los resultados iniciales de la Batería de Rendimiento Intelectual indicaron adecuadas propiedades psicométricas en su fase de prueba.

Sin embargo, los estudios que nos sirven de antecedente tienen algunas limitaciones de orden práctico y metodológico.

1. El tamaño de la muestra por edades fue pequeño.
2. Los niños con CI bajo estaban pobremente representados en la muestra; por lo cual la sensibilidad y la especificidad estaban distorsionadas.
3. La prueba empleada como regla de oro era una prueba no verbal de inteligencia fluida, por lo que no disponíamos de medidas de inteligencia verbal.

4. La prueba de inteligencia cristalizada (Analogías Verbales) resultó psicométricamente apropiada para evaluar los niños más pequeños, pero demasiado fácil para niños mayores de 10 años.

El presente estudio es la continuación de la investigación iniciada en 2011 y pretende aportar nuevos datos que permitan superar total o parcialmente estas limitaciones; y se propone el siguiente objetivo general:

1. Completar y validar la Batería de Rendimiento Intelectual para la evaluación de la inteligencia fluida y cristalizada en niños cubanos de edad escolar.

Y los siguientes objetivos específicos:

1. Validar la Escala de Capacidad Intelectual utilizando como referencia el WISC abreviado, de reconocida utilidad en esta área.
2. Diseñar y probar un instrumento para la evaluación de la inteligencia cristalizada (Analogías Verbales, AV) para niños de 5to y 6to grado, con edades entre 10 y 11 años.

Metodología

Para responder a los objetivos de la investigación, se realizaron dos estudios con muestras independientes.

Estudio I

1.1 Participantes

En el estudio se incluyeron 370 niños escolares desde 2do a 6to grado, con edades comprendidas entre 7 y 11 años (238 varones), provenientes de varias escuelas regulares y especiales de las provincias de La Habana, Matanzas y Cienfuegos. La muestra total se dividió en dos sub-muestras independientes N1 y N2. La muestra N1 constituye el grupo de referencia y por consiguiente se seleccionó en la enseñanza regular. Con esta muestra se calculó el punto de corte por edades. La muestra N2 se seleccionó de ambas enseñanzas (regular y especial) para validar los puntos de corte.

Muestra N1: En la muestra N1 se incluyeron 270 escolares (167 niños). Aunque no se empleó un método de estratificación por provincias, que hubiera sido deseable, se tomaron sujetos de tres provincias del país: La Habana (68%), Matanzas (23%) y Cienfuegos (7%). La descripción de la muestra se observa en la Tabla 1. Se seleccionaron al azar dos aulas de cada grado escolar evaluado.

Muestra N2: En la muestra N2 se incluyeron 100 escolares (71 niños) provenientes de escuelas regulares y especiales de Matanzas (49%) y Cienfuegos (51%). Esta se describe detalladamente en la Tabla 2. Se incluyen ambos tipos de enseñanza para enriquecer la muestra y tener una representación de niños con coeficiente intelectual normal, niños con CI bajo y niños con discapacidad intelectual leve. En la enseñanza regular, se seleccionaron niños con evaluación de mal y regular en las materias escolares fundamentales de lectura, escritura y matemática según el criterio de los maestros, puesto que estos niños presentan mayor riesgo de tener deficiencias intelectuales. En la enseñanza especial se seleccionaron niños con diagnóstico de discapacidad intelectual leve.

1.2 Instrumentos

Escala de Capacidad Intelectual (ECI): Prueba diseñada para la evaluación de la inteligencia fluida en niños cubanos de edades comprendidas entre los 7 y 11 años (Amor, 2011). Está constituida por 36 problemas con un diseño similar al Test de Matrices Progresivas Coloreadas de Raven (Raven, Court & Raven, 1993). Cada uno de ellos consiste en una figura incompleta que el niño debe completar, seleccionando una de seis alternativas posibles. Por cada respuesta correcta se otorga un punto y se recoge una puntuación total en la ejecución de la prueba (36 puntos).

Escala de Inteligencia de Wechsler para niños (WISC-R): Forma abreviada del WISC-R (Brooker y Cyt, 1986) que permite obtener el CI general del niño, a través de la aplicación de dos sub-tests. Por la escala ejecutiva se presenta el subtest de *Diseño de Cubos*; y por la escala verbal, el sub-test de *Vocabulario*.

La clasificación de las puntuaciones del WISC-R se realizó según los criterios de Flanagan y Kaufman (2009): las puntuaciones < 85 puntos (alteración intelectual) y puntuaciones ≥ 85 puntos (sin alteración intelectual).

Tabla 1.
Descripción de la muestra N1

Grado	Total	Provincias			Género		Edad	
		Habana	Cienfuegos	Matanzas	F	M	Media	DE
2do	51	32	4	15	21	30	7,2	0,25
3ro	55	38	3	14	14	41	8,1	0,23
4to	59	33	7	19	20	39	9,1	0,22
5to	49	34	5	10	19	30	10,2	0,25
6to	56	49	1	6	29	27	11,1	0,24
Total	270	186	20	64	103	167	9,1	1,4

Nota: DE (desviación estándar), F (Género Femenino), M (Género Masculino).

Tabla 2.
Descripción de la muestra N2

Grado	Total	Provincias		Enseñanza		Género		Edad	
		Cienfuegos	Matanzas	Especial	Regular	F	M	Media	DE
2do	20	8	12	7	13	1	19	7,1	0,22
3ro	20	9	11	14	6	9	11	8,0	0,15
4to	20	13	7	9	11	4	16	9,1	0,20
5to	20	11	9	10	10	9	11	10,1	0,22
6to	20	10	10	11	9	6	14	11,0	0,18
Total	100	51	49	51	49	29	71	9,09	1,4

Nota: DE (desviación estándar), F (Género Femenino), M (Género Masculino).

1.3 Procedimientos

Se aplicó la prueba ECI primeramente en la muestra N1. Posteriormente, fueron aplicadas ambas pruebas (ECI y WISC-R) en la muestra N2, en este mismo orden. La aplicación se realizó de forma individual, en las respectivas escuelas donde estudian los niños, en un local tranquilo, con iluminación y temperatura adecuada. La ECI fue aplicada en la pantalla de un ordenador.

La aplicación del WISC-R abreviado se realizó siguiendo las instrucciones del manual correspondiente (Brooker y Cyt, 1986).

1.4 Análisis estadístico

- Se realizó un análisis descriptivo (media, desviación estándar, valores mínimos y máximos) de los puntajes de la ECI en la muestra total (370 niños) con el objetivo de analizar las puntuaciones de los niños por grupo.
- Cálculo de puntos de corte.

Se calcularon puntos de corte para cada edad en la muestra N1 a través de un método de anclaje a la media. Según el procedimiento del método de anclaje (Agresti, 1996), inicialmente se calculó la media del puntaje total y la DE para cada edad. Posteriormente, se calculó los posibles puntos de corte: punto de corte 1 (restando 2 DE a la media), punto de corte 2 (restando 1 DE a la media), punto de corte 3 (restando 0,5 DE a la media) y punto de corte 4 (selección de la media).

Con estos cuatro posibles puntos de corte se calcularon los indicadores de validez en la muestra N2 usando los resultados y la clasificación del WISC-R abreviado como regla de oro.

La selección del punto de corte dependió de los indicadores de validez en la muestra N2: sensibilidad, especificidad, valor predictivo negativo y positivo. En caso de que los indicadores no superen el valor de 0,7, se selecciona otro punto de corte. Si el punto de corte seleccionado no supera el 0,7, se repite nuevamente el procedimiento hasta obtener el punto de corte de mejor ajuste estadístico según la regla de oro. Según Swets (1988), los valores entre 0,7 y 0,9 indican que la nueva prueba puede ser útil para clasificar al niño positivo y negativo, mientras que los valores entre 0,50 y 0,69 indica baja exactitud para clasificar.

1.5 Propiedades psicométricas

Confiabilidad: Se obtuvo el coeficiente Alfa de Cronbach de la prueba ECI en la muestra total (370 niños) mediante el método de multipartición, cada ítem es una sub-prueba. Adicionalmente, se analizó el alfa desde el funcionamiento de los ítems: alfa cuando el ítem es eliminando y la correlación ítem – total.

Validez: Se obtuvo mediante un análisis de correlación entre la ECI y el CI total del WISC-R abreviado en la muestra N2. Además, se analizaron los valores de sensibilidad, especificidad, el valor predictivo negativo y positivo del instrumento con relación a la regla de oro.

Estudio II

2.1 Participantes

En el estudio se seleccionaron 36 escolares (15 niños), 18 de 5to grado y 18 niños de 6to grado con edades comprendidas entre 10 y 11 años provenientes de una escuela regular en la provincia de Cienfuegos. Esta se describe detalladamente en la Tabla 3. La selección de los niños fue al azar. La Dirección Provincial del Ministerio de Educación en Cienfuegos autorizó la entrada a las escuelas. Asimismo, se solicitó el consentimiento informado a los padres de los niños seleccionados en el estudio.

Tabla 3.
Descripción de la muestra de Estudio II

Grados	Total	Género		Edad	
		F	M	Media	DE
5to	18	11	7	10,1	,57
6to	18	8	10	11,3	,69
Total	36	19	15	10,4	,79

2.2 Instrumentos

Diseño de ítems: Se diseñó una prueba de analogía verbales cuyo formato clásico es A:B:C:D, o sea, A es a B como C es a D (Sternberg & Nigro, 1980). Ejemplo, “Árbol es a Pájaro como Mar es a Pez”. Se construyeron 20 ítems según las categorías de Sternberg & Nigro (1980): por **relación funcional** (ejemplo: agua: beber; pan: comer), por **relación lineal** (ejemplo: oso: cueva; gorrion: nido) y por **relación categorial** (ejemplo: martillo: herramienta; sofá: mueble). Se utilizaron este tipo de categorías por su nivel de complejidad con respecto a las analogías por sinonimia y antonimia. Adicionalmente, los autores comentan que en la edad escolar las analogías por relación funcional y categorial son más complejas con respecto a las analogías por relación lineal.

Desde esta perspectiva, y con la intención de crear una prueba de analogías verbal más compleja, se construyó una cantidad superior de ítems por relación funcional y categorial (16 ítems) y, en menor cantidad, ítems por relación lineal (6 ítems). De los 20 ítems, se diseñaron 6 basados en relación

lineal (1, 2, 3, 4, 19 y 20), 4 basados en relación categorial (5, 6, 7 y 8) y 10 ítems basados en la relación funcional (9, 10, 11, 12, 13, 14, 15, 16, 17 y 18).

La cantidad de opciones de respuestas también complejiza el ítem. De esta forma, se diseñaron 16 ítems con 4 opciones de respuesta (entre el ítem 1 y el 16) y 4 ítems con 6 opciones de respuesta (ítems 17, 18, 19 y 20) con el objetivo de diferenciar los niños de mayor habilidad.

En el caso de los distractores, se diseñaron 3 tipos de distractores: opciones repetidas, opciones de cercanía semántica y opciones imposibles. En las opciones repetidas se le ofrece al niño la misma palabra a relacionar, por tanto, es una opción fácil de desechar, ejemplo "Guitarra es a cuerda como Piano es a: (1) Tecla, (2) Música, (3) Piano, (4) Tambor". En este caso, Piano es el distractor por opción repetida. En la opción de cercanía semántica, se espera que sea la opción de mayor confusión, pues es una opción parecida a la respuesta correcta, ejemplo "Cuchillo es a cortar como cuchara es a: (1) Recoger, (2) Comer, (3) Tenedor, (4) Cuchara. En este caso, Tenedor es el distractor por opción cercanía semántica. La opción de respuesta imposible, se se aleja de la respuesta correcta de manera obvia, por ejemplo "Cuchillo es a cortar como cuchara es a: (1) Recoger, (2) Comer, (3) Tenedor, (4) Cuchara. En este caso, Recoger es el distractor por opción imposible. Las opciones imposibles y repetidas permiten al investigador detectar niños que están respondiendo al azar y cuyos resultados no son confiables.

2.3 Procedimientos.

La aplicación se realizó de forma individual en la escuela, en un local tranquilo, con iluminación y temperatura adecuada. Primero se aplicó la prueba de Vocabulario y luego la prueba de AV. Se comenzó con los niños de 5to grado y después con los niños de 6to grado.

En la prueba de Vocabulario se siguió el procedimiento y las instrucciones del manual correspondiente a la prueba (Brooker y Cyt, 1986).

2.4 Análisis estadísticos.

Se realizó un análisis de los ítems de la prueba: índice de discriminación, índice de dificultad y funcionamiento de los distractores. El índice de discriminación se clasificó de la siguiente manera: Bueno (valores mayores o igual a .20), Aceptable (valores entre .10 y .19) y Mala (valores menores o igual a .10). El índice de dificultad se clasificó: Baja (valores mayores o igual a .80), Media (valores entre .51 y .79) y Alta (valores menores o igual a .50).

En relación al funcionamiento de distractores, se realizó un análisis de las tablas de frecuencia de selección de los diferentes distractores para cada ítem, con el objetivo de detectar:

- Distractores que son seleccionados por más del 90% de los niños que se equivocan en el ítem, lo cual indica que el distractor es demasiado parecido a la respuesta correcta y debe ser modificado.
- Distractores de respuesta imposible que son seleccionados por los niños en más del 25% de su ejecución total, lo cual indica que el niño podría estar respondiendo al azar y sus resultados no son confiables.

Además, se realizó un análisis de medias para muestras independientes con el objetivo de determinar diferencias significativas en el rendimiento de la prueba AV entre los niños de 10 años y los niños de 11 años. Este análisis permitió determinar si AV discrimina entre los niños de 10 y 11 años.

2.5 Propiedades psicométricas.

Confiabilidad: La confiabilidad se obtuvo a través del coeficiente Alfa de Cronbach utilizando el método de multipartición, o sea cada ítem de la prueba fue considerado una sub-prueba. Además, se analizó el alfa desde el funcionamiento de los ítems: alfa cuando el ítem es eliminando y la correlación ítem – total.

Validez: La validez se determinó a través del coeficiente de correlación entre la prueba AV y el sub-test de Vocabulario (WISC-R).

Resultados

Estudio I

3.1 Análisis de la Estadística descriptiva

Los valores de la media indican que en la medida que aumenta el grado escolar, los niños aumentan sus puntuaciones en la ECI. Con respecto a la desviación estándar, fue necesario eliminar 4 sujetos que funcionaban como outlayers (2 niños de 5to y 2 niños de 6to grado). Los valores de la estadística descriptiva se muestran en la Tabla 4.

Tabla 4.

Estadística descriptiva de los valores de la ECI

Grados	N total (370 niños)				
	N	Media	DE	Mín	Máx
2do	71	17,08	4,84	6	33
3ro	75	18,58	5,77	5	30
4to	79	19,69	5,22	8	32
5to	69	21,98	6,33	9	35
6to	76	24,26	7,09	10	35

Nota: DE (desviación estándar), Mín (mínimo), Máx (máximo).

3.2 Propiedades psicométricas de la ECI

Confiabilidad: El instrumento ECI mostró un índice de consistencia interna Alfa de Cronbach de .87. Por otra parte, el alfa cuando el ítem es eliminado, indicó valores entre 0,85 y 0,87.

El análisis correlación ítem-total indicó valores entre 0,20 y 0,51. Según el análisis, 20 ítems mostraron excelentes correlaciones con la varianza total de la prueba (ítems 6, 8, 9, 10, 11, 18, 19, 20, 21, 22, 23, 24, 28, 29, 30, 31, 32, 33, 34, 35), 9 ítems con buenas correlaciones (4, 5, 7, 12, 16, 17, 26, 27 y 36) y 7 ítems con regulares correlaciones (ítems 1, 2, 3, 13, 14, 15 y 25).

Validez: La correlación entre la ECI y el CI total del WISC-R abreviado fue de 0,78, ($p < 0,01$).

Puntos de Corte (indicadores de sensibilidad y especificidad):

- **Punto de corte 1:** El método de anclaje se utilizó para calcular los puntos de corte por edad en la N1 (Ver Tabla 5). El punto de corte 1 se obtuvo restando 2 DE a la media por edad, como se muestra en la Tabla 5. El punto obtenido presenta valores de sensibilidad por debajo del azar (0,7), como se muestra en la Tabla 6.
- **Punto de corte 2:** Se re-analizaron los datos restando 1 DS a la media como se muestra en la Tabla 5 para obtener el punto de corte 2. En la Tabla 6 se muestran los indicadores de validez calculados. Los valores de sensibilidad son bajos en cada edad con valores entre 0,52 y 0,62, los cuales no superan los valores de 0,70.
- **Punto de corte 3:** Se re-analizaron los datos restando 0,5 DS a la media para obtener el punto de corte 3 como se muestra en la Tabla 5. En la Tabla 6 se muestran los indicadores de validez, especialmente la sensibilidad muestra valores adecuados en las edades 8, 9 y 10 años (0,75, 0,81 y 0,70) respectivamente. No obstante, en el resto de las edades (7 y 11 años), los valores son bajos (por debajo de 0,70). La coincidencia al clasificar los niños entre el WISC y el ECI es adecuada (75%), como se muestra en la Figura 1.
- **Punto de corte 4:** El siguiente paso fue seleccionar como punto de corte la media. En la Tabla 6 se muestra los indicadores de validez, especialmente la sensibilidad muestra valores por encima de 0,7 en cada edad. Además, la coincidencia al clasificar entre la ECI y la regla de oro es adecuada y superior a los demás puntos de corte.

La coincidencia entre la ECI y el WISC es 89% como se muestra en la Figura 1. De manera general, la sensibilidad es de 0,90 y la especificidad de 0,81. Se encontraron 76 casos Verdaderos Positivos (76%); 13 casos Verdaderos Negativos (13%); 3 casos Falsos Positivos (3%) y 8 casos Falsos Negativos (8%). El valor predictivo positivo (VPP) fue de .96 y el valor predictivo negativo (VPN) de .61. A los 7 años, la ECI clasifica el 100% de los niños en iguales condiciones a la regla de oro, a los 8 y 9 años clasifica el 90%, a los 10 años clasifica el 85% y a los 11 años, el 80%.

		Regla de oro		
		RI Bajo	RI Normal	Total
ECI	RI Bajo	59 (76)	1 (3)	60 (79)
	RI Normal	24 (8)	16 (13)	40 (21)
Total		83 (84)	17 (16)	100 (100)

Figura 1. Clasificación del WISC vs ECI tomando el punto de corte 3 y 4: Valores de punto de corte 3 y punto de corte 4, estos últimos en paréntesis.

Tabla 5.
Puntos de corte por edad en N1

Edad	Punto de Corte 1 (menos 2 DE)	Punto de Corte 2 (menos 1 DE)	Punto de Corte 3 (menos 0,5 DE)	Punto de Corte 4 (Media)
7 años	8,42	13,30	15,74	18,19
8 años	8,77	14,31	17,08	19,85
9 años	11,21	15,93	18,29	20,66
10 años	11,74	17,53	20,42	23,32
11 años	15,92	21,02	23,57	26,12

Nota: DE (desviación estándar).

Tabla 6.
Valores de sensibilidad y especificidad en todos los puntos de corte

Grupo de Edades	Punto de Corte 1		Punto de Corte 2		Punto de Corte 3		Punto de Corte 4	
	Sens.	Esp.	Sens.	Esp.	Sens.	Esp.	Sens.	Esp.
7 años	0,06	1	0,55	1	0,64	1	1	1
8 años	0,05	1	0,62	1	0,75	1	0,94	0,66
9 años	0,18	1	0,62	1	0,81	1	0,87	1
10 años	0,22	1	0,52	1	0,70	1	0,93	0,66
11 años	0,55	1	0,62	0,75	0,68	0,75	0,92	0,75
Total	0,22	1	0,59	0,94	0,71	0,94	0,90	0,81

Nota: Sens. (sensibilidad), Esp. (especificidad).

Estudio II

4.1 Análisis de los ítems

El 80% de los ítems (16 ítems de 20) son buenos para discriminar al niño con alta y baja habilidad (ítems 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 14, 15, 17, 18, 19 y 20) y un 20% (4 ítems de 20) de los ítems son aceptables (ítems 7, 11, 13, 16). La prueba no mostró ítems con bajo índice de discriminación.

El 45% (9 ítems de 20) de los ítems tiene un índice de dificultad media. El 35% (7 ítems de 20) de los ítems tiene una dificultad alta y el 20% (4 ítems de 20) una dificultad baja.

La organización inicial de los ítems fue modificada según el índice de dificultad y discriminación. La versión final de la prueba quedó conformada con un nivel de complejidad ascendente: 3, 4, 6, 9, 10, 12, 14, 1, 2, 5, 7, 8, 11, 15, 16, 18, 13, 17, 19 y 20.

Análisis de los distractores: Se analizó si algún distractor en especial fue seleccionado por el 90% o más de los niños que se equivocan en el ítem, lo cual indica que tal distractor es demasiado parecido a la respuesta correcta y su funcionamiento no es adecuado. En este caso se encontró el distractor número 4 del ítem 14 con 100% de los niños que se equivocan, el distractor número 1 en el ítem 15 con 90%. Estos distractores deben ser modificados en próximas revisiones de la prueba AV. Igualmente, el distractor número 1 en el ítem 18 con 87%; el distractor número 1 en el ítem 13 con un 85% y el distractor número 4 en el ítem 16 con 73% de los niños que se equivocan, podrían ser revisados y mejorados.

Por último, se analizó la cantidad de niños que escogen los distractores claramente imposibles más del 25% de las veces durante la prueba. Los resultados en este análisis mostraron que ningún niño de la muestra respondió al azar.

4.2 Propiedades psicométricas

Confiabilidad: El Alfa de Cronbach mostró indicadores aceptables, con un valor de 0,76. El alfa cuando el ítem es eliminado indicó valores entre .77 y .80. Según el alfa cuando el ítem es eliminado, los ítems 7 y 13 deben ser mejorados o eliminados para aumentar el alfa a .80. Igualmente, la correlación ítem –total, indica que los ítems 7 y 13 presentan bajas correlaciones con la varianza total de la prueba (.12 y .14 respectivamente), por tanto, se deben analizar el diseño de estos ítems que podrían responder al aumento de los valores de consistencia interna de la prueba. Las correlaciones ítem – total indicaron valores entre 0,12 a 0,62.

Validez: El análisis de correlación entre la AV y el sub-test de Vocabulario del WISC-R fue de .73, ($p < 0.01$).

4.3 Análisis de medias entre los grupos de 5to y 6to grado

El análisis de medias para muestras independiente mostró diferencias significativas entre los grupos de 5to y 6to grado (prueba t (36) = 2,769, $p = .009$).

Discusión

En un primer estudio, se validó la ECI para medir la inteligencia fluida en niños escolares (2do – 6to grado) con edades comprendidas entre 7 y 11 años. En un segundo estudio, se presentó el diseño de la prueba AV para medir la inteligencia cristalizada en niños escolares de 5to y 6to grado con edades comprendidas entre 10 y 11 años. Ambos estudios responden a la necesidad de contar con pruebas de pesquisa validadas en nuestro contexto cultural para identificar al niño escolar con riesgo intelectual. Los resultados indican que la ECI puede ser utilizada como una medida de inteligencia fluida para detectar el riesgo intelectual en la edad escolar. Además, los resultados iniciales de la AV muestran adecuada validez y confiabilidad en la evaluación de la inteligencia cristalizada en los niños escolares de 5to y 6to grado.

Estudio de la validación de la Escala de Capacidad Intelectual (ECI)

Dos propiedades psicométricas resultan de esencial relevancia en la construcción de pruebas psicológicas: la confiabilidad y la validez. El alfa de Cronbach es actualmente la medida de consistencia interna más frecuente para evaluar la confiabilidad. Normalmente, los índices de confiabilidad reportados por los autores de pruebas de inteligencia se encuentran entre .70 y .90 (Sánchez, Santamaría & Abad, 2015; Cerda, Pérez, Melipillán & Ortega-Ruiz, 2015; Rosa & Cerir, 2015; Brenlla et al, 2013).

No obstante, una revisión de los principales test de inteligencia con matrices indica que los índices de confiabilidad se encuentran entre .80 y .95 (Sánchez, Santamaría & Abad, 2015; Cerda, Pérez, Melipillán & Ortega-Ruiz, 2015; Elosua, Mujika, Almeida & Hermosilla, 2014; Sánchez - Sánchez & Arribas, 2014). En conjunto, los resultados emitidos en la literatura sugieren que las pruebas de inteligencia con matrices son válidas y confiables para medir la inteligencia en niños y adolescentes.

En el presente estudio, y en coincidencia con otros estudios donde proponen pruebas de inteligencia con matrices (Sánchez, Santamaría & Abad, 2015; Cerda, Pérez, Melipillán & Ortega-Ruiz, 2015; Elosua, Mujika, Almeida & Hermosilla, 2014; Sánchez - Sánchez & Arribas, 2014), el índice de confiabilidad de la ECI resultó adecuado con un valor de .87. Con respecto al alfa cuando el ítem es eliminado, los valores obtenidos estuvieron entre .85 y .87. Estos valores de confiabilidad son similares a los obtenidos anteriormente en el estudio de diseño de la prueba (Amor, 2011), lo cual indica adecuada consistencia interna. En el estudio del 2011, la ECI mostró un alfa de .84 con una muestra de 100 niños de la provincia de la Habana, en esta ocasión, se utilizó una muestra independiente con 270 niños de tres provincias y los valores del alfa mantienen índices apropiados. Los resultados de ambas investigaciones indican que los valores de consistencia interna de la ECI son estables en el tiempo e independientes de la muestra utilizada.

Con relación a la validez, la estrategia más frecuente es el cálculo de la correlación con una medida reconocida del constructo de interés. Así, corre-

laciones positivas y significativas se interpretan como evidencias de una adecuada validez (Carretero - Dios & Pérez, 2005). Desde esta perspectiva, las escalas de Wechsler son usadas con frecuencia como regla de oro para analizar la validez en pruebas de inteligencia (Riveros, Sepúlveda, Figueroa & Rosas, 2015; Villasenor-Lozano, et al, 2014). Según Riveros y colaboradores (2015), una tarea de atención y otra de memoria son suficientes para obtener medidas rápidas del factor *g* en adolescentes. Los autores obtuvieron correlaciones significativas entre la nueva prueba de inteligencia y la escala de Wechsler (.85, $p < 0.01$), indicando apropiados valores de validez. En igual sentido, Villasenor-Lozano y colaboradores (2014), proponen una escala corta de inteligencia para medir el rendimiento intelectual en niños entre 6 y 8 años, la cual mantiene una significativa correlación con el CI total del WISC-IV (.82, $p < 0.01$). En el presente estudio, la ECI mostró una correlación positiva y significativa con el CI total del WISC-R abreviado (.78, $p < 0.01$). De esta forma, se esperaba correlaciones por encima de .80, sin embargo, el hecho de utilizar una escala abreviada del WISC-R a través de dos sub-pruebas podría influir en la correlación, ya que los demás autores utilizaron todas las pruebas de las dos escalas. Además, una de las sub-prueba del WISC-R que empleamos para construir el CI total que funcionó como regla de oro en nuestro trabajo, mide la inteligencia cristalizada, y esto podría influir en el resultado final, pues la ECI mide sólo la inteligencia fluida. Aun así, Hernández, Fernández & Baptista (2010) indican que los valores por encima de .70 son valores significativos y considerables en un análisis de correlación.

Las altas correlaciones entre dos pruebas psicométricas constituyen un criterio necesario, pero no suficiente para demostrar la validez de la nueva prueba. Los autores recomiendan, además, que ambas pruebas coincidan al clasificar el trastorno de interés en un alto por ciento. En este sentido, se encontró que la coincidencia en la clasificación de los niños entre la ECI y el WISC-R abreviado es de 89%. Específicamente, del total de niños clasificados por el WISC-R abreviado con CI bajo fueron identificados con riesgo intelectual por la ECI el 90%. Asimismo, del total de niños clasificados con CI normal por el WISC-R abreviado fueron descartados sin riesgo por la ECI el 81%. Estos resultados indican que la ECI constituye un instrumento capaz de detectar efectivamente el riesgo intelectual en la edad escolar.

Los valores de sensibilidad se encuentran entre .87 y 1. El valor más bajo de sensibilidad se encontró a la edad de 9 años (.87). En relación a la especificidad, o capacidad de descartar al niño con valores por encima de 85 puntos, los valores generales se encuentran entre .66 y 1. Los valores de especificidad fueron óptimos en las edades de 7 y 9 años, es decir, el 100% de los niños con CI típico fueron descartados por la ECI. En las edades de 8, 10 y 11 años la especificidad es más baja. En estas edades, el instrumento detecta con riesgo niños que muestran un rendimiento intelectual típico, lo que provoca el empleo de recursos materiales y humanos adicionales en la evaluación completa de niños que realmente no presentan problemas intelectuales. Este indicador deberá ser mejorado en fases posteriores de la investigación. Sin embargo, es necesario considerar que la función de este instrumento es garantizar la detección de todos los niños en riesgo de presentar algún grado de afectación intelectual en la población escolar general. De modo que es preferible evaluar niños que realmente no presentan problemas antes que dejar de evaluar a los que los presenten.

Llama la atención que la mejor discriminación se obtiene ubicando como punto de corte la propia media de la muestra N1. En primer lugar, la meta de detectar todos los niños con bajo rendimiento intelectual nos ha hecho adoptar un criterio que maximice la sensibilidad, aunque se deteriore ligeramente la especificidad. Sin embargo, este criterio extremo no es realmente necesario en una prueba de pesquisa y podría funcionar bien con un punto de corte menos exigente que ronda el 70 %.

En segundo lugar, la clasificación del niño como positivo o no que funciona como regla de oro se obtuvo con una versión reducida del test original de Wechsler, lo cual trae como consecuencia que sea un indicador rápido pero sólo aproximado al puntaje real de CI. Además, este CI está construido tomando en cuenta tanto el resultado del subtest de vocabulario como el subtest de cubos que responden a dos áreas diferentes del rendimiento intelectual, mientras que nuestras pruebas responden solo a una de estas áreas. Quizás si empleamos pruebas específicas para habilidades verbales y razonamiento espacial por separado, los resultados serían diferentes.

Por último, la muestra N1 puede estar sesgada. Recuérdese que los niños con rendimiento normal bajo están en la enseñanza general y en muchas ocasiones también permanecen en ella los niños con retraso mental leve, como medida conservadora para no enviarlos a la enseñanza especial. Sería importante revisar la composición de esta muestra de referencia.

Finalmente, en otros estudios se han encontrado este tipo de procedimiento (aumentar la sensibilidad aun cuando baje la especificidad ligera-

mente) (Riveros, Sepúlveda, Figueroa & Rosas, 2015; Ramírez, Jiménez-Morales & Díaz, 2015). Riveros y colaboradores (2015) indican valores de sensibilidad de .85 en una prueba breve de inteligencia (test de atención y test de memoria), mientras que la especificidad fue de .83 con respecto a la regla de oro (escala de Wechsler). Asimismo, informan que la intención no es diagnosticar la discapacidad intelectual, más bien detectar los problemas intelectuales en esta población para después profundizar en un posible diagnóstico a través de una prueba estandarizada. Con iguales intenciones, Ramírez y colaboradores (2015) proponen un punto de corte para detectar problema en el razonamiento analógico, y en menor medida para diagnosticar discapacidad. Los autores maximizaron la sensibilidad a un valor de .89 y se permitieron bajar el valor de la especificidad a .66. En general, los resultados del ECI indican que, aunque es necesario mejorar su especificidad, este constituye una medida válida y confiable de *Gf* y puede ser usado como instrumento de pesquisa.

Diseño de la prueba Analogías Verbales (AV)

El diseño de un nuevo test se analiza desde las propiedades psicométricas de validez, confiabilidad y funcionamiento de ítems. Según Ledesma, Molina & Valero (2002), valores superiores a .70 constituyen adecuados niveles de confiabilidad. En este sentido, los resultados iniciales mostraron valores apropiados de confiabilidad: entre .77 - .80.

Resultados similares a estos valores, entre .80 y .81, son reportados por Sánchez - Sánchez & Arribas (2014) cuando proponen una sub-prueba de analogías verbales, incluida en la BAT-7 (Batería de Aptitudes), para medir la inteligencia cristalizada en niños y adolescente españoles entre 10 y 18 años. En igual sentido, Elosua y colaboradores (2014) mostraron valores de .88 en una sub-prueba de analogías verbales, incluida en la batería de prueba de razonamiento, aplicadas a 186 niños españoles entre 8 y 18 años. Con respecto a este último estudio, la diferencia entre los valores del alfa, obtenidos por este estudio y los alcanzados en la presente investigación, podría estar justificada por el tamaño de la muestra, así como por la presencia de otras edades en el análisis. Aun así, los valores del alfa por encima de .70 son apropiados para un estudio preliminar de la prueba AV.

En relación a la validez, se usó el sub-test de Vocabulario del WISC-R como medida reconocida de la inteligencia cristalizada (Villasenor-Lozano y col, 2014, Akshoomoff y col, 2013) para la validación de la prueba AV. Los resultados obtenidos en la investigación indicaron que la AV muestra correlaciones positivas y significativas con el sub-test de Vocabulario (.73, $p < 0.01$). Así, AV muestra una alta correlación puesto que ambas tareas están midiendo la inteligencia cristalizada. Igualmente informan Gulini & Muglia (2009) en una muestra de niños entre 10 y 15 años, los cuales obtuvieron una significativa correlación (.70, $p < 0.01$) entre el sub-test de Vocabulario del WISC-III y el sub-test de Analogías verbales del WJ-III. Asimismo, Richland & Burchinal (2013), indicaron que las puntuaciones del sub-test Vocabulario en niños de 10 años es un buen predictor de las puntuaciones del sub-test Analogías verbales a los 15 años. Los autores se apoyan en las significativas correlaciones entre Vocabulario y Analogías (.46 $p < 0.001$).

Por otra parte, el análisis de ítems permitió conocer el funcionamiento de la prueba en términos de dificultad de las analogías y su capacidad para discriminar la alta y baja habilidad en los niños. En este sentido, el 100% de los ítems de la AV tienen un índice de discriminación bueno o aceptable. Dicho resultado indica que los ítems de la prueba se pueden conservar, pues diferencia a los niños con altas y bajas calificaciones.

Respecto al índice de dificultad, AV incluye ítems de los tres niveles de dificultad, como es deseable para evaluar niños con diferente nivel de habilidad. Sin embargo, en el caso específico de nuestro instrumento, recuérdese que la intención de su construcción es completar un instrumento anterior, capaz de evaluar adecuadamente niños de 1ro a 4to grado, pero no niños de 5to y 6to grado. De modo que, en el caso de AV, la investigación se propone diseñar ítems con índice de dificultad medio y alto en primera instancia.

De esta manera, la AV contiene 13 ítems con índice de dificultad medio y alto para medir la inteligencia cristalizada y 7 ítems con dificultad baja. Estos últimos no serán excluidos, más bien se conservan para aplicarse en las primeras 7 analogías a resolver, pues así el niño avanza en la prueba de menor complejidad a mayor complejidad.

Por otra parte, la prueba es capaz de diferenciar el rendimiento de los niños con respecto al grado escolar. El análisis de media indicó diferencias significativas entre el grupo de 5to y el grupo de 6to grado, es decir, los niños de 10 años muestran puntajes inferiores significativamente a los obtenidos por los niños de 11 años, lo que sugiere que la AV es capaz de discriminar niños de ambas edades.

De manera general, los resultados iniciales de la AV indican que el instrumento presenta propiedades psicométricas apropiadas para explorar la inteligencia cristalizada en niños de 10 - 11 años de edad.

Finalmente, el presente estudio cumple con un objetivo práctico al proponer dos instrumentos de pesquisa para evaluar la inteligencia fluida y la inteligencia cristalizada en el niño escolar. Ambos instrumentos responden a la necesidad de contar con pruebas de inteligencia validadas en escuelas cubanas, las cuales son apropiadas para detectar alteraciones intelectuales, así como economizar recursos personales y temporales, por ser fáciles de aplicar y de rápida ejecución.

Conclusiones

1. El ECI resulta una prueba apropiada para detectar al niño escolar con riesgo intelectual. Esta afirmación se sustenta en la sensibilidad general de 89% obtenida con un punto de corte óptimo y conservador. Considerando que esta no es una prueba diagnóstica, sino solamente de pesquisa de riesgo de trastornos intelectuales, la concordancia con la clasificación del WISC-R resulta más que satisfactoria.
2. El diseño de la prueba de Analogías Verbales presenta apropiadas propiedades psicométricas (validez y confiabilidad) para explorar el razonamiento analógico verbal en los niños 10 y 11 años. Las correlaciones significativas entre la prueba Analogías verbales y la sub-prueba de Vocabulario del WISC-R, así como el análisis de funcionamiento de ítems lo corroboran.

Recomendaciones

1. Mejorar los indicadores de validez de la ECI (especificidad y valor predictivo negativo) con el objetivo de identificar con mayor precisión al niño sano y diferenciarlo del niño con riesgo intelectual.
2. Unificar los instrumentos actuales de Analogías Verbales para crear un solo instrumento que pueda utilizarse en edades entre 7 – 11 años.
3. Emplear otros modelos y métodos para calcular los puntos de corte y la clasificación de los sujetos, como las Curvas ROC o el Item Response Theory.
4. Tomar como regla de oro instrumentos que midan por separado habilidades verbales y no verbales, y no instrumentos compuestos como el WISC-R que se empleó en este estudio, y que pudieran estar sesgando los valores de clasificación.

Referencias

- Ackerman, P. L. (1996). A theory of intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22, 227–257.
- Amor Díaz V. (2011). Diseño de instrumentos de evaluación de la capacidad intelectual en la edad escolar. Tesis en opción al título académico de Máster en Neurociencias Mención: Neurociencias Cognitivas. Centro de Neurociencias de Cuba.
- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: Wiley.
- Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen S. (2013). NIH Toolbox Cognitive Function Battery (CFB): Composite Scores of Crystallized, Fluid, and Overall Cognition. *Monogr Soc Res Child Dev*. 78(4): 119–132.
- Arribas Águila D. (2013). La evaluación de las aptitudes intelectuales. Padres y Maestros. N° 354 Diciembre: 36 -39.
- Archibald LMD, Oram Cardy J, Joanisse MF, Ansari D (2013) Language, Reading, and Math Learning Profiles in an Epidemiological Sample of School Age Children. *PLoS ONE*. 8(10): e77463. doi:10.1371/journal.pone.0077463
- Brenlla M E, Labin A, Steinberg LB, Soffita Y, D'Anna A, Salvino CV, Gottheil B, Molina MA, Abbona G (2013). Guía para una interpretación integral del WISC- V. Editorial Paidós. Buenos Aires. Argentina.
- Brooker BH, Cyt JJ. (1986). Tables for clinicians to use to convert WAIS-R short forms. *Journal of Clinical Psychology*. 42: 983.
- 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.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54: 1–22.
- Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.
- Cerda, G., Pérez, C., Melipillán, R., & Ortega-Ruiz, R. (2015). Examen psicométrico del IQ Test como herramienta de discriminación de individuos normales y talentosos en la población escolar chilena. *Universitas Psychologica*, 14(3), 899-912. <http://dx.doi.org/10.11144/Javeriana.upsy14-3.epiq>
- Elosua P, Mujika J, Almeida LS, Hermosilla D. (2014). Procedimientos analítico-razonales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento. *Rev Latinoam Psicol*. 46(2):117-126.
- Flanagan, D. P., & Kaufman, A. S. (2009). Claves para la evaluación con WISC-IV. Editorial El Manual Moderno.
- Gómez-Benito J, Dolores Hidalgo M, Guilera G. (2010). El sesgo de los instrumentos de medición: tests justos. *Papeles del Psicólogo*. 31(1): 75-84.
- Gulini CM, Muglia WS. (2009). Escala de inteligencia WISC-III y Batería de Habilidades cognitivas Woodcock Johnson –III. *Avaliação Psicológica*, 2009, 8(3), pp. 313-324.
- Hernández Sampieri R, Fernández Collado C, Baptista Lucio MP. (2010). *Metodología de la investigación*. (5ta Edición). Editorial Mc Graw Hill. México.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual* (pp. 197–232). Chicago: Riverside.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Ledesma, R., Molina, G. & Valero, P. (2002). Análisis de consistencia interna mediante Alfa de Cronbach: un paquete basado en gráficos dinámicos. *Psico-USF*. 7(2): 143-152.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10.
- Pina V, Fuentes LJ, Castillo A, Diamantopoulou S. (2014). Disentangling the effects of working memory, language, parental education, and non-verbal intelligence on children's mathematical abilities. 5 (415): 1-12. doi: 10.3389/fpsyg.2014.00415
- Ramírez Benítez Y, Jiménez-Morales RM, Díaz Bringas M. (2015). Matrices progresivas de Raven: punto de corte para preescolares 4 - 6 años. *Evaluar*. 15: 123–133.
- Raven, J. C., Court, J. H. & Raven, J. (1992). *Manual for the Raven's Progressive Matrices and Vocabulary Scales*. Oxford Psychologists Press. USA
- Ritchie SJ, Bates TC, Plomin R. (2015). Does Learning to Read Improve Intelligence? A Longitudinal Multivariate Analysis in Identical Twins From Age 7 to 16. *Child Development*. 86 (1): 23–36.
- Riveros R, Sepúlveda A, Figueroa P, Rosas R. (2015). Memoria de trabajo y atención pueden ser usados como un nuevo test de screening de inteligencia. *Papeles de investigación*. 4: 1-16.
- Richland LE, Burchinal MR. (2013). Early Executive Function Predicts Reasoning Development. *Psychological Science*. 24(1) 87–92.
- Rosas R, Cerir F. (2015). ¿Pruebas Tradicionales o Evaluación Invisible a Través del Juego? *Psykhē*. 24 (1): 1-11. doi:10.7764/psykhe.23.2.724
- Roid, G. H. (2003). *Stanford–Binet Intelligence Scales* (5th ed.) Itasca, IL: Riverside Publishing.
- Sánchez - Sánchez, F., Santamaría, P. y Abad, F. J. (2015). Matrices. Test de Inteligencia General. Madrid: TEA Ediciones.
- Sánchez-Sánchez, F; Arribas Águila D. (2014). BAT-7, Batería de Aptitudes de TEA: Descripción y datos psicométricos. *International Journal of Developmental and Educational Psychology*. 2 (1): 353-364.
- Spearman, C. (1904). General intelligence, IQ objectively determined and measured. *American Journal of Psychology*. 15: 201–293.
- Stevenson CE, Bergwerff CE, Heiser WJ, Resing WCM. (2014) Working Memory and Dynamic Measures of Analogical Reasoning as Predictors of Children's Math and Reading Achievement. *Infant and Child Development*. 23: 51–66.
- Sternberg, R. Ferrandiz, C; Hernandez, D., & Ferrando, M. (2011). Inteligencia Exitosa y Alta Habilidad. En M. D. Prieto Sánchez (Eds.), *Psicología de la Excepcionalidad*. Madrid: Editorial Síntesis.

- Sternberg, R. J. & Nigro, G. (1980). Developmental patterns in the solution of verbal analogies. *Child Development*, 51: 27-38.
- Swets JA. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240: 1.285-1.293.
- Tong F, Fu T. (2013). Meta-Analysis of Fluid Intelligence Tests of Children from the Chinese Mainland with Learning Difficulties. *PLoS ONE*, 8(11): e78311. doi:10.1371/journal.pone.0078311
- Van Bergen E, Jong PF, Maassen B, Krikhaar E, Plakas A, Van der Leij A. (2013). IQ of Four-Year-Olds Who Go On to Develop Dyslexia. *Journal of Learning Disabilities*. <http://jdx.sagepub.com/content/early/2013/03/14/0022219413479673>
- Van Bergen E, de Jong P F, Plakas A, Maassen B, & Van der Leij A. (2012). Child and parental literacy levels within families with a history of dyslexia. *Journal of Child Psychology and Psychiatry*, 53(1), 28–36.
- Villaseñor-Lozano CG, Calderón Hernández J, Morales Villegas R, Sánchez-Armás O, Díaz-Barriga F. (2014). Validez concurrente del formato corto BIA de la batería III por correlación con el formato corto de Crawford CIT del WISC-IV. *Rev Mex Neuroci*, 15(3): 133-137
- Vukovic RK, Lesaux NK. (2013). The relationship between linguistic skills and arithmetic knowledge. *Learning and Individual Differences*, 23, February: 87–91
- Wechsler, D. (2003). Wechsler Intelligence Scale for Children (4th ed.). San Antonio, TX: The Psychological Corporation.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Cognitive Abilities. Rolling Meadows, IL: Riverside Publishing.