

## VALIDEZ DIFERENCIAL Y SESGO EN LA PREDICTIVIDAD DE LAS PRUEBAS DE ADMISIÓN A LAS UNIVERSIDADES CHILENAS (PSU)

DIFFERENTIAL VALIDITY AND SLANT ON THE PREDICTABILITY OF THE ADMISSION TESTS TO THE CHILEAN UNIVERSITIES (PSU)

*Jorge Manzi, Angélica Bosch, David Bravo, Guido del Pino, Graciela Donoso y Raúl Pizarro*

Revista Iberoamericana de Evaluación Educativa 2010 - Volumen 3, Número 2

<http://www.rinace.net/riee/numeros/vol3-num2/art2.pdf>

Fecha de recepción: 4 de abril de 2010.

Fecha de dictaminación: 30 de mayo de 2010.

Fecha de aceptación: 30 de mayo de 2010.

La sostenida brecha en los resultados entre establecimientos públicos y privados en Chile ha generado un permanente debate acerca de la equidad y del posible sesgo socioeconómico de las Pruebas de Selección Universitaria (PSU). Sin embargo, este debate no es exclusivo de la experiencia chilena ni del contexto de admisión universitaria. Las pruebas estandarizadas a gran escala suelen revelar diferencias en los puntajes obtenidos por ciertas minorías étnicas, socioeconómicas o de género. No obstante, Zwick (2007) plantea que las brechas en los resultados no representan necesariamente un sesgo de las pruebas y sostiene, por otro lado, que “la equidad de una prueba está inextricablemente ligada a su validez”.

En este sentido, la inequidad podría manifestarse a través del sesgo de medición del instrumento (que ocurre cuando aparecen diferencias sistemáticas en los puntajes obtenidos por algunos grupos, que no son atribuibles a diferencias en la habilidad que se desea medir), o bien, a través de un uso de los puntajes que perjudica de manera consistente a ciertos grupos, en relación al objetivo para el cual fue diseñada la prueba.

En el caso de las pruebas de selección universitaria, la evidencia más pertinente acerca de su validez es el grado en que los puntajes permiten predecir el éxito de los estudiantes en la enseñanza universitaria. En rigor, se espera que el ordenamiento de los postulantes según los puntajes obtenidos en una prueba de selección coincida con el ordenamiento que tendrían respecto de su potencial desempeño universitario. Una condición necesaria para que esto se cumpla es que la capacidad predictiva de la prueba sea invariante respecto de las características sociodemográficas de los examinados. Ello no ocurre si, para un mismo resultado en la prueba, algún subgrupo de la población obtiene sistemáticamente mejores o peores resultados en su desempeño académico. De ser así, los puntajes de las pruebas de selección estarían siendo usados de un modo que resulta inequitativo para ciertos grupos de la población examinada (Young, 2004).

La constante preocupación por la equidad en medición educacional ha llevado a desarrollar diversos procedimientos para evaluar los riesgos de sesgo asociados a las características sociodemográficas de los examinados. Dos técnicas que se utilizan para establecer si la capacidad predictiva de un instrumento de medición es equivalente entre distintos grupos de examinados son los análisis de *Validez Diferencial* y de *Predicción Diferencial*. Los estudios de Validez Diferencial se basan en comparar la magnitud de la relación entre el resultado de la prueba de selección y el rendimiento en la universidad, para diferentes grupos. Cuando el coeficiente de correlación estimado para cada subgrupo es sustancialmente distinto en alguno de ellos, entonces se habla de Validez Diferencial. La evaluación de Predicción Diferencial, por otra parte, consiste en analizar si la relación entre los factores de selección y el rendimiento académico puede describirse a través de un modelo de regresión único para todos los subgrupos examinados. Si los residuos de la ecuación de regresión empleada para la predicción son desiguales entre grupos, entonces se produce una subestimación (o sobreestimación) del rendimiento académico de alguno de éstos. En otras palabras, los alumnos de dicho grupo tienden a tener un desempeño que es mejor (o peor) al que el modelo de regresión común predice.

Para el caso norteamericano, Young (2004) realizó una revisión de los estudios llevados a cabo desde 1974 sobre validez y predicción diferencial en diferentes pruebas de admisión universitaria. La mayor parte de los 49 estudios incluidos en este trabajo analizan la relación entre puntajes de pruebas de selección y notas de primer año en la universidad, para diferentes grupos étnicos o raciales y/o entre mujeres y hombres. Aún cuando los resultados varían entre estudios, en general se observan correlaciones más bajas para los grupos minoritarios conformados por negros e hispanos, en comparación

con los estudiantes blancos, mientras que para las comparaciones de género, estas correlaciones suelen ser mayores en el caso de las mujeres. En términos de predicción diferencial, se observa una sobrepredicción de las notas de primer año para los grupos que conforman las minorías étnicas y raciales, y una subpredicción del rendimiento de las mujeres.

De las pruebas de admisión usadas actualmente en Estados Unidos, el SAT es una de las de mayor cobertura en términos del número de examinados anual (Young, 2004). El SAT es una prueba de razonamiento que se compone de tres secciones: redacción (SAT-W), lectura crítica (SAT-CR) y matemáticas (SAT-M), en las cuales la mayoría de las preguntas son de selección múltiple. Respecto a esta prueba, el estudio más reciente sobre validez y predicción diferencial fue llevado a cabo por Mattern et al. (2008) en base a una muestra de 155.316 alumnos, provenientes de 110 instituciones. En su trabajo, los autores encuentran un patrón de resultados que es consistente con el conocido hasta entonces:

En los análisis de validez diferencial, para todos los factores de selección obtienen correlaciones con el rendimiento académico que son superiores en el grupo de estudiantes blancos versus los grupos de negros e hispanos. En las comparaciones por género, obtienen correlaciones entre notas en la universidad y el SAT que son más altas para mujeres que para hombres, mientras que las correlaciones con las notas de colegio resultan ser más parejas. (Los principales resultados de este análisis se resumen en la Tabla 1 que se presenta a continuación).

**TABLA 1. VALIDEZ DIFERENCIAL DEL SAT**

Grupo Sociodemográfico	SAT-CR	SAT-M	SAT-W	Notas Colegio
Hombres	.44	.45	.47	.52
Mujeres	.52	.53	.54	.54
Blancos	.48	.45	.51	.56
Negros	.40	.40	.43	.44
Hispanos	.43	.41	.46	.46

(Fuente: Mattern et al, 2008)

En cuanto a la predicción diferencial, los autores concluyen que existe una sobrepredicción del rendimiento académico de los hombres y una subpredicción del rendimiento académico de las mujeres. En el caso de estudiantes blancos versus minorías raciales y étnicas, observan un patrón contrario: una leve subpredicción del rendimiento para alumnos blancos, una sobrepredicción del rendimiento para alumnos negros y una sobrepredicción algo menor para alumnos hispanos. (Los resultados, que se presentan en la Tabla 2 se interpretan de la siguiente manera: valores negativos indican una sobrepredicción del rendimiento y valores positivos, una subpredicción.)

**TABLA 2. PREDICCIÓN DIFERENCIAL DEL SAT**

Grupo Sociodemográfico	SAT-CR	SAT-M	SAT-W	NEM
Hombres	-.14	-.20	-.11	-.08
Mujeres	.12	.17	.10	.07
Blancos	.04	.05	.04	.06
Negros	-.30	-.26	-.26	-.32
Hispanos	-.17	-.16	-.16	-.27

(Fuente: Mattern et al, 2008)

El presente estudio tiene como objetivo analizar la validez diferencial y el sesgo de predicción en los factores de selección utilizados por las 25 universidades que conforman el Consejo de Rectores de las Universidades Chilenas (CRUCH). Estos factores de selección incluyen dos pruebas obligatorias (Lenguaje

y Comunicación y Matemáticas) y dos pruebas opcionales (Ciencias e Historia y Ciencias Sociales), además de las Notas de Enseñanza Media (NEM).<sup>1</sup>

La batería de pruebas, denominada Prueba de Selección Universitaria (PSU), reemplazó a partir de la admisión 2004, a las pruebas precedentes o Pruebas de Aptitud Académica (PAA). En este contexto de cambio, el Consejo de Rectores ha propiciado la realización de un conjunto de estudios técnicos que permitan contar con antecedentes sobre el funcionamiento de las PSU. A la fecha, esta tarea ha sido encomendada al Comité Técnico Asesor para las Pruebas de Selección y Actividades de Admisión del Consejo de Rectores de las Universidades Chilenas (CTA). El año 2006 el Comité Técnico Asesor entregó el primer estudio respecto de la validez predictiva de la PSU. En dicha oportunidad se consideró importante comparar la capacidad predictiva de las nuevas pruebas con la de aquéllas que habían regido hasta la admisión del año anterior. Por lo tanto, el estudio contempló el análisis de la validez predictiva de las baterías de admisión para los alumnos ingresados los años 2003 y 2004, considerando en el rendimiento de estos estudiantes al término de su primer año de estudios en la universidad. Las técnicas de análisis incluyeron correlaciones simples para cada uno de los factores de selección y regresiones lineales para explicar el aporte conjunto de los diferentes factores de selección. El año 2008 se publicó una actualización del estudio anterior, incorporando en esta versión a los alumnos de las admisiones 2005 y 2006. Los resultados reportados en ambos informes revelan que las pruebas con mayor capacidad predictiva promedio en las PSU corresponden a las pruebas de Matemática y Ciencias, las que muestran valores predictivos individuales semejantes o superiores a la validez predictiva de las Notas de Enseñanza Media. Por su parte, las pruebas de Lenguaje y Ciencias Sociales presentan valores promedio positivos, pero marcadamente más bajos. Finalmente, ambos estudios concluyen que la comparación entre las dos baterías de selección es favorable a las nuevas pruebas en la gran mayoría de las universidades, carreras y áreas de estudio. (Manzi, et al., 2006, 2008)

El trabajo que se presenta a continuación constituye un complemento a los estudios recién citados, al incorporar los análisis de sesgo de validez y predictividad de los factores de selección universitaria. Las características de agrupación aquí consideradas son el género del estudiante y la Dependencia<sup>2</sup> del establecimiento de egreso de la Educación Media.

## 1. ASPECTOS METODOLÓGICOS

### 1.1. Bases de Datos

Los estudios de validez predictiva se basan en analizar la relación que existe entre ciertas variables predictoras (los factores de selección) y alguna variable de criterio, normalmente relacionada con el desempeño académico. En este trabajo, la variable criterio utilizada corresponde al Promedio Ponderado

<sup>1</sup> El Sistema de Selección se basa en la asignación de un puntaje que combina estos factores con ponderadores diferentes dependiendo de cada carrera.

<sup>2</sup> En Chile, los establecimientos escolares se clasifican en 3 Dependencias, de acuerdo al tipo de administración y financiamiento. Los Particulares Pagados son de financiamiento privado, mientras que los Particulares Subvencionados y Municipales reciben subvención del Estado. Estos últimos se diferencian en que los Particulares Subvencionados son de administración privada mientras que los Municipales son establecimientos públicos. En Chile, dada la alta segregación educativa que existe, la Dependencia suele considerarse como proxy del nivel socioeconómico del estudiante.

Acumulado (PPA) de los estudiantes al término del primer año universitario. Las variables predictoras consideradas incluyen las pruebas de selección obligatorias de Matemática y Lenguaje y Comunicación (PSUM/PAAM y PSUL/PAAV), el puntaje por notas de educación media (NEM), y el Puntaje de Selección (PSEL) con el cual cada alumno es admitido a su carrera.

El dato de rendimiento académico fue obtenido a partir de las bases entregadas por las 25 universidades del Consejo de Rectores para los estudios de Validez Predictiva llevados a cabo por el CTA. Esta información fue complementada con las bases de datos del DEMRE<sup>3</sup>, para obtener los antecedentes relativos a los resultados en los factores de selección (puntajes de pruebas y NEM) y a las características de género y dependencia del establecimiento de egreso de estos estudiantes. Los Puntajes de Selección fueron computados a partir de los puntajes y de los ponderadores de los factores de selección, según la carrera a la que es admitido cada alumno.

En total, se contó con 214.259 datos válidos de notas y puntajes de selección para las admisiones 2003 (última PAA) a 2007. La distribución de alumnos por año de ingreso según género y dependencia se presenta en la Tabla 3.

TABLA 3. DISTRIBUCIÓN DE ALUMNOS CONSIDERADOS EN EL ESTUDIO

Grupo	Año admisión					Total
	2003	2004	2005	2006	2007	
Hombre	22.955	20.554	21.576	22.445	23.645	111.175
	53.7%	51.2%	52.2%	51.5%	50.9%	
Mujer	19.770	19.606	19.758	21.130	22.820	103.084
	46.3%	48.8%	47.8%	48.5%	49.1%	
Particular Pagado	8.745	7.824	7.733	7.868	7.999	40.169
	20.5%	19.5%	18.7%	18.1%	17.3%	
Particular Subvencionado	16.555	16.421	17.564	18.900	2.1125	90.565
	38.9%	41.0%	42.6%	43.5%	45.6%	
Municipal	17.297	15.805	15.944	16.707	17.223	82.976
	40.6%	38.5%	38.7%	38.4%	37.1%	
Total	42.725	40.160	41.334	43.575	46.465	214.259

*Nota:* En las bases de datos empleadas, el dato de género estaba presente para todos los casos, no así el de dependencia del establecimiento de egreso. Es por ello que el número de casos por dependencia es algo menor que el número de casos totales. El porcentaje es sobre el total de casos con dato de dependencia.

## 2. ASPECTOS ANALÍTICOS

### 2.1. Validez Diferencial

Para determinar la relación entre cada una de las variables predictoras y el rendimiento académico, se calcularon correlaciones simples (correlación Producto-Momento de Pearson) separadamente para cada subgrupo de estudiantes. Con esta metodología, la validez diferencial de los factores de selección se manifiesta cuando los coeficientes de correlación son sustancialmente diferentes entre ciertos subgrupos.

<sup>3</sup> El DEMRE es un organismo dependiente de la U. de Chile que está encargado del desarrollo y construcción de las PSU y de la administración del Sistema de Selección a la Educación Superior.

Es importante destacar que, de modo de asegurar condiciones de homogeneidad para la variable criterio, se decidió emplear a cada carrera como unidad de análisis básica. Con ello, valores diferentes en los coeficientes de correlación calculados revelarían la existencia de validez diferencial y no otras diferencias en las experiencias educacionales que podrían darse entre carreras.

Por otra parte, dado que el número de estudiantes válidos por carrera es en algunos casos muy pequeño, se establecieron límites inferiores de manera de evitar que las estimaciones de la capacidad predictiva de los factores de selección fueran imprecisas. Así, para las estimaciones de los coeficientes de correlación se exigieron al menos 10 casos válidos por subgrupo en cada carrera. En consecuencia, el número de carreras sobre el cual se calcula el sesgo es inferior al total de carreras disponibles para el análisis.

Los resultados que se presentan corresponden a promedios de estas correlaciones a nivel de carrera. La metodología para promediar correlaciones se basa en el método de Hedges-Olkin, que consiste en calcular el promedio ponderado (en función del tamaño de cada muestra) de las transformaciones Z para las correlaciones individuales. Este resultado es reconvertido en un índice de correlaciones mediante la transformación inversa tangente hiperbólica ( $\bar{r} = \tanh(\bar{Z})$ ).

Finalmente, es importante advertir que, dado que los factores de selección son usados para tomar decisiones de admisión, se produce una inevitable restricción en el rango de los valores de ellos, en la medida que sólo se dispone de información para los postulantes que fueron aceptados a las universidades. Esta condición lleva a una subestimación de la capacidad predictiva de estos factores, pues solo se analiza la porción de su variabilidad que queda disponible una vez hecha la selección. No obstante lo anterior, los coeficientes de correlación pueden ser utilizados como base para realizar inferencias comparativas entre poblaciones distintas (Wainer, Saka & Donoghue, 1993).

Alternativamente, es posible usar algún método estadístico que permita corregir la restricción de rango en las correlaciones. El método más usado en estudios de Validez Predictiva corresponde a la corrección multivariada de Pearson-Lawley, conocida también como la corrección de Gulliksen (Gulliksen, 1950). Éste consiste en estimar las correlaciones en la población restringida (la muestra de estudiantes universitarios) y corregir posteriormente estas correlaciones, considerando las desviaciones estándar de los factores de selección para la población completa de examinados. La fórmula utilizada en este método se presenta a continuación:

$$r_{corr} = \frac{\sigma_x r_{xy}}{\sqrt{\sigma_x^2 (1 - r_{xy}^2) + \sigma_x^2 r_{xy}^2}}$$

Donde:

$r_{corr}$  = Correlación corregida por restricción de rango.

$r_{xy}$  = Correlación original.

$\sigma_x$  = Desviación estándar de puntajes en población restringida (universitarios).

$\sigma_x$  = Desviación estándar en población no restringida (todos los examinados).

Las correlaciones presentadas en el trabajo de Mattern et. al. (2008) incluyen este tipo de corrección. Por lo tanto, y con el objeto de poder comparar los resultados para las PSU con la evidencia internacional, en el presente estudio se incluyen ambos resultados: coeficientes de correlación con y sin correcciones por restricción de rango.

## 2.2. Predicción diferencial

Para analizar el sesgo predictivo, se ajustó una regresión lineal común para todos los estudiantes de una misma carrera, y se consideró una ecuación para cada predictor en forma separada: Prueba de Matemática, Prueba de Lenguaje y Comunicación, NEM y Puntaje de Selección. Posteriormente se comparó el promedio de los residuos de estas regresiones (es decir, la diferencia entre las notas obtenidas y las predichas por el modelo) para cada subgrupo. Sólo se consideraron para el promedio, carreras con al menos 10 casos en cada grupo.

Para facilitar la combinación de resultados, previo al ajuste de los modelos, las notas de primer año se estandarizaron con media 0 y desviación estándar 1. De esta forma, los residuos de cada regresión estimada también se encuentran en esta escala estandarizada. Calculados de esta manera, la interpretación de los resultados es bastante sencilla: el promedio de los residuos para la muestra completa es igual a cero, por lo tanto, cuando el promedio de los residuos para algún subgrupo es distinto de cero, el instrumento presenta predicción diferencial. Así por ejemplo, si el promedio de los residuos es positivo para alguno de los grupos, entonces la prueba tiende a subpredecir el rendimiento académico de ese grupo. En otras palabras, los alumnos de dicho grupo tienden a tener un desempeño mejor al que el modelo de regresión común predice. Análogamente, si el promedio de los residuos es negativo para alguno de los grupos, entonces la prueba tiende a sobrepredecir el rendimiento académico de dicho grupo. Por otra parte, el sesgo predictivo, calculado como la diferencia entre los promedios de los residuos para cada grupo, tiene valores que están en términos de las desviaciones estándar de las notas de primer año. Por lo tanto, este método tiene una interpretación directa en cuanto a evaluar el sesgo en la capacidad predictiva del instrumento, lo que representa una ventaja respecto a la comparación de las correlaciones entre desempeño y factores de selección.

## 3. RESULTADOS

### 3.1. Descriptivos

Las tablas siguientes muestran los promedios por género y dependencia del establecimiento de egreso para los puntajes obtenidos en los diferentes factores de selección

#### 3.1.1. Puntajes según Género

En la prueba de Lenguaje se observan resultados muy parejos entre hombres y mujeres para todas las admisiones. Distinto es el caso de la Prueba de Matemáticas donde aparece sistemáticamente una brecha en los resultados a favor de los hombres. Es importante advertir que las diferencias de puntajes entre años no son comparables directamente ya que sólo a partir de la admisión 2005 comenzaron a estandarizarse los puntajes de las PSU<sup>4</sup>. En términos de las desviaciones estándar de cada prueba, las diferencias varían relativamente poco, entre un 25% y un 31% y además, estas diferencias son muy similares a las que se observan en la población completa de examinados en cada admisión. Por otra parte, en el puntaje por NEM, son las mujeres las que obtienen mejores resultados, con una ventaja de

<sup>4</sup> A partir del año 2005 los puntajes son convertidos a escalas normalizadas, con media 500 y desviación estándar 110. Los años 2003 y 2004 las desviaciones estándar para la prueba de matemática fueron de 140 y 109, respectivamente, y para la de lenguaje, de 122 en ambos años de admisión.

entre 30 y 40 puntos promedio. Pese a estas diferencias en algunos factores, una vez combinados, los promedios en puntajes de selección son bastante homogéneos para todas las admisiones (con una diferencia a favor de los hombres de 4 puntos como máximo).

**TABLA 4. PROMEDIOS EN FACTORES DE SELECCIÓN SEGÚN GÉNERO. ADMISIONES 2003 A 2007**

Admisión	N carreras	Género	N alumnos	Lenguaje		Matemática		NEM		PSEL	
				Media	D.E.	Media	D.E.	Media	D.E.	Media	D.E.
2003	776	Hombres	22955	611	90	665	96	601	96	629	67
		Mujeres	19770	612	85	630	97	645	85	625	64
2004	816	Hombres	20554	598	99	611	88	612	91	611	72
		Mujeres	19606	603	93	577	80	648	81	609	65
2005	850	Hombres	21576	591	85	615	84	616	92	613	71
		Mujeres	19758	592	81	583	78	649	83	609	65
2006	890	Hombres	22445	590	85	616	85	615	91	612	72
		Mujeres	21130	594	81	583	78	646	82	609	65
2007	941	Hombres	23645	597	83	619	82	615	91	616	70
		Mujeres	22820	599	80	590	75	644	83	612	65

### 3.1.2. Puntajes según Dependencia

Al comparar los resultados según dependencia de egreso, se observa que los alumnos provenientes de establecimientos Particulares Pagados obtienen, en promedio, mejores puntajes en todos los factores de selección.

**TABLA 5. PROMEDIOS EN FACTORES DE SELECCIÓN SEGÚN DEPENDENCIA. ADMISIONES 2003 A 2007.**

Admisión	N carreras	Dependencia	N alumnos	Lenguaje		Matemática		NEM		PSEL	
				Media	D.E.	Media	D.E.	Media	D.E.	Media	D.E.
2003	178	Particular Pagado	8745	661	78	714	83	648	90	674	62
		Particular Subvencionado	16555	606	83	643	91	621	92	623	60
		Municipal	17297	592	87	621	96	608	94	608	62
2004	149	Particular Pagado	7824	659	82	660	86	657	85	661	67
		Particular Subvencionado	16421	597	90	586	75	628	88	604	62
		Municipal	15805	575	97	571	80	618	87	591	63
2005	145	Particular Pagado	7733	648	80	665	82	663	84	664	67
		Particular Subvencionado	17564	586	76	591	73	627	89	604	62
		Municipal	15944	570	79	577	77	621	88	593	62
2006	154	Particular Pagado	7868	653	80	671	82	662	85	668	69
		Particular Subvencionado	18900	586	76	591	72	626	88	603	62
		Municipal	16707	570	78	577	77	620	87	592	62
2007	148	Particular Pagado	7999	660	76	674	78	661	84	671	66
		Particular Subvencionado	21125	593	75	598	71	625	87	608	61
		Municipal	17223	576	78	580	74	619	87	595	61



Las mayores diferencias se dan entre establecimientos Particulares Pagados y Municipales, en la Prueba de Matemática (alrededor de 90 puntos, poco más de 80% de una desviación estándar de las PSU) y en la de Lenguaje y Comunicación (alrededor de 80 puntos, 75% de una desviación estándar). Esta brecha disminuye en el caso de las NEM. Las diferencias entre Particulares Subvencionados y Municipales, si bien son estables en el tiempo, son bastante menores: poco menos de 20 puntos en el caso de las pruebas obligatorias y de 10 puntos en el caso de las NEM.

En ambas comparaciones por dependencia, las brechas presentan las mismas tendencias que las observadas en la población completa de examinados que rinden las PSU. Sin embargo, las magnitudes de estas brechas son algo menores<sup>5</sup> en el caso de los alumnos admitidos a universidades del Consejo de Rectores.

Para el caso del Puntaje de Selección, se mantienen los mismos patrones anteriores (con una diferencia de 70 puntos entre Particulares Pagados y Municipales y una diferencia de 15 puntos entre Particulares Subvencionados y Municipales).

### 3.2. Validez Diferencial

#### 3.2.1. Correlaciones Simples con Notas de Primer Año (PPA)

La siguiente tabla muestra las correlaciones promedio entre factores de selección y notas de primer año (PPA). Los resultados, que coinciden con los presentados en Estudios de Validez Predictiva de las PSU, muestran una mayor capacidad predictiva promedio para la PSU de Matemática, la que presenta correlaciones semejantes a la validez predictiva de las notas de enseñanza media. La Prueba de Lenguaje también presenta valores promedio positivos, pero marcadamente más bajos. En ambos subsectores las PSU presentan correlaciones con el PPA superiores a las obtenidas para la última PAA (Manzi et al, 2006).

TABLA 6. CORRELACIONES PROMEDIO DE LOS FACTORES DE SELECCIÓN CON PPA. ADMISIONES 2003 A 2007.

Año Admisión	N carreras	N alumnos	Lenguaje	Matemática	NEM	PSEL
2003	766	42725	0.02	0.13	0.25	0.30
2004	777	40160	0.08	0.29	0.27	0.39
2005	824	41334	0.11	0.25	0.26	0.38
2006	851	43575	0.10	0.24	0.24	0.36
2007	918	46465	0.10	0.25	0.25	0.36

#### 3.2.2. Validez Diferencial según Género

En el caso de las comparaciones por género, no se observan diferencias importantes en las correlaciones calculadas para ninguno de los factores. Los resultados son similares a los obtenidos al considerar la muestra completa en cada carrera (tabla 4).

<sup>5</sup> Las brechas entre Particulares Pagados y Municipales se reducen en alrededor de 20 a 35 puntos, (lo que se traduce en reducciones que varían entre un 20 y 30% dependiendo del año y prueba), mientras que las brechas entre Particulares Subvencionados y Municipales se reducen en alrededor de 10 puntos, (lo que se traduce en reducciones que varían entre un 10 y 40% dependiendo del año y prueba)

**TABLA 7. CORRELACIONES PROMEDIO DE LOS FACTORES DE SELECCIÓN CON PPA SEGÚN GÉNERO. ADMISIONES 2003 A 2007.**

Año admisión	N carreras	Género	N alumnos	Lenguaje	Matemática	NEM	PSEL
2003	519	Hombres	17825	0.02	0.15	0.23	0.31
		Mujeres	15093	0.03	0.14	0.26	0.31
2004	469	Hombres	15447	0.09	0.30	0.26	0.40
		Mujeres	14131	0.08	0.29	0.28	0.39
2005	483	Hombres	15560	0.12	0.27	0.26	0.39
		Mujeres	14438	0.12	0.25	0.26	0.37
2006	534	Hombres	16883	0.10	0.25	0.23	0.36
		Mujeres	15938	0.10	0.27	0.24	0.37
2007	572	Hombres	17836	0.10	0.26	0.24	0.36
		Mujeres	17422	0.11	0.26	0.24	0.36

### 3.2.3. Validez Diferencial según Dependencia

En el caso de la PSU de Lenguaje y Comunicación, no pareciera haber importantes diferencias en las magnitudes de las correlaciones con el PPA entre grupos de estudiantes egresados de distintas dependencias.

En Matemática, en cambio, se observan correlaciones relativamente más altas para Particulares Subvencionados y Municipales que para Particulares Pagados, en los dos primeros años de las PSU (2004 y 2005). Las diferencias se atenúan los siguientes dos años de admisión. Para la última PAA de Matemática (2003) tampoco se aprecian diferencias en las correlaciones aunque éstas son marcadamente más bajas que las observadas para las PSU.

**TABLA 8. CORRELACIONES PROMEDIOS DE LOS FACTORES DE SELECCIÓN CON PPA SEGÚN DEPENDENCIA. ADMISIONES 2003 A 2007.**

Año admisión	N carreras	Dependencia	Lenguaje	Matemática	NEM	PSEL
2003	178	Particular Pagado	0.06	0.13	0.31	0.37
		Particular Subvencionado	0.00	0.14	0.30	0.32
		Municipal	0.05	0.13	0.20	0.31
2004	149	Particular Pagado	0.09	0.23	0.36	0.41
		Particular Subvencionado	0.06	0.28	0.28	0.39
		Municipal	0.07	0.27	0.23	0.38
2005	145	Particular Pagado	0.10	0.18	0.37	0.38
		Particular Subvencionado	0.10	0.26	0.28	0.39
		Municipal	0.10	0.24	0.19	0.35
2006	154	Particular Pagado	0.11	0.20	0.33	0.40
		Particular Subvencionado	0.07	0.22	0.23	0.34
		Municipal	0.07	0.24	0.18	0.35
2007	148	Particular Pagado	0.10	0.22	0.36	0.41
		Particular Subvencionado	0.08	0.22	0.23	0.34
		Municipal	0.06	0.24	0.15	0.32

En cuanto a las NEM, se observa un patrón diferente. En la última admisión de las PAA, las correlaciones para el grupo de Particulares Pagados y Particulares Subvencionados son del orden de 0.3, mientras que para los Municipales la correlación promedio con el PPA resulta bastante más baja (de 0.2). Durante las admisiones siguientes (baterías PSU), se observa una diferencia marcada entre los tres grupos: las correlaciones más altas se dan para Particulares Pagados, seguidas por las del grupo de Particulares Subvencionados y, finalmente, los Municipales que presentan correlaciones bastante más bajas.

En el Puntaje de Selección, se observa una tendencia de correlaciones levemente más altas para el grupo de Particulares Pagados que para Particulares Subvencionados y Municipales.

### 3.2.4. Correlaciones Corregidas por Restricción de Rango

Las siguientes tablas muestran las correlaciones de las pruebas obligatorias y NEM, corregidas por la restricción de rango. En el caso del Puntaje de Selección no es posible llevar a cabo esta corrección ya que sólo es posible determinar su varianza para los alumnos que fueron efectivamente admitidos a alguna carrera.

En términos generales, al comparar con las correlaciones sin corrección (tabla 9), notamos que el efecto de la restricción de rango es más importante en las pruebas de Lenguaje y Matemática que en el caso de las NEM.

**TABLA 9. CORRELACIONES PROMEDIOS DE LOS FACTORES DE SELECCIÓN CON PPA, CORREGIDAS POR RESTRICCIÓN DE RANGO. ADMISIONES 2003 A 2007.**

Año Admisión	N carreras	N alumnos	Lenguaje	Matemática	NEM
2003	766	42725	0.04	0.33	0.37
2004	777	40160	0.15	0.60	0.42
2005	824	41334	0.23	0.55	0.41
2006	851	43575	0.21	0.55	0.39
2007	918	46465	0.22	0.57	0.40

### 3.2.5. Validez Diferencial según Género, con correcciones por Restricción de Rango

Una vez corregidas por la restricción de rango, las diferencias en las correlaciones entre hombres y mujeres aumentan un poco en algunos años, pero las NEM son el único factor que presenta una tendencia más clara en todas las admisiones, siendo las correlaciones con las notas de primer año algo mayores para las mujeres. Este efecto tiende a disminuir para los últimos años.

Estos resultados son distintos a los presentados por Mattern et. al. (2008), donde se obtienen correlaciones con las secciones del SAT que son más altas para las mujeres, mientras que las correlaciones con las notas del colegio son relativamente similares (ver tabla 1)

**TABLA 10. CORRELACIONES PROMEDIOS DE LOS FACTORES DE SELECCIÓN CON PPA SEGÚN GÉNERO, CORREGIDAS POR RESTRICCIÓN DE RANGO. ADMISIONES 2003 A 2007**

Año admisión	N carreras	Género	N alumnos	Lenguaje	Matemática	NEM
2003	519	Hombres	17825	0.06	0.39	0.35
		Mujeres	15093	0.09	0.33	0.44
2004	469	Hombres	15447	0.19	0.61	0.41
		Mujeres	14131	0.17	0.61	0.48
2005	483	Hombres	15560	0.26	0.57	0.40
		Mujeres	14438	0.26	0.54	0.44
2006	534	Hombres	16883	0.23	0.56	0.38
		Mujeres	15938	0.23	0.59	0.42
2007	572	Hombres	17836	0.23	0.59	0.40
		Mujeres	17422	0.26	0.58	0.42

### 3.2.6. Validez Diferencial según Dependencia, con correcciones por Restricción de Rango

Al comparar por dependencia también se obtiene un patrón sistemático en las diferencias en las correlaciones entre notas de primer año y las NEM. En este caso, las diferencias son más importantes que en el caso de las comparaciones por género y mantienen la misma tendencia que en el caso de las correlaciones sin corrección por restricción de rango: las más altas se dan en el grupo de Particulares Pagados, seguido por el de Particulares Subvencionados y finalmente, las correlaciones más bajas se dan para el grupo proveniente de establecimientos Municipales. Este resultado es concordante con el patrón encontrado por Mattern et al. (2008) para las minorías raciales y étnicas (en todos los factores de selección). Sin embargo, en el caso de las NEM, las diferencias son algo mayores que las reportadas en el estudio de Mattern et al. (2008) para las notas de colegio: correlación para estudiantes blancos de 0.56, para Hispanos de 0.46 y para negros de 0.44 (ver tabla 1).

En la Prueba de Matemática, el año 2003 (PAA) se observa una mayor correlación para Particulares Pagados y correlaciones similares para los otros dos grupos de dependencia. En cambio, para las PSU de Matemática se observa un efecto contrario: mayores correlaciones para Particulares Subvencionados y Municipales que para Particulares Pagados, aunque este efecto tiende a atenuarse los últimos dos años.

Respecto a la Prueba de Lenguaje y Comunicación, los resultados son menos concluyentes y sólo pareciera darse una tendencia (que sería similar a la de los SAT y PAAM) en las últimas dos admisiones: correlaciones algo mayores para Particulares Pagados que para Particulares Subvencionados y Municipales.

**TABLA 11. CORRELACIONES PROMEDIOS DE LOS FACTORES DE SELECCIÓN CON PPA SEGÚN DEPENDENCIA, CORREGIDAS POR RESTRICCIÓN DE RANGO. ADMISIONES 2003 A 2007**

Año admisión	N carreras	Dependencia	Lenguaje	Matemática	NEM
2003	178	Particular Pagado	0.14	0.42	0.53
		Particular Subvencionado	0.02	0.32	0.45
		Municipal	0.11	0.31	0.29
2004	149	Particular Pagado	0.20	0.51	0.62
		Particular Subvencionado	0.12	0.57	0.45
		Municipal	0.16	0.55	0.36
2005	145	Particular Pagado	0.21	0.40	0.64
		Particular Subvencionado	0.19	0.54	0.46
		Municipal	0.21	0.51	0.28
2006	154	Particular Pagado	0.25	0.47	0.60
		Particular Subvencionado	0.16	0.49	0.39
		Municipal	0.16	0.53	0.29
2007	148	Particular Pagado	0.22	0.50	0.65
		Particular Subvencionado	0.18	0.50	0.41
		Municipal	0.17	0.54	0.26

## 3.3. Predicción diferencial

### 3.3.1. Predicción Diferencial según Género

A partir de una regresión común para cada carrera, se estimaron los residuos estandarizados para cada alumno, como la diferencia entre las notas obtenidas y las predichas por cada modelo de regresión. En la tabla 12 se presentan los promedios de estos residuos para cada uno de los factores, según género.

Se observan, para las pruebas obligatorias, promedios negativos para los estudiantes hombres y promedios positivos para las mujeres. Esto se traduce en un sesgo predictivo en favor del género masculino, que en el caso de la prueba de Matemática es de 0.2 desviaciones estándar y en el caso de Lenguaje y Comunicación es de al menos 0.12 desviaciones estándar. En ambos subsectores los resultados son bastante estables entre admisiones, excepto para la prueba de lenguaje del año 2003 (última versión de la PAA Verbal). En dicha admisión se observa un sesgo predictivo que es levemente superior a los obtenidos posteriormente con las PSU de Lenguaje y Comunicación (0.17 vs. 0.12 a 0.14 desviaciones estándar).

Estos resultados son consistentes con los patrones encontrados en el estudio de Mattern et al. (2008), aunque las magnitudes en los sesgos en el caso de las PSU son un poco inferiores a los reportados en dicho trabajo para el SAT (en la tabla 2, la diferencia absoluta entre las predicciones diferenciales de hombres y mujeres es de 0.37 desviaciones estándar para el SAT Matemáticas, de 0.26 para el SAT-CR y de 0.22 para el SAT-W; todos estos sesgos son a favor de los estudiantes hombres).

En el caso de las Notas de Enseñanza Media también se observa un sesgo predictivo a favor de los hombres, sin embargo, en este caso, la magnitud es sustancialmente inferior y no supera las 0.04 desviaciones estándar.

Finalmente, el Puntaje de Selección presenta, en promedio, un sesgo predictivo a favor de los hombres con un patrón muy similar a la prueba de Lenguaje y Comunicación.

**TABLA 12. PROMEDIOS DE RESIDUOS ESTANDARIZADOS POR GÉNERO. ADMISIONES 2003 A 2007**

Año admisión	N carreras	Género	N alumnos	Lenguaje	Matemática	NEM	PSEL
2003	519	Hombres	17825	-0.08	-0.09	-0.02	-0.08
		Mujeres	15093	0.09	0.11	0.02	0.09
2004	469	Hombres	15447	-0.06	-0.09	0.00	-0.06
		Mujeres	14131	0.06	0.10	0.00	0.06
2005	483	Hombres	15560	-0.06	-0.09	-0.01	-0.06
		Mujeres	14438	0.07	0.10	0.01	0.06
2006	534	Hombres	16883	-0.06	-0.10	-0.01	-0.06
		Mujeres	15938	0.07	0.10	0.02	0.06
2007	572	Hombres	17836	-0.07	-0.10	-0.02	-0.07
		Mujeres	17422	0.07	0.10	0.02	0.07

### 3.3.2. Predicción Diferencial según Género, por área del conocimiento

Las siguientes tablas presentan la magnitud total de la predicción diferencial entre hombres y mujeres, por área de conocimiento, para las admisiones de las baterías PSU. Los resultados, aunque varían un poco en algunos años, presentan tendencias bastante estables.

En el caso de la prueba de Matemática, son las áreas de Ciencias Sociales, Economía y Administración y Educación las que presentan mayor sesgo de predicción a favor de los hombres. Las áreas con menor sesgo son las de Arte, Ingeniería y Ciencias.

En la Prueba de Lenguaje y Comunicación, también son las áreas de Ciencias Sociales y Economía y Administración las que presentan mayor sesgo de predicción a favor de los hombres, mientras que las de Arte, Ingeniería y Ciencias, son las que tienen menores sesgos.

**TABLA 13. DIFERENCIAS EN LOS PROMEDIOS DE LOS RESIDUOS ESTANDARIZADOS ENTRE HOMBRES Y MUJERES, POR ÁREA. ADMISIONES 2003 A 2007. CASO PSU MATEMÁTICA.**

Área del Conocimiento	Año de Admisión			
	2004	2005	2006	2007
Agronomía-Forestal	-0.24	-0.18	-0.21	-0.20
Arte	-0.12	-0.04	-0.09	-0.12
Ciencias	-0.13	-0.12	-0.21	-0.16
Ciencias Sociales	-0.31	-0.30	-0.27	-0.34
Economía-Administración	-0.26	-0.35	-0.36	-0.30
Educación	-0.23	-0.27	-0.29	-0.24
Humanidades	-0.27	-0.16	-0.22	-0.21
Ingeniería	-0.14	-0.11	-0.13	-0.16
Salud	-0.20	-0.22	-0.16	-0.23
Tecnología	-0.14	-0.18	-0.17	-0.24

**TABLA 14. DIFERENCIAS EN LOS PROMEDIOS DE LOS RESIDUOS ESTANDARIZADOS ENTRE HOMBRES Y MUJERES, POR ÁREA. ADMISIONES 2003 A 2007. CASO PSU LENGUAJE.**

Área del Conocimiento	Año de Admisión			
	2004	2005	2006	2007
Agronomía-Forestal	-0.10	-0.07	-0.12	-0.07
Arte	-0.10	-0.01	-0.04	-0.09
Ciencias	0.00	-0.02	-0.12	-0.06
Ciencias Sociales	-0.28	-0.28	-0.23	-0.31
Economía-Administración	-0.20	-0.30	-0.30	-0.24
Educación	-0.16	-0.22	-0.23	-0.20
Humanidades	-0.23	-0.13	-0.19	-0.19
Ingeniería	-0.02	-0.01	0.00	-0.06
Salud	-0.17	-0.20	-0.12	-0.21
Tecnología	-0.13	-0.11	-0.06	-0.14

En el caso de las NEM los sesgos predictivos por área son más bajos que en el caso de las PSU, y no todas son a favor de los hombres. Las áreas de Ciencias Sociales y de Economía y Administración son las que presentan mayor sesgo a favor de los hombres mientras que el área de las Ingenierías presenta un leve sesgo a favor de las mujeres.

**TABLA 15. DIFERENCIAS EN LOS PROMEDIOS DE LOS RESIDUOS ESTANDARIZADOS ENTRE HOMBRES Y MUJERES, POR ÁREA. ADMISIONES 2003 A 2007. CASO NEM.**

Área del Conocimiento	Año de Admisión			
	2004	2005	2006	2007
Agronomía-Forestal	0.06	0.07	0.00	0.00
Arte	0.04	0.06	0.06	0.00
Ciencias	0.12	0.10	-0.02	0.06
Ciencias Sociales	-0.13	-0.15	-0.12	-0.19
Economía-Administración	-0.06	-0.15	-0.14	-0.09
Educación	-0.02	-0.10	-0.10	-0.11
Humanidades	-0.06	0.04	-0.04	-0.05
Ingeniería	0.08	0.07	0.07	0.03
Salud	-0.05	-0.10	-0.05	-0.12
Tecnología	0.04	-0.01	-0.02	-0.10

Finalmente, respecto de los Puntajes de Selección, se observa un patrón de resultados similar al de la PSU de Lenguaje, donde las áreas de Ciencias Sociales y de Economía y Administración son las que presentan

mayor sesgo de predicción a favor de los hombres, mientras que las de Arte, Ingeniería y Ciencias, son las que tienen menores sesgos.

**TABLA 16. DIFERENCIAS EN LOS PROMEDIOS DE LOS RESIDUOS ESTANDARIZADOS ENTRE HOMBRE Y MUJERES, POR ÁREA. ADMISIONES 2003 A 2007. CASO PUNTAJE DE SELECCIÓN.**

Área del Conocimiento	Año de Admisión			
	2004	2005	2006	2007
Agronomía-Forestal	-0.11	-0.09	-0.11	-0.10
Arte	-0.06	0.00	-0.05	-0.08
Ciencias	0.00	-0.02	-0.12	-0.07
Ciencias Sociales	-0.28	-0.30	-0.22	-0.28
Economía-Administración	-0.18	-0.26	-0.26	-0.20
Educación	-0.16	-0.21	-0.21	-0.20
Humanidades	-0.23	-0.10	-0.18	-0.18
Ingeniería	0.00	0.01	-0.02	-0.05
Salud	-0.16	-0.18	-0.12	-0.20
Tecnología	-0.08	-0.08	-0.03	-0.10

### 3.3.3. Predicción Diferencial según Dependencia

Los promedios por dependencia de los residuos estandarizados revelan una subpredicción en el rendimiento de estudiantes provenientes de establecimientos particulares pagados y una sobrepredicción del rendimiento para alumnos de colegios municipales en la Prueba de Lenguaje, aunque con un sesgo predictivo relativamente bajo a favor de los establecimientos municipales (de 0.08 desviaciones estándar como máximo).

Los sesgos predictivos son aún más bajos en la Prueba de Matemáticas y en el Puntaje de Selección (diferencia que no supera las 0.04 desviaciones estándar), y no se observa una tendencia clara a favor o en contra de alguno de los grupos de dependencia.

**TABLA 17. PROMEDIOS DE RESIDUOS ESTANDARIZADOS POR DEPENDENCIA. ADMISIONES 2003 A 2007**

Año admisión	N carreras	Dependencia	Lenguaje	Matemática	NEM	PSEL
2003	178	Particular Pagado	0.01	-0.01	0.04	-0.02
		Particular Subvencionado	0.01	0.01	0.01	0.02
		Municipal	-0.02	-0.01	-0.02	0.00
2004	149	Particular Pagado	0.05	0.00	0.07	0.01
		Particular Subvencionado	0.01	0.01	0.01	0.01
		Municipal	-0.03	-0.01	-0.04	-0.02
2005	145	Particular Pagado	0.04	0.00	0.07	0.01
		Particular Subvencionado	0.01	0.01	0.02	0.02
		Municipal	-0.03	-0.01	-0.05	-0.02
2006	154	Particular Pagado	0.03	-0.02	0.07	0.00
		Particular Subvencionado	0.00	0.01	0.01	0.01
		Municipal	-0.02	0.00	-0.04	-0.01
2007	148	Particular Pagado	0.04	-0.02	0.08	0.00
		Particular Subvencionado	0.01	0.01	0.02	0.02
		Municipal	-0.03	-0.01	-0.05	-0.02

Las NEM constituyen el factor que presenta mayor sesgo predictivo (de hasta 0.13 desviaciones estándar), con un sesgo positivo a favor de los establecimientos municipales. Esto significa que, en promedio, un alumno egresado de colegio municipal tendrá un menor desempeño relativo en la universidad en relación con un alumno de colegio Particular Pagado con igual NEM. Estas diferencias podrían explicarse por mecanismos diferentes de asignación de notas entre colegios de distinta dependencia. Sin embargo, en términos de equidad, estos resultados se traducen en una ventaja a favor de los alumnos de establecimientos municipales, ya que incrementa en forma artificial sus puntajes de postulación a la universidad.

#### 4. CONCLUSIONES

En un sistema de admisión universitario es primordial contar con un mecanismo de selección que sea equitativo para todos los postulantes. En Chile, la selección de estudiantes a las universidades del Consejo de Rectores se realiza en base a los resultados obtenidos en una batería compuesta de pruebas estandarizadas (PSU, desde el año 2004) y las notas de la Enseñanza Media.

Para los años de admisión considerados en este estudio (2003 a 2007), las brechas en resultados en las pruebas de selección, entre los estudiantes admitidos a las universidades del Consejo de Rectores, siguen las mismas tendencias que las que se observan en el universo de postulantes que las rinden. Esto es, un mejor desempeño relativo de hombres versus mujeres en la Prueba de Matemáticas y mejores desempeños relativos de alumnos provenientes de establecimientos Particulares Pagados versus Particulares Subvencionados y Municipales, en ambas pruebas obligatorias.

Sin embargo, y más allá de las diferencias en resultados, en términos de equidad del sistema de admisión, lo que se espera es que los factores de selección sean capaces de discriminar correctamente a aquellos alumnos que tienen una mayor probabilidad de éxito en su desempeño académico universitario. Para que ello ocurra, la capacidad predictiva de los instrumentos de selección debe ser insesgada respecto a características sociodemográficas de los postulantes. De no ser así, se corre el riesgo de que los puntajes de selección estén siendo usados de un modo que, en definitiva, resulta consistentemente inequitativo para ciertos grupos.

El objetivo de este estudio era evaluar posibles sesgos en la capacidad predictiva de los factores de selección obligatorios a las universidades del Consejo de Rectores (Prueba de Matemática, Prueba de Lenguaje y Comunicación (Verbal) y NEM), además del Puntaje de Selección de cada carrera. Se emplearon el género y dependencia de los establecimientos educacionales como variables de agrupamiento de los estudiantes.

En los resultados obtenidos de los análisis de Validez Diferencial, que se basaron en la comparación de las correlaciones entre los factores de selección y el rendimiento académico universitario de primer año, **no se evidencian diferencias importantes en las comparaciones por género para ninguno de los factores.** Al corregir por restricción de rango, las diferencias en las correlaciones entre hombres y mujeres aumentan un poco en algunos años, pero es el puntaje por NEM el único factor que presenta una tendencia más clara en todas las admisiones, con correlaciones un poco mayores para las mujeres. Este efecto, sin embargo, tiende a disminuir para los últimos años de admisión.

En relación a las comparaciones por dependencia, aunque se observan diferencias en las magnitudes de las correlaciones con las pruebas obligatorias en algunos años, **no se encuentra ningún patrón**



**sistemático que evidencie la existencia de validez diferencial.** En el caso de las NEM ocurre algo diferente. En la última admisión del Sistema PAA, las correlaciones para el grupo de Particulares Pagados y Particulares Subvencionados son similares y algo mayores que para el grupo de Municipales. Sin embargo, durante las admisiones siguientes (baterías PSU), se observa una diferencia marcada entre los tres grupos: las correlaciones más altas se dan para Particulares Pagados, seguidas por las del grupo de Particulares Subvencionados, y finalmente, los Municipales, que presentan correlaciones bastante más bajas. Estos patrones se mantienen bastante similares después de corregir por la restricción de rango.

En términos de la predicción diferencial de las pruebas obligatorias, los resultados obtenidos aportan evidencia equivalente a los patrones observados en estudios norteamericanos. Respecto a las diferencias predictivas por dependencia, se observa una leve sobrepredicción del rendimiento de estudiantes provenientes de establecimientos municipales en la Prueba de Lenguaje, lo que se traduce en un sesgo de 0,07 desviaciones estándar de las notas de primer año a favor de este grupo, en relación a los estudiantes de establecimientos Particulares Pagados. Por otra parte, los sesgos predictivos son muy bajos en la Prueba de Matemáticas y no se observa una tendencia clara a favor de alguno de los grupos de dependencia. En cuanto a las comparaciones por género, existe una subpredicción del rendimiento de las mujeres, que es del orden de un 13% de la desviación estándar para la prueba de Lenguaje y de un 20% para la prueba de Matemática. Estos resultados entre hombres y mujeres son algo menores a los encontrados por Mattern et al. en el último estudio de validez diferencial del SAT (de 26% para el SAT-CR y 37% para el SAT-M), sin embargo, confirman las tendencias observadas en este tipo de pruebas. Al realizar los análisis de género por área de conocimiento, también se confirman estos resultados, aunque las magnitudes varían un poco: los mayores sesgos predictivos a favor de los hombres están en las áreas de Ciencias Sociales y Economía y Administración (sobre 30% en Matemáticas y sobre 25% en Lenguaje para la mayoría de las admisiones) y los menores sesgos en las áreas de Arte, Ingeniería y Ciencias (poco más de 10% en Matemáticas y bajo 5% en Lenguaje para la mayoría de las admisiones),

Por otra parte, las Notas de Enseñanza Media presentan una tendencia similar a la de las pruebas pero tienen comportamientos diferentes en relación a la magnitud relativa de los sesgos. En las comparaciones por género también se observa un sesgo predictivo a favor de los hombres pero, en este caso, la magnitud es sustancialmente inferior a la de las pruebas y no supera las 0.04 desviaciones estándar. En cambio, en las comparaciones por dependencia, son las NEM las que constituyen el factor que presenta mayor sesgo predictivo (de hasta 0.13 desviaciones estándar). En términos de equidad, estos resultados se traducen en una ventaja a favor de los establecimientos municipales.

Por último, cabe señalar que los Puntajes de Selección, que se calculan como una combinación ponderada de los factores de selección, tienen comportamientos coherentes con los resultados de validez y predicción diferencial observadas para los factores que lo constituyen.

Entre la literatura dedicada a este tema, diferentes hipótesis se han planteado para explicar los fenómenos de sesgo en las capacidades predictivas de los factores de selección.

Una explicación que se ha dado, especialmente en el contexto norteamericano, hace referencia a la validez en el uso de las notas de la universidad como criterio de análisis, especialmente cuando los alumnos tienen la posibilidad de elegir sus cursos. Si las asignaturas escogidas por algún grupo son sistemáticamente más fáciles que las escogidas por los demás, entonces es posible que estos alumnos obtengan relativamente mejores notas, y por tanto, los análisis realizados en estas condiciones podrían sugerir, en forma errónea, predicción diferencial de los factores de selección. Para minimizar este riesgo,

en el presente trabajo se decidió utilizar como criterio de comparación sólo las notas de primer año (donde existe poca libertad para la elección de ramos) y, como unidad de análisis básica cada carrera (de modo de controlar por las diferencias en las experiencias educacionales que podrían darse entre carreras). Con ello, la variabilidad de ramos cursados por alumnos en una misma carrera es bastante baja, haciendo más comparable los rendimientos académicos entre estudiantes. En consecuencia, la hipótesis anterior resulta poco plausible para explicar los sesgos de validez y predicción el caso abordado en este estudio.

Por otra parte, y suponiendo que se acepta como válido el uso de las notas universitarias como variable de criterio, los estudios de predicción diferencial pueden servir para detectar la presencia de sesgo de medición en las pruebas. En este sentido, predicciones diferenciales del rendimiento podrían deberse a diferencias sistemáticas en los puntajes obtenidos por ciertos grupos, las que no serían atribuibles a diferencias en el constructo (habilidad, capacidad) que la prueba está tratando de medir. No obstante, la mayoría de las pruebas estandarizadas de altas consecuencias, como la PSU, realizan análisis del funcionamiento diferencial de sus ítems (DIF) como forma de detectar la existencia de sesgo de medición en la prueba. Por tanto, aún en ausencia de sesgo de medición, es posible encontrar evidencia de validez diferencial en las pruebas de selección universitaria.

El caso de los puntajes asociados a las NEM es un poco distinto ya que este es el único factor de selección que no constituye un instrumento estandarizado de medición y que, por tanto, no representa una escala común de aprendizajes. Por el contrario, las NEM dependen directamente del sistema de evaluación que posea el establecimiento educacional de origen. Este hecho podría explicar por qué, a pesar de que este factor tiene una alta capacidad predictiva promedio (de 0.25, sin corrección de rango y de 0.4, con corrección), es que se observan diferencias importantes en las magnitudes de estas correlaciones según la dependencia de origen. Por otra parte, el sesgo de predicción a favor de los establecimientos municipales revelaría en qué forma operan estos mecanismos diferenciales de asignación de notas, ya que los resultados indican que, en promedio, un alumno egresado de un colegio municipal tendría un menor desempeño relativo en la universidad comparado con un alumno de un colegio Particular Pagado con igual NEM.

Para concluir, es importante mencionar que existen hipótesis alternativas que intentan explicar el sesgo en las capacidades predictivas de los factores de selección. Estas suponen entre otras cosas, que existen expectativas, propias o del entorno, que pueden afectar el desempeño académico de ciertas minorías, o que, en el contexto universitario, ciertos grupos enfrentan opciones o poseen hábitos o actitudes diferentes a las de sus pares, los que se traducen en un rendimiento académico diferencial.

Dentro de estas hipótesis está la que se conoce como "amenaza de estereotipo", que señala que, estereotipos culturalmente compartidos sobre el desempeño de cierto grupo, y que son evidenciados en un contexto que apela a dicho estereotipo, pueden afectar el desempeño de un individuo que se identifica con ese grupo (Steele y Aronson, 1995, 1997). Dicho en otras palabras, cuando un individuo está consciente de las bajas expectativas de rendimiento asociadas a su grupo, la tensión o ansiedad que dicha conciencia produce, puede llevar a rendir por debajo de su potencial, confirmando las expectativas.

Si bien esta hipótesis podrían explicar el comportamiento diferencial de ciertos grupos socioeconómicos, como minorías étnicas y raciales en EEUU y dependencia educacional de origen, en Chile, la evidencia respecto a los resultados diferenciales en predicción del rendimiento entre hombres y mujeres no son consistentes con dicha hipótesis (pues en este caso, las mujeres demuestran un mejor rendimiento que el previsto según sus puntajes en las pruebas de admisión).

En definitiva, no contamos aun con buenas explicaciones para los resultados detectados en los estudios de validez diferencial (ver Zwick, 2006). El hecho que los resultados de este estudio sean consistentes con los norteamericanos (sobrepredicción de grupos socialmente desventajados y subpredicción del rendimiento de las mujeres, sugiere una cierta consistencia en distintos contextos culturales, lo que debiera motivar esfuerzos para comprender los mecanismos que explican tales resultados. Los sesgos detectados no son importantes en su magnitud (especialmente los asociados a la posición social –como grupo étnico en EEUU y dependencia educacional en Chile–), pero su consistencia en el tiempo y entre culturas los debiera constituir en aspectos a ser monitoreados sistemáticamente. Asimismo, se debieran establecer programas de investigación que permitan establecer sus orígenes o al menos sus correlatos.

## REFERENCIAS BIBLIOGRÁFICAS

- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. (1999). *The Standards for Educational and Psychological Testing*. Washington, D.C.: Autor
- Bravo, D., G. Del Pino, G. Donoso, G. Hawes, J. Manzi y M. Martínez (2008), "Resultados de la Aplicación de Pruebas de Selección Universitaria Admisión 2006 a 2008", Comité Técnico Asesor, Honorable Consejo de Rectores de las Universidades Chilenas, Serie Documentos Técnicos, Septiembre. Disponible en [www.cta-psu.cl](http://www.cta-psu.cl)
- Burton, N. and Ramist, L.. Predicting success in college: SAT studies of classes graduating since 1980. New York: College Entrance Examination Board.
- Camara, Wayne, and Gary Echternacht. 2000. "The SAT I and High School Grades: Utility in Predicting Success in College." The College Board Research Notes RN-10:1-10.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons.
- Manzi, Jorge, David Bravo, Guido Del Pino, Graciela Donoso, Manuel Martínez, y Raúl Pizarro. 2006. "Estudio acerca de la Validez Predictiva de los Factores de Selección a las Universidades del Consejo de Rectores." Comité Técnico Asesor del CRUCH.
- Manzi, Jorge, David Bravo, Guido Del Pino, Graciela Donoso, Manuel Martínez, y Raúl Pizarro. 2008. "Estudio acerca de la Validez Predictiva de los Factores de Selección a las Universidades del Consejo de Rectores." Comité Técnico Asesor del CRUCH.
- Mattern, Krista D., Brian F. Patterson, Emily J. Shaw, Jennifer L. Kobrin, and Sandra M. Barbuti. "Differential Validity and Prediction of the SAT" (College Board Research Report No. 2008-4). New York: The College Board.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Wainer, H., & Steinberg, L. S. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bidirectional validity study, *Harvard Educational Review*, 62, 323–36.

Young, J. W. (2001). Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis (College Board Research Report No. 2001-6). New York: The College Board.

Zwick, R. (2006). Higher Education Admissions Testing, en R. Brennan (Ed.), Educational measurement (4th ed., pp. 647-679). National Council on Measurement in Education Greenwood Press., West Port, CT.

