

Diferencias en el desempeño académico y en los predictores de éxito universitario por escuela de procedencia

Rafael J. Colorado Laguna, Ph.D.

Facultad de Ciencias Naturales
Universidad de Puerto Rico
Recinto de Río Piedras
rafaeljcolorado@gmail.com

Lillian Corcino Marrero, Ed.D.

Facultad de Educación
Universidad de Puerto Rico
Recinto de Río Piedras
l_corcino19@hotmail.com

RESUMEN

El propósito de esta investigación fue explorar posibles diferencias entre estudiantes procedentes de escuela pública y escuela privada respecto a las puntuaciones en las pruebas estandarizadas de aptitud y de aprovechamiento utilizadas como criterios de admisión a las instituciones universitarias en Puerto Rico, el conjunto de predictores de éxito universitario y la utilidad de predicción de cada predictor, individualmente y en conjunto. Los resultados fueron los siguientes: los estudiantes de escuela privada obtuvieron puntuaciones promedio mayores que los estudiantes de escuela pública en las pruebas de aptitud y en las pruebas de aprovechamiento; los coeficientes de correlación entre predictores y criterios, promedio de primer año de estudios universitarios y promedio de graduación, fueron mayores para estudiantes procedentes de escuela privada, y los predictores de éxito universitario de estudiantes procedentes de ambos tipos de escuela fueron similares. Se incluyen posibles explicaciones para las diferencias en el desempeño académico de los estudiantes procedentes de ambos tipos de escuela.

Palabras clave: escuelas públicas, escuelas privadas, pruebas estandarizadas de aptitud, pruebas estandarizadas de aprovechamiento, criterios de admisión universitaria, predictores de éxito universitario

ABSTRACT

The purpose of this research was to explore possible differences between public and private school students with respect to: scores on standardized aptitude and achievement tests used as admission criteria to higher education institutions in Puerto Rico, the set of college success predictors, and the relative predictive value of each predictor, individually and in combination with other predictors. Major study findings were: private school students' aptitude and achievement tests scores were higher than those of public school students; correlation coefficients between selected predictors and criteria, freshman's GPA and graduation's GPA, were higher for private school students, and the set of college success predictors for both school types were similar. Possible explanations for the differences in academic performance are considered.

Keywords: public schools, private schools, standardized aptitude tests, standardized achievement tests, college admissions criteria, college success predictors

Fecha de recepción: agosto 2013

Fecha de aceptación: mayo 2014

Introducción

La mayoría de las instituciones universitarias en los Estados Unidos y Puerto Rico utiliza un conjunto de criterios de admisión para determinar el ingreso de aquellos estudiantes egresados de escuela superior que aspiran a proseguir estudios universitarios. Tradicionalmente, tales criterios han sido de carácter académico. Hawkins y Clinedinst (2006) recogen los principales criterios de admisión en los Estados Unidos: el promedio académico de escuela superior (GPA), el promedio académico en cursos de escuela superior preparatorios para estudios universitarios, las puntuaciones en pruebas estandarizadas de aptitud y de aprovechamiento, y el rango percentil del estudiante respecto a su clase al momento de graduarse de escuela superior.

En Puerto Rico, a partir de 1963, el College Board (CB) ha confeccionado y administrado pruebas estandarizadas de aptitud y de aprovechamiento académico que las instituciones de educación postsecundaria han utilizado como criterios de admisión. El Programa de Evaluación y Admisión Universitaria (PEAU) del CB conforma una batería de pruebas entre las que se incluyen dos exámenes principales: la Prueba de Aptitud Académica (PAA) y las Pruebas de Aprovechamiento Académico (PACH). La PAA consiste de dos componentes: razonamiento verbal y razonamiento matemático. Por su parte, las PACH consisten de tres componentes: español, matemáticas e inglés.

Menéndez (1995) informa que, desde principios de los años 1960, la Universidad de Puerto Rico (UPR) ha utilizado diversas fórmulas de admisión, consistentes en promedios ponderados que incorporan los siguientes criterios: el promedio académico de escuela superior, las puntuaciones en la PAA y las puntuaciones en las PACH. A partir de 1995, la fórmula de admisión a la UPR, conocida como Índice General de Solicitud (IGS), ha consistido de un promedio ponderado de los siguientes tres componentes: el promedio de escuela superior (50%), la puntuación en la Prueba de Razonamiento Verbal (25%) y la puntuación en la Prueba de Razonamiento Matemático (25%). A pesar de que en algún momento lo hiciera, la fórmula de admisión actual no incluye ninguna de las puntuaciones en las Pruebas de Aprovechamiento Académico. Sin embargo, estas puntuaciones se utilizan como criterios de ubicación en los cursos de español, matemáticas e inglés de primer año de estudios universitarios.

El procedimiento empleado para evaluar la efectividad de los criterios de admisión respecto a indicadores de éxito universitario se conoce con el nombre de *estudio de validez*. El concepto de *validez*, dentro del contexto de pruebas educativas y psicológicas, así como de otras medidas de evaluación, es un concepto abarcador que incluye múltiples dimensiones. Messick (1994) lo define como: un juicio valorativo del grado en el que la evidencia empírica y el razonamiento teórico apoyan la adecuación y la pertinencia de las interpretaciones y las acciones basadas en puntuaciones en pruebas o en otros modos de evaluación (p. 6). Advierte, además, que la validez no es una atributo de la prueba, sino más bien, del significado de las puntuaciones en la prueba (p. 6).

Los *Standards for Educational and Psychological Testing* [SEPT] (AERA, APA & NCME, 1999) definen el concepto de validez como el grado en el que la evidencia y la teoría apoyan las interpretaciones de las puntuaciones en pruebas, interpretaciones que son el resultado del uso sugerido para ellas (p. 9). Al igual que Messick (1994), los SEPT aclaran que son las interpretaciones de las puntuaciones, requeridas por los usos sugeridos, las que son evaluadas, no las pruebas en sí mismas (p. 9).

Tradicionalmente, se ha distinguido entre tres diferentes tipos de validez: validez de constructo, validez de contenido y validez de criterio. Sin embargo, los SEPT, adoptando la línea de pensamiento de Messick (1994), sostienen que el concepto de validez es de naturaleza unitaria, por lo que entienden más correcto distinguir entre diferentes

fuentes de evidencia, en lugar de diferentes tipos de validez (AERA et al., 1999). Establecida esta discrepancia de enfoque, los SEPT incluyen las siguientes fuentes de evidencia para la validez de una prueba: evidencia basada en el contenido de la prueba, evidencia basada en el proceso de respuestas, evidencia basada en la estructura interna de la prueba, evidencia basada en la relación con otras variables y evidencia basada en posibles consecuencias de la prueba.

En la medida en que interesa evaluar la validez de predicción de pruebas de aptitud y de aprovechamiento respecto a algún criterio de éxito universitario, corresponde examinar concretamente la cuarta fuente de evidencia: la evidencia fundamentada en la relación entre las puntuaciones en estas pruebas y variables externas a las pruebas. Los SEPT subdividen dichas variables en tres tipos: criterios que las pruebas procuran predecir (como, por ejemplo, alguna medida de éxito universitario o de desempeño en el trabajo); puntuaciones en otras pruebas que, hipotéticamente, miden los mismos constructos, y puntuaciones en otras pruebas que miden constructos parecidos o diferentes. La evidencia sobre la relación de las pruebas con variables externas puede recolectarse tanto mediante estudios de tipo experimental, como de estudios correlacionales.

Para referirse a la evidencia de la relación entre las puntuaciones de las pruebas y algún criterio de relevancia, los SEPT utilizan el término “relaciones pruebas-criterios”. Para ellos, la pregunta medular en estos casos es: ¿con cuánta precisión predicen las puntuaciones en las pruebas el desempeño en el criterio? Históricamente, se ha distinguido entre dos tipos de diseño al evaluar las relaciones pruebas-criterios, distinción fundamentada en el momento en el tiempo en el que se recolecta la información del predictor o predictores y del criterio. Un *estudio concurrente* obtiene la información en momentos temporalmente cercanos, mientras que un *estudio predictivo* persigue determinar con cuánta precisión las puntuaciones en las pruebas permiten predecir resultados en el criterio obtenidos posteriormente.

La evidencia de la validez de las puntuaciones en las pruebas de aptitud y de aprovechamiento, además del promedio de escuela superior, que perseguía recolectar este estudio, se circunscribió a evidencia basada en la relación de las puntuaciones con otras variables, específicamente, con la capacidad de predecir criterios de éxito universitario (“relaciones pruebas-criterios”). Consistió, por tanto, en un estudio de validez de carácter predictivo, y las técnicas estadísticas utilizadas fueron de tipo correlacional.

En el contexto de un estudio experimental, cuyo propósito primordial es determinar relaciones de causalidad, distinguimos entre variable *independiente* y variable *dependiente*. La primera es aquella variable manipulada por el experimentador con el propósito de determinar si existe o no una relación de causalidad con la variable dependiente. De otra parte, en el contexto de un estudio de predicción, distinguimos entre variable *predictor* y variable *criterio*. La primera es aquella variable que utilizamos para predecir la segunda. La variable predictor es análoga a la variable independiente; sin embargo, no es manipulada en forma alguna. Similarmente, la variable criterio es análoga a la variable dependiente, pero la intención no es determinar causalidad entre el predictor y el criterio, sino más bien magnitud de relación entre estas variables. En el caso de este estudio, los predictores utilizados fueron los resultados de las PAA y las PACH, así como el promedio de escuela superior. Las variables criterio fueron el promedio académico al finalizar el primer año de estudios universitarios y al completar el grado de bachillerato.

Siendo este un estudio de naturaleza correlacional, incorporó las siguientes técnicas estadísticas: el cómputo de *coeficientes de correlación simple* entre cada predictor y el criterio ($-1 \leq r_{xy} \leq 1$), la generación de ecuaciones o *modelos de regresión lineal múltiple* y el cómputo del *coeficiente de correlación múltiple* ($0 \leq R \leq 1$). Este último no es otra cosa que el coeficiente de correlación simple entre los valores del criterio *predichos* por la ecuación de regresión y los valores *observados* del criterio. El cuadrado del coeficiente de correlación simple (r^2) o del coeficiente de correlación múltiple (R^2), denominado *coeficiente de determinación*, representa una medida del tamaño del efecto (“effect size”) del predictor o predictores sobre el criterio, y expresa la proporción o porcentaje en la variación del criterio predecible por la variación en el predictor (r^2) o en los predictores (R^2).

Al comparar las magnitudes de dos coeficientes de correlación, debemos tomar en cuenta, como señalan Hinkle, Wiersma y Jurs (1994), que la escala de este coeficiente es de tipo *ordinal*, por lo que no es correcto afirmar, por ejemplo, que si un coeficiente de correlación es igual a 0.80 y otro a 0.40, el primero es el doble del segundo. En lo que respecta a la interpretación de dicha magnitud, Hinkle y colaboradores (1994) presentan una tabla, que reproducimos como Tabla 1, y que recoge la regla general para clasificar la magnitud de este coeficiente.

Según expresa Punch (2004), los objetivos de los *modelos de regresión lineal múltiple* son: por un lado, estimar qué porcentaje de la varia-

Tabla 1

Regla general para interpretar la magnitud del coeficiente de correlación

Magnitud de la correlación	Interpretación
0.90 a 1.00 (-0.90 a -1.00)	Correlación positiva (negativa) muy alta
0.70 a 0.89 (-0.70 a -0.89)	Correlación positiva (negativa) alta
0.50 a 0.69 (-0.50 a -0.69)	Correlación positiva (negativa) moderada
0.30 a 0.49 (-0.30 a -0.49)	Correlación positiva (negativa) baja
0.00 a 0.29 (-0.00 a -0.29)	Correlación positiva (negativa) muy baja

ción en el criterio puede ser explicado por un conjunto de dos o más predictores y, por otro, comparar el peso o contribución relativa de cada uno de los predictores para explicar la variabilidad del criterio. Mientras mayor sea el porcentaje de variación del criterio explicado por el conjunto de predictores, mayor será la utilidad de predicción conjunta de los predictores. De igual forma, mientras mayor sea el peso o contribución relativa de cada predictor, mayor será la utilidad de predicción de este respecto al resto de los predictores.

Como toda técnica estadística, el análisis correlacional tiene sus limitaciones. Para empezar, es de esperar que diferentes muestras generen distintos valores para los coeficientes de correlación y diferentes ecuaciones de regresión. De otra parte, Borg y Gall (1983) advierten sobre un problema del que padece todo estudio de validez predictiva: el problema de *reducción* o *contracción* (“shrinkage”). Este consiste en la tendencia a la disminución en la magnitud del coeficiente de correlación múltiple cuando repetimos un estudio de predicción. La restricción será mayor, según los autores, cuando la muestra original incluya un número limitado de sujetos y cuando el número de predictores sea grande.

Sobre el fenómeno de restricción, Borg y Gall (1983) comentan que la reducción se debe, principalmente, al hecho de que, cuando inicialmente validamos nuestras medidas, algunas de ellas producirán correlaciones significativas por pura casualidad. En otras palabras, características inherentes al grupo de sujetos examinado tienden a producir una validez de predicción máxima para algunas de las medidas de predicción (p. 622). Para contrarrestar el efecto de restricción recomiendan que, como no es posible hacer predicciones a base de correlaciones que solo aplican a una muestra de sujetos, es siempre

aconsejable llevar a cabo una validación cruzada de los predictores antes de utilizarlos en situaciones prácticas de predicción (p. 622).

Hinkle y colaboradores (1994) establecen la necesidad de mantener un balance prudente entre el número de predictores (k) y el tamaño de la muestra (n). A medida que el número de predictores se acerca al tamaño de la muestra ($k \rightarrow n$), el coeficiente de correlación múltiple (R), aumenta. De hecho, cuando $k + 1 = n$, el coeficiente de correlación múltiple es igual a 1.0, aun cuando ninguno de los predictores correlacione con el criterio. Sin embargo, señalan que, cuando el tamaño de la muestra es grande ($n > 100$) y se utilizan tan solo cinco o seis predictores —como es el caso de este estudio—, no debemos preocuparnos por este fenómeno. Borg y Gall (1983) recomiendan, como regla general, aumentar el tamaño de la muestra por lo menos en 15 sujetos, cada vez que se incorpore un nuevo predictor a la ecuación de regresión múltiple.

Borg y Gall (1983) destacan la importancia que puede tener examinar subgrupos de la muestra en un estudio de validez. Con frecuencia, la ecuación de regresión para un subgrupo no es la misma que la de la totalidad de la muestra. Esto muestra diferencias en los coeficientes de correlación y en la utilidad de predicción relativa de cada predictor y de todos los predictores en conjunto. Aquella variable que, al utilizarse para dividir una muestra en subgrupos, genera coeficientes de correlación y ecuaciones de regresión diferentes para ellos, se conoce con el nombre de variable *moderadora*.

Entre las variables moderadoras utilizadas con mayor frecuencia en los estudios de validez llevados a cabo en los Estados Unidos están: raza o grupo étnico, género y nivel socioeconómico (Bridgeman, McCamley-Jenkins & Ervin, 2000; Burton & Ramist, 2001; Ramist, Lewis & McCamley-Jenkins, 1993; Zwick, 2007). Recientemente, se han considerado otras variables moderadoras, tales como la presencia de algún tipo de discapacidad (Zwick, 2007).

Tradicionalmente, ha existido preocupación sobre la validez de predicción de las pruebas de aptitud y de aprovechamiento como criterios de admisión a estudios postsecundarios, dada la posibilidad de que puedan discriminar a favor de un segmento de la población y en contra de otro. Naturalmente, este es un asunto muy serio, puesto que el acceso a estudios postsecundarios es la puerta a una mejor preparación académica y a mejores oportunidades de empleo y calidad de vida. La relevancia del asunto es aún mayor cuando se trata de admisión a una institución universitaria pública.

Los estudios de validez llevados a cabo en los Estados Unidos, por lo general, no consideran el tipo de escuela de procedencia como variable moderadora. En su lugar, Geiser y Studley (2002), así como Geiser y Santaelices (2007) utilizaron un índice de desempeño académico conocido como el Academic Performance Index (API). El API, desarrollado por el Departamento de Educación de California, es un indicador de la excelencia académica de una escuela superior, estrechamente vinculado a diferencias socioeconómicas, étnicas y demográficas. El estudio de Geiser y Studley (2002) utilizó el API para poner a prueba la hipótesis de que, en escuelas con un API bajo, las pruebas de aptitud cobraban relevancia como predictores de promedio de primer año universitario. De otra parte, Geiser y Santaelices (2007) hizo lo mismo, pero utilizando como criterio el promedio acumulativo de cuarto año.

En un estudio de validez llevado a cabo en la UPR, Menéndez (1996) incorporó las siguientes variables moderadoras: área o disciplina de estudios, género (femenino o masculino), tipo de escuela superior de procedencia (pública o privada), ingreso familiar, nivel de educación de los padres y ocupación de los padres. La Tabla 2 recoge los coeficientes de correlación que encontrara Menéndez entre los predictores y cada uno de los dos criterios, promedio de primer año (FGPA) y promedio de graduación (GGPA), para estudiantes del Recinto de Río Piedras de la UPR, por escuela de procedencia.

Los coeficientes de correlación presentados en la Tabla 2 pueden resumirse de la siguiente manera:

1. Para el promedio de primer año, todos los coeficientes de correlación de la escuela privada fueron mayores que los coeficientes correspondientes de la escuela pública. Para el promedio de graduación, el patrón fue similar, excepto para las pruebas de matemáticas, para las que se invirtió la tendencia.
2. El promedio de escuela superior registró el mayor coeficiente de correlación para ambos criterios y ambos tipos de escuela.
3. Para el promedio de primer año, el segundo coeficiente mayor, para ambos tipos de escuela, correspondió a la prueba de aprovechamiento en español. Por otro lado, el coeficiente de correlación de la prueba de razonamiento verbal aumentó para el promedio de graduación para ambos tipos de escuela.
4. Las pruebas de razonamiento matemático y de aprovechamiento en matemáticas registraron coeficientes de correlación compa-

Tabla 2

Coefficientes de correlación entre cada predictor y cada criterio, promedio de primer año y promedio de graduación, por escuela de procedencia, en el estudio de Menéndez (1996)

Criterio	FGPA		GGPA	
	Pública	Privada	Pública	Privada
HSGPA	0.3706	0.4486	0.3542	0.3826
PAPTV	0.2083	0.2935	0.2994	0.3645
PAPTM	0.1370	0.2258	0.1791	0.1203
PAPRE	0.3034	0.3669	0.2708	0.4062
PAPRM	0.2076	0.3004	0.1869	0.1693
PAPRI	0.1367	0.1923	0.2356	0.2728

FGPA = Promedio de primer año universitario

GGPA = Promedio de graduación de universidad

HSGPA = Promedio de escuela superior

PAPTV = Puntuación en la prueba de razonamiento verbal

PAPTM = Puntuación en la prueba de razonamiento matemático

PAPRE = Puntuación en la prueba de aprovechamiento en español

PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas

PAPRI = Puntuación en la prueba de aprovechamiento en inglés

rativamente menores, excepto en el caso de la correlación de la prueba de aprovechamiento en matemáticas con el promedio de primer año para escuelas privadas, que registró el tercer mayor coeficiente de correlación.

5. El coeficiente de correlación de la prueba de aprovechamiento en inglés, a pesar de ser uno de los menores, registró un aumento con respecto al promedio de graduación, para ambos tipos de escuela, sobrepasando al coeficiente de correlación de ambas pruebas matemáticas.

La Tabla 3 recoge los modelos de regresión con los mejores tres predictores de éxito universitario, a corto y a largo plazos, por escuela de procedencia, que obtuviera Menéndez (1966). Los resultados pueden resumirse de la siguiente manera:

1. El mejor predictor de éxito universitario a corto y largo plazos resultó ser el promedio de escuela superior.
2. La prueba de razonamiento verbal ocupó la segunda posición como mejor predictor en tres de los cuatro modelos de regresión, y la tercera posición en el cuarto.
3. La prueba de razonamiento matemático no aparece entre los mejores tres predictores de ninguno de los modelos, y la prueba

Tabla 3

Modelos de regresión con los mejores tres predictores para promedio de primer año y promedio de graduación, por escuela de procedencia, en el estudio de Menéndez (1996)

Criterio	FGPA		GGPA	
	Predictor	R	Predictor	R
Pública	HSGPA		HSGPA	
	PAPRE	0.404	PAPTV	0.442
	PAPTV		PAPRI	
Privada	HSGPA		HSGPA	
	PAPTV	0.511	PAPTV	0.526
	PAPRM		PAPRE	

FGPA = Promedio de primer año universitario

GGPA = Promedio de graduación de universidad

HSGPA = Promedio de escuela superior

PAPTV = Puntuación en la prueba de razonamiento verbal

PAPTM = Puntuación en la prueba de razonamiento matemático

PAPRE = Puntuación en la prueba de aprovechamiento en español

PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas

PAPRI = Puntuación en la prueba de aprovechamiento en inglés

de aprovechamiento en matemáticas solamente aparece en tercera posición respecto al promedio de primer año para escuelas privadas.

- La prueba de aprovechamiento en español aparece en dos de los cuatro modelos, en segunda posición para promedio de primer año en escuelas públicas y en tercera posición para promedio de graduación en escuelas privadas.
- La prueba de aprovechamiento en inglés solo aparece en tercera posición para promedio de graduación en escuelas públicas.
- Los coeficientes de correlación múltiple para las escuelas privadas, para ambos criterios (FGPA y GGPA), fueron mayores para las escuelas privadas que para las escuelas públicas. Según la clasificación establecida por Hinkle y colaboradores (1994, Véase Tabla 1), los coeficientes de correlación múltiple para las escuelas privadas fueron moderados ($R = 0.511$ y $R = 0.526$), mientras que los respectivos coeficientes para las escuelas públicas fueron bajos ($R = 0.404$ y $R = 0.442$).

Colorado y Corcino (2013) examinaron la validez de predicción de los criterios de admisión al Recinto de Río Piedras de la UPR utilizando como predictores el promedio de escuela superior, las puntuaciones en las pruebas de aptitud y las puntuaciones en las pruebas de aprovechamiento. Como hiciera Menéndez (1996), emplearon como criterios de éxito universitario el promedio académico de primer año de estudios universitarios y el promedio académico de graduación. El presente estudio es una continuación de dicho trabajo, e incorpora como variable moderadora el tipo de escuela de procedencia (pública o privada) del estudiante admitido al Recinto de Río Piedras de la UPR.

Este estudio buscaba responder las siguientes tres preguntas de investigación:

1. ¿Qué diferencia, si alguna, existe en el desempeño de los estudiantes en las PAA (razonamiento verbal y razonamiento matemático) y en las PACH (español, matemáticas e inglés) que administra el CB, de acuerdo al tipo de escuela de procedencia?
2. ¿Qué diferencia, si alguna, existe entre los coeficientes de correlación entre predictores y criterios de éxito universitario para los estudiantes procedentes de escuela pública y escuela privada?
3. ¿Qué diferencia existe, si alguna, en la utilidad de predicción de éxito universitario de los diferentes predictores empleados como criterios de admisión para los estudiantes procedentes de escuela pública y escuela privada?

Método

Para contestar la primera pregunta de investigación, sobre posibles diferencias en el desempeño de los estudiantes en las PAA y las PACH por escuela de procedencia, este estudio utilizó resultados suministrados por la Oficina del College Board de Puerto Rico para los años de 1985 a 2013. De otra parte, para contestar las últimas dos preguntas de investigación, se utilizaron datos recolectados de la cohorte de estudiantes admitidos al Recinto de Río Piedras de la UPR en el año 2000. El conjunto original de datos de esta cohorte consistió de un total de 3,116 estudiantes.

Los criterios de admisión utilizados como predictores en este estudio fueron: el promedio académico de escuela superior, las puntuaciones en las PAA del CB (razonamiento verbal y razonamiento matemático) y las puntuaciones en las PACH del CB (aprovechamiento en español, matemáticas e inglés como segundo idioma). A pesar de que la fórmula de admisión actual no incluye las puntuaciones en las

pruebas de aprovechamiento, este trabajo las consideró como posibles predictores en el ánimo de determinar su utilidad de predicción respecto a los demás componentes de la fórmula. Las puntuaciones en las pruebas de aptitud y aprovechamiento fluctúan entre un mínimo de 200 puntos y un máximo de 800.

Los criterios de éxito universitario seleccionados fueron dos: el promedio académico del primer año de estudios universitarios y el promedio académico de graduación. El promedio académico, en ambos casos, fluctúa entre un mínimo de 0.0 y un máximo de 4.0 puntos. Colorado y Corcino (2013) elaboran sobre la justificación por la que, en estudios de validez de predicción, comúnmente se han utilizado estos dos criterios.

No todos los estudiantes admitidos a cursar estudios universitarios completan su primer año. De aquellos que lo completan, no todos aprueban el mismo número de créditos. Por ello, promedios iguales no necesariamente representan igual esfuerzo o aprovechamiento. Tampoco los cursos que toman los estudiantes en su primer año de estudios universitarios son necesariamente los mismos, aun tratándose de la misma materia.

Para minimizar el posible efecto de estas diferencias, solamente aquellos estudiantes que aprobaron al menos 24 créditos al completar su primer año fueron considerados en el cómputo del promedio de primer año. Esto representa, en promedio, 12 créditos por semestre, lo que equivale a una carga académica completa para el estudiante. Una vez removidos de la cohorte original aquellos que no completaron su primer año o que aprobaron menos de 24 créditos, la cantidad de estudiantes se redujo a 2,124.

Al momento de llevar a cabo el estudio, algunos estudiantes se habían dado de baja de la universidad, se habían transferido a otra institución universitaria o aún no habían completado su grado de bachillerato. Una vez removidos estos alumnos, un total de 1,653 estudiantes de la cohorte original había completado su grado de bachillerato, por lo que fueron incluidos en el cómputo del promedio académico de graduación.

Los resultados del estudio fueron generados en el siguiente orden: primero, estadísticas descriptivas de cada predictor y cada uno de los dos criterios; luego, coeficientes de correlación lineal entre cada predictor y cada uno de los dos criterios, y por último, ecuaciones de regresión lineal múltiple entre la totalidad de los predictores y cada uno de los dos criterios con sus correspondientes coeficientes de correlación múltiple.

Resultados

Para contestar la primera pregunta de investigación sobre posibles diferencias en los resultados en las pruebas de aptitud y de aprovechamiento, por tipo de escuela de procedencia, la Oficina del College Board en Puerto Rico suministró las puntuaciones promedios de estas pruebas para los años de 1985 a 2013. A continuación, se incluyen dos diagramas de línea (Diagrama 1 y Diagrama 2) que ilustran las diferencias entre las puntuaciones promedio de las pruebas de razonamiento verbal y razonamiento matemático, respectivamente. Seguido, se incluyen otros tres diagramas de línea (Diagrama 3, Diagrama 4 y Diagrama 5) que ilustran las diferencias entre las puntuaciones promedio en las tres pruebas de aprovechamiento: español, matemáticas e inglés como segundo idioma.

Como puede apreciarse en los diagramas de línea, las puntuaciones promedio de los estudiantes de escuela privada en las dos pruebas de aptitud y en las tres pruebas de aprovechamiento fueron mayores que las puntuaciones promedio correspondientes de los estudiantes de escuela pública. Para facilitar la interpretación de estas diferencias, la Tabla 4 recoge, por escuela de procedencia, las puntuaciones promedio en las pruebas de aptitud y de aprovechamiento, las diferencias promedio, las diferencias menores, las diferencias mayores y la amplitud de las diferencias.

La información recogida en la Tabla 4 permite hacer las siguientes observaciones:

1. Según se refleja en los diagramas, para las cinco pruebas, las puntuaciones promedio de las escuelas privadas fueron mayores que las puntuaciones promedio correspondientes de las escuelas públicas.
2. Las diferencias promedio mayores fueron en las pruebas de aprovechamiento en inglés (107.4 puntos), de aprovechamiento en matemáticas (90.7 puntos) y de razonamiento matemático (85.0 puntos). Las diferencias promedio menores fueron en la prueba de razonamiento verbal (74.5 puntos) y las pruebas de aprovechamiento en español (64.3 puntos).
3. Mientras las puntuaciones promedio de las escuelas públicas, en las cinco pruebas, se han mantenido relativamente similares a través de los años, las puntuaciones promedio de las escuelas privadas, excepto por las puntuaciones en la prueba de aprovechamiento en inglés, han reflejado una tendencia descendente.

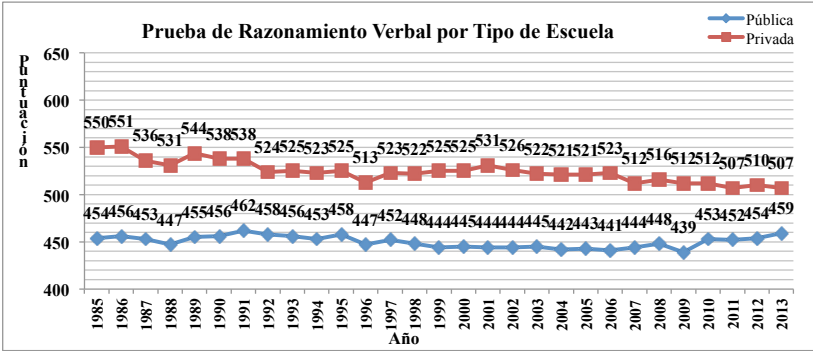


Diagrama 1. Puntuación promedio en la prueba de Razonamiento Verbal para los años del 1985 al 2013, por escuela de procedencia

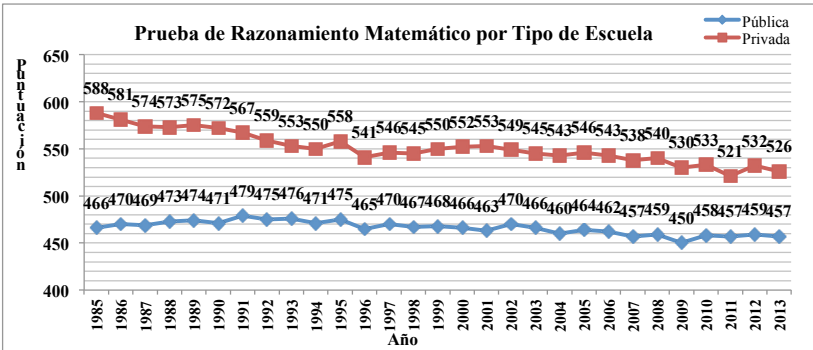


Diagrama 2. Puntuación promedio en la prueba de Razonamiento Matemático para los años del 1985 al 2013, por escuela de procedencia

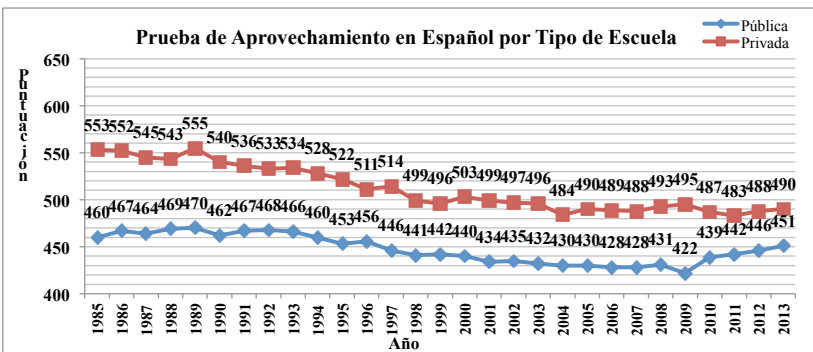


Diagrama 3. Puntuación promedio en la prueba de Aprovechamiento en Español para los años del 1985 al 2013, por escuela de procedencia

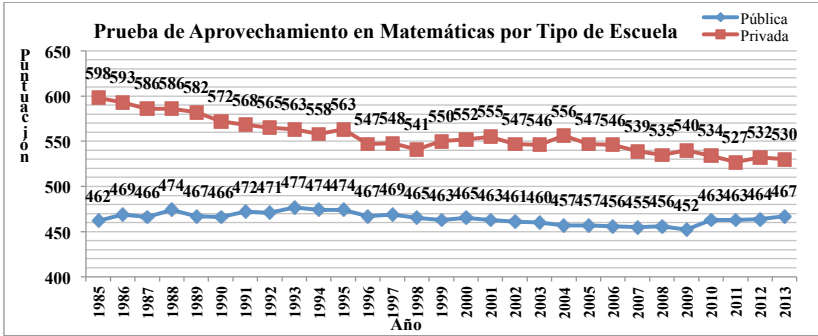


Diagrama 4. Puntuación promedio en la prueba de Aprovechamiento en Matemáticas para los años del 1985 al 2013, por escuela de procedencia

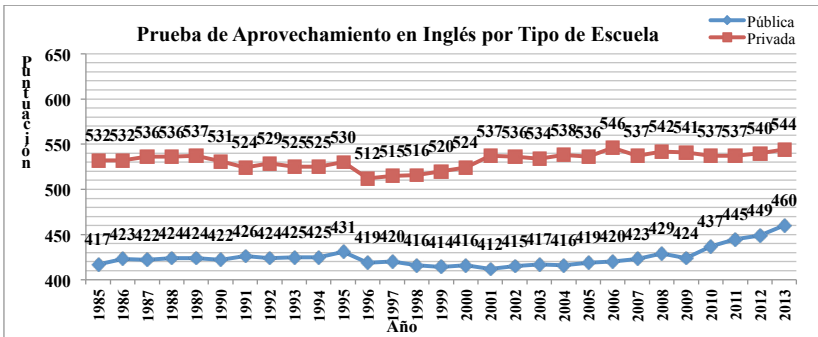


Diagrama 5. Puntuación promedio en la prueba de Aprovechamiento en Inglés como segundo idioma para los años del 1985 al 2013, por escuela de procedencia

El resultado neto de la combinación de estas dos tendencias es que la brecha entre las puntuaciones promedio de las escuelas se ha ido cerrando. La brecha mayor la exhibe la prueba de aprovechamiento en inglés.

Con el propósito de determinar posibles diferencias estadísticamente significativas entre las medias de las puntuaciones en las pruebas de aptitud y de aprovechamiento y las medias en el promedio de escuela superior por escuela de procedencia, se llevó a cabo una prueba de hipótesis. Los datos utilizados fueron suministrados por la Oficina de Admisiones de la Administración Central de la UPR y corresponden a la población de solicitantes al Recinto de Río Piedras de la UPR para el primer semestre del año escolar 2014-2015.

Para determinar si las medias de las puntuaciones en las pruebas de aptitud y de aprovechamiento y del promedio de escuela superior

Tabla 4
Puntuaciones promedio y diferencias promedio de las pruebas de aptitud y las pruebas de aprovechamiento entre las escuelas privadas y las escuelas públicas, para los años de 1985 al 2013

Prueba	Puntuación Promedio		Diferencia		Diferencia		Amplitud de las Diferencias
	Pública	Privada	Promedio	Menor	Mayor		
PAPTV	450.1	524.6	74.5	48	96		48
PAPTM	466.1	551.1	85.0	64	122		58
PAPRE	447.6	511.8	64.3	39	93		54
PAPRM	464.7	555.4	90.7	63	136		73
PAPRI	424.6	532.0	107.4	84	126		42

PAPTV = Puntuación en la prueba de razonamiento verbal

PAPTM = Puntuación en la prueba de razonamiento matemático

PAPRE = Puntuación en la prueba de aprovechamiento en español

PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas

PAPRI = Puntuación en la prueba de aprovechamiento en inglés

de las escuelas públicas era menor que las medias correspondientes de las escuelas privadas, se llevó a cabo una prueba de hipótesis unidireccional con hipótesis nula: $H_0: \mu_{Pública} \geq \mu_{Privada}$. La hipótesis alternativa correspondiente es: $H_1: \mu_{Pública} < \mu_{Privada}$. La estadística de prueba es:

$$z = \frac{\mu_{Pública} - \mu_{Privada}}{\sqrt{\frac{\sigma^2_{Pública}}{N_{Pública}} + \frac{\sigma^2_{Privada}}{N_{Privada}}}}$$

En la Tabla 5 se incluye la información sobre estas poblaciones y los resultados de las pruebas de hipótesis.

Tabla 5

Resultados de las pruebas de hipótesis para la diferencia entre las medias de las puntuaciones en las pruebas de aptitud y de aprovechamiento y en el promedio de escuela superior, por escuela de procedencia ($\alpha = 0.01$)

Prueba	Escuela	n	Media	Desviación	Z-test	
PAA	RV	Pública	1306	550.45	92.334	-11.08
		Privada	2035	585.08	81.159	
	RM	Pública	1306	541.18	104.584	-17.91
		Privada	2035	607.18	102.986	
PACH	ING	Pública	1296	537.14	110.719	-20.99
		Privada	2029	612.28	82.468	
	MAT	Pública	1296	545.77	100.552	-17.21
		Privada	2029	607.01	99.348	
	ESP	Pública	1296	531.47	92.853	-8.891
		Privada	2030	559.98	85.839	
HSGPA	Pública	1301	3.5563	0.44603	0.7744	
	Privada	2043	3.5445	0.40239		

PAA = Pruebas de Aptitud
 RV = Prueba de razonamiento verbal
 RM = Prueba de razonamiento matemático
 PACH = Pruebas de Aprovechamiento Académico
 ING = Prueba de aprovechamiento en inglés
 MAT = Prueba de aprovechamiento en matemáticas
 ESP = Prueba de aprovechamiento en español
 HSGPA = Promedio de escuela superior

Con un nivel de significancia $\alpha = 0.01$ y una prueba de hipótesis de cola de izquierda, el valor crítico correspondiente en la distribución normal es -2.33 . Note que, para todas las pruebas, la estadística de prueba resultó mucho menor que el valor crítico. Por tanto, rechazamos la hipótesis nula y podemos aseverar que las medias de las puntuaciones de las escuelas públicas resultaron ser estadísticamente menores que las correspondientes de las escuelas privadas en todas las pruebas. Sin embargo, la prueba de hipótesis para la diferencia de las medias de los promedios de escuela superior reveló que no existe diferencia significativa entre ellas. Este hallazgo resulta contradictorio, pues esperaríamos una correlación alta entre las puntuaciones en las pruebas de aptitud y aprovechamiento y el promedio de escuela superior.

La Tabla 6 y la Tabla 7 recogen las estadísticas descriptivas de la cohorte de este estudio sobre los predictores de promedio de primer año y promedio de graduación, respectivamente. Las últimas dos columnas (a la derecha de cada tabla) corresponden al coeficiente de correlación de cada predictor con el criterio correspondiente para escuelas públicas y escuelas privadas.

Este estudio llevó a cabo una prueba de hipótesis para todos los coeficientes de correlación (12 en total) entre los predictores y el promedio de primer año de estudios universitarios (6 coeficientes) y entre los predictores y el promedio de graduación (6 coeficientes). Para una prueba de hipótesis sobre la magnitud del coeficiente de correlación simple, la hipótesis nula de cero correlación es: $H_0 : \rho = 0$, donde ρ representa el coeficiente de correlación en la población. La fórmula para la estadística de prueba es:

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

con $n-2$ grados de libertad (Hinkle et al., 1994, p. 206-215).

Aunque los coeficientes de correlación fluctuaron entre 0.221 y 0.415, todas las pruebas de hipótesis, de dos colas, resultaron en un rechazo de la hipótesis nula de cero correlación, utilizando un nivel de significancia $\alpha = 0.01$. Nótese que la fórmula para la estadística de prueba t aumenta a medida que el tamaño de la muestra también aumenta. Haber utilizado muestras de tamaño grande en este estudio puede explicar por qué todos los coeficientes de correlación resultaron significativamente diferentes de cero.

Tabla 6

Estadísticas descriptivas de los predictores con respecto al promedio de primer año (FGPA) por escuela de procedencia

Variable	Media		Desviación Estándar		Coeficiente con FGPA	
	Pública	Privada	Pública	Privada	Pública	Privada
FGPA	3.083	3.174	0.531	0.490	1.000	1.000
HSGPA	3.592	3.503	0.357	0.404	0.398	0.447
PAPTV	574.80	599.49	74.566	68.461	0.328	0.334
PAPTM	587.92	626.46	87.663	87.555	0.246	0.288
PAPRE	542.48	566.97	65.620	65.427	0.315	0.346
PAPRM	579.95	620.20	88.367	85.220	0.265	0.328
PAPRI	515.60	591.86	102.520	95.173	0.241	0.297

FGPA = Promedio de primer año universitario

HSGPA = Promedio de escuela superior

PAPTV = Puntuación en la prueba de razonamiento verbal

PAPTM = Puntuación en la prueba de razonamiento matemático

PAPRE = Puntuación en la prueba de aprovechamiento en español

PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas

PAPRI = Puntuación en la prueba de aprovechamiento en inglés

Total de estudiantes de escuelas públicas = 994

Total de estudiantes de escuelas privadas = 1,229

Además de llevar a cabo una prueba de hipótesis para determinar si los coeficientes de correlación simples de la Tabla 6 y la Tabla 7 eran diferentes de cero, también se llevaron a cabo pruebas de hipótesis para determinar si los coeficientes de correlación correspondientes de escuela pública y privada eran iguales. La hipótesis nula establece que no hay diferencia entre dos coeficientes de correlación, es decir, $H_0: \rho_1 - \rho_2$. En este caso, la estadística de prueba es diferente a la de la prueba de hipótesis para un solo coeficiente de correlación (Hinkle et al., 1994, p. 262-266).

Tratándose de 12 pares de coeficientes de correlación (6 para cada criterio), se llevaron a cabo 12 pruebas de hipótesis, de dos colas, utilizando un nivel de significancia de $\alpha = 0.01$. Para los coeficientes de correlación de los predictores con el FGPA (Tabla 6), no se rechazó

Tabla 7

Estadísticas descriptivas de los predictores con respecto al promedio de graduación (GGPA) por escuela de procedencia

Variable	Media		Desviación Estándar		Coeficiente con GGPA	
	Pública	Privada	Pública	Privada	Pública	Privada
GGPA	3.335	3.341	0.332	0.348	1.000	1.000
HSGPA	3.608	3.527	0.386	0.406	0.250	0.415
PAPTV	574.92	604.53	76.454	68.081	0.320	0.330
PAPTM	587.86	630.21	90.844	88.659	0.221	0.249
PAPRE	541.80	570.23	69.060	66.204	0.302	0.334
PAPRM	580.93	625.17	91.454	86.057	0.239	0.258
PAPRI	516.11	598.90	104.071	94.005	0.275	0.264

GGPA = Promedio de graduación universitaria

HSGPA = Promedio de escuela superior

PAPTV = Puntuación en la prueba de razonamiento verbal

PAPTM = Puntuación en la prueba de razonamiento matemático

PAPRE = Puntuación en la prueba de aprovechamiento en español

PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas

PAPRI = Puntuación en la prueba de aprovechamiento en inglés

Total de estudiantes de escuelas públicas = 736

Total de estudiantes de escuelas privadas = 918

la hipótesis nula en ninguna de las pruebas. Para los coeficientes de correlación de los predictores con el GGPA (Tabla 7), solo se rechazó la hipótesis nula para los coeficientes del HSGPA (valor $p = 0.0001$). Es decir, el coeficiente de correlación del HSGPA respecto al GGPA para escuela privada ($r = 0.415$) resultó ser diferente al coeficiente correspondiente para escuela pública ($r = 0.250$).

Por último, se llevó a cabo una prueba de hipótesis para posibles diferencias en el FGPA y el GGPA de escuela pública y escuela privada, utilizando la misma estadística de prueba (z) que para comparar las medias en las puntuaciones de las pruebas. Para una prueba de hipótesis de dos colas, con $\alpha = 0.01$, encontramos que, para las diferencias en las medias del FGPA, se rechazó la hipótesis nula ($t = -4.158$, para un valor crítico = ± 2.58), mientras que, para las diferencias en las medias

del GGPA, no se rechazó la hipótesis nula ($t = -0.357$, para un valor crítico = ± 2.58).

Las estadísticas presentadas en la Tabla 6 y la Tabla 7 pueden resumirse de la siguiente manera:

1. A pesar de que las diferencias entre las puntuaciones promedio en las pruebas de aptitud y de aprovechamiento persisten, estas son menores que en la población de solicitantes, puesto que ya ha ocurrido el proceso de selección al ser admitidos a la UPR. Por la misma razón, el promedio de HSGPA de escuela pública y el de escuela privada son similares.
2. A pesar de que una prueba de hipótesis comprobó que hay una diferencia significativa entre las medias del FGPA de las escuelas públicas y las privadas, una prueba de hipótesis similar para las medias del GGPA no rechazó la hipótesis nula. Es decir, los promedios de graduación de los estudiantes de escuela pública y de escuela privada resultaron ser estadísticamente similares.
3. El promedio de escuela superior exhibió mayor desgaste como predictor de éxito universitario a largo plazo para las escuelas públicas (de $R = 0.398$ a $R = 0.250$) que para las escuelas privadas (de $R = 0.447$ a $R = 0.415$).
4. Los coeficientes de correlación de cada uno de los predictores, con cada uno de los dos criterios, fueron mayores para las escuelas privadas que para las escuelas públicas, con una sola excepción: el coeficiente de correlación de la prueba de aprovechamiento en inglés con el promedio de graduación.
5. Sin embargo, las pruebas de hipótesis sobre posible diferencia entre los coeficientes de correlación no reflejaron diferencia significativa, excepto en el caso del coeficiente de correlación entre el HSGPA y el GGPA.
6. Los coeficientes de correlación de la prueba de razonamiento verbal y la prueba de aprovechamiento en español con respecto a ambos criterios fueron mayores que los coeficientes correspondientes de las dos pruebas matemáticas y la prueba de aprovechamiento en inglés, para ambos tipos de escuela.
7. Las pruebas de razonamiento verbal y de aprovechamiento en español evidenciaron menos desgaste como predictores de criterios de éxito a largo plazo que las pruebas matemáticas.
8. Para escuelas privadas, la prueba de aprovechamiento en inglés perdió utilidad de predicción respecto a criterios de éxito a largo plazo. Por el contrario, para escuelas públicas, la misma

prueba ganó utilidad de predicción respecto a criterios de éxito a largo plazo.

La Tabla 8 incluye los resultados de los modelos de regresión, por escuela de procedencia, incorporando todos los predictores para cada uno de los dos criterios de éxito universitario. Los coeficientes beta (destacados en negritas) representan un nivel de significación $\alpha = 0.01$.

Al igual que para el coeficiente de correlación simple (r_{xy}), podemos llevar a cabo una prueba de hipótesis para la magnitud del coeficiente de correlación múltiple (R). En este caso, la hipótesis nula es: $H_0: R_{pop} = 0$. La estadística de prueba es:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

donde k representa el número de predictores. La distribución subyacente es una distribución F , con k grados de libertad para el numerador y $(n - k - 1)$ grados de libertad para el denominador (Hinkle et al., 1994, p. 460). Este estudio llevó a cabo una prueba de hipótesis para los cuatro coeficientes de correlación múltiple incluidos en la Tabla 8. En todos los casos, se rechazó la hipótesis nula de cero correlación, utilizando un nivel de significancia de $\alpha = 0.01$.

Los resultados de los modelos de regresión ilustrados en la Tabla 8 pueden resumirse de la siguiente manera:

1. El promedio de escuela superior resultó ser el mejor predictor en todos los modelos, excepto para la predicción a largo plazo de las escuelas públicas, en la que fuera desplazado del primer lugar al segundo por la prueba de razonamiento verbal. Esto es resultado de que, para escuela pública, la prueba de razonamiento verbal obtuvo el mayor coeficiente de correlación con el GGPA, $r = 0.320$, mientras que el correspondiente coeficiente de correlación del HSGPA fue de $r = 0.250$. Por ello, la prueba de razonamiento verbal entró como primer predictor en el cómputo de la ecuación de regresión múltiple.
2. La prueba de razonamiento verbal resultó ser el segundo mejor predictor en dos de los modelos, el mejor predictor para escuela pública a largo plazo y el tercer mejor predictor para escuela privada a corto plazo.
3. La prueba de aprovechamiento en español mostró utilidad de predicción solamente a corto plazo para ambos tipos de escuela.

Tabla 8

Modelos de regresión, por escuela de procedencia, con los coeficientes beta de cada uno de los predictores respecto al promedio de primer año de estudios universitarios y al promedio de graduación de bachillerato

Criterio	Escuela	HSGPA	PAPTV	PAPTM	PAPRE	PAPRM	PAPRI	R	R ²	N
FGPA	Pública	0.315	0.165	0.015	0.090	0.026	0.033	0.476	0.221	994
	Privada	0.362	0.105	0.005	0.072	0.095	0.124	0.540	0.288	1,129
GGPA	Pública	0.141	0.156	-0.028	0.076	0.051	0.116	0.382	0.139	736
	Privada	0.329	0.137	0.055	0.073	-0.021	0.096	0.490	0.235	918

FGPA = Promedio de primer año universitario
 GGPA = Promedio de graduación de universidad
 HSGPA = Promedio de escuela superior
 PAPTV = Puntuación en la prueba de razonamiento verbal
 PAPT M = Puntuación en la prueba de razonamiento matemático
 PAPRE = Puntuación en la prueba de aprovechamiento en español
 PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas
 PAPRI = Puntuación en la prueba de aprovechamiento en inglés

4. La prueba de razonamiento matemático no evidenció utilidad de predicción en ninguno de los cuatro modelos. Por su parte, la prueba de aprovechamiento en matemáticas solo evidenció utilidad de predicción para las escuelas privadas en la predicción de promedio de primer año.
5. La prueba de aprovechamiento en inglés evidenció utilidad de predicción en los modelos de las escuelas privadas, tanto a corto como a largo plazo. Para las escuelas públicas y para criterio a corto plazo, esta prueba no evidenció utilidad de predicción; sin embargo, lo hizo para criterios a largo plazo.
6. Los coeficientes de regresión múltiple de los modelos de regresión de las escuelas privadas fueron mayores y moderados ($R = 0.540$ y $R = 0.490$), mientras que los coeficientes correspondientes de las escuelas públicas fueron bajos ($R = 0.476$ y $R = 0.382$). Estos resultados indican que, para las escuelas privadas, el conjunto de predictores explica un porcentaje mayor de la variación en el criterio que para las escuelas públicas. Además, aunque los predictores, para ambos tipos de escuela, pierden utilidad de predicción a largo plazo, los de la escuela privada pierden menos.

La Tabla 9 recoge los modelos de regresión, por escuela de procedencia, con los mejores tres predictores para promedio de primer año y promedio de graduación.

Al igual que para los coeficientes de correlación múltiple de la Tabla 8, este estudio llevó a cabo una prueba de hipótesis para los cuatro coeficientes de correlación múltiple incluidos en la Tabla 9. En todos los casos, se rechazó la hipótesis nula de cero correlación, utilizando un nivel de significancia $\alpha = 0.01$.

Los resultados de la Tabla 9 pueden ser resumidos de la siguiente manera:

1. El mejor predictor, en tres de los cuatro modelos, fue el promedio de escuela superior. Solamente para escuelas públicas a largo plazo, el promedio de escuela superior ocupó el segundo lugar, después de la prueba de razonamiento verbal.
2. La prueba de razonamiento verbal emergió entre los mejores tres predictores en todos los modelos de regresión. Ocupó la primera posición en el modelo de escuelas públicas para criterio a largo plazo, la segunda posición en los modelos de escuela pública a corto plazo y de escuela privada a largo plazo, y la tercera posición en el modelo de escuela privada a corto plazo.

Tabla 9

Modelos de regresión, por escuela de procedencia, con los mejores tres predictores para promedio de primer año y promedio de graduación, con sus coeficientes beta correspondientes

Criterio	Escuela	Predictor	Coefficiente beta	R	R ² Ajustado	N
FGPA	Pública	HSGPA	0.324	0.473	0.222	994
		PAPTV	0.188			
		PAPRE	0.105			
	Privada	HSGPA	0.388	0.530	0.279	1,129
		PAPRI	0.177			
		PAPTV	0.164			
GGPA	Pública	PAPTV	0.197	0.377	0.139	736
		HSGPA	0.165			
		PAPRI	0.137			
	Privada	HSGPA	0.344	0.485	0.233	918
		PAPTV	0.178			
		PAPRI	0.128			

FGPA = Promedio de primer año universitario

GGPA = Promedio de graduación de universidad

HSGPA = Promedio de escuela superior

PAPTV = Puntuación en la prueba de razonamiento verbal

PAPTM = Puntuación en la prueba de razonamiento matemático

PAPRE = Puntuación en la prueba de aprovechamiento en español

PAPRM = Puntuación en la prueba de aprovechamiento en matemáticas

PAPRI = Puntuación en la prueba de aprovechamiento en inglés

3. La prueba de aprovechamiento en inglés apareció entre los mejores tres predictores en tres de los cuatro modelos de regresión. Los resultados apuntan a que esta prueba adquiere validez de predicción para criterios de éxito universitario a largo plazo.
4. Ninguna de las dos pruebas de matemáticas —ni la de razonamiento matemático ni la de aprovechamiento en matemáticas— figuró entre los mejores tres predictores en ninguno de los cuatro modelos de regresión.
5. Al igual que ocurriera con los resultados de la Tabla 8, el coeficiente de correlación múltiple de los modelos de regresión de las escuelas privadas fue mayor ($R = 0.530$ [moderado] para FGPA y $R = 0.485$ [bajo] para GGPA) que el coeficiente corres-

pondiente de los modelos de regresión de las escuelas públicas ($R = 0.473$ [bajo] para FGPA y $R = 0.377$ [bajo] para GGPA).

Discusión

En lo que atañe a la primera pregunta de investigación de este estudio, las medias de las puntuaciones en las pruebas de aptitud y de aprovechamiento de las escuelas privadas, durante los años de 1985 a 2013, fueron mayores que las medias correspondientes de las escuelas públicas. La prueba de hipótesis sobre la diferencia entre estas medias para la población de solicitantes al Recinto de Río Piedras de la UPR para el primer semestre del año académico 2014-15 reveló que las medias de las puntuaciones de las escuelas públicas, en todas las pruebas, eran estadísticamente menores que las correspondientes de las escuelas privadas. Sin embargo, la brecha entre las medias se ha ido cerrando como resultado de que las medias de las puntuaciones de las escuelas privadas manifiestan una tendencia decreciente, mientras que las medias de las puntuaciones de las escuelas públicas se han mantenido relativamente similares.

Por otro lado, la prueba de hipótesis para la diferencia entre las medias del promedio de escuela superior para la población de solicitantes al Recinto de Río Piedras de la UPR para el primer semestre del año académico 2014-15 no reveló diferencia significativa. Esto resulta contradictorio, pues anticiparíamos una alta correlación entre las puntuaciones en las pruebas de aptitud y aprovechamiento, y el promedio de escuela superior. Sin embargo, una vez admitidos al Recinto, la prueba de hipótesis para la diferencia de los promedios académicos de graduación (GGPA) de los estudiantes de escuela pública y privada no reveló diferencia significativa. Estos resultados parecen indicar que, una vez admitido a estudios universitarios, el efecto inicial que puedan tener las diferencias en las pruebas de aptitud y de aprovechamiento sobre el estudiante se disipa con el tiempo, dado que entran en juego otras variables relacionadas con el esfuerzo, el compromiso y la dedicación hacia sus responsabilidades académicas.

Dadas las diferencias en las puntuaciones promedio de las pruebas, resulta natural buscarles explicación. Una de las fuentes de evidencia de la validez de una prueba, como indicáramos anteriormente, es aquella basada en sus consecuencias. Los SEPT advierten la importancia de distinguir entre evidencia directamente relacionada con la validez de la prueba y evidencia que puede informar decisiones de política social, que cae fuera del ámbito de su validez. Sobre este particular, los SEPT

afirman que distinguir entre asuntos de validez y asuntos de política social resulta particularmente importante en aquellos casos en los que consecuencias diferenciables del uso de una prueba son observadas para diferentes grupos (AERA et al., 1999, p. 16).

Cuando se trata de buscar evidencia directamente vinculada a la validez de las pruebas, debemos referirnos a cuestiones tales como la *baja representación del constructo* (“construct underrepresentation”) y la *varianza irrelevante del constructo* (“construct irrelevance variance”). El primer concepto expresa el grado en el que una prueba no incluye aspectos importantes del constructo, como, por ejemplo: algún tipo de contenido, procesos psicológicos y maneras de responder vinculadas a este. De otra parte, el segundo concepto denota el grado en el que la puntuación en una prueba se ve afectada por procesos foráneos al constructo de interés. Es decir, las puntuaciones pueden ser sistemáticamente influenciadas por componentes que no forman parte del constructo.

Estos dos conceptos, la baja representación del constructo y la varianza irrelevante del constructo, vinculados directamente a la validez de la prueba, podrían ser la explicación de las diferencias entre las puntuaciones promedio de las escuelas públicas y las privadas. Sobre este particular, los SEPT advierten que es de particular preocupación el grado en el que la baja representación del constructo o la varianza irrelevante del constructo pueden concederle ventaja o desventaja a uno o más subgrupos examinados (AERA et al., 1999, p. 12).

Dado que un porcentaje considerable de instituciones de educación postsecundaria utilizan las pruebas de aptitud como criterios de admisión, es necesario recolectar toda la evidencia posible sobre su validez para garantizar que, efectivamente, discriminan en la selección de aquellos estudiantes que tienen mayor probabilidad de éxito en sus estudios. Es por esto que las organizaciones que confeccionan tales pruebas periódicamente recogen evidencia sobre su validez. La responsabilidad de la recolección de evidencia debe ser compartida, tanto por las organizaciones que confeccionan las pruebas, como de las instituciones que las utilizan (AERA et al., 1999).

Asumiendo que la evidencia sobre la validez de la prueba es aceptable, las diferencias entre las puntuaciones promedio podrían atribuirse a factores tales como: la preparación y capacitación del personal magisterial de las escuelas, la calidad de la enseñanza, las metas y el tipo de orientación de las escuelas privadas, el apoyo institucional que

recibe el estudiante, y factores socioeconómicos, tales como el ingreso familiar y el nivel de educación de los padres.

Para enseñar en las escuelas públicas de Puerto Rico, se requiere que el maestro o maestra obtenga un certificado que le faculte para hacerlo. En cambio, este requisito no existe para las escuelas privadas. Por tanto, la preparación y competencias del personal docente y la calidad de la enseñanza parecerían no ser factores determinantes en la explicación del fenómeno de las diferencias en las puntuaciones. En todo caso, tenderían a favorecer a la escuela pública.

Por otro lado, la gran mayoría de las escuelas privadas están orientadas hacia la capacitación de sus estudiantes para proseguir estudios universitarios. Siendo así, es natural que ofrezcan todo el apoyo institucional posible, incluyendo personal, materiales, equipo e instalaciones, para la consecución de dicha meta. Estas escuelas, conscientes de la importancia de las pruebas de aptitud y aprovechamiento para lograr acceso a estudios universitarios, ofrecen orientaciones sobre estas pruebas y, en ocasiones, adiestran y ofrecen práctica en ellas. Además, existe evidencia a los efectos de que el porcentaje de estudiantes de escuela privada que repite estas pruebas es mayor que el correspondiente a las escuelas públicas. Esto permite que los estudiantes de escuela privada tengan oportunidad de mejorar su puntuación.

Una explicación para la diferencia tan marcada en las puntuaciones promedio en la prueba de aprovechamiento en inglés podría ser que las escuelas privadas enfatizan más la enseñanza del inglés; de hecho, en algunas instancias, la enseñanza es enteramente en inglés. También, por lo general, disponen de una mayor oferta de cursos preparatorios para estudios universitarios, especialmente en ciencias y matemáticas.

Un estudio longitudinal llevado a cabo en 2007 en los Estados Unidos por el Center on Educational Policy (CEP), consideró diversas variables para explicar y comparar el aprovechamiento académico de una muestra de 1,003 estudiantes procedentes de diferentes tipos de escuelas públicas y privadas (tanto laicas como religiosas). El trabajo utilizó técnicas estadísticas muy similares a las que sigue un estudio de validez de predicción. Los criterios de éxito seleccionados fueron las puntuaciones obtenidas en las pruebas de razonamiento verbal y de razonamiento matemático, conocidas en los Estados Unidos como el SAT I (Scholastic Aptitude Test I). Estas pruebas son muy similares a las pruebas de aptitud (PPA) que administra el CB en Puerto Rico. Los resultados de este estudio revelaron que los mejores dos predictores de estas pruebas fueron: el aprovechamiento académico evidenciado en

pruebas similares tomadas en el octavo grado y los factores de carácter socioeconómico.

Todo parece indicar que, además del aprovechamiento académico anterior, los factores que más influyen sobre el aprovechamiento académico de los estudiantes son de naturaleza socioeconómica. Padres de un nivel socioeconómico y profesional más alto disponen de mejores y mayores recursos que les permiten proveer a sus hijos de oportunidades educativas a las que otros, de nivel más bajo, no tienen acceso. Por ejemplo, ubicar a sus hijos en escuelas que gozan de prestigio académico, proveerles oportunidades de viajes, visitas a museos y otras actividades extracurriculares, que inciden sobre su formación académica. De igual manera, tienen a su disposición la posibilidad de servicios profesionales, tales como preparación y repaso para las pruebas de aptitud y de aprovechamiento, tutorías, servicios psicológicos, entre otros. También, los padres de nivel socioeconómico más alto tienden a involucrarse más en la educación de sus hijos y cooperar, directa o indirectamente, con la escuela. El estudio del CEP (2007) denomina estos factores de carácter socioeconómico como “capital económico” y “capital cultural”.

Por último, las escuelas públicas tienen una responsabilidad social que no tienen las escuelas privadas. De entrada, estas últimas, por el mero hecho de cobrar por sus servicios, resultan ser selectivas, pues solo aquellos estudiantes que dispongan de los recursos económicos necesarios podrán ser admitidos. Además, pueden ser selectivas en su admisión, utilizando criterios como puntuaciones de exámenes de entrada. Las escuelas públicas, por el contrario, no pueden ser selectivas: están obligadas, por ley, a ser inclusivas. Finalmente, las escuelas privadas tienen la opción de expulsar a cualquier estudiante que no muestre cierto nivel de aprovechamiento académico o que exhiba una conducta social indeseable. Las escuelas públicas no pueden responder de esa manera.

No por esto las escuelas públicas deben renunciar a su responsabilidad social de proveer a todos sus estudiantes una educación de excelencia, que les permita aspirar, en igualdad de condiciones, a proseguir estudios postsecundarios. Por el contrario, deben procurar todos los recursos y apoyo que estén a su alcance para fortalecer sus ofrecimientos y subsanar la brecha existente, evidenciada en las diferencias en las puntuaciones de las pruebas de aptitud y aprovechamiento. No obstante, a la luz de la tendencia reflejada en los últimos años, todo parece indicar que la brecha, paulatinamente, se está reduciendo.

Tocante a la segunda pregunta de investigación, los resultados de este estudio revelan que los coeficientes de correlación entre los predictores y los criterios de éxito universitario, tanto a corto como a largo plazos, fueron mayores para las escuelas privadas que para las escuelas públicas, con tan solo una excepción: el coeficiente de correlación de la prueba de aprovechamiento en inglés con respecto al promedio de graduación. Los resultados obtenidos en este estudio, para una cohorte de estudiantes admitidos al Recinto de Río Piedras de la UPR en el año 2000, fueron similares a los resultados obtenidos por Menéndez (1996), para una cohorte de estudiantes admitidos a la UPR en el año 1989.

Al igual que en el estudio de Menéndez (1996), el coeficiente de correlación múltiple (R), computado para los modelos de regresión generados para ambos criterios, resultó mayor para las escuelas privadas que para las escuelas públicas. Es decir, el mismo conjunto de predictores demostró tener mayor utilidad de predicción para las primeras que para las segundas. Además, la disminución en la utilidad de predicción del conjunto de predictores para criterios a largo plazo fue menor para las escuelas privadas.

Respecto a la tercera pregunta de investigación, los modelos de regresión con los mejores tres predictores generados para ambos criterios y ambos tipos de escuela revelaron semejanzas en lo que concierne a la utilidad de predicción de sus predictores. El promedio de escuela superior emerge como el mejor predictor de éxito universitario en todos los casos, excepto para las escuelas públicas respecto al promedio de graduación, en donde quedara en segunda posición después de la puntuación en la prueba de razonamiento verbal. Estos resultados confirman resultados obtenidos por Menéndez (1996). Colorado y Corcino (2013) obtuvieron resultados similares, pero para toda la cohorte del 2000, es decir, sin considerar la variable moderadora de escuela de procedencia.

Muy de cerca en utilidad de predicción, le siguió la prueba de razonamiento verbal, que se ubicara entre los tres mejores predictores de éxito universitario a corto y a largo plazo, para ambos tipos de escuela. Para Menéndez (1996), dicha prueba ocupó la tercera posición como predictor de éxito universitario a corto plazo y la segunda posición a largo plazo. Para Colorado y Corcino (2013), esta prueba ocupó la segunda posición como predictor para ambos criterios, para toda la cohorte de 2000.

De otra parte, en este estudio, las pruebas de razonamiento matemático y aprovechamiento en matemáticas no figuraron entre los mejores tres predictores en ninguno de los modelos de regresión. Tampoco figuraron entre los primeros tres predictores de éxito universitario a corto o a largo plazos para Menéndez (1996), ni para Colorado y Corcino (2013), para toda la cohorte de 2000. Podemos especular que las pruebas matemáticas reflejarían mayor utilidad de predicción si consideráramos el área de estudio como variable moderadora. Es razonable anticipar que, para áreas de estudio de alto contenido matemático, como matemáticas, ciencias, ingeniería y administración de empresas, la utilidad de predicción de estas pruebas aumente. Para examinar esta conjetura, sería recomendable llevar a cabo un estudio similar al presente, incorporando el área de estudio como variable moderadora.

La situación de la prueba de aprovechamiento en inglés resultó particularmente interesante. A pesar de que correlacionó muy poco con ambos criterios de éxito universitario, para ambos tipos de escuela, adquirió utilidad de predicción para el criterio a largo plazo, emergiendo como el tercer mejor predictor de promedio de graduación para ambos tipos de escuela, por encima de las puntuaciones en las pruebas matemáticas. Estos resultados coinciden con los obtenidos por Colorado y Corcino (2013), para toda la cohorte de 2000, para los que la prueba de aprovechamiento en inglés no solo emergió como el tercer mejor predictor para criterios a largo plazo, sino también para criterios a corto plazo.

Al igual que hiciéramos con las pruebas matemáticas, podríamos anticipar que la prueba de aprovechamiento en inglés refleje mayor utilidad de predicción en áreas de estudios como matemáticas, ciencias, ingeniería y administración de empresas, y naturalmente en la concentración en inglés. En estas disciplinas, el estudiante, con frecuencia, utiliza libros de texto y otros materiales didácticos en inglés, además de redactar informes, monografías y otros trabajos en dicho idioma.

Los resultados de este estudio evidenciaron que los mejores predictores de éxito universitario, tanto a corto como a largo plazos, para ambos tipos de escuela, son el promedio de escuela superior y las pruebas de razonamiento verbal y de aprovechamiento en inglés. La posición preponderante del promedio de escuela superior como predictor de éxito justifica el peso que se le ha asignado (50%) en la fórmula de ingreso vigente. Sin embargo, parecería conveniente considerar la

posibilidad de incluir, en la fórmula de ingreso, como se hiciera en el pasado, las puntuaciones en las pruebas de aprovechamiento, especialmente la puntuación en la prueba de aprovechamiento en inglés.

Finalmente, dado que la cohorte que utilizara este estudio es del 2000, sería recomendable llevar a cabo un nuevo estudio de validez con una cohorte más reciente, para examinar si el comportamiento de los predictores y la utilidad de predicción de cada uno de ellos y en conjunto ha cambiado. Además, un nuevo estudio de validez podría considerar la inclusión de predictores de tipo socioeconómico, como el ingreso familiar y la preparación académica de los padres de los solicitantes. Actualmente, algunos de estos datos están disponibles para estudiantes que solicitan ingreso a la UPR, puesto que, al momento de hacerlo, llenan un cuestionario que incluye algunos de estos datos.

REFERENCIAS

- American Educational Research Association [AERA], American Psychological Association [APA] & National Council of Measurement in Education [NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA. Recuperado de <https://law.resource.org/pub/us/cfr/ibr/001/aera.standards.1999.pdf>
- Borg, R. B. & Gall, M. D. (1983). *Educational research: An introduction* (4ta. ed.). New York: Longman.
- Bridgeman, W., McCamley-Jenkins, L. & Ervin, N. (2000). *Prediction of freshman grade point average from the revised and recentered SAT I: Reasoning Test*. College Board Research Report No. 2000-1. New York: College Entrance Examination Board. Recuperado de <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2000-1-predictions-freshman-gpa-revised-recentered-sat-reasoning.pdf>
- Burton, N. W. & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980*. College Board Research Report No. 2001-2. New York: College Entrance Examination Board. Recuperado de <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2001-2-predicting-college-success-sat-studies.pdf>
- Center for Educational Policy [CEP]. (2007). *Are private high schools better academically than public high schools?* Washington, D.C.: Autor. Recuperado de: www.edline.com/uploads/pdf/PrivateSchoolsReport.pdf
- College Board [CB]. (s.f.). Programa de Evaluación y Admisión Universitaria (PEAU). Recuperado de <http://opr1a.collegeboard.com/ptorico/pr/program/peau.html>

- Colorado, R. & Corcino, L. (2013). Validez de predicción de los criterios de admisión a la Universidad de Puerto Rico. *Revista Pedagogía*, 46(1), 66-92.
- Geiser, S. & Santelices, M. (2007). Validity of high-school grades in predicting student success beyond freshman year: High-school record vs. standardized tests as indicators of fourth-year college outcomes. *Research & Occasional Paper Series: CSHE.6.07*. Recuperado de http://cshe.berkeley.edu/publications/docs/ROPS.GEISER_SAT_6.12.07.pdf
- Geiser, S. & R. Studley. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26.
- Hawkins, D.A. & Clinedinst, M. (2006). *State of College Admission*. Alexandria: Virginia: National Association for College Admission Counseling. Recuperado de http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/NACAC_US/N060508H.pdf
- Hinkle, D., Wiersma, W. & Jurs, S. (1994). *Applied statistics for the behavioral sciences* (3ra. ed.). Boston, Mass.: Houghton Mifflin Company.
- Menéndez, A. (1995). Estudio sobre el proceso de admisiones a la Universidad de Puerto Rico. Informe presentado a la Administración Central de la Universidad de Puerto Rico.
- Menéndez, A. (1996). *The predictive validity of admissions measures at the University of Puerto Rico* (Disertación doctoral). Rutgers University.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Punch, K. F. (2004). *Introduction to social research: Quantitative and qualitative approaches*. Londres: Sage Publications.
- Ramist, L., Lewis, C. & McCamley-Jenkins, L. (1993). Student group differences in predicting college grades: Sex, language and ethnic groups. *College Board Research Report No. 93-1*. New York: College Entrance Examination Board. Recuperado de <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-1993-1-student-group-differences-predicting-college-grades.pdf>
- Zwick, R. (2007). *College Admissions Testing*. Report for the National Association for College Admission Counseling (NACAC). Recuperado de <http://www.nacacnet.org/research/PublicationsResources/Marketplace/Documents/TestingWhitePaper.pdf>