



Propiedades Psicométricas de una prueba de admisión universitaria

Psychometric Properties of a Higher Education Admission Test

Tania Elena Moreira-Mora ¹ *

1 - Departamento de Orientación y Psicología, Comité Exámen de Admisión. Instituto Tecnológico de Costa Rica, Cartago.

Recibido: 30/09/2020 **Revisado:** 17/11/2020 **Aceptado:** 08/12/2020

Introducción
Método
Resultados
Discusión
Referencias

Resumen

Esta investigación pretende demostrar las propiedades psicométricas de la prueba de admisión del Instituto Tecnológico de Costa Rica y presentar las evidencias de contenido, constructo y criterio para el modelo de medida. En el estudio no experimental y correlacional participaron estudiantes que realizaron la prueba en 2019, de los cuales 57% eran hombres y 43% mujeres, 74% provenían de colegios públicos y el resto de colegios privados y subvencionados. Los hallazgos demostraron que la tabla de especificaciones proporciona las primeras evidencias asociadas al contenido, cuyos indicadores matemáticos y verbales se asociaron significativamente con el rasgo latente. También se comprobó el valor predictivo de estos indicadores y del promedio de la educación diversificada con respecto al rendimiento del estudiantado en el primer periodo universitario. Se concluyó que la validación debe ser un proceso permanente y riguroso para consolidar el diseño de una prueba y garantizar una interpretación y uso adecuado de sus resultados.

Palabras clave: *evidencias de validez, prueba estandarizada, aptitud académica, modelo de ecuaciones estructurales, modelo predictivo*

Abstract

The objective of this study is to demonstrate the psychometric properties of the Aptitude Academic Test of the Instituto Tecnológico of Costa Rica, and it presents evidences about the content, construct and criteria. On this non-experimental and correlational study participated the students who applied for this test on 2019, 57% of them were men and 43% women, 74% came from public high schools and the rest, from private and subsidized schools. The findings demonstrate that the specification table provides the first evidence associated with the test content, whose mathematical and verbal indicators are significantly associated with the trait of academic aptitude. Additionally, the predictive value of these test indicators and the grades of diversified education were verified with the performance of the students in the first period. It is concluded that validation should be a permanent and rigorous process to consolidate the design of a standardized test and guarantee the proper use and interpretation of its results.

Keywords: *evidence of validity, standardized test, academic aptitude, structural equation model, predictive model*

*Correspondencia a: tmoreira@tec.ac.cr

Introducción

Las instituciones de educación superior han venido utilizando sistemas, estrategias e instrumentos para la selección y admisión de sus estudiantes. En el caso de la universidad estatal, y específicamente en el del Instituto Tecnológico de Costa Rica (TEC), se ha utilizado la Prueba de Aptitud Académica (PAA) desde su apertura en 1973 para elegir a los candidatos con mayores probabilidades de éxito académico. Desde entonces, con algunas variantes en el diseño, el modelo de medida de la prueba se mantuvo enfocado tanto en la medición de habilidades matemáticas relacionadas con: operaciones básicas, ecuaciones, expresiones algebraicas, porcentajes, geometría, entre otras; como en habilidades verbales relativas a la comprensión lectora en textos cortos y largos, antónimos y el razonamiento mediante analogías y silogismos.

En el 2009, los responsables de la construcción de la prueba consideraron necesaria una revisión teórica y técnica del modelo de medida de la PAA con el objetivo de actualizarlo de acuerdo al perfil de entrada de los estudiantes, a las nuevas demandas cognoscitivas de las carreras y al compromiso ético de garantizar una medición equitativa acorde con las propiedades psicométricas requeridas para una prueba con altas consecuencias sociales. [Bringas-Benavídez y Pérez-Mejía \(2014\)](#) señalan que este tipo de pruebas, además de cumplir con los propósitos específicos de un examen de admisión y de seleccionar a los aspirantes a primer ingreso; pueden proporcionar un diagnóstico de las competencias desarrolladas en la formación previa, informar de las capacidades académicas de los aspirantes, y brindar insumos para generar acciones compensatorias con la finalidad de mejorar el rendimiento académico.

Gracias a esta revisión del modelo se logró establecer un conjunto de habilidades verbales y

matemáticas básicas que, para estudiantes y docentes, eran relevantes en el perfil de entrada de los estudiantes en el TEC. Estas han sido la base para iniciar una etapa de transición hacia un nuevo modelo de medición de la PAA a partir de un riguroso proceso de validación.

De esta manera, en el presente estudio se pretende demostrar las propiedades psicométricas de la Prueba de Aptitud Académica del Instituto Tecnológico de Costa Rica y analizar cuál es el grado de relación entre las variables observables de las habilidades verbales y matemáticas del modelo de medida de la PAA y el rasgo latente de aptitud académica, y cuál es el grado de validez predictiva de los componentes de la nota de admisión en el rendimiento académico de los estudiantes en su primer año de ingreso.

Por ello, en esta investigación se presentan los análisis más recientes de las propiedades psicométricas de la prueba, relativos a las evidencias de validez asociadas al contenido, criterio y constructo, desde la perspectiva de la teoría clásica de los test (TCT). Tales análisis se sustentan en la premisa de que cuantas más evidencias de validez se obtengan de una prueba, más sólido es el modelo de medida y más apropiado el uso de sus resultados.

Desde la perspectiva unificada propuesta por [Messick](#), la validez encierra

...un juicio integrado y evaluativo del grado en que la evidencia empírica y las razones teóricas apoyan lo adecuado y lo apropiado de las interpretaciones y las acciones basadas en las puntuaciones de las pruebas u otras formas de evaluación ([1995, p. 5](#)).

En los estándares para las pruebas educativas y psicológicas establecidos por la [American Educational Research Association \(AERA\)](#), la [American Psychological Association \(APA\)](#) y el [National Council on Measurement in Education](#)

(NCME), el proceso de validación es un asunto del grado en que la evidencia acumulada proporciona una base científica para la interpretación de las puntuaciones de la prueba y de su relevancia (2014). Además, señalan que existen cinco fuentes de evidencias de validez, basadas en el contenido, los procesos de respuesta, la estructura interna, la relación con otras variables y las consecuencias de los usos de la prueba.

La validez de contenido es la primera característica a analizar en una prueba y, como lo destacan Juárez-Hernández y Tobón (2018), es un componente importante de la validez de constructo, ya que se relaciona con el grado en que los ítems son relevantes y pertinentes al constructo de interés.

La construcción de los ítems de la PAA se fundamenta en las tareas cognitivas descritas en una tabla de especificaciones, lo que permite garantizar una adecuada conceptualización y operacionalización del constructo; es decir, especificar previamente las dimensiones a medir y sus indicadores, como lo destacan Escobar-Pérez y Cervo-Martínez (2008). Además, es el uso riguroso de esta tabla de especificaciones lo que garantiza las evidencias de validez basadas en el contenido (Rojas-Torres & Ordóñez-Gutiérrez, 2019).

También se pretende obtener evidencias del grado de relación entre las variables observables relativas a las habilidades verbales y matemáticas del modelo de medida de la PAA y el rasgo latente de *aptitud académica*. Como lo explica Cea-D'Ancona (2002), estas evidencias se pueden generar a partir de una teoría, de generalizaciones empíricas o de la estructura latente que el investigador espera encontrar en los datos.

Usualmente, la investigación en ciencias sociales ha usado los modelos de ecuaciones estructurales (SEM) para evaluar tanto las propiedades psicométricas del modelo de medida como para obtener evidencias de validez de constructo,

puesto que permiten relacionar las variables latentes con indicadores o variables observables.

Específicamente, en un análisis realizado sobre los indicadores del constructo de un modelo de admisión para predecir el rendimiento académico en postgrado, resultó que los predictores más potentes fueron las habilidades intelectuales como razonamiento verbal o numérico, dependiendo de la magnitud del coeficiente de validez predictiva, y de las características de los postgrados (Guglietta & Delgado-Álvarez, 2010).

En cuanto a las evidencias asociadas a los coeficientes de validez predictiva entre una medida criterio, usualmente el promedio del primer año universitario, y los predictores, como la puntuación de las variables de la batería de selección, se han usado modelos de regresiones mínimo cuadráticas (Contreras, Gallegos, & Meneses 2009; Young, 2001). Particularmente, se espera que el ordenamiento de los postulantes por nota de examen coincida con el ordenamiento que tendrían respecto de su potencial desempeño universitario (Manzi et al., 2010).

En estudios predictivos con pruebas de ingreso como predictores del rendimiento, se ha encontrado una correlación significativa entre los predictores y los resultados obtenidos por los estudiantes en los primeros semestres (García-Domínguez, 2016; López, Echazarreta, Pech, & Gómez, 2010).

Particularmente, Vergara-Díaz y Peredo-López (2017) encontraron que los puntajes de las pruebas de selección universitaria tienen una relación más débil con el rendimiento académico estudiantil en comparación con el promedio de nota de la educación media, que lo explica mejor al presentar una mejor correlación en el primer semestre y una más fuerte en el segundo. De acuerdo con Bleyaert (2010), esta tendencia también se ha encontrado en algunos estudios donde la calificación promedio (Grade Point Average,

GPA) de escuelas secundarias ha mostrado una mayor contribución predictiva que las puntuaciones de la prueba universitaria americana (American College Testing, ACT); aunque en otros estudios ACT ha sido un mejor predictor que el GPA universitario. Finalmente, la autora concluye que la combinación más eficaz para predecir el éxito universitario es el GPA de secundaria y la nota de ACT.

Una limitación en estos análisis predictivos es que los promedios de las calificaciones obtenidas en la universidad y secundaria no son buenos criterios, por la variación de un profesor a otro y de una institución a otra (Backhoff, Tirado, & Larrazolo, 2001). La otra limitación es el problema del sesgo en la selección de la muestra, conocido como restricción de rango, ya que los sujetos de evaluación son aquellos estudiantes que logran ingresar al sistema universitario (Contreras et al., 2009).

A modo de síntesis, destaca Moreira-Mora (2016) que los estudios de pruebas estandarizadas se han centrado en dos vertientes. Una es la psicométrica, fundamentada en los estándares para pruebas psicológicas y educativas establecidas por la AERA, la APA y el NCME, que competen a la calificación, confiabilidad y validez, procesos administrativos y de construcción (Kobrin & Kimmel, 2006; Lawrence, Rigor, Van Essen, & Jackson, 2003; Martínez-Arias, 2005; Young, 2001). La segunda está más orientada hacia los modelos cognitivos y la conceptualización del constructo, que se relacionan con las habilidades o dimensiones cognoscitivas abordadas por estas pruebas (Gierl, Tan, & Wang, 2005; Milewski, Johnsen, Glazer, & Kubota, 2005). No obstante, a pesar de que estos paradigmas se han desarrollado separadamente, en la actualidad existen propuestas que plantean posibles integraciones (Cliff & Montero, 2010).

En relación con el rasgo latente de *razona-*

miento general medido en la prueba de admisión del TEC, la teoría psicológica respalda teóricamente que un rasgo es evaluable en una única medición, puesto que los factores que determinan la capacidad general de una persona no deberían experimentar modificaciones notorias en un período relativamente breve, por lo tanto, esa estabilidad favorece las predicciones del desempeño (Donoso-Díaz & Hawes-Barrios, 2001). Al respecto, Contreras, Bravo y Sanhueza (2001) destacan que estas pruebas están referidas principalmente a rasgos estables a partir de los cuales pueden hacerse predicciones sobre desempeños futuros de los sujetos y que permiten clasificar y ordenar a los postulantes sobre una base científicamente sustentada.

Este rasgo de razonamiento general ha sido medido con dos factores: *razonamiento matemático* y *verbal*. El primero es una habilidad que involucra la resolución de problemas a partir del aprendizaje de conocimientos y del desarrollo de estrategias cognitivas y metacognitivas. Para la Organización para la Cooperación y Desarrollo Económico (OCDE), la *competencia matemática* es la capacidad del estudiantado para razonar y emplear conceptos, procedimientos, hechos y herramientas matemáticas para describir, explicar y predecir fenómenos de diverso tipo (Instituto Nacional de Evaluación Educativa, 2019).

Una de las categorías de este factor es la *resolución de problemas*, que implica el uso de conocimientos básicos, definir y ejecutar una estrategia de solución. La resolución de un problema, como señalan Marino y Rodríguez (2009), demanda “una complejidad cognitiva mayor, en tanto que el alumno debe elaborar su propio método de resolución, apelando a sus conocimientos previos, estableciendo nuevas relaciones entre ellos y, además, empleando diversos procedimientos, tanto algorítmicos como heurísticos” (p. 161).

La categoría de *razonamiento deductivo*

“implica procedimientos matemáticamente bien definidos para extraer las consecuencias que se siguen con certeza o necesidad de algún tipo de evidencia” (Lassiter & Goodman, 2015, p. 124). Mediante la deducción, se conduce de forma sistemática de un grupo de proposiciones a otro. En tanto que, mediante la inducción, se pueden crear nuevos conceptos analizando semejanzas o diferencias; este razonamiento se basa en “operaciones como clasificar, completar series, hacer analogías y comparaciones con diferentes tipos de símbolos (verbales, figuras, entre otros), que permiten llegar a hacer inferencias para definir esos nuevos conceptos y posteriormente aplicarlos y evaluarlos” (Iriarte-Díaz-Granados et al., 2010, p. 42). Finalmente, el *razonamiento con figuras* está más relacionado con la habilidad de percibir formas y transformarlas mentalmente, mediante la manipulación mental de formas y objetos percibidos.

En cuanto al *razonamiento verbal*, ha sido definido como la capacidad para utilizar el lenguaje verbal en el análisis semántico e inferencial en la lectura de diversos textos. Según el *College Board* (2014) las pruebas de aptitud académica se enfocan, por un lado, en la habilidad para reconocer las relaciones entre las distintas partes de una oración y reconocer el significado de las palabras en contexto y, por otro lado, en la habilidad para analizar, inferir y establecer relaciones analógicas, así como la habilidad para la síntesis de la información y la comparación entre las partes de un mismo texto o entre textos o lecturas diferentes.

En el caso de la PAA del TEC, estos indicadores matemáticos y verbales se han ido especificando en los últimos años, a partir de la inclusión de distintas tareas cognitivas que permiten medir de manera más precisa el rasgo latente de razonamiento general.

Método

El enfoque es cuantitativo por el interés en demostrar las propiedades psicométricas de la PAA, por tal razón, este estudio se diseñó como no experimental y correlacional y se realizó con los resultados de la prueba de admisión aplicada en el 2019.

Participantes

Para el análisis factorial confirmatorio se incluyó a los 16285 examinados de las tres convocatorias ordinarias del 2019, de los cuales 74.1% provenían de colegios públicos, 17.1% de establecimientos privados y 8.3% de subvencionados. De estos, 57.2% eran hombres y 42.8% mujeres. En cuanto al lugar de procedencia, la provincia con más alto porcentaje de población fue San José, a donde pertenecía el 28.8% de los examinados, seguida de Alajuela con 20.9%, Cartago con un 20.6% y Heredia con un 8.4%; en tanto que el 21.1% restante era de las provincias costeras de Guanacaste, Limón y Puntarenas. Finalmente, de acuerdo con el Índice de Desarrollo Social (IDS), 76.7% de los examinados provenía de distritos con un índice igual o superior a 60 y 23.1% menor a 60 (niveles más bajos de desarrollo). El IDS es calculado por el Ministerio de Planificación Nacional y Política Económica (MIDEPLAN, 2017) a partir de un conjunto de índices socioeconómicos distribuidos en cinco dimensiones: económica, participación social, salud, educación y seguridad, para clasificar los distritos de Costa Rica por su nivel de desarrollo. De esta manera, un valor de 100 indica que, en promedio, el distrito tiene los mejores indicadores en relación con los demás.

La muestra para los análisis predictivos es-

tuvo conformada por los estudiantes que se matricularon en el primer semestre del 2020: 66% hombres y 34% mujeres, de los cuales 62% estudiaba en colegios públicos, 23% en colegios privados y 15% en colegios subvencionados. Además, el 85.6% provenía de las provincias de San José, Cartago, Alajuela y Heredia, el grupo restante era de las provincias costeras de Guanacaste, Puntarenas y Limón.

Instrumentos

La fuente de las notas del componente matemático y verbal de la PAA fue el Comité Examen de Admisión. Este equipo técnico construye y valida cada año la PAA, la cual está estructurada con 80 preguntas de selección única: 50 de matemáticas y 30 de razonamiento verbal. El Departamento de Admisión y Registro (DAR) proporcionó la nota de la educación diversificada y el promedio simple del primer semestre del 2020.

La ponderación de la nota de admisión en el TEC corresponde en un 60% a los componentes de matemática y verbal de la PAA y en un 40% a la nota de la educación diversificada informada por el colegio al DAR. En el 2019, esta nota se calculó como un promedio de las calificaciones obtenidas en español, matemática, estudios sociales, educación cívica, lengua extranjera (inglés o francés) y ciencia (biología, química o física) en el último ciclo de la educación media.

Procedimiento

A partir de la identificación del conjunto de habilidades matemáticas y verbales que requieren los estudiantes de primer ingreso del TEC en el

estudio exploratorio-descriptivo realizado en el 2009 se logró rediseñar la tabla de especificaciones de la PAA. Desde ese año se han recolectado evidencias de los ítems, las tareas cognitivas y las categorías de habilidades para lograr una mejor precisión en la medición de los contenidos de acuerdo con el siguiente procedimiento:

1. Determinación de las categorías de razonamiento matemático y razonamiento verbal, con sus correspondientes tareas cognitivas, que son medidas en la prueba.
2. Diseño de la tabla de especificaciones como una matriz con los contenidos en las filas y los niveles de complejidad en las columnas.
3. Asignación de la proporción de ítems por cada habilidad, según el valor predictivo con respecto al criterio evaluado, la pertinencia con el perfil de entrada de los estudiantes de primer ingreso del TEC y su grado de relación con el constructo de la aptitud académica.
4. Una vez elaborada la tabla de especificaciones se procede con la construcción de la prueba.
5. Selección de los ítems por parte del equipo técnico, quien evalúa el grado en que cada ítem representa una categoría de razonamiento y sus tareas cognitivas.

Análisis de datos

En el análisis descriptivo de los datos fue utilizado el programa estadístico IBM SPSS Statistics 19 (IBM Corporation, 2010).

Con el propósito de obtener evidencias del grado de relación entre los indicadores de las habilidades verbales y matemáticas del modelo de medida de la PAA y el rasgo latente de razonamiento general, se realizó un análisis factorial confirmatorio (AFC), conforme al siguiente pro-

cedimiento:

1. Identificación del modelo: La regla de los grados de libertad (gl) es la más frecuente y se calcula con la fórmula $k * (k + 1) / 2$ (Cupani, 2012), donde k son los indicadores. Según el número de indicadores matemáticos (3) y verbales (2) incluidos en la ecuación, se tiene un total de 15 elementos conocidos en la matriz de covarianza y se ha especificado un total de cinco parámetros. Al restar este número, el modelo queda con 10 grados de libertad; por lo tanto, es un modelo sobreidentificado, lo que asegura que el modelo sea tan generalizable como sea posible.

2. Evaluar la identificación del modelo: Una recomendación para tratar de que todos los parámetros estén identificados es utilizar al menos tres indicadores por variable latente e igualar la métrica de cada variable latente con uno de sus indicadores, lo que se consigue fijando arbitrariamente al valor 1 el peso de uno de los indicadores (Cea-D'Ancona, 2002). En el caso de este modelo se incluyeron cinco indicadores para el constructo de *razonamiento general*.

3. Seleccionar las medidas, con base en la recolección, descripción y depuración de datos: las mediciones se basaron en las puntuaciones de los ítems de la PAA del 2019, las cuales se agruparon según la categorización de las habilidades matemáticas y verbales descritas en la tabla de especificaciones.

4. Estimar el modelo con el método de máxima verosimilitud, ya que es la técnica más empleada en estos análisis.

5. Comprobar el ajuste del modelo de acuerdo con los índices de ajuste global, de parsimonia y comparativos.

Una vez concluida la etapa de la estimación del modelo AFC, se interpretan las relaciones de los indicadores con la variable latente medida en

la PAA con la finalidad de comprobar el grado de relación entre ambos factores. Para una descripción más detallada de este procedimiento se recomienda consultar a Escobedo, Hernández, Estebané y Martínez (2016), quienes apuntan que el software EQS (abreviatura de *Equations*), creado por Bentler, ha tenido gran aceptación entre los investigadores que trabajan con modelos de ecuaciones estructurales. En suma, con este procedimiento se pretendió comprobar el siguiente modelo de un factor de razonamiento general, como se muestra en la Figura 1.

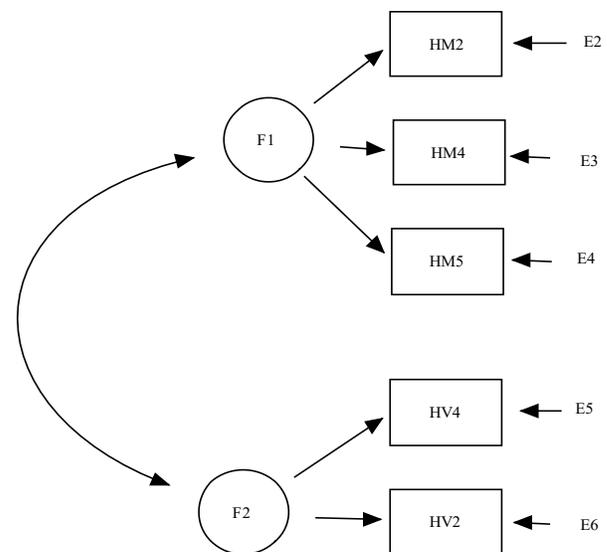


Figura 1

Modelo de medida de la Prueba de Aptitud Académica del 2019.

El análisis predictivo se realizó con base en un modelo de regresión múltiple lineal usando el programa IBM SPSS 19 (IBM Corporation, 2010), donde se incluyeron las variables predictoras de la nota de admisión matemática, verbal y nota de presentación de la educación diversificada, en tanto que el criterio fue promedio del primer semestre del 2020 de los estudiantes de primer ingreso al TEC. El modelo que se busca probar es el siguiente:

$$Y = \beta_0 + \beta_1 X_{\text{matemática}} + \beta_2 X_{\text{verbal}} + \beta_3 X_{\text{ciclo diversificado}} + \epsilon_i$$

Y: es la variable respuesta, el promedio del primer periodo universitario del 2020.

β_0 : es la intersección que determina el valor de Y cuando X es cero.

β : es la pendiente que determina la cantidad en la que cambia Y cuando X se incrementa en una unidad.

X: son las variables explicativas del modelo: notas de matemática, verbal y educación diversificada.

ϵ_i : errores aleatorios.

De acuerdo con Carrasquilla-Batista et al. (2016) este modelo es usado cuando la variable dependiente depende linealmente de cada una de las variables explicativas y un regresor no basta para explicar suficientemente la variabilidad de la variable respuesta.

Resultados

Validez de contenido

Como se indicó en la sección metodológica, el diseño de la tabla de especificaciones de la PAA se ha ajustado año tras año a partir de las evidencias psicométricas de los ítems, las tareas cognitivas y las categorías de habilidades. Como se muestra en la Tabla 1, para el 2019 se seleccionaron tres *habilidades matemáticas*, principalmente por sus aportes predictivos durante los últimos años.

En cuanto al *razonamiento deductivo*, medido tradicionalmente con silogismos, resulta necesario aclarar que antes de 2016 pertenecía al componente verbal de la prueba. No obstante, por su alta correlación con las habilidades matemáticas, se creó una nueva categoría que combinó los silogismos con ítems de *razonamiento inductivo*. A partir de este cambio, entre otros, en el 2019 se midieron solo dos habilidades verbales como se

Tabla 1

Especificaciones de las habilidades matemáticas medidas en la PAA del 2019.

Habilidades matemáticas	Peso	Tareas cognitivas
Razonamiento deductivo e inductivo	32%	A) Realizar inferencias o deducciones a partir de cierta información dada. B) Identificar premisas que permitan obtener una conclusión dada. C) Obtener la forma general de una sucesión o algunos de sus términos.
Resolución de problemas	34%	A) Realizar representaciones gráficas para resolver un problema. B) Interpretar la información dada en un problema para plantear la solución. C) Aplicar reglas básicas de conteo. D) Aplicar definiciones o teoremas nuevos en la resolución de problemas.
Razonamiento con figuras	34%	A) Identificar patrones geométricos. B) Analizar secuencias de figuras. C) Reconocer dos o más elementos que conforman un todo. D) Reconocer un objeto por medio de sus partes y combinar características de varios objetos para realizar representaciones.

resume en la Tabla 2.

Las categorías de *razonamiento verbal* también han mostrado una relación positiva estadísticamente significativa con el desempeño en el primer año universitario, a pesar de ser las más cambiantes en el diseño de la tabla de especificaciones. Por ejemplo, el razonamiento inferencial se medía en dos subcategorías, una con textos cortos y otra con textos largos, pero en 2014 se integró en una sola categoría. También, por unos años se incluyó la categoría de *razonamiento dis-*

Tabla 2

Especificaciones de las habilidades verbales medidas en la PAA del 2019.

Habilidades verbales	Peso	Tareas cognitivas
Razonamiento inferencial	50%	A) Establecer relaciones entre ideas de un texto. B) Sintetizar información de distintos textos. C) Comprender instrucciones. D) Comparar ideas de textos distintos. E) Extraer conclusiones.
Razonamiento semántico	50%	A) Reconocer las relaciones semánticas entre palabras en un texto. B) Identificar el significado de las palabras en contexto.

cursivo, sin embargo, por no mostrar valor predictivo fue excluida a partir de 2018.

Finalmente es necesario destacar que, en el proceso de juzgamiento de los ítems, es crucial el cumplimiento de los estándares de escritura, el criterio de expertos en el análisis de sensibilidad de los ítems para asegurar que no haya contenidos irrelevantes que favorezcan o perjudiquen a determinado grupo de examinados, la aplicación piloto de ítems experimentales y análisis psicométricos como el de funcionamiento diferencial de los ítems, dificultad y discriminación. Estos análisis permiten garantizar los estándares psicométricos de los ítems antes de ser seleccionados para el banco, de manera que no se altera el modelo de medida de la PAA al cumplir con los parámetros establecidos en el diseño de cada año.

Modelo logístico de dos parámetros

Con el propósito de ampliar las evidencias de la calidad psicométrica de los ítems de la PAA,

se estimó el modelo logístico de dos parámetros de la Teoría de Respuesta al Ítem (TRI) para cada uno de los factores de la PAA, con el uso de STATA 14 (StataCorp, 2015). También se utilizó la función de información (FI) para determinar en qué niveles del constructo se proporcionaban mediciones más precisas.

La medida de adecuación muestral KMO para el componente de *matemática* fue de .97 y la prueba de esfericidad de Bartlett arrojó un valor de 117741.9 ($gl = 946$; $p < .001$) y para el componente *verbal* fue de .93 y en la prueba de esfericidad de Bartlett se obtuvo un valor de 36518.5 ($gl = 300$; $p < .001$). Estos valores confirmaron la posibilidad de aplicar un análisis de componentes principales a la matriz de datos y tener una aproximación aceptable al supuesto de unidimensionalidad requerido para los modelos de la TRI, como se muestra en los gráficos de sedimentación de la Figura 2.

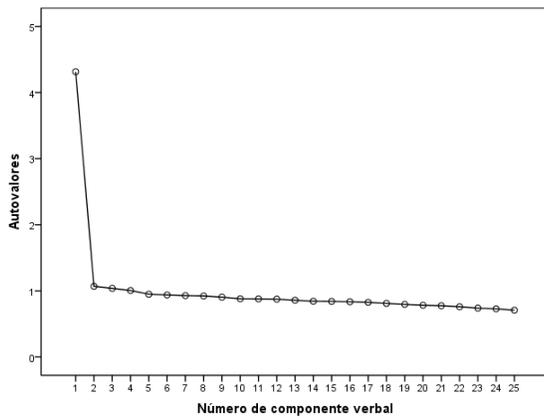
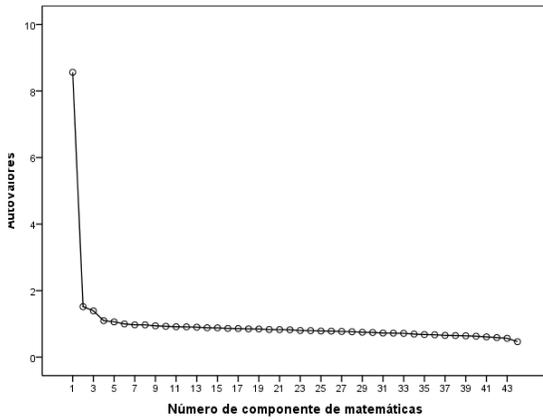
En cuanto a la función de información de cada componente, que se puede observar en la Figura 3, ambos aportan una medición más precisa en los niveles intermedios de θ , lo cual es consistente con el diseño de la PAA. En el caso del componente *verbal*, se presenta un ligero aumento en la precisión en niveles fáciles de habilidad.

Además de la FI relacionada con la confiabilidad de la medida, también se calculó el coeficiente alfa de Cronbach, que resultó en .92. Para completar el análisis de los ítems, se presentan los parámetros de dificultad y discriminación en las Tablas 3 y 4.

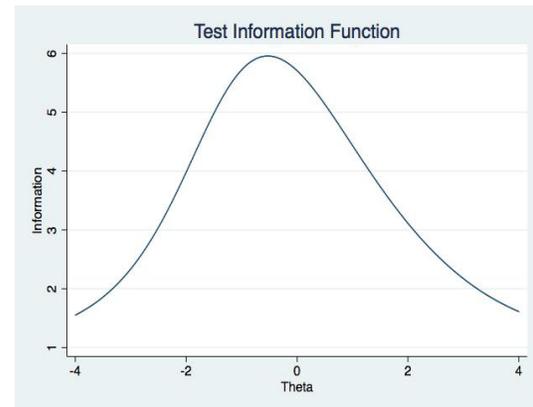
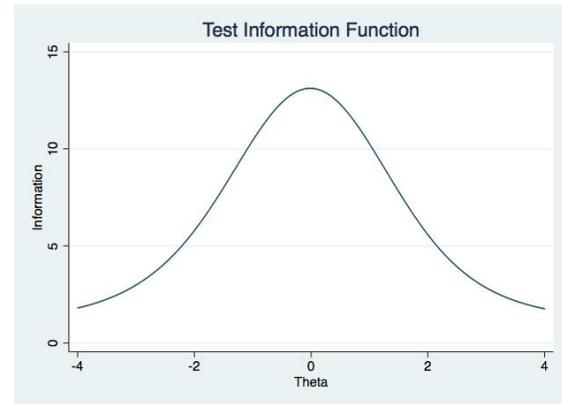
A partir de estos parámetros se puede observar que solo dos ítems de *matemática* (43 y 44) no discriminan entre los examinados de alto y bajo nivel de habilidad; en tanto que el ítem 74 de *razonamiento verbal* resultó con valores extremos en la dificultad y el error estándar. Estos resultados del modelo TRI complementan los análisis realizados con la TCT y demuestran el

Figura 2

Gráficos de sedimentación de los factores de matemática y verbal de la PAA 2019.

**Figura 3**

Función de información del factor de matemática y de razonamiento verbal de la PAA 2019.



cumplimiento de los parámetros de dificultad y discriminación por los ítems de la PAA del 2019, con excepción de los tres mencionados.

Evidencias de validez de constructo

En este análisis multivariado se utilizó el paquete estadístico EQS 6.2 para Windows ([Multivariate Software Inc, 2012](#)) para estimar el modelo SEM con una distribución normal. De acuerdo con [Pardo y Ruiz \(2002\)](#), con estos modelos se busca ajustar las covarianzas entre las variables y minimizar la diferencia entre las covarianzas ob-

servadas de la muestra y las covarianzas estimadas por el modelo estructural.

Los indicadores del factor *matemático* (F1) del modelo de medida de la PAA son los que más explican la varianza en el constructo: 67% razonamiento deductivo e inductivo (HM2), 74% resolución de problemas (HM4) y 66% razonamiento con figuras (HM5). Los indicadores del factor de *razonamiento verbal* (F2) denominados razonamiento inferencial (HV2) y semántico (HV4) también evidencian un buen ajuste con el modelo, con 61% y 62% de varianza explicada. En general, cuanto más alta sea la contribución de los indicadores, más alta es la correlación del

Tabla 3

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente de *matemática* de la PAA 2019.

Parámetros	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rm3 Discrim	1.60	.03	46.93	0.001	1.53 1.67
Diff	-.74	.02	-44.45	0.001	-.77 -.71
rm4 Discrim	1.25	.03	44.07	0.001	1.20 1.31
Diff	-.78	.02	-39.14	0.001	-.82 -.74
rm5 Discrim	.93	.02	42.47	0.001	.89 .98
Diff	.22	.02	10.68	0.001	.18 .26
rm6 Discrim	1.17	.03	44.88	0.001	1.12 1.23
Diff	-.51	.02	-27.83	0.001	-.55 -.48
rm7 Discrim	1.42	.03	47.87	0.001	1.34 1.48
Diff	-.47	.02	-29.38	0.001	-.50 -.44
rm8 Discrim	1.12	.02	44.96	0.001	1.07 1.17
Diff	-.33	.02	-18.50	0.001	-.37 -.30
rm9 Discrim	1.27	.03	46.58	0.001	1.21 1.32
Diff	-.43	.02	-25.41	0.001	-.47 -.40
rm10 Discrim	.67	.02	32.80	0.001	.63 .71
Diff	-.68	.03	-21.87	0.001	-.74 -.62
rm11 Discrim	.83	.02	39.65	0.001	.79 .87
Diff	.29	.02	12.58	0.001	.24 .33
rm12 Discrim	.54	.02	29.28	0.001	.51 .58
Diff	.38	.03	11.30	0.001	.31 .44
rm13 Discrim	.94	.02	42.37	0.001	.89 .98
Diff	.67	.02	27.73	0.001	.62 .72
rm14 Discrim	.82	.02	39.43	0.001	.78 .86
Diff	.37	.02	15.81	0.001	.33 .42
rm15 Discrim	.41	.02	22.87	0.001	.37 .44
Diff	1.26	.07	19.11	0.001	1.14 1.39
rm16 Discrim	1.01	.02	40.29	0.001	.97 1.05
Diff	-.85	.02	-35.12	0.001	-.90 -.80
rm19 Discrim	1.25	.03	46.91	0.001	1.20 1.30
Diff	-.33	.02	-19.81	0.001	-.37 -.30
rm20 Discrim	.95	.02	42.85	0.001	.91 1.00
Diff	.10	.02	5.01	0.001	.06 .14
rm21 Discrim	1.11	.02	46.14	0.001	1.07 1.16
Diff	.14	.02	7.78	0.001	.10 .17
rm22 Discrim	1.03	.02	44.37	0.001	.99 1.08
Diff	-.04	.02	-1.97	0.049	-.07 -.00
rm23 Discrim	.91	.02	40.10	0.001	.86 .95
Diff	-.50	.02	-22.50	0.001	-.54 -.45
rm24 Discrim	1.52	.03	49.32	0.001	1.46 1.58
Diff	-.39	.01	-25.87	0.001	-.42 -.36

Tabla 3 (Cont.)

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente de *matemática* de la PAA 2019.

Parámetros	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rm25 Discrim	1.58	.03	51.27	0.001	1.52 1.64
Diff	-.08	.01	-5.63	0.001	-.11 -.05
rm26 Discrim	1.30	.03	48.71	0.001	1.25 1.35
Diff	-.05	.02	-3.02	0.003	-.08 -.02
rm27 Discrim	.90	.02	39.87	0.001	.86 .95
Diff	-.48	.02	-21.80	0.001	-.53 -.44
rm28 Discrim	.27	.02	16.00	0.001	.24 .31
Diff	.99	.08	11.78	0.001	.83 1.16
rm29 Discrim	1.35	.03	49.84	0.001	1.29 1.40
Diff	.19	.02	11.87	0.001	.16 .22
rm30 Discrim	1.10	.02	46.09	0.001	1.05 1.14
Diff	.31	.02	16.26	0.001	.27 .34
rm31 Discrim	.89	.02	40.62	0.001	.85 .94
Diff	-.16	.02	-7.87	0.001	-.20 -.12
rm32 Discrim	1.66	.03	52.46	0.001	1.60 1.72
Diff	.22	.01	15.28	0.001	.19 .25
rm33 Discrim	1.44	.03	50.78	0.001	1.38 1.49
Diff	.22	.02	13.78	0.001	.19 .25
rm36 Discrim	1.40	.03	40.62	0.001	1.33 1.46
Diff	-1.24	.02	-49.94	0.001	-1.29 -1.20
rm37 Discrim	.76	.02	35.91	0.001	.72 .80
Diff	-.58	.03	-22.01	0.001	-.64 -.53
rm38 Discrim	.83	.02	38.98	0.001	.79 .87
Diff	-.17	.02	-7.66	0.001	-.21 -.12
rm39 Discrim	.89	.02	37.88	0.001	.84 .94
Diff	-.89	.03	-32.56	0.001	-.94 -.83
rm40 Discrim	.65	.02	33.07	0.001	.61 .69
Diff	-.22	.03	-8.02	0.001	-.27 -.16
rm41 Discrim	.98	.02	43.45	0.001	.94 1.03
Diff	.09	.02	4.62	0.001	.05 .13
rm42 Discrim	.70	.02	34.14	0.001	.66 .74
Diff	-.53	.03	-19.09	0.001	-.59 -.48
rm43 Discrim	1.16	.02	46.86	0.001	1.11 1.21
Diff	.01	.02	0.32	0.750	-.03 .04
rm44 Discrim	.95	.02	42.65	0.001	.91 1.00
Diff	.02	.02	0.77	0.441	-.02 .05
rm45 Discrim	.59	.02	30.17	0.001	.55 .62
Diff	1.46	.05	27.66	0.001	1.36 1.56
rm46 Discrim	1.38	.03	50.26	0.001	1.32 1.43
Diff	.39	.02	22.95	0.001	.35 .42

Tabla 3 (Cont.)

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente de *matemática* de la PAA 2019.

Parámetros	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rm47 Discrim	.60	.02	30.22	0.001	.56 .63
Diff	1.64	.06	28.82	0.001	1.52 1.75
rm48 Discrim	1.78	.03	52.37	0.001	1.71 1.84
Diff	.60	.02	37.86	0.001	.57 .63
rm49 Discrim	1.42	.03	47.34	0.001	1.36 1.48
Diff	1.24	.02	51.85	0.001	1.19 1.29
rm50 Discrim	1.65	.03	47.71	0.001	1.58 1.72
Diff	1.26	.02	56.86	0.001	1.21 1.30

indicador con la aptitud académica. Como se observa en la Figura 4, el gráfico de la solución estandarizada muestra el grado de asociación y los errores de medición.

Particularmente, el diagrama facilita la interpretación de los resultados al basarse en unidades estandarizadas que favorecen la comparabilidad entre los coeficientes de correlación. Según el análisis, el indicador matemático con el mayor grado de asociación a la variable latente exógena fue *resolución de problemas* (HM4), fijado con valor 1, mientras que el indicador de *razonamiento semántico* (HV4), mostró la más alta correlación. Este modelo incluye los indicadores observables, los errores que afectan a las mediciones y las relaciones entre el constructo, específicamente, se consideran:

- Variables observadas exógenas (Lambda x): HM2, HM4, HM5, HV2 y HV4.

- Variable latente exógena (ξ): razonamiento general.

- Errores de medición en variables observadas exógenas (δ): De .510 a .621.

- Coeficientes de correlación entre la variable latente con sus respectivos indicadores (λ): Los valores oscilan entre .78 y .86.

Este modelo de ecuaciones estructurales

proporcionó evidencias de la existencia de factores comunes que explicaban la variabilidad de un conjunto de indicadores (HMx y HVx) y la de un factor general de *razonamiento* que explicaba la variabilidad de estos factores comunes, altamente correlacionados (.79) del modelo de medida de la PAA. Finalmente, se comprobó un buen ajuste del modelo de medida de la PAA, tal como se muestra en la Tabla 5, cuyos valores se analizan de acuerdo con los criterios descritos por Ruiz, Pardo y San Martín (2010).

A partir de estos datos, se puede afirmar que existe un ajuste óptimo entre el rasgo latente de *aptitud académica* y los indicadores, al obtenerse un valor aceptable en el índice RMSEA (inferior a .08), y también valores aceptables de ajuste global del modelo propuesto (GFI) de ajuste incremental (NFI) y de ajuste de parsimonia (AGFI). De acuerdo con Cupani (2012), los valores superiores a .95 sugieren un ajuste óptimo y, en lo que respecta al chi cuadrado, explica que debe ser no significativo para indicar un buen ajuste; sin embargo, al ser un estadístico muy sensible al tamaño muestral debe interpretarse con precaución. Por ello, usualmente se interpreta que la razón del χ^2 sobre los grados de libertad con valores inferiores a 2 indica un buen ajuste. Por otra parte, también debe resaltarse que todos los indicadores reflectivos de los componentes *matemático* y *verbal*

Tabla 4

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente *verbal* de la PAA 2019.

Verbal	Coficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rv53 Discrim	1.34	.03	39.16	0.001	1.27 1.41
Diff	-1.33	.03	-48.34	0.001	-1.38 -1.27
rv54 Discrim	1.09	.03	41.00	0.001	1.03 1.14
Diff	-.65	.02	-30.42	0.001	-.69 -.61
rv55 Discrim	1.08	.03	42.07	0.001	1.03 1.13
Diff	-.09	.02	-4.88	0.001	-.12 -.05
rv56 Discrim	.82	.02	34.66	0.001	.78 .87
Diff	-1.09	.03	-32.42	0.001	-1.16 -1.03
rv57 Discrim	.91	.02	38.28	0.001	.86 .96
Diff	-.48	.02	-21.41	0.001	-.53 -.44
rv58 Discrim	.61	.02	29.81	0.001	.57 .65
Diff	-.38	.03	-12.67	0.001	-.44 -.32
rv59 Discrim	.75	.02	34.45	0.001	.71 .79
Diff	-.34	.02	-13.58	0.001	-.39 -.29
rv60 Discrim	1.21	.03	43.82	0.001	1.16 1.27
Diff	-.15	.02	-9.03	0.001	-.19 -.12
rv61 Discrim	.94	.02	39.02	0.001	.89 .99
Diff	.43	.02	19.87	0.001	.39 .47
rv62 Discrim	.81	.02	35.07	0.001	.76 .85
Diff	.83	.03	27.94	0.001	.77 .88
rv63 Discrim	.85	.03	33.53	0.001	.80 .90
Diff	1.50	.04	35.65	0.001	1.42 1.58
rv64 Discrim	.82	.02	32.97	0.001	.77 .87
Diff	1.48	.04	34.52	0.001	1.40 1.57
rv65 Discrim	1.58	.04	40.76	0.001	1.50 1.65
Diff	-1.18	.02	-52.24	0.001	-1.22 -1.13
rv69 Discrim	1.20	.03	41.29	0.001	1.14 1.26
Diff	-.89	.02	-39.41	0.001	-.93 -.84
rv70 Discrim	1.30	.03	43.82	0.001	1.25 1.36
Diff	-.56	.02	-30.92	0.001	-.59 -.52
rv71 Discrim	.58	.02	28.57	0.001	.54 .62
Diff	-.45	.03	-13.83	0.001	-.51 -.39
rv72 Discrim	.86	.02	37.23	0.001	.81 .90
Diff	-.43	.02	-18.81	0.001	-.48 -.39
rv73 Discrim	.67	.02	31.76	0.001	.62 .71
Diff	-.40	.03	-14.35	0.001	-.46 -.35
rv74 Discrim	.04	.02	2.00	0.046	.00 .08
Diff	24.07	12.05	2.00	0.046	.46 47.69
rv75 Discrim	.84	.02	37.00	0.001	.79 .88
Diff	-.11	.02	-4.85	0.001	-.15 -.06

Tabla 4 (Cont.)

Estimación de los niveles de discriminación y dificultad del modelo de dos parámetros del componente *verbal* de la PAA 2019.

Verbal	Coefficiente	Error estándar	z	P > z	[Intervalo de confianza 95%]
rv76Discrim	1.23	.03	43.91	0.001	1.17 1.28
Diff	-.22	.02	-12.88	0.001	-.25 -.18
rv77Discrim	.96	.02	38.47	0.001	.91 1.01
Diff	.79	.03	31.37	0.001	.74 .84
rv78Discrim	.76	.02	32.62	0.001	.71 .80
Diff	1.26	.04	31.51	0.001	1.18 1.34
rv79Discrim	.57	.03	22.07	0.001	.52 .62
Diff	3.06	.13	23.77	0.001	2.81 3.31
rv80Discrim	.90	.02	37.94	0.001	.86 .95
Diff	.60	.02	25.17	0.001	.56 .65

resultaron coeficientes significativos; lo cual resulta muy positivo para efectos del modelamiento de la PAA y la confirmación de la hipótesis de que existe una relación entre el modelo de medida y el hipotetizado.

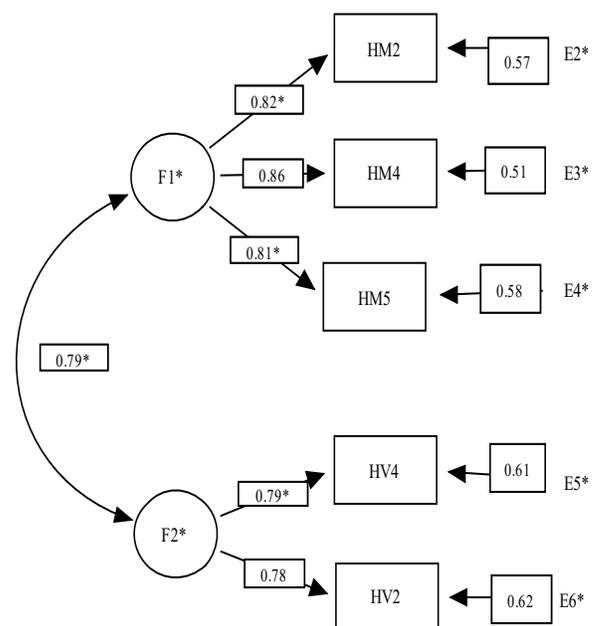
Evidencias de validez predictiva

El modelo de regresión lineal múltiple se ejecutó con el software SPSS Statistics versión 19 (IBM Corporation, 2010), con los examinados que realizaron la PAA en el 2019 y se matricularon en el primer semestre del curso lectivo del 2020. Específicamente, en el modelo clásico se incluyeron 1769 estudiantes de primer ingreso, cuyos resultados se presentan en la Tabla 6.

Con respecto a los supuestos del modelo de regresión lineal múltiple, no se encontraron correlaciones altas entre estos predictores, cuyos coeficientes Pearson oscilaron entre .231 y .303, por lo tanto, estas estimaciones no están afectadas por la colinealidad. El supuesto de normalidad no se cumplió, según la prueba de Kolmogorov-Smirnov; como tampoco el supuesto de homocedasticidad, comprobado con el estadístico Breusch-Pa-

gan; por tal razón, se estimó un segundo modelo de regresión robusta con R Core Team (2018), como se muestra en la Tabla 7.

Estas son las evidencias predictivas más recientes del modelo de medida de la prueba de admisión del TEC. Para el análisis correspondien-

**Figura 4**

EQS 6 modelo PAA 2019_Chi Sq.= 236.00 $p < 0.001$ CFI = 0.99 RMSEA = 0.06.

Tabla 5
Estadísticos de bondad de ajuste del modelo SEM de la PAA 2019.

Estadístico	Abreviatura	Criterio	Valores del modelo
Chi-cuadrado	χ^2	Significación > .05	.000
Razón chi cuadrado	χ^2/df		2.370
Bentler-Bonett Normed Fit Index	NFI	≥ .95	.994
Bentler-Bonett Non-Normed Fit Index	NNFI	≥ .95	.986
McDonald's fit index	MFI	≥ .90	.993
Bollen's fit index	IFI	≥ .90	.994
Índice de bondad de ajuste comparativo	CFI	≥ .95	.994
Joreskog-Sorbom's GFI	GFI	≥ .95	.994
Joreskog-Sorbom's AGFI	AGFI	≥ .95	.978
Raíz del residuo cuadrático promedio	RMR	Próximo a cero	.136
Raíz del residuo cuadrático promedio de aproximación	RMSEA	< .08	.060
90% Intervalo de confianza de RMSEA			(.053; .066)

te de cada uno de los predictores, se registraron los resultados de los coeficientes de regresión no estandarizados y el nivel de significancia ($\alpha = .05$). Estos coeficientes indican el cambio que corresponde a la nota del promedio simple por cada unidad de cambio en la puntuación de cada componente, todas medidas en una escala 0-100. De manera que la nota del ciclo de educación diversificada (EDciclo100) es la que muestra una mayor contribución en ambos modelos. Esto significa que el incremento en el promedio simple de .58 puntos (vg. modelo robusto) se produce independientemente de la puntuación de los otros dos componentes *matemático* y *verbal*.

A partir de los coeficientes estandarizados se puede comparar el poder explicativo de cada predictor en la ecuación. Al ser directamente comparables, se observa que en ambos modelos la nota de matemática y educación diversificada resulta con el más alto coeficiente; en tanto que el razonamiento verbal no resulta tan relevante; pero sí ha mostrado una relación estadísticamente significativa con el promedio simple en el primer año desde el 2016.

La proporción de la variabilidad de los promedios simples explicada por los tres predictores, esto es, el coeficiente de determinación, fue de .095; en tanto que en años anteriores para el

Tabla 6
Modelo de regresión clásico de los componentes de la nota de admisión con el promedio simple del primer semestre del 2020.

Modelo	Coeficientes no estandarizados		Coeficientes Beta	t	Sig.
	B	Error típ.			
(Constante)	-13.656	7.743		-1.764	.078
EDciclo100	.749	.080	.211	9.325	.000
Mate100	.360	.038	.228	9.526	.000
Verbal100	-.079	.037	-.051	-2.119	.034

Nota. Modelo 1 promedio simple del primer semestre del 2020.

Tabla 7

Modelo de regresión robusta de los componentes de la nota de admisión con el promedio simple primer semestre de 2020.

Predictores	Coefficientes	Error estándar	Beta	Error estándar	Estadístico	<i>p</i>
(Intercepto)	11.22	5.55	.22	.02	13.85	.043
EDciclo100	0.58	0.06	.17	.02	9.92	< .001
Mate100	0.24	0.03	.15	.02	9.12	< .001
Verbal100	-0.02	0.02	-.02	.02	-0.98	.327

Nota. Promedio simple del primer semestre de 2020.

mismo periodo fue de .174 y .124 con la PAA aplicada en el 2018 y 2017 respectivamente. Esta disminución en la varianza explicada, como en el valor predictivo del componente *verbal*, se presume que se relaciona con el cambio radical de la modalidad presencial hacia la virtualidad de todos los cursos universitarios durante el primer periodo del 2020, a causa de los efectos de la pandemia en el país. Por otra parte, tales coeficientes podrían mejorar si se incluyen otras variables independientes, tales como tipo de financiamiento del colegio, sexo y provincia de procedencia entre otros. Sin embargo, para los propósitos de esta investigación lo que interesa es encontrar evidencias del grado de validez predictiva de estos tres componentes de la nota de admisión con respecto al rendimiento de los estudiantes en su primer año de ingreso.

Discusión

La validación debe ser un proceso permanente: cuantas más evidencias de validez se obtengan de una prueba, más sólido es el diseño y más razonables los usos y las interpretaciones de sus resultados, especialmente por sus altas consecuencias sociales, como es el caso de la PAA para el TEC. Como lo resaltan [Muñiz y Fonseca-Pedrero \(2019\)](#), siempre habrá que aportar evidencias empíricas de la fiabilidad y la validez, para

garantizar que los instrumentos de medida evalúan de forma objetiva y rigurosa.

Por ello, el primer paso para validar el modelo de medida es analizar las evidencias de validez de contenido a partir de una herramienta básica: la tabla de especificaciones. Esta permite a los constructores de ítems tener una guía básica sobre el tipo de tareas que deben medir los ítems; así como garantizar la distribución estándar del nivel de dificultad en la prueba: 25% fáciles, 50% intermedias y 25% difíciles. En este sentido, [Sireci y Faulkner-Bond \(2014\)](#) consideran que las evidencias de validez de contenido son críticas para hacer inferencias acerca de los examinados con respecto al dominio de interés. Además, en esta etapa es muy importante el criterio de los expertos para analizar el grado de representatividad de los ítems de cada indicador o categoría de razonamiento y sus tareas cognitivas.

Una vez que se obtienen evidencias de contenido, como la proporcionada por la tabla de especificaciones, es fundamental comprobar el grado de asociación entre las variables observadas exógenas (x) de los indicadores matemáticos y verbales por una parte, y el constructo de aptitud académica, la variable latente exógena (ξ) por otra parte, para realizar los ajustes necesarios. Precisamente, cada año el análisis de ese conjunto de indicadores ha permitido tomar decisiones para mejorar la precisión de la medida, tales como la eliminación de algunos indicadores (analogías,

antónimos, análisis discursivo) y la integración de algunas habilidades, como el razonamiento inferencial con textos cortos y largos; o bien, realizar ajustes en las tareas cognitivas como en los silogismos, el conteo o la resolución de problemas con operaciones aritméticas, entre otros.

De esta manera, es muy importante apuntar que se estima un modelo SEM entre otros posibles para medir el razonamiento general. La evidencia obtenida de este modelo cumple satisfactoriamente con los distintos criterios de ajuste y significación estadística, por lo que permite afirmar que existe una relación significativa entre los indicadores matemáticos y verbales de la prueba y el rasgo latente de razonamiento general. Evidentemente, este es un rasgo complejo que, al ser medido por cinco indicadores, resulta un posible modelo entre otros. Al respecto, [Donoso-Díaz y Hawes-Barríos \(2001\)](#) señalan que el reducido ámbito de las dimensiones de inteligencia de las pruebas que consideran solo la aptitud verbal y matemática; si bien es relevante, es insuficiente para predecir acertadamente el éxito de un estudiante ante un proceso de formación, que en la actualidad demanda otras dimensiones de inteligencia (espacial, kinésica, interpersonal, intrapersonal, etc).

Finalmente, resulta de suma importancia la obtención de evidencias del valor predictivo de los componentes de la nota de admisión, puesto que el propósito de la PAA es seleccionar a los candidatos con mayor probabilidad de éxito en el TEC. De manera que las evidencias de asociación de las notas de matemática, verbal y de la educación diversificada con el rendimiento de los estudiantes en el primer año universitario permite afirmar que sí aportan evidencias predictivas. Como destacan [Guglietta y Delgado-Álvarez \(2010\)](#), las pruebas estandarizadas de habilidad verbal y numérica se comportan como predictores válidos del rendimiento. Por supuesto que

este modelo resulta limitado por considerar únicamente los predictores de la nota de admisión. Al respecto, existen múltiples estudios de factores asociados al rendimiento en el primer año, como lo señala [García-Domínguez \(2016\)](#), por lo que resultaría interesante incorporar otros datos de naturaleza personal y socioeconómica como variables que pudieran incidir en el rendimiento de los estudiantes.

Una limitación de estos análisis predictivos es la variabilidad de criterios para calificar los cursos universitarios y de las asignaturas de la educación diversificada. Como señalan [Contreras et al. \(2009\)](#), estos criterios reflejan distintas políticas de asignación de notas, niveles de exigencia y, eventualmente, ‘inflación’ de notas. No obstante, la inclusión de la nota de educación diversificada en el modelo de regresión se debe a que resulta un buen predictor del rendimiento de los estudiantes durante su primer año en el TEC.

Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Autor.
- Backhoff, E., Tirado, F., & Larrazolo, N. (2001). Ponderación diferencial de reactivos para mejorar la validez de una prueba de ingreso a la universidad. *Revista Electrónica de Investigación Educativa*, 3(1), 1-10. Recuperado de <https://redie.uabc.mx/redie>
- Bleyaert, B. (2010). ACT and college success. *Education Partnerships*. ERIC N° ED537914. Recuperado de <https://eric.ed.gov/?id=ED537914>
- Bringas-Benavides, M. R., & Pérez-Mejía, J. (2014). El examen de ingreso al nivel superior. ¿Admisión o de-

- cepción? *Ra Ximhai*, 10(5), 103-114. Recuperado de <https://www.redalyc.org/pdf/461/46132134007.pdf>
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde-Cerdas, J., & Guerrero-Barrantes, M. (2016). Regresión lineal simple y múltiple: Aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Tecnología en Marcha*, 29(8), 33-45. doi: 10.18845/tm.v29i8.2983
- Cea-D'Ancona, M. A. (2002). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid, España: Síntesis.
- Cliff, A., & Montero, E. (2010). El balance entre excelencia y equidad en pruebas de admisión: Contribuciones de experiencias en Sudáfrica y Costa Rica. *Revista Iberoamericana de Evaluación Educativa*, 3(2), 7-28. Recuperado de <https://revistas.uam.es/index.php/rie>
- College Board. (2014). *Guía de estudio para presentar la Prueba de Aptitud Académica*. Recuperado de <https://latam.collegeboard.org>
- Cupani, M. (2012). Análisis de ecuaciones estructurales: Conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis*, 2(1), 186-199. Recuperado de <https://revistas.unc.edu.ar/index.php/tesis>
- Contreras, D., Bravo, D., & Sanhueza, C. (2001). PAA, ¿una prueba de inteligencia? *Revista Perspectivas*, 4(2), 233-247. Recuperado de <http://www.dii.uchile.cl/~revista>
- Contreras, D., Gallegos, S., & Meneses, F. (2009). Determinantes de desempeño universitario: ¿Importa la habilidad relativa? *Calidad en la Educación*, 30, 18-48. doi: 10.31619/caledu.n30.172
- Donoso-Díaz, S., & Hawes-Barrios, G. (2001). El constructo de inteligencia del sistema de selección de alumnos a las universidades chilenas. Discusión de sus fundamentos y lineamientos para el cambio. *Revista Cubana de Educación Superior*, 21(3), 57-80.
- Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27-36. Recuperado de <http://www.humanas.unal.edu.co/psicometria/revista/volumenes/volumen-6>
- Escobedo, M. T., Hernández, J. A., Estebané, V., & Martínez, G. (2016). Modelos de ecuaciones estructurales: Características, fases, construcción, aplicación y resultados. *Ciencia & Trabajo*, 18(55), 16-22. Recuperado de <https://scielo.conicyt.cl/pdf/cyt/v18n55/art04.pdf>
- García-Domínguez, L. A. (2016). Pruebas de selección como predictores del rendimiento académico de estudiantes de Medicina. *Investigación en Educación Médica*, 5(18), 88-92. Recuperado de <http://riem.facmed.unam.mx>
- Gierl, M. J., Tan, X., & Wang, C. (2005). *Identifying content and cognitive dimensions on the SAT*. College Board Research (Reporte N° 2005-11). Recuperado de <https://files.eric.ed.gov>
- Guglietta, L., & Delgado-Álvarez, C. (2010). Validez de constructo de un modelo de admisión a postgrado. Un análisis de ruta. *Revista Galego-Portuguesa de Psicología e Educación*, 18(1), 227-237. Recuperado de <https://ruc.udc.es/dspace/handle/2183/6562>
- IBM Corporation. (2010). IBM SPSS Statistics for Windows (19.0). [Software de cómputo]. Armonk, NY: Autor.
- Instituto Nacional de Evaluación Educativa. (2019). *Pisa 2018. Informe español*. Madrid, España: Ministerio de Educación y Formación Profesional. Recuperado de <https://www.educacionyfp.gob.es/inee/portada.html>
- Iriarte-Díaz-Granados, F., Espeleta-Maya, A., Zapata-Zapata, E., Cortina-Peñaranda, L., Zambrano-Ojeda, E., & Fernández-Candama, F. (2010). El razonamiento

- lógico en estudiantes universitarios. *Zona Próxima*, 12. Recuperado de <http://rescientificas.uninorte.edu.co/index.php/zona>
- Juárez-Hernández, L. G., & Tobón, S. (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Revista Espacios*, 39(53), 23. Recuperado de <http://www.revistaespacios.com>
- Kobrin, J. L., & Kimmel, E. W. (2006). Test development and technical information on the writing section of the SAT reasoning test. *Research Notes*, 25. Recuperado de <https://files.eric.ed.gov>
- Lassiter, D., & Goodman, N. (2015). How many kinds of reasoning? Inference, probability and natural language semantics. *Cognition*, 136, 123-134. doi: 10.1016/j.cognition.2014.10.016
- Lawrence, I. M., Rigol, G. W., Van Essen, T., & Jackson, C. A. (2003). *A historical perspective on the content of the SAT*. College Board (Report N° 2003-3). doi: 10.1002/j.2333-8504.2003.tb01902.x
- López, I., Echazarreta, C., Pech, S., & Gómez, B. (2010). Selección y permanencia en la educación superior: El caso de la Universidad Autónoma de Yucatán. *Revista Iberoamericana de Evaluación Educativa*, 3(2), 90-102. Recuperado de http://rinace.net/rie/rie_home.html
- Manzi, J., Bosch, A., Bravo, D., del Pino, G., Donoso, G., & Pizarro, R. (2010). Validez diferencial y sesgo en la predictividad de las pruebas de admisión a las universidades chilenas (PSU). *Revista Iberoamericana de Evaluación Educativa*, 3(2), 29-48. Recuperado de http://rinace.net/rie/rie_home.html
- Marino, T., & Rodríguez, M. (2009). Un estudio exploratorio sobre heurísticas en estudiantes de un curso de matemática de nivel pre-universitario. *Paradigma*, 30(2), 159-178. Recuperado de <http://revistaparadigma.online/ojs/index.php/paradigma>
- Martínez-Arias, M. R. (2005). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid, España: Síntesis.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8. doi: 10.1111/j.1745-3992.1995.tb00881.x
- MIDEPLAN. (2017). *Costa Rica Índice de Desarrollo Social (IDS)*. Recuperado de <https://www.mideplan.go.cr>
- Milewski, G. B., Johnsen, D., Glazer, N., & Kubota, M. (2005). *A survey to evaluate the alignment of the New SAT®. Writing and critical reading sections to curricula and instructional practices* (Research report N° 2005-1). College Board. doi: 10.1002/j.2333-8504.2005.tb01984.x
- Moreira-Mora, T. E. (2016). Eficacia de la prueba de aptitud académica para pronosticar el rendimiento académico: Un reto de las universidades públicas. *VIII Congreso del Consejo Superior Universitario Centroamericano (CSUCA)*, Ciudad de Panamá. Recuperado de <https://repositoriotec.tec.ac.cr>
- Multivariate Software Inc. (2012). EQS for Windows (6.2). [Software de cómputo]. Recuperado de <http://www.mvsoft.com/eqsdownload.htm>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. Recuperado de <http://www.psicothema.com>
- Pardo, A., & Ruiz, M. A. (2002). *SPSS II. Guía para el análisis de datos*. Madrid, España: McGraw-Hill.
- R Core Team. (2018). R: A language and environment for statistical computing. [Software de cómputo]. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <http://www.R-project.org>
- Rojas-Torres, L., & Ordóñez-Gutiérrez, G. (2019). Proceso de construcción de pruebas educativas: El caso de la

Prueba de Habilidades Cuantitativas. *Revista Evaluar*, 19(2). doi: [10.35670/1667-4545.v19.n2.25080](https://doi.org/10.35670/1667-4545.v19.n2.25080)

Ruiz, M. A., Pardo, A., & San Martín, R. (2010). Modelos de ecuaciones estructurales. *Papeles del Psicólogo*, 31(1), 34-45. Recuperado de <http://www.papeles-delpsicologo.es>

StataCorp. (2015). Stata Statistical Software: Release 14. [Software de cómputo]. College Station, TX: Autor.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100-107. Recuperado de <http://www.psicothema.com>

Vergara-Díaz, G., & Peredo-López, H. (2017). Relación del desempeño académico de estudiantes de primer año de universidad en Chile y los instrumentos de selección para su ingreso. *Revista Educación*, 41(2), 95-104. doi: [10.15517/revedu.v41i2.21514](https://doi.org/10.15517/revedu.v41i2.21514)

Young, J. W. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis*. College Board Research (Reporte N° 2001-6). Recuperado de <https://files.eric.ed.gov>