






Prueba de tamizaje de razonamiento cuantitativo en estudiantes universitarios de una universidad pública colombiana

Felipe Ramírez Cortázar – Universidad de Cundinamarca
 Maricel Barrera Patiño – Universidad de Cundinamarca
 Rafael Leonardo Cortes Lugo – Universidad de Cundinamarca
 Adrián David Galindo Ubaque – Universidad de Cundinamarca
 Harvey Iovani Hernández Yomayusa – Universidad de Cundinamarca

 0000-0001-6650-5579
 0009-0002-4120-503X
 0000-0001-5938-9215
 0000-0002-1125-9400
 0000-0001-7849-4133

Recepción: 18.11.2024 | Aceptado: 29.11.2024

Correspondencia a través de **ORCID**: Felipe Ramírez Cortázar

 **0000-0001-6650-5579**

Citar: Ramírez Cortázar, F, Barrera Patiño, M, Cortes Lugo, RL, Galindo Ubaque, AD, & Hernández Yomayusa, HI (2024). Prueba de tamizaje de razonamiento cuantitativo en estudiantes universitarios de una universidad pública colombiana. *REIDOCREA*, 13(44), 651-660.

Financiación: V convocatoria interna de financiación de proyectos de investigación de la Universidad de Cundinamarca

Estudio de investigación de psicología educativa
 Área o categoría del conocimiento: Psicología

Resumen: Este estudio desarrolló y validó una prueba de tamizaje en razonamiento cuantitativo destinada a identificar estudiantes con dificultades en esta competencia en la Universidad de Cundinamarca. La prueba se basa en las competencias de interpretación y representación, formulación y ejecución, y argumentación del examen Saber Pro, que es obligatorio para estudiantes universitarios en Colombia. Tras la construcción inicial de 35 ítems, se llevó a cabo una evaluación por jueces expertos, análisis de consistencia interna mediante el coeficiente Alfa de Cronbach y validación psicométrica usando teoría clásica de tests (TCT) y el modelo de Rasch. Los resultados indicaron que la versión final de la prueba, con 20 ítems, presenta propiedades adecuadas de dificultad y discriminación, y una fiabilidad aceptable. Además, el análisis de funcionamiento diferencial por género no mostró sesgos, lo que garantiza la equidad en la medición. Este instrumento permite una detección temprana de estudiantes en riesgo y aporta una herramienta de mejora pedagógica para instituciones de educación superior en Colombia.

Palabra clave: Razonamiento cuantitativo

Quantitative reasoning screening test in university students from a Colombian public university

Abstract: This study developed and validated a quantitative reasoning screening test aimed at identifying students with difficulties in this skill at the University of Cundinamarca. The test is based on the competencies of interpretation and representation, formulation and execution, and argumentation in the Saber Pro exam, which is mandatory for university students in Colombia. After the initial construction of 35 items, the test underwent expert judge evaluation, internal consistency analysis using Cronbach's alpha coefficient, and psychometric validation through classical test theory (CTT) and the Rasch model. Results indicated that the final 20-item version of the test has suitable properties of difficulty and discrimination, with acceptable reliability. Additionally, differential item functioning analysis by gender showed no bias, ensuring equity in measurement. This instrument allows for early detection of students at risk and provides a pedagogical improvement tool for higher education institutions in Colombia.

Keyword: Quantitative reasoning

Introducción

Los exámenes estandarizados *de salida* son elementos fundamentales en la educación superior para entender la preparación de los estudiantes y cómo se han desarrollado las habilidades básicas y específicas necesarias para su competencia profesional (El-Hassan et al., 2024). En Colombia, desde el año 2009, los estudiantes de pregrado deben realizar un examen elaborado por el Instituto Colombiano para la Evaluación de la Educación (ICFES) como requisito obligatorio para obtener su título profesional (Congreso de Colombia, 2009; Presidencia de la República de Colombia, 2009). Este examen, denominado Saber PRO, desde el 2014, evalúa cinco módulos genéricos para

todos los programas académicos en el país: lectura crítica, razonamiento cuantitativo, competencias ciudadanas, inglés y comunicación escrita (Instituto Colombiano para la Evaluación de la Educación, 2023a).

Este examen también incluye hasta tres módulos específicos, los cual varían según los programas académicos. Por ejemplo, los programas de lenguas modernas, literatura, lingüística y afines solo presentan los cinco módulos genéricos, mientras programas como ingeniería de sistemas, telemática y afines presentan los módulos genéricos más tres específicos: Formulación de Proyectos de Ingeniería, Diseño de Software y Pensamiento Científico Matemáticas y Estadística. Esta distinción se realiza para evaluar competencias que aplican a las características particulares de cada profesión en Colombia (Instituto Colombiano para la Evaluación de la Educación, 2024). En este contexto, los resultados del examen Saber PRO no solo son un indicador del desempeño individual de los estudiantes, sino que, también, permiten evaluar la calidad de los programas académicos y de las instituciones de educación superior en general.

El análisis detallado de los resultados del examen Saber PRO ha revelado que el promedio global del examen tuvo una disminución en los últimos ocho años y que el promedio en competencias generales disminuyó entre 2022 y 2023, excepto en la competencia de comunicación escrita (Instituto Colombiano para la Evaluación de la Educación, 2023b). Además, se observan efectos significativos asociados a características de las instituciones de educación superior (IES) que diferencian los resultados obtenidos en las competencias genéricas y específicas. Entre las variables que se han encontrado que influye es si la universidad es pública o privada, la ubicación o la región de la sede de la institución y que cuente con la acreditación de alta calidad del programa de formación lo que permite que tenga puntaje más altos en los resultados de las prueba (Mendoza-Mendoza et al., 2024; J. L. E. Pérez et al., 2023; L. A. Pérez et al., 2023).

Otras variables que se destacan y pueden incidir en los resultados de la prueba, tanto en competencias específicas como generales, son, por ejemplo, el género donde las mujeres tienden a tener puntajes menores en competencias como razonamiento cuantitativo e inglés, pero tienen mejores puntajes en comunicación escrita (Angulo et al., 2023; Gallego & Casadiego, 2023). Esta distinción es importante dado que, la competencia de razonamiento cuantitativo se refiere a un conjunto de habilidades que usan las personas para comprender, argumentar, analizar y en general solucionar problemas que contengan información matemática, estadística y en general numérica, en los diversos ámbitos de la vida personal (Martínez Solano et al., 2022). El examen, el razonamiento cuantitativo se divide en tres competencias puntuales: i) interpretación y representación, ii) formulación y ejecución y iii) argumentación (Instituto Colombiano para la Evaluación de la Educación, 2018). La primera se refiere a la capacidad de comprender y analizar información proveniente de distintos formatos (e.j. gráficos o tablas); la segunda es la habilidad de diseñar planes de acciones y ejecutar soluciones teniendo en cuenta información numérica. La tercera, es la evaluación y sustento de acciones tomadas en respuesta a un problema.

Aunque lo anterior hace un énfasis en las competencias matemáticas, el razonamiento cuantitativo también es necesario para generar herramientas para la participación activa en contextos sociales, culturales, políticos, administrativos, económicos, educativos y laborales (Mendoza et al., 2021). Además, de lo anterior Mendoza-Mendoza et al., (2024) señalaron que esta competencia es una de la que permite distinguir entre programas académicos de alta calidad. Junto a esto, el estudio de Lombana & Cabeza (2024) analizó diferentes tipos de factores que influyen en el desempeño del razonamiento cuantitativo en la prueba, donde destacaron variables personales,

familiares, instituciones y de vivienda. Entre las que resaltan, una vez más, el género de la persona, el carácter de la universidad, el valor pagado por la matrícula, educación de los padres y acceso a internet, entre otros.

Todo lo anterior pone de manifiesto, en el marco de entender la evaluación educativa, que se debe tener un abordaje integral y que las IES reflexionen sobre sus prácticas pedagógicas, tomen estrategias para mejorar los procesos de enseñanza-aprendizaje y mejorar el valor agregado, que se refiere al aporte pedagógico que se debe, netamente, a la influencia de la enseñanza universitaria, excluyen la influencia de variables externas (Rodríguez-Revilla & Vallejo-Molina, 2022); estos mismos autores encontraron que, aunque las instituciones públicas tienen mejores puntajes que las instituciones privadas en Colombia, en términos de evaluación de valor agregado de las competencias genéricas la tendencia puede ser contraria debido a que, las instituciones públicas reciben, en su mayor parte, estudiantes con mayor desarrollo de esta competencia.

En este sentido, teniendo en cuenta la necesidad de apoyar el valor agregado de la competencia de razonamiento cuantitativo, la presente investigación tiene como objetivo desarrollar una serie de pruebas de tamizaje como una herramienta complementaria con el fin de identificar, tempranamente, a los estudiantes con dificultades en razonamiento cuantitativo. De este modo, se pueden implementar intervenciones pedagógicas dirigidas a mejorar sus competencias antes de presentar el examen Saber PRO, lo que permitiría contribuir a la mejora general de la calidad educativa y del valor agregado de la institución, sobre todo, la implementación de esta herramienta dentro de la Universidad de Cundinamarca, que es una IES pública colombiana donde se realizó este proyecto. De acuerdo con lo anterior, esta estrategia permitiría identificar estudiantes en riesgo de bajo desempeño en el Saber PRO, lo que permitiría, a la institución, adoptar medidas y acciones de mejora oportunas e impulsar el desarrollo de la competencia. Además, se espera que esta serie de pruebas aporten al debate sobre la evaluación de la calidad de la educación superior y a la mejora continua de los procesos académicos en Colombia.

Objetivo General

Desarrollar una prueba de tamizaje para identificar, tempranamente, a los estudiantes con dificultades en razonamiento cuantitativo en la Universidad de Cundinamarca, que cuente con indicadores psicométricos de calidad que sustenten su uso en el marco de la evaluación del valor agregado.

Objetivos Específicos

1. Desarrollar ítems según la competencia de razonamiento cuantitativo de Saber Pro que compongan una prueba de tamizaje específica para la evaluación temprana en estudiantes de pregrado.
2. Evaluar la consistencia interna de la prueba de tamizaje y sus propiedades psicométricas evaluando la dificultad y discriminación de los ítems.
3. Analizar las diferencias según variables sociodemográficas encontradas por la prueba de tamizaje de razonamiento cuantitativo.

Método

El proceso de desarrollo de la prueba de tamizaje inició con una revisión de las tres competencias de razonamiento cuantitativo evaluadas en el Saber PRO que fueron antes señaladas: i) interpretación y representación, ii) formulación y ejecución, y iii) argumentación. Posteriormente, se procedió a la elaboración de un conjunto inicial de

treinta y cinco ítems, todos ellos con cuatro opciones de respuesta con una única respuesta correcta. Este conjunto se sometió a una evaluación de dos jueces expertos, en medición educativa y en psicología educativa, quienes revisaron el problema presentado, las opciones de respuesta, la competencia a la que pertenecía el ítem y se realizó una evaluación del ítem en términos de claridad, coherencia y relevancia (Escobar-Pérez & Cuervo-Martínez, 2008). Para el cálculo del acuerdo Inter jueces, en primera instancia, se analizaron los estadísticos descriptivos de cada una de las puntuaciones asignadas por ítem, estas iban de 1 a 4 donde 4 implicaban una mayor claridad, coherencia o relevancia. Posteriormente, se calculó el índice V de Aiken, que permite evaluar el nivel del concordancia de los jueces expertos (Ventura-León, 2019). Los ítems que no alcanzan el umbral de aceptabilidad fueron modificados o eliminados, lo que aseguró que solo aquellos con alta calidad fueran incluidos en la versión final de la prueba.

Una vez completado el proceso de evaluación por jueces, se llevó a cabo una prueba piloto con una muestra de estudiantes del programa de psicología de la Universidad de Cundinamarca. Los datos recolectados se utilizaron para realizar un análisis de consistencia interna mediante el coeficiente alfa de Cronbach para evaluar la fiabilidad de la prueba. Adicionalmente, se efectuó un análisis de dificultad y discriminación de los ítems y sus opciones de respuesta con el uso de técnicas clásicas de la teoría clásica de las pruebas y un análisis posterior de dificultad según la teoría de respuesta al ítem (TRI). Este análisis permitió identificar la capacidad de cada ítem para discriminar entre estudiantes, con diferentes niveles de habilidad en razonamiento cuantitativo, así como ajustar los ítems para maximizar la precisión de la medición en futuros estudios. El proceso metodológico, aquí descrito, garantiza que la prueba de tamizaje desarrollada no solo cumpla con los más altos estándares de validez y fiabilidad, sino que también esté alineada con las necesidades y contextos de la educación superior en Colombia ya que facilita la identificación temprana de estudiantes en riesgo y promoviendo mejoras pedagógicas en las instituciones educativas.

Análisis de la información

El análisis de aplicación de la prueba de tamizaje se centró en la evaluación de los índices de dificultad y discriminación de los ítems, en la que se utilizó tanto la teoría clásica de los test (TCT) como el modelo de Rasch en el marco de la Teoría de Respuesta al Ítem (TRI) (Stemler & Naples, 2021). En el contexto de la TCT, el índice de dificultad (p) se calcula como la proporción de estudiantes que respondieron correctamente el ítem, donde un valor de p cercano a 0 indica una mayor dificultad y un valor próximo a 1 sugiere que el ítem es bastante sencillo de responder para la población participante; los valores aceptables se sugieren entre 0.20 y 0.80 (Khairani & Shamsuddin, 2016). Por otro lado, el índice de discriminación (D) evalúa la capacidad del ítem para diferenciar las personas con un nivel de habilidad/atributo o desempeño alto el cual se calcula realizando la diferencia entre las proporciones de personas que aciertan el ítems y su pertenencia al grupo superior e inferior; se considera aceptable un valor de a superior a 0.20 (Khairani & Shamsuddin, 2016).

Complementariamente, se aplicó el modelo de Rasch, una técnica de análisis de ítems que utiliza una función logística para estimar la habilidad del estudiante y la dificultad del ítem de manera simultánea. Este modelo se basa en la premisa de que la probabilidad de que un estudiante responda correctamente a un ítem es una función de entre su nivel de habilidad y la dificultad del ítem (von Davier, 2016). El modelo también produce índices de ajuste, como el de infit y el de outfit, que indican en qué medida, los ítems, se alinean con las expectativas del modelo (Aryadoust et al., 2021). Además, se calculó el funcionamiento diferencial de los ítems (DIF) utilizando el método de Mantel-

Haenszel (Herrera Rojas et al., 2001). Este procedimiento permite detectar ítems que funcionan de manera distinta para distintos grupos, en este caso por género, lo que asegura que la medición sea equitativa y que no existan sesgos en la medición. Así, la combinación de estos permite garantizar que los ítems seleccionados ayuden la identificación temprana de estudiantes con dificultades en razonamiento cuantitativo, dado la estimación de su habilidad.

Los análisis psicométricos se llevaron a cabo en el entorno RStudio. En el caso de la Teoría Clásica de los Tests (TCT), se empleó la función `tab_itemscale` del paquete `sjPlot` (Lüdecke, 2024) para calcular los índices de dificultad y discriminación de los ítems, y así proporcionar una medida precisa de su desempeño. Para el modelo de Rasch se utilizaron los paquetes `mirt` (Chalmers, 2012) y `ggmirt` (Masur, 2022), lo cual facilitó tanto la estimación de los parámetros del modelo como la visualización de los resultados bajo el enfoque de la Teoría de Respuesta al Ítem (TRI). Adicionalmente, se implementó el análisis de funcionamiento diferencial de los ítems con el método Mantel-Haenszel a través de la función `difMH` incluida en el paquete `difR` (Magis et al., 2010) para identificar el Funcionamiento Diferencial de los Ítems (DIF), en función de la agrupación por sexo.

Tipo de estudio

El presente estudio se inscribe en el ámbito de la investigación instrumental (Ato & López, 2013) dado que se construye y analizan las propiedades psicométricas de un instrumento de medida. En este caso, también se busca obtener evidencias de validez que sustenten el uso del instrumento diseñado, para lo cual se siguieron los estándares de validación de test (AERA et al., 2014). En este contexto, el estudio aborda la elaboración sistemática de ítems basados en competencias específicas de razonamiento cuantitativo, con el fin de evaluar su consistencia interna, dificultad y discriminación.

Participantes

En la presente investigación participaron 249 estudiantes del programa de Psicología de la Universidad de Cundinamarca. Esto corresponde al 72.3% de la población total de estudiantes de Psicología en la institución, que eran para el momento de la aplicación 344 estudiantes, lo cual es importante decir que corresponde al total de estudiantes entre el primer y sexto semestre del programa, dado que los estudiantes pueden optar a presentar el examen una vez tienen el 75% de avance de carrera, porcentaje que alcanzan generalmente al finalizar el séptimo semestre pues, según el pensum vigente, este pregrado tiene una duración de 9 semestres. La muestra estaba compuesta por un porcentaje de 77.9% de mujeres, que es una tendencia consistente con la población total del programa de Psicología, y 22% hombres. Los participantes diligenciaron consentimientos informados de acuerdo con los lineamientos éticos de la Ley 1090 de 2006 de la profesión de Psicología en Colombia.

Resultados

El análisis inicial de los ítems evidenció que, 15 de los 35 ítems tenían una discriminación negativa (ítems 8, 26 y 32) o demasiado cercana a 0 (ítems 19, 20, 23, 24 y 35). Esto hacía, además, que el Alpha de Cronbach inicial fuera bastante bajo (0.572). Posteriormente, se eliminaron aquellos ítems que su eliminación aumentara el índice Alpha teniendo todos estos una discriminación menor a .15 (ítems 1, 4, 7, 15, 30, 33 y 34), aquellos tenían correlación con los otros ítems. Finalmente quedaron un total de 20 ítems, cuyos resultados principales se muestran en la tabla 1. De esta versión destacan 3 ítems que son bastante difíciles, que son los ítems 6, 21 y 29, mientras los ítems 5,

16, 21, 28 y 29 presentaron índices de discriminación menores a .20. Sin embargo, se optó por mantenerlos dado que su eliminación disminuía la consistencia interna, que obtuvo finalmente un índice 0.652

Tabla 1.
Estadísticas psicométricas de los 20 ítems que conforman la versión final.

Ítem	Dificultad	Discriminación	X ²	p-valor
2	0.39	0.29	8.022	0.627
3	0.57	0.21	6.864	0.810
5	0.33	0.16	17.593	0.037
6	0.16	0.21	13.370	0.270
9	0.63	0.20	14.964	0.133
10	0.34	0.25	6.720	0.821
11	0.46	0.27	14.709	0.196
12	0.63	0.25	15.347	0.120
13	0.78	0.31	16.477	0.087
14	0.58	0.26	20.798	0.036
16	0.36	0.19	10.574	0.480
17	0.43	0.31	9.553	0.571
18	0.41	0.38	9.853	0.453
21	0.13	0.14	8.505	0.667
22	0.58	0.21	7.892	0.723
25	0.66	0.21	4.095	0.943
27	0.56	0.29	7.752	0.735
28	0.35	0.18	7.311	0.773
29	0.17	0.16	14.713	0.196
31	0.39	0.18	12.256	0.268

El ajuste al modelo de Rasch de los ítems también se puede observar en la tabla 1, donde se realizó una prueba de chi cuadrado de ajuste con el modelo. Por lo tanto, un valor p menor a 0.05 significaría un ajuste no aceptable. De los 20 ítems, se encuentran que, únicamente, los ítems 5 y 14 presentan problemas. Sin embargo, en la figura 1, que ilustra el ajuste por criterios de infit y outfit, estos ítems presentan valores acordes, por lo tanto, se mantuvieron también para el análisis final. Por último, se presentan los resultados principales del modelo de Rasch en las figuras 2 y 3, donde se puede observar que, la mayoría de la habilidad de los estudiantes están entre -1 y 1 logit de atributo, teniendo en cuenta que el 0 se asume como el nivel promedio de habilidad esto nos indica realmente los estudiantes tienen un nivel medio de habilidad y que esto puede mantenerse a lo largo del pregrado. En cambio, cuando se analiza la información de los ítems, se encuentra que hay una distribución más amplia de la medición de la dificultad del razonamiento cuantitativo. En este sentido se puede encontrar que la prueba mide en general ítems con una dificultad baja, media y alta, que sería lo esperado para la prueba de tamizaje

Figura 1.
Gráfico de ajuste de los ítems para el modelo de Rasch.
Item Infit and Outfit Statistics

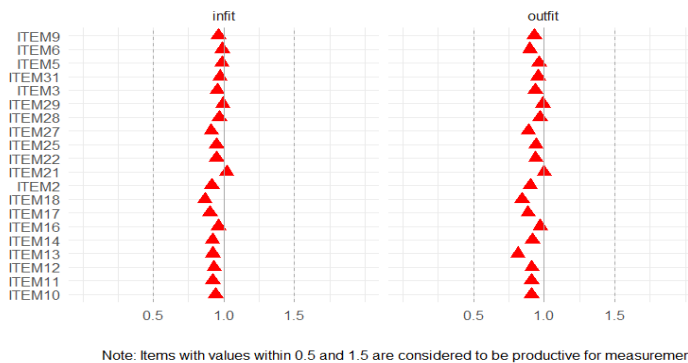


Figura 2.
Wright Map del modelo ajustado.

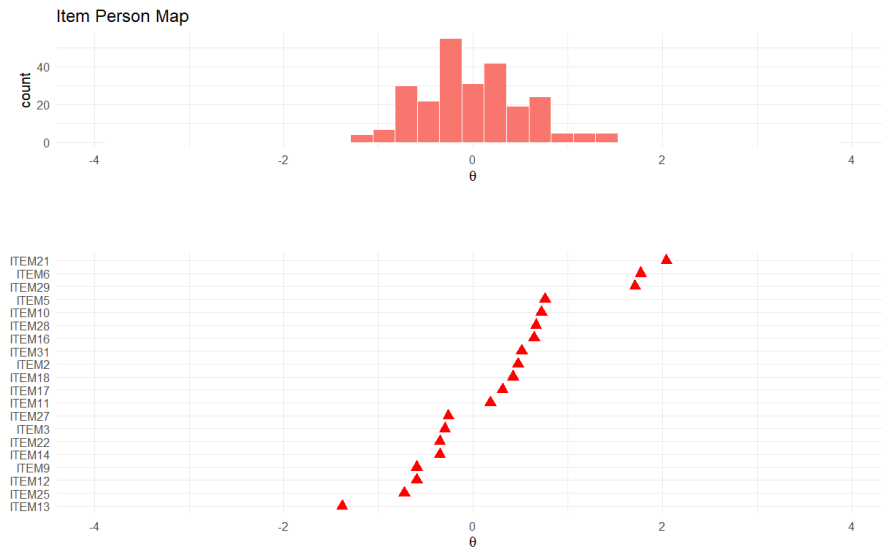
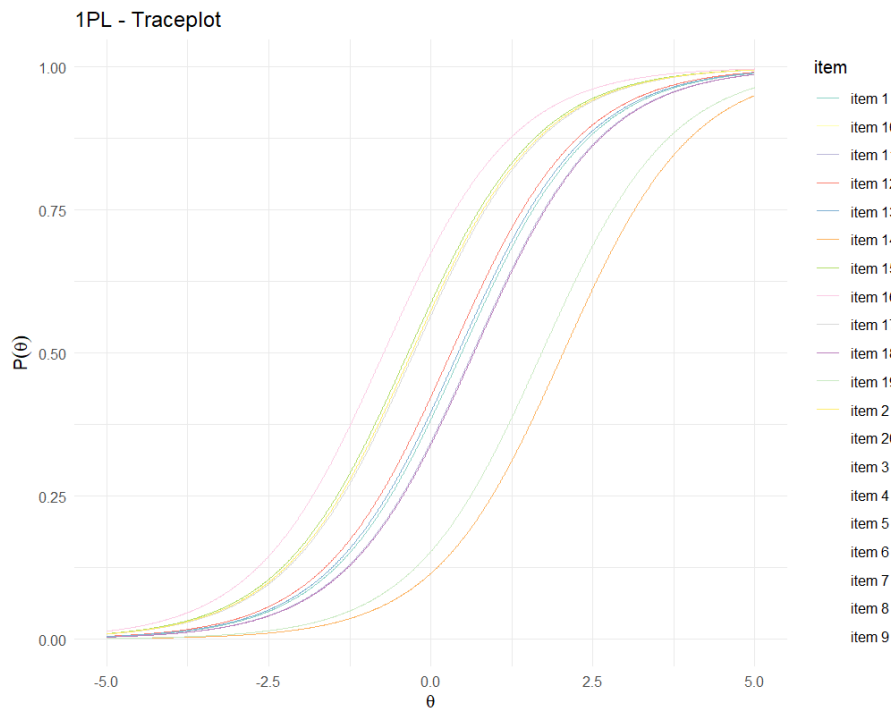


Figura 3.
Curva característica de los ítems.



Por último, la Tabla 2 presenta los resultados del análisis de Funcionamiento Diferencial del Ítem (DIF) mediante el método de Mantel-Haenszel. Los valores de significancia ajustados se calcularon con el procedimiento de Benjamini-Hochberg para controlar el error de tipo I en múltiples comparaciones. De los ítems analizados, únicamente el ítem 6 mostró una significancia estadística antes del ajuste ($p=0.0204$), sugiriendo la posible presencia de DIF. Sin embargo, tras la corrección, todos los ítems obtuvieron valores de p ajustados superior a 0.05, lo cual indica que ninguno de los ítems presenta DIF. Esto implica que los ítems son estadísticamente equitativos, tanto para el grupo de hombres como el de mujeres. De igual manera, para finalizar se calculó el puntaje total, para

hombres y mujeres, a través de la prueba U-Mann-Whitney, sin encontrarse diferencias significativas por sexo ($W = 5985.5$, $p\text{-value} = 0.1663$).

Tabla 2.
Resultados del análisis de Funcionamiento Diferencial del Ítem (DIF) usando el método de Mantel-Haenszel.

Ítem	Estadístico	p-valor	p-valor ajustado
ITEM2	1.4929	0.2218	0.7303
ITEM3	0.0663	0.7968	0.9488
ITEM5	0.0057	0.9398	0.9488
ITEM6	5.3739	0.0204	0.4088
ITEM9	1.3833	0.2395	0.7303
ITEM10	0.0138	0.9066	0.9488
ITEM11	0.0856	0.7698	0.9488
ITEM12	2.4712	0.1160	0.7303
ITEM13	0.0041	0.9488	0.9488
ITEM14	0.4170	0.5184	0.9488
ITEM16	0.0066	0.9351	0.9488
ITEM17	0.3319	0.5645	0.9488
ITEM18	2.6894	0.1010	0.7303
ITEM21	1.5884	0.2076	0.7303
ITEM22	0.0113	0.9154	0.9488
ITEM25	0.0105	0.9184	0.9488
ITEM27	0.2824	0.5952	0.9488
ITEM28	1.2923	0.2556	0.7303
ITEM29	0.0292	0.8643	0.9488
ITEM31	0.2423	0.6226	0.9488

Discusión

Los resultados obtenidos en la presente investigación contribuyen con el desarrollo de herramientas de tamizaje para identificar las necesidades en razonamiento cuantitativo de estudiantes universitarios del programa de Psicología. A partir de investigaciones previas, se ha reconocido que el sexo puede ser una variable diferencial para tener en cuenta (Angulo et al., 2023; Gallego & Casadiego, 2023; Lombana & Cabeza, 2024). Sin embargo, esta tendencia no se encontró en el programa de Psicología, lo que podría deberse a la predominancia, mayormente, de mujeres en el grupo y el aseguramiento de una medición sin análisis diferencial, lo que evita encontrar diferencias. Sin embargo, como parte del ingreso a la Universidad de Cundinamarca se presentan pruebas de ingreso, lo que también podría asegurar similitud en las habilidades de las personas, lo cual puede, de manera conjunta, certificar un grupo homogéneo. Sin embargo, se destaca que la creación de la prueba de tamizaje en la Universidad de Cundinamarca representa un avance, dado que es un insumo para diseñar estrategias educativas que pueden mitigar las dificultades en el razonamiento cuantitativo de los estudiantes.

El análisis psicométrico inicial reveló que varios ítems no discriminaban adecuadamente entre estudiantes con diferentes niveles de competencia. La selección final de 20 ítems con un coeficiente alfa de Cronbach de 0.652 es aceptable para una herramienta de tamizaje, teniendo en cuenta su fácil aplicación a diferentes grupos y usarla como una herramienta de detección de personas que tienen una habilidad menor al promedio del programa de psicología. Aunque destacamos que este valor indica que el instrumento podría beneficiarse de mejoras en términos de consistencia interna. Aunque, un aporte importante de la presente investigación es la implementación de los análisis según el modelo de Rasch, el cual mostró que la mayoría de los ítems presentaron un ajuste adecuado, indicaría que la prueba es capaz de diferenciar de manera efectiva entre estudiantes con diversos niveles de habilidad cuantitativa y se tienen ítems en una gama amplia de los niveles de razonamiento cuantitativo.

Sobre el análisis de funcionamiento diferencial del ítem (DIF) no evidenció variaciones significativas en función del género, lo que sugiere que el instrumento evalúa

equitativamente a estudiantes de ambos. Esta es una característica esencial para garantizar la equidad en las evaluaciones y es un aporte importante. Toda vez, que este instrumento es un insumo necesario para impulsar la competencia de manera efectiva, que recordando lo señalado por Mendoza-Mendoza et al., (2024) sirve para reforzar la calidad educativa y en la construcción de la distinción académica del programa de psicología de la Universidad de Cundinamarca, teniendo en cuenta un objetivo importante a nivel de institución, como es la acreditación de alta calidad.

A pesar de los resultados positivos, este estudio presenta ciertas limitaciones. Aunque la muestra es representativa de estudiantes de psicología en la Universidad de Cundinamarca, se recomienda extender la evaluación a estudiantes de otras disciplinas y en diferentes regiones del país para validar la generalización del instrumento. Dado, que las investigaciones previas han señalado factores contextuales adicionales que afectan el desempeño en razonamiento cuantitativo, como el nivel socioeconómico y características familiares, que en la presente investigación no se analizaron. Estos factores pueden ofrecer una comprensión más completa y detallada de las habilidades cuantitativas de los estudiantes universitarios y podrían ser relevantes en el diseño de intervenciones más específicas. Además, queda el vacío, de si el instrumento encuentra diferencias tanto en funcionamiento diferencial como de resultado de total obtenido.

Es importante señalar que el presente instrumento de tamizaje no solo aporta en el contexto de la Universidad de Cundinamarca, sino que también puede implementarse en otras instituciones en Colombia, donde se espera que este sea un insumo de uso interno, pero con una mira de enfoque proactivos para que la institución apoye el proceso pedagógico de sus estudiantes, más allá del enfoque disciplinar de la formación en pregrado o universitaria. De esta manera, la investigación contribuye a la mejora continua de la calidad educativa, alineándose con los estudios previos sobre el impacto del razonamiento cuantitativo en el desempeño académico y profesional de los estudiantes universitarios.

Referencias

- AERA, APA, & NCME. (2014). Standards for Educational and Psychological Testing: National Council on Measurement in Education.
- Angulo, G, Mercado Mejia, M, Aguilera-Prado, M, & Rincón Moreno, M (2023). Gap in math by gender. Case: university students from Colombia. *Revista Brasileira de Estudos de Populacao*, 40, 1–21. <https://doi.org/10.20947/S0102-3098a0239>
- Aryadoust, V, Ng, LY, & Sayama, H (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6–40. <https://doi.org/10.1177/0265532220927487>
- Ato, M, & López, JJ (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038–1059. <https://www.redalyc.org/pdf/167/16728244043.pdf>
- Chalmers, RP (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48. <https://doi.org/10.18637/jss.v048.i06>
- Congreso de Colombia (2009). Ley 1324 de 2009. 13 De Julio De 2009, 9. http://www.secretariassenado.gov.co/senado/basedoc/ley_1324_2009.html
- El-Hassan, H, Issa, A, Hamouda, MA, Maraqa, MA, & El-Maaddawy, T (2024). Continuous Improvement of an Exit Exam Tool for the Effective Assessment of Student Learning in Engineering Education. *Trends in Higher Education*, 3(3), 560–577. <https://doi.org/10.3390/higheredu3030033>
- Escobar-Pérez, J, & Cuervo-Martínez, Á (2008). Validez de contenido y juicio de expertos : Una aproximación a su utilización. *Avances En Medición*, 6, 27–36.
- Gallego, LE, & Casadiego, MA (2023). A Gender Gap Analysis on Academic Performance in Engineering Students on Admission and Exit Standardized Tests. *Ingeniería e Investigación*, 43(3), 1–12. <https://doi.org/10.15446/ing.investig.103276>
- Herrera Rojas, AN, Sánchez Pedraza, R, & Gómez Benito, J (2001). Funcionamiento diferencial de los ítems : una revisión conceptual y metodológica. In *Acta Colombiana de Psicología* (Issue 5, pp. 41–61). <https://actacolombianapsicologia.ucatolica.edu.co/article/view/589>
- Instituto Colombiano para la Evaluación de la Educación. (2018). Razonamiento Cuantitativo Saber Pro (p. 1).
- Instituto Colombiano para la Evaluación de la Educación. (2023a). • Saber Pro •
- Instituto Colombiano para la Evaluación de la Educación. (2023b). Resultados de los exámenes Saber TyT y Saber Pro.

- Instituto Colombiano para la Evaluación de la Educación. (2024). Oferta de combinatorias para el examen Saber Pro. 28. <https://www2.icfes.gov.co/documents/39286/1258809/Oferta+de+combinatorias+Saber+Pro+2022-2.pdf>
- Khairani, AZ, & Shamsuddin, H (2016). Assessing Item Difficulty and Discrimination Indices of Teacher-Developed Multiple-Choice Tests. In *Assessment for Learning Within and Beyond the Classroom* (pp. 417–426). <https://doi.org/10.1007/978-981-10-0908-2>
- Lombana, J, & Cabeza, L (2024). Factores que Afectan el Desempeño de Estudiantes en Competencias de Razonamiento Cuantitativo en Pruebas Estandarizadas. 22(3), 47–67.
- Lüdecke, D (2024). Package "sjPlot" Title Data Visualization for Statistics in Social Science. 1, 107.
- Magis, D, Béland, S, Tuerlinckx, F, & de Boeck, P (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847–862. <https://doi.org/10.3758/BRM.42.3.847>
- Martínez Solano, R, Cervantes Campo, G, & Jiménez Blanco, G (2022). Quantitative Reasoning, Language and Mathematics. *Zona Próxima*, 36, 76–92. <https://doi.org/10.14482/zp.36.510.71>
- Masur, PK (2022). ggmirt: Plotting functions to extend "mirt" for IRT analyses. R package version 0.1.0.
- Mendoza-Mendoza, A, Mendoza-Casseres, D, & De La Hoz-Domínguez, E (2024). Comparison of industrial engineering programs in Colombia based on standardized test results. *Evaluation and Program Planning*, 103(February), 102415. <https://doi.org/10.1016/j.evalprogplan.2024.102415>
- Mendoza, SL, Ramírez LI, P, & María Serpa, A.J (2021). Errores y Dificultades Vinculadas Al Razonamiento Cuantitativo Entre Estudiantes De Nuevo Ingreso En La Carrera De Ingeniería Errors and Related Difficulties To Quantitative Reasoning Among New Students in Engineering. *Revista Boletín Redipe*, 10(11), 379–399.
- Pérez, JLE, Rincón, EHH, & Peñuela, CLJ (2023). State Tests for Higher Education and Medical Education in Colombia. *Revista Cubana de Medicina General Integral*, 39(1).
- Pérez, LA, Hurtado Orejuela, M, & Murillo Vargas, G (2023). Calidad y desempeño estudiantil en programas de administración de empresas en Colombia. *Revista Venezolana de Gerencia*, 28(Especial 9), 810–830. <https://doi.org/10.52080/rvgluz.28.e9.50>
- Presidencia de la República de Colombia (2009). Decreto 3963 de 2009. *Diario Oficial*, 4.
- Rodríguez-Revilla, R, & Vallejo-Molina, RD (2022). Valor agregado y las competencias genéricas de los estudiantes de educación superior en Colombia. *Revista Iberoamericana de Educación Superior*, 13(36), 44–62. <https://doi.org/10.22201/iisue.20072872e.2022.36.1183>
- Stemler, SE, & Naples, A (2021). Rasch Measurement v. Item Response Theory: Knowing When to Cross the Line. *Practical Assessment, Research and Evaluation*, 26, 1–16. <https://doi.org/10.7275/v2gd-4441>
- Ventura-León, J (2019). De regreso a la validez basada en el contenido. *Adicciones*, xx(01), 2018–2020. <https://doi.org/10.20882/adicciones.1213>
- von Davier, M (2016). Rasch Model. In W. J. van der Linden (Ed.), *Handbook of item response theory* (Vol. 17, Issue 1, pp. 59–76).