

Compartiendo experiencias: Análisis psicométrico de una prueba de diagnóstico matemático en universidades de Costa Rica y Uruguay
Sharing experiences: Psychometric análisis of a mathematical diagnostic test in universities from Costa Rica and Uruguay
Partilha de experiências: análise psicométrica de uma prova diagnóstica matemática em universidades da Costa Rica e do Uruguai

José Andrey Zamora-Araya
Universidad Nacional, Costa Rica
ROR id <https://ror.org/01t466c14>
Heredia, Costa Rica
Correo: jzamo@una.ac.cr
<http://orcid.org/0000-0001-6050-5850>

Matías Salomone-Cayaffa
Universidad Tecnológica, Uruguay
ROR id
Durazno, Uruguay
Correo: matias.salomone@utec.edu.uy
<https://orcid.org/0009-0007-0654-2942>

Alejandro Duarte Furiati
Universidad Tecnológica, Uruguay
ROR id
Durazno, Uruguay
Correo: alejandro.duarte@utec.edu.uy
<https://orcid.org/0009-0004-0637-4684>

Pierina Costa-Alanís
Universidad Tecnológica, Uruguay
ROR id
Durazno, Uruguay
Correo: pierina.costa@utec.edu.uy
<https://orcid.org/0009-0006-6476-9769>

Rosibel Tatiana Vallejos-Brenes
Universidad Nacional, Costa Rica
ROR id <https://ror.org/01t466c14>
Heredia, Costa Rica
Correo: rosibel.vallejos.brenes@una.ac.cr
<https://orcid.org/0009-0002-9910-8208>

Resumen

Introducción. El bajo rendimiento académico en matemática que presenta el estudiantado de nuevo ingreso a la universidad muestra la dificultad que existe en la transición entre la educación secundaria y la universidad. Por ello, es importante contar con instrumentos que permitan identificar debilidades y fortalezas, para así acompañar de mejor manera el proceso educativo en esta etapa de la vida universitaria. **Objetivo** Analizar las propiedades psicométricas de los ítems que conforman la prueba diagnóstica en matemática (PruDiMa), aplicada en 2025 en Costa Rica y Uruguay. **Metodología** El estudio tiene un enfoque cuantitativo de corte transversal y tipo instrumental. Se aplicó un análisis psicométrico basado en la Teoría Clásica de los Test a una prueba diagnóstica aplicada a 93 y 158 estudiantes de dos universidades, una de Costa Rica y otra de Uruguay. Se calcularon para ambas instituciones los índices de dificultad, discriminación y confiabilidad y se compararon los resultados obtenidos en la PruDiMa. **Resultados.** Las alfas de Cronbach fueron de 0,63 y 0,69, entre 8 y 10 ítems tuvieron índices de discriminación bajos y niveles de dificultad variados. **Conclusión.** No se aprecian diferencias significativas en el promedio de las pruebas en las universidades, aunque la prueba presenta tanto baja confiabilidad como discriminación en algunos ítems, pero con ítems de dificultad variada en niveles bajo, medio y alto.

Palabras claves: Prueba Diagnóstica; Matemática; Psicometría, Educación Superior, Teoría Clásica de los Test

Abstract

Introduction. The low academic performance in mathematics presented by students new to university shows the difficulty that exists in the transition between secondary education and university. Therefore, it is important to have instruments that allow us to identify weaknesses and strengths, in order to better accompany the educational process at this stage of university life. **Objective** To analyze the psychometric properties of the items that make up the diagnostic test in mathematics (PruDiMa), applied in 2025 in Costa Rica and Uruguay. **Methodology** The study has a cross-sectional and instrumental quantitative approach. A psychometric analysis based on Classical Test Theory was applied to a diagnostic test administered to 93 and 158

students from two universities, one from Costa Rica and the other from Uruguay. The difficulty, discrimination and reliability indices were calculated for both institutions and the results obtained in the PruDiMa were compared. **Results.** Cronbach's alphas were 0.63 and 0.67, between 8 and 10 items had low discrimination indices and varied levels of difficulty. **Conclusion.** There are no significant differences in the average of the tests in the universities, although the test presents both low reliability and discrimination in some items, but with items of varying difficulty at low, medium and high levels.

Keywords: Diagnostic test; Mathematics; Psychometric; Higher Education ; Classical Test Theory

Resumo

Introdução. O baixo desempenho acadêmico em matemática apresentado pelos alunos que ingressam na universidade mostra a dificuldade que existe na transição entre o ensino secundário e a universidade. Por isso, é importante dispor de instrumentos que permitam identificar fragilidades e pontos fortes, para melhor acompanhar o processo educativo nesta fase da vida universitária. **Objectivo** Analisar as propriedades psicométricas dos itens que compõem o teste de diagnóstico em matemática (PruDiMa), aplicado em 2025 na Costa Rica e no Uruguai. **Metodologia** O estudo tem uma abordagem quantitativa transversal e instrumental. Uma análise psicométrica baseada na Teoria Clássica dos Testes foi aplicada a um teste de diagnóstico aplicado a 93 e 158 estudantes de duas universidades, uma da Costa Rica e outra do Uruguai. Foram calculados os índices de dificuldade, discriminação e fiabilidade para ambas as instituições e comparados os resultados obtidos no PruDiMa. **Resultados.** Os alfas de Cronbach foram de 0,63 e 0,69, entre 8 e 10 itens apresentaram baixos índices de discriminação e níveis de dificuldade variados. **Conclusão.** Não existem diferenças significativas nas médias dos testes nas universidades, embora o teste apresente baixa fiabilidade e discriminação em alguns itens, mas com itens de dificuldade variada nos níveis baixo, médio e alto.

Palavras-chaves: Teste diagnóstico; Matemática; Psicométrico; Ensino Superior; Teoria Clássica dos Testes

Introducción

El rendimiento mostrado por el estudiantado de nuevo ingreso a las instituciones de educación superior (IES) en el área de matemática, en particular en el periodo post pandemia, ha decaído notablemente lo que deja en evidencia el rezago educativo que muestra esta asignatura en aquellas personas que deciden continuar estudios a nivel superior, incluso en aquellas que se decantan por profesiones relacionadas a las áreas STEM (Alqabandi & Durugbo, 2025).

Aunque existen varios factores que pueden relacionarse con el bajo desempeño académico en matemática, el grado de conocimientos previos adquiridos en la educación secundaria es uno de los más relevantes (Hewson, 2011). Esta problemática se refleja en los cursos iniciales, pues investigaciones muestran la dificultad que existe en la transición entre la educación secundaria y la universidad (Hatfield et al., 2022; Hed & Norqvist; 2025; Landgärds-Tarvoll, 2024).

Es así como, las pruebas diagnósticas son una valiosa herramienta para determinar el nivel de conocimientos matemáticos con que ingresa el estudiantado de nuevo ingreso y, a su vez, pueden ser utilizadas en la toma de decisiones que involucren diferentes aspectos del currículo, así como de los apoyos educativos que las IES pueden ofrecer a su población. Una prueba técnicamente bien elaborada puede brindar información no solo sobre la dificultad de los reactivos sino sobre su capacidad de diferenciar las áreas en las que el estudiantado presenta mayores dificultades.

Contexto de la investigación

En Costa Rica, tanto la Universidad de Costa Rica (UCR), la Universidad Nacional (UNA) como el Instituto Tecnológico de Costa Rica (TEC); realizan pruebas diagnósticas de matemática, y los resultados en todas ellas son desalentadores causando presión en los sistemas universitarios tanto por el bajo rendimiento académico en temas clave como álgebra y funciones y el rezago asociado que obliga a la apertura de cursos adicionales debido a los altos niveles de repitencia (Barquero & Ramírez, 2014; Marín- Ballón et al., 2023; Rojas-Torres & Jiménez-Alfaro, 2025; Zamora-Araya & Moreira-Mora, 2025).

La UNA es una institución pública creada en 1974 que recibe alrededor de 4000 estudiantes de nuevo ingreso cada año, caracterizada por su misión humanista, en la cual históricamente se les ha dado preponderancia a las carreras de educación y ciencias sociales; pero que actualmente está ampliando su oferta educativa hacia carreras STEM.

Además, la UNA tiene un sistema de admisión que, mediante acciones afirmativas, favorece el ingreso de estudiantes con menores oportunidades educativas, al estratificar por colegio de procedencia da mayores oportunidades de ingreso a la institución a estudiantes de colegios nocturnos, colegios rurales, educación a distancia en comparación con otras universidades (Zamora-Araya, 2012).

En el caso de la Universidad Tecnológica de Uruguay (UTEC), es la segunda universidad pública del país, de perfil tecnológico creada en 2012 y orientada a la investigación e innovación, con una matrícula de alrededor de 1800 estudiantes cada año, tiene por objetivo brindar acceso equitativo a propuestas de formación terciaria universitaria (Universidad Tecnológica del Uruguay [UTEC], 2026).

Ambos países poseen un sistema de educación secundaria gratuito y en los últimos años han incorporado de manera gradual un enfoque tecnológico, aunque con retos en cuanto a Internet gratuito y equipamiento tecnológico.

Además, ambos sistemas de educación secundaria tienen un enfoque por competencias y, en el área de matemática, comparten contenidos en áreas como álgebra, funciones, modelización orientados a desarrollar el razonamiento lógico-matemático, por lo que afrontan retos similares (Administración Nacional de Educación Pública [ANEP], 2026; Ministerio de Educación Pública [MEP], 2012).

Por estas razones, el equipo investigador decide construir, validar y aplicar una prueba diagnóstica de matemática a estudiantes de nuevo ingreso durante el primer semestre del año 2025 que pueda orientar futuras acciones tendientes a mejorar el rendimiento universitario en esta materia.

De ahí surge la interrogante, ¿en qué medida la prueba diagnóstica en matemática (PruDiMa) presenta propiedades psicométricas adecuadas para identificar los conocimientos matemáticos previos del estudiantado universitario de

nuevo ingreso, a partir de su aplicación en dos universidades una de Costa Rica y otra de Uruguay?

Por ello, el objetivo de la presente investigación es analizar las propiedades psicométricas de los ítems que conforman la prueba diagnóstica en matemática (PruDiMa), aplicada en 2025 en Costa Rica y Uruguay, mediante la Teoría Clásica de los Test, a partir del estudio de sus índices de dificultad y discriminación, con el propósito de valorar su capacidad diagnóstica y orientar la mejora del instrumento.

Por otra parte, al ser una prueba conjunta, se desea comparar el rendimiento académico y las características psicométricas de la prueba diagnóstica de matemáticas en estudiantes de nuevo ingreso de ambos países, mediante el análisis estadístico de resultados y de indicadores de confiabilidad, con el fin de identificar el grado de dominio de las temáticas evaluadas y establecer diferencias significativas entre los contextos educativos.

Antecedentes

1. Evaluación diagnóstica en educación universitaria

Es común que muchas universidades soliciten como requisito de ingreso alguna prueba de admisión o de ejecución con el fin de evaluar conocimientos, habilidades o capacidades de razonamiento que se consideren útiles en la vida universitaria.

Por ejemplo, las carreras de artes suelen realizar pruebas de dibujo, teatro, ejecución de algún instrumento o habilidades sobre temáticas particulares, adicional a la prueba de ingreso general de admisión y, no siempre, los puntajes obtenidos (en cualquiera de las pruebas) se asocian con el rendimiento académico obtenido una vez se ingresa a la universidad (García-Potayo et al., 2026; Vergara-Palma, 2025).

Sin embargo, no es sencillo diseñar e implementar pruebas diagnósticas de alta calidad técnica debido al grado de complejidad, el tiempo y los costos asociados, por lo que en ocasiones resultan ser meras extensiones de las pruebas de admisión, dejando de lado habilidades o conocimientos puntuales de interés (Cruz et al., 2024). En general, la evaluación diagnóstica es una herramienta que posibilita el diseño de estrategias de aprendizaje, identificar debilidades y fortalezas del

estudiantado en temas específicos antes de iniciar con los procesos de instrucción (Keria et al., 2025).

2. Pruebas diagnósticas en matemática

En el caso de las asignaturas de carreras STEM y, particularmente en matemáticas, la prueba escrita ha sido el estándar para evaluar los conocimientos sobre la materia, por lo que en la mayoría de los cursos de esta disciplina el porcentaje asignado a este tipo de pruebas es alto y, el bajo resultado en ellas es una de las razones de la repitencia y el abandono estudiantil (Zamora-Araya & Moreira-Mora, 2025).

Asimismo, la dificultad para aprobar esta asignatura no es nueva para las IES y puede resumirse en el llamado “*problema de las matemáticas*”, que se caracteriza por una falta de conocimientos sobre ciertos contenidos como medidas, álgebra, funciones, geometría o probabilidad, e incluso, estudiantes con buenas calificaciones en estos temas presentan dificultades para aplicarlos en la resolución de problemas en contextos reales (Hewson, 2011).

Esta problemática tiene un carácter multifactorial en donde intervienen aspectos como tener un pensamiento excesivamente procedimental, falta de capacidad para traducir el significado matemático a contextos reales, poca práctica de los contenidos, baja sentido de autoeficacia, bajo nivel de dominio de temáticas abordados en niveles previos, necesarios para comprender los conceptos y procedimientos que se desarrollan en los cursos (Cajacuri, 2022; Hatfield et al., 2022; Hewson, 2011).

Al respecto, Hed & Norqvist (2025) señalan que para estudiantes de las áreas STEM el cambio de secundaria a universidad no es solo académico, sino que representa un salto cualitativo hacia ambientes más abstractos, rigurosos y demandantes que, dependiendo de la formación previa, puede ser el factor determinante entre una carrera exitosa o un abandono prematuro de los estudios superiores.

2.1 Investigaciones sobre pruebas diagnósticas en matemática

Ante el bajo rendimiento en los cursos iniciales de matemática, las IES han diseñado pruebas diagnósticas al estudiantado admitido que debe matricular cursos de matemática, con el objetivo de medir su grado de conocimiento sobre temas relevantes para los diferentes contextos educativos, lo que ha llevado a encontrar desajustes entre lo que el estudiantado recibe en su educación previa y lo que la universidad espera al momento de iniciar sus estudios universitarios (Landgärds-Tarvoll, 2024).

Por ejemplo, en su investigación Hyland & O'Shea (2022) realizaron una encuesta a todas las universidades de Irlanda para conocer cuántas de ellas utilizaban pruebas diagnósticas y cuáles áreas y temas de matemática evaluaban. En total identificaron 14 universidades, que representan el 70% de las IES, 11 de ellas las aplican a gran escala en los cursos de primer año y siete utilizan el formato de opción múltiple, las demás incluyen secciones de respuesta abierta.

La población encuestada se centra en las “matemáticas de servicio”, es decir, de carreras como ingeniería, negocios o educación y los principales temas evaluados fueron números reales y álgebra con 43% y 35% de los ítems, seguidos de geometría, trigonometría y funciones con 8%, 7% y 4% respectivamente. Los resultados muestran deficiencias en las habilidades matemáticas básicas y el estudio recomienda que las instituciones colaboren en el diseño de un banco compartido de pruebas de diagnóstico.

Por otra parte, Keria et al. (2025) realizaron un estudio en Malasia con 38 estudiantes (20 mujeres y 18 hombres) de primer año de ingeniería civil que efectuaron una prueba diagnóstica de 10 preguntas sobre fundamentos de matemática que determinó las fortalezas y debilidades en las áreas. Luego, basado en los resultados de la prueba se realizó una intervención de 14 semanas que incluyó ejercicios semanales, tutorías y retroalimentación constante.

Los resultados que comparan la nota final del curso con la prueba diagnóstica muestran una mejoría. En el caso de las mujeres las puntuaciones pasaron de un promedio de 48,3% a 70,1% y un 75% mejoró su nota. En contraste, los hombres pasaron de una puntuación de 44,5% a 62%. Se concluye que el rendimiento

depende no solo del conocimiento inicial, sino de una combinación de factores en donde las intervenciones tempranas y la práctica continua son determinantes.

También se encuentran pruebas diagnósticas que no solo evalúan conocimiento sino también habilidades procedimentales y de resolución de problemas como es el caso de Marín-Ballón et al. (2023), quienes administraron en Perú una prueba de 18 ítems (50% de tipo procedimental y 50% de resolución de problemas) a 374 estudiantes de primer ingreso en carreras de ingeniería, contabilidad y farmacia.

Los resultados muestran que la nota media en escala de 0 a 10 en habilidades procedimentales fue de 4,12 y en resolución de problemas de 1,63. Además, se encontraron diferencias de género a favor de los hombres, entre carreras y tipo de colegio, siendo las ingenierías y las instituciones privadas los mejores posicionados.

2.2 Pruebas diagnósticas de matemática en Uruguay y Costa Rica

En Uruguay se identificaron varios de estudios sobre pruebas diagnósticas. En el primero Álvarez et al. (2007) aplicaron un instrumento de 30 ítems de opción múltiple para cinco facultades de dos universidades (una pública y otra privada) en temas relacionados con aspectos conceptuales, algorítmicos y de aplicación de álgebra, funciones y cálculo.

Los resultados muestran que el estudiantado de la facultad de Física-Matemática obtiene los mejores resultados con una media de 53,19% de respuestas correctas. Se sugiere cambiar el diseño curricular para que exponga a la totalidad de estudiantes al desarrollo de habilidades y conocimientos matemáticos, no solo a los colegios de orientación científica.

En la segunda investigación Rodríguez (2017) evaluó las competencias en matemáticas en números, álgebra, geometría y estadística a 1380 estudiantes de nuevo ingreso en las sedes regionales de la Universidad de la República. La prueba clasificó al estudiantado en tres niveles: insuficiente, suficiente y avanzado, donde solo el 22% superó la suficiencia.

Un tercer estudio aplicó una prueba de 37 preguntas 12 de física, 10 de química y 15 de matemática a 1243 estudiantes de la facultad físico matemáticas

de la Universidad de la República y concluye que la mayoría del estudiantado necesita orientación y apoyo académico, pues solo el 4,5% supera el nivel de suficiencia (Loureiro & Miguez, 2016).

Por otro lado, en Costa Rica, Barquero & Ramírez (2014) realizaron un informe de validez predictiva de la prueba de diagnóstico compuesta de 17 ítems en temas de números, álgebra, funciones y geometría a 306 estudiantes que matricularon el curso de matemática general en el año 2010. Los resultados muestran que solo el 16,19% obtuvo notas superiores a 70 (en escala de 0 a 100) y mediante un modelo de regresión múltiple determinaron que el porcentaje de varianza explicada fue de tan solo 29,2%.

Asimismo, Zamora-Araya (2012) realizó una validación de la prueba diagnóstica de matemática en la UNA a una muestra de 417 estudiantes por medio de un modelo de ecuaciones estructurales. Mediante un análisis de Rasch determinó una alta dificultad de los ítems comparada con el nivel de habilidad de los sujetos en temas de trigonometría, álgebra y funciones. El porcentaje de aprobación de la prueba fue de tan solo el 1,8%.

Por último, Rojas-Torres & Jiménez-Alfaro (2025) aplicaron una prueba diagnóstica de opción múltiple de 30 reactivos sobre conocimientos en álgebra, funciones y temas específicos de precálculo, en modalidad virtual, para estudiantes que ingresan a cursos de cálculo en la UCR, a una muestra de 2634 personas.

Los resultados revelaron una deficiencia en la preparación académica del estudiantado de primer ingreso, que no los capacita para afrontar con éxito un curso de cálculo, pues el 75% de las personas participantes obtuvieron un rendimiento inferior al 50%. Además, se sugiere una revisión de los contenidos del curso de precálculo, dado que el personal docente que imparte los cursos de cálculo considera que existen contenidos que pueden obviarse.

En resumen, se aprecia que la mayoría de las pruebas diagnósticas en matemática son relativamente cortas (entre 15 y 30 ítems), de opción múltiple centradas en temas de álgebra, funciones y geometría. En general, los resultados no son alentadores y evidencian una necesidad de mejorar el nivel de conocimientos

previos del estudiantado de nuevo ingreso y, así, mejorar los índices de aprobación de los cursos en las carreras del área STEM.

Es así como, el presente trabajo incorpora a estos antecedentes una colaboración conjunta entre universidades de Costa Rica y Uruguay, que no solo aporta validez externa a la prueba, sino que ayuda a complementar visiones sobre problemáticas comunes. En adición, el estudio presenta una aplicación diagnóstica a una población estudiantil que se vio expuesta durante sus primeros años de colegio a una disminución en el número de contenidos, debido a las condiciones que provocó la pandemia de COVID-19. Por ello, es relevante analizar cuál es el nivel de conocimiento en las áreas estipuladas en el currículum escolar.

Marco Teórico

3. . La Teoría Clásica de los Test (TCT)

Como ya se ha mencionada, las pruebas de diagnóstico son aplicadas por muchos departamentos de matemáticas a estudiantes de nuevo ingreso con el fin, por un lado, de que el personal docente conozca el nivel de habilidades matemáticas de sus futuros estudiantes y, por otro, informar al estudiantado sobre el nivel de conocimientos matemáticos en determinados temas de interés (Akveld & Kinnear, 2024).

3.1 Fundamentación Teórica

En cuanto al diseño de este tipo de pruebas se puede optar por dos opciones, la tradicional basada en la teoría clásica de los test (TCT) o la teoría de respuesta a los ítems (TRI). La TCT asume que la puntuación observada (X) de una persona es el resultado de la suma de su puntuación verdadera (V) y un error de medición aleatorio (e), es decir, $X=V + e$.

En otras palabras, las puntuaciones que las personas obtienen en el test, en realidad son la suma de la puntuación verdadera más un error. Sin embargo, el modelo no permite conocer ni cuál es su puntuación verdadera ni el error contenido en esa puntuación, ya que el error cometido al medir alguna variable con un test puede deberse a variadas causas, que pueden estar en la propia persona, en el contexto que la rodea, o en el mismo test (Nunnally & Berstein, 1995).

Según Muñiz (2002), el modelo clásico tiene los siguientes supuestos:

1. Se define la puntuación verdadera (V) como la esperanza matemática de la puntuación empírica, $V = E(X)$.
2. Se asume que no existe correlación entre las puntuaciones verdaderas de las personas y el tamaño de los errores que afectan a esas puntuaciones, $\rho(v, e) = 0$. Es decir, no hay razón para pensar que el tamaño del error vaya a afectar la puntuación verdadera, lo que significa que puede haber puntuaciones verdaderas altas con errores bajos, o altos.
3. Si se aplican correctamente los test, los errores de medida de las personas obtenidos en un test j no correlacionan con los obtenidos en un test distinto k , es decir, $\rho(e_j, e_k) = 0$, pues los errores serían aleatorios en cada aplicación y no habría razón para pensar que covaríen.

3.2 Consistencia Interna e Índices de Dificultad y Discriminación

Existen medidas que permiten determinar las características psicométricas de un test. Una de las más utilizadas, es el alfa de Cronbach, desarrollado en 1951 y que, según Muñiz (2002), “refleja el grado en el que covarían los ítems que constituyen el test, es, por tanto, un indicador de la consistencia interna del test” (p.54). Aunque no existe un valor mínimo, pues la confiabilidad es una cuestión de grado, se suele considerar valores de 0,70 o superiores como aceptables y la consistencia interna se utiliza para medir el grado de solidez de los resultados de los ítems de una escala y en cuánto se correlacionan entre sí (Nunnally & Bernstein, 1995).

Por su parte, el índice de dificultad de un ítem en la TCT se entiende como la proporción de personas que aciertan el ítem entre la totalidad de personas que intentaron resolverlo, es decir, valores altos se asocian con ítems sencillos y valores bajos con ítems difíciles (Nunnally & Bernstein, 1995). Otro índice utilizado para analizar la calidad de un ítem es el índice de discriminación. De acuerdo con Muñiz (2002), un ítem tiene un poder discriminativo en la medida que es capaz de distinguir entre las personas que obtienen puntuaciones altas en el test y las que tienen puntuaciones bajas.

Se suele definir en términos de la correlación entre las puntuaciones de las personas en el ítem y sus puntuaciones totales en el test, Por lo general, en pruebas

de conocimiento donde se acierta o falla el ítem, se utiliza la correlación biserial (Muñiz, 2002).

3.3 Limitaciones de la TCT

No obstante, no siempre se puede verificar empíricamente el cumplimiento de los anteriores supuestos. Además, de acuerdo con Rusch et al. (2017) la TCT presenta los siguientes problemas:

1. Asume una relación lineal entre la variable latente y los puntajes observados, lo que empíricamente es difícil de sostener.
2. El puntaje verdadero no puede estimarse directamente o solo se puede haciendo fuertes suposiciones.
3. Los parámetros como la confiabilidad, la discriminación, ubicación o cargas factoriales dependen de la muestra obtenida, es decir, no existe invarianza de las propiedades de los test respecto de las personas utilizadas para estimarlas.

En contraste la TRI ofrece una alternativa que permite asumir relaciones no lineales entre la variable latente y las puntuaciones del test, una estimación más apropiada para la puntuación verdadera y una estimación independiente de la muestra utilizada (Rusch et al.,2017).

No obstante, la TRI también presenta limitaciones y la principal es el tamaño de muestra, pues según Schroedersy & Gnamb (2025), los tamaños mínimos aceptados para los modelos de 1, 2 y 3 parámetros van de 100 a 250, de 500 a 1000 y mayores a 1000, respectivamente. Esto hace que, a pesar de las limitaciones, la TCT se siga utilizando cuando se disponen de muestras relativamente pequeñas, como es el caso de la presente investigación.

Métodos

Esta investigación tiene un enfoque cuantitativo con diseño no experimental de corte transversal y de tipo instrumental, pues como lo señala Hernández et al. (2014) se hace uso de instrumentos de medición y se realiza de forma empírica en un único momento.

Participantes

En el estudio participaron estudiantes de primer ingreso, 93 de la UNA en carreras como administración, economía, enseñanza de las ciencias e ingenierías y 158 de la UTEC matriculados en Ingeniería en Energías Renovables, Ingeniería Agroambiental e Ingeniería en Agua y Desarrollo Sostenible. En la UNA la muestra estuvo constituida por estudiantes que recibieron el curso de elementos del álgebra, geometría y las funciones impartido por la instancia de la vicerrectoría de docencia la UNATEPREPARA, que ofrece este curso como repaso de los temas de secundaria antes del inicio de clases.

La matrícula es abierta y la mayoría de los cursos son virtuales, la muestra resultante se obtuvo de los grupos que se ofrecieron en modalidad presencial. En el caso de la UTEC, al realizarse en la primera semana de clases, la muestra resultante comprende a la totalidad de la matrícula de las carreras participantes.

Instrumento

El instrumento utilizado en esta investigación, inicialmente fue diseñado por Costa et al. (2023) que consistía en una prueba de opción múltiple de 18 ítems agrupados en tres bloques de seis preguntas cada uno: álgebra, evalúa expresiones algebraicas, ecuaciones y sistema de ecuaciones, modelización, tiene que ver con preguntas que sobre la resolución de problemas asociados a distintos modelos matemáticos y cálculo de funciones, con preguntas sobre los modelos de función lineal, cuadrático y exponencial.

A partir de la colaboración realizada con la UNA, la prueba incorporó un cuarto bloque relacionado con preguntas de razonamiento matemático. En esta nueva versión, los ítems fueron elaborados por el equipo investigador y fueron validados por expertos en las diferentes áreas de ambos países, por ejemplo, para el bloque de modelización se seleccionaron jueces que tenían proyectos de investigación y publicaciones en esa área y para los ítems de razonamiento profesores con experiencia en olimpiadas internacionales de matemática.

Cada bloque fue evaluado por tres jueces, para un total de 12 personas que validaron los ítems de acuerdo con los criterios sugeridos por Escobar-Pérez & Cuervo-Martínez (2008) que son: (a) suficiencia, el ítem es adecuado para cumplir con el propósito, (b) claridad, el ítem se comprende fácilmente, es decir, su

redacción y gramática son adecuadas, (c) coherencia, el ítem tiene relación lógica con el propósito que está midiendo y (d) relevancia, el ítem es esencial o importante, es decir debe ser incluido.

La rúbrica de evaluación se envió por correo electrónico en un archivo de Excel a cada persona evaluadora, según el bloque asignado. La escala de evaluación para cada criterio era de 1 a 4, donde 1 representaba muy en desacuerdo, 2 en desacuerdo, 3 de acuerdo y 4 muy de acuerdo. El indicador utilizado para medir el grado de concordancia de los jueces por bloque fue el V de Aiken y en todos los casos se obtuvieron valores superiores o iguales a 0,89.

Procedimientos

La investigación se dividió en tres fases. La primera fue la elaboración de los ítems por parte del equipo investigador, basados en una tabla de especificaciones con temas comunes evaluados en educación secundaria en ambos países. La segunda fase consistió en la validación de los ítems por parte de las personas expertas en cada bloque y la tercera fue la aplicación de la prueba en las respectivas universidades. La aplicación, en la UNA se realizó de manera presencial durante la última semana de enero y primera de febrero, y en la UTEC se efectuó en la segunda semana de marzo. En ambos casos la prueba tuvo una duración de una hora y no se permitió el uso de ningún tipo de calculadora.

Técnicas

La prueba diagnóstica (PruDima), debido al tamaño de las muestras en cada universidad, fue analizada en el contexto de la TCT. Se establecieron los índices de dificultad y de discriminación para cada uno de los ítems y para determinar el nivel de confiabilidad se utilizó el alfa de Cronbach. Además, para comparar las notas en ambas IES, se efectuó la prueba de Mann-Whitney, pues la nota no sigue una distribución normal según la prueba Shapiro- Wilk (Valores p de 0,01 y 0,02). Los análisis se realizaron con el software R versión 4.5.2.

Aspectos éticos

Cada prueba venía acompañada de una instrucción que explicaba sus objetivos y se estipula la confidencialidad y anonimato de las respuestas. El manejo de la información se realizó acorde con los criterios establecidos por las

regulaciones de cada universidad y, en el caso de Costa Rica, sigue los lineamientos estipulados en la ley 8968: Ley de Protección de la Persona frente al tratamiento de sus datos personales. En cuanto al uso de inteligencia artificial (IA), esta se utilizó como asistencia en la escritura, así como para la búsqueda y resumen de bibliografía sobre la temática del estudio.

Por último, los datos necesarios para reproducir los resultados del estudio, incluida la PruDiMa, junto con el código utilizado para generar los resultados de esta investigación pueden ser consultados en el siguiente enlace: <https://n9.cl/ajjgil>

Análisis de resultados y discusión

En esta sección los resultados se estructuran de la siguiente forma. Primero se realiza un análisis descriptivo de los ítems por universidad, luego se brinda los resultados correspondientes a los índices de confiabilidad, dificultad y discriminación. Finalmente se ilustran algunos ítems de la prueba.

Los resultados de la PruDiMa se pueden describir según la nota obtenida, así como por las propiedades psicométricas de los ítems que la componen. La Tabla 1 presenta un resumen de los resultados de la prueba, por bloque en cada universidad.

Tabla 1: Puntuaciones en la PruDima por bloque según universidad. I semestre 2025

Módulo ^a	Costa Rica (UNA) N=93				Uruguay (UTEC) N=158			
	Media	moda	Mediana	Des. Est	Media	moda	Mediana	Des. Est
Álgebra	2,39	2	2	1,42	2,2	1	2	1,36
Modelización	2,48	3	3	1,09	2,52	3	3	1,17
Funciones	2,52	3	3	1,19	2,53	3	3	1,38
Razonamiento	3,87	4	4	1,27	3,24	3	3	1,5
Puntaje Total	11,26	13	11	3,28	10,49	9	10	3,8
Nota	46,91	54	46	13,67	43,72	38	42	15,81

Nota: a. Los puntajes de los módulos se dan en escala de 1 a 6, cada uno tiene 6 preguntas y la nota está en escala de 0 a 100.

Fuente: Elaboración propia.

En general los resultados de la PruDima no fueron alentadores, ya que las notas en las dos universidades en promedio están por debajo de 50 y la media de respuestas correctas por módulo, salvo el de razonamiento, es inferior a 3 (ver Tabla 1). Aunque la UNA obtuvo una nota promedio ligeramente superior a su contra parte uruguaya, la diferencia no es significativa (prueba de Mann-Whitney con $p=0,13$).

Además, los resultados sugieren un comportamiento similar en ambas IES, pues se logra apreciar como los puntajes van aumentando progresivamente en los módulos desde álgebra, que tiene el menor puntaje, hasta razonamiento que la que tiene la mejor calificación. Por otro lado, la cantidad de personas que obtuvieron una nota de 60 (en escala de 0 a 100) es del 18,28% y 15,19% en la UNA y UTEC, respectivamente; lo que evidencia un bajo nivel de dominio de los contenidos evaluados.

En lo que respecta a la confiabilidad, los resultados para la UNA recomendaron eliminar los ítems 20, 23 y 24 del módulo de razonamiento. El análisis de confiabilidad de los 21 ítems restantes arrojó un alfa de Cronbach de 0,63 que, aunque bajo, para efectos de explorar el comportamiento de los ítems y mejorar a futuro la prueba, se considera suficiente.

En el caso de UTEC, el análisis sugirió eliminar los ítems 4 y 9 el alfa de Cronbach al eliminarlos fue de 0,69 que al igual que en el caso costarricense es suficiente para los propósitos del estudio. Esto sugiere una revisión de los ítems de la prueba con el fin de mejorar este indicador.

Además, se hace necesario determinar cuáles ítems tienen un funcionamiento adecuado o, en su defecto, cuáles deben ser revisados. Una forma de visualizar los ítems que requieren revisión es calcular sus índices de dificultad y discriminación.

Los índices de dificultad, así como los de discriminación varían entre ítems e instituciones (ver Tabla 2). Por ejemplo, el ítem más difícil fue el 14 y el más fácil el 18 (ambos del bloque de funciones) para ambas universidades. La Figura 1, ilustra estos ítems. Con respecto a la capacidad de discriminación de los ítems, en la UNA nueve de ellos pueden considerarse de bajo poder discriminativo y 6 en el caso de la UTEC.

Tabla 2: Índices de dificultad y discriminación en la PruDiMa por ítem según universidad. I semestre 2025

Ítems	Costa Rica (UNA)		Uruguay (UTEC)	
	Índice de Dificultad	Índice de Discriminación	Índice de Dificultad	Índice de Discriminación
I1_al	0,48	0,45	0,52	0,46
I2_al	0,56	0,32	0,52	0,56
I3_al	0,22	0,35	0,27	0,46
I4_al	0,45	0,48	NA	NA
I5_al	0,38	0,23	0,37	0,35
I6_al	0,30	0,29	0,27	0,35
I7_mod	0,12	0,23	0,22	0,25
I8_mod	0,56	0,32	0,59	0,52
I9_mod	0,13	0,06	NA	NA
I10_mod	0,66	0,32	0,59	0,31
I11_mod	0,19	0,26	0,23	0,19
I12_mod	0,83	0,39	0,75	0,27
I13_fun	0,58	0,48	0,53	0,52
I14_fun	0,08	0,19	0,08	0,10
I15_fun	0,26	0,26	0,32	0,29
I16_fun	0,30	0,39	0,28	0,65
I17_fun	0,45	0,32	0,54	0,52
I18_fun	0,85	0,23	0,78	0,35
I19_raz	0,84	0,16	0,74	0,17
I20_raz	NA	NA	0,54	0,31
I21_raz	0,76	0,32	0,53	0,52
I22_raz	0,65	0,48	0,57	0,33
I23_raz	NA	NA	0,39	0,52
I24_raz	NA	NA	0,47	0,44

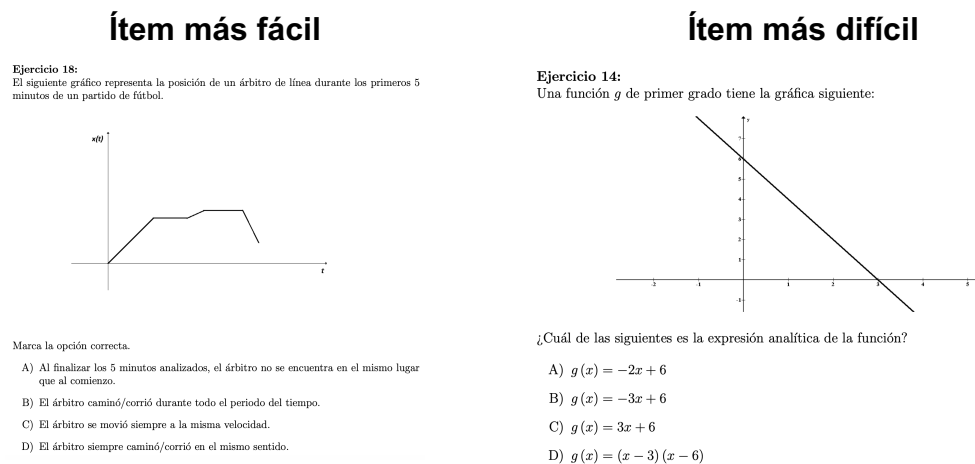
Nota: Los números expresados con negrita identifican a los ítems con baja discriminación.

Fuente: Elaboración propia

Asimismo, los ítems 7, 11, 14, 15 y 19 tienen baja discriminación en las dos universidades. En la UNA se agregan los ítems 5, 6 de álgebra, 9 de modelización y 18 de funciones, en el caso de UTEC se adiciona el ítem 12 de modelización (ver Tabla 2).

Como puede apreciarse, el comportamiento en cuanto poder discriminativo de los ítems es variado, no obstante, comparten la mayoría de los ítems con baja discriminación, el número de respuestas correctas por bloque es similar (ver Tabla 1), incluso comparten el ítem más fácil y difícil de la prueba (ver Figura 1). Esto sugiere que a pesar de tener contextos diferentes, el estudiantado de ambas universidades muestra estadísticas e indicadores psicométricos similares en la PruDiMa.

Figura 1. Ítems de mayor y menor dificultad de la PruDiMa



Fuente: Elaboración propia

Nótese como el ítem 14 requiere para su resolución, además del componente procedimental que podría ser aplicar una fórmula (como la pendiente y el intercepto), un componente de análisis conceptual o interpretativo que va más allá de la simple aplicación de algoritmos, lo que denota una falencia en este aspecto en las personas que participaron de la prueba. En contraste, el ítem 18 no requiere de procedimientos o fórmulas para su resolución, solo analizar la gráfica en función de los enunciados propuestos.

Además de su dificultad, el ítem 14 presenta un bajo nivel de discriminación en ambas instituciones, lo que sugiere una revisión del reactivo o su sustitución por otro con menor dificultad y mayor poder discriminativo.

Conclusiones

La PruDiMA tiene como fin determinar el nivel de conocimientos previos del estudiantado de primer ingreso en temas de matemática que se abordan en la educación secundaria. Los resultados evidencian un bajo nivel de conocimientos matemáticos en la mayor parte del estudiantado, en particular en el tema de álgebra, la cual es la base para los cursos introductorios de primer año.

El objetivo de este estudio fue evaluar la calidad técnica de la PruDiMa en cuanto a sus propiedades psicométricas y comparar los resultados entre la UNA y la UTEC. AL respecto, la conclusión es que no existen diferencias significativas en el resultado promedio de la prueba entre la UNA y la UTEC, por lo que este resultado constituye una evidencia de validez externa de la prueba. Tampoco se aprecian brechas importantes en el nivel de dificultad por bloque ni en la mayoría de los ítems con baja discriminación, pero sí en la consistencia interna.

Esto podría deberse, entre otras causas, al contexto educativo, al tamaño de la muestra y al momento en que se realizó la prueba, pues se puede conjeturar que refleja las deficiencias derivadas de los ajustes metodológicos de los años anteriores debidos a la pandemia de COVID-19.

Por ello, se recomienda realizar más estudios que puedan profundizar en esta hipótesis que puedan obtener muestras mayores y condiciones de aplicación más homogéneas y, así, poder realizar análisis más sofisticados como los propuestos en la TRI. También es importante que futuras investigaciones aborden el tema del bajo nivel de conocimientos previos con metodologías que permitan profundizar en sus causas, ya sea con diseños cualitativos o mixtos que complementen los resultados de pruebas diagnósticas.

Con respecto a las características psicométricas, los resultados sugieren que se debe mejorar la consistencia interna y el poder discriminativo de varios ítems de la PruDiMa, dado que en la UNA y UTEC la prueba presenta tanto baja confiabilidad (alfas de 0,62 y 0,69, respectivamente) como baja discriminación en algunos ítems.

Esto es un indicador de que la prueba debe calibrarse, es decir, revisar aquellos ítems con baja discriminación y replantearlos o sustituirlos por nuevos en

futuras aplicaciones. También es prudente revisar los ítems de baja y alta dificultad, con el fin de obtener una prueba equilibrada.

A pesar de ello, la prueba presenta ítems de dificultad variada tanto baja, media y alta; lo que es apropiado para un test con fines diagnósticos. Además, se logran observar ítems con valores altos de discriminación, lo que puede guiar la construcción de nuevos ítems.

Limitaciones

Entre las limitaciones que afrontó la investigación es que no se cuenta en ninguna de las universidades con un proceso de evaluación diagnóstica consolidado, no existen fechas en el calendario ni presupuesto para este fin y la aplicación de las pruebas se debe hacer con las unidades académicas que deseen participar, es decir, es de carácter voluntario. Otra limitante es el tamaño de muestra, pues en el caso de UTEC la matrícula estuvo restringida a tres carreras y en el caso de la UNA la mayoría de la población, al momento de la aplicación, recibía clases en modalidad virtual. Además, al ser estudiantes de primer ingreso una buena parte de esta población no está familiarizada con estos procesos y, dado que la prueba hoy en día no tiene ninguna repercusión, no siempre se realiza con el debido compromiso académico que se esperaría.

Referencias

- Administración Nacional de Educación Pública. (2026). *Programa de Matemática: Educación Media Superior. 3.º año*. Dirección General de Educación Secundaria. <https://www.dges.edu.uy/propuesta-educativa/programas-de-asignatura>
- Akveld, M., & Kinnear, G. (2024). Improving mathematics diagnostic tests using item analysis. *International Journal of Mathematical Education in Science and Technology*, 55(10), pp.2478–2505. <https://doi.org/10.1080/0020739X.2023.2167132>
- Alqabandi, F., & Durugbo, C. M. (2026). Examining university instructors' continuance intentions for online assessment: the role of safeguards,

- openness and cognitive reactions. *Assessment & Evaluation in Higher Education*, 51(2), 263–280. <https://doi.org/10.1080/02602938.2025.2591557>
- Álvarez, W.; Lacuéz, E., Pagano, M., Lacués, E., Czerwonogora, A., Isolabella, G. & Leymonié, J. (2007). La matemática al ingreso en la universidad. Un estudio comparativo de cuatro Facultades en el Uruguay. *Revista Iberoamericana de Educación*, 42(4), 1-9. <https://rieoei.org/RIE/article/view/2409/3414>
- Barquero, S., & Ramírez, A. (2014). Análisis de las Pruebas de Diagnóstico en Matemática del Instituto Tecnológico de Costa Rica. *Revista Digital: Matemática, Educación E Internet*, 11(2). <https://doi.org/10.18845/rdmei.v11i2.1957>
- Cajacuri Garcilazo, J. N. (2023). *Conocimientos Previos de matemática y su relación con el Rendimiento académico del curso Cálculo Diferencial en estudiantes del II ciclo de la Facultad de Ingeniería Civil de la Universidad Nacional Federico Villarreal, 2020.*[Tesis de Maestría]. Universidad Nacional de Educación. <https://repositorio.une.edu.pe/handle/20.500.14039/8114>
- Costa, P., Rodríguez, E., Duarte, A. & Salomone, M. (2024). Desde lo cognitivo a lo afectivo: la evaluación en matemática como medio. En R. del Valle (Ed.), *Actas XII Congreso Latinoamericano sobre el Abandono en Educación Superior (CLABES)* (pp. 493–500). Universidad Católica de Temuco. https://www.redguia.org/_files/ugd/31655e_ca0f7f94a4be4ef1ba5df110fee4b572.pdf
- Cruz, S.; Jimenez, D.; Sun, Y.; Kaiser, G. & Varas, L. (2024) Design and validation of initial diagnostic tests for preservice teachers as a tool for teacher education effectiveness, *Journal of Curriculum Studies*, 56(4), 392-412 <https://doi.org/10.1080/00220272.2024.2322490>
- 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.

https://gc.scalahed.com/recursos/files/r161r/w25645w/Juicio_de_expertos_u4.pdf

García-Potayo, J; Arregui Eaton, I.G., & Espinoza Galindo, A. F. (2026). Diseño del Examen de Egreso de la Licenciatura en Artes Plásticas en la Universidad Autónoma de Baja California. *CPU-e, Revista de Investigación Educativa*, (42). <https://doi.org/10.25009/cpue.v1i42.2897>

Hatfield, N.; Brown, N. & Topaz, C. (2022). Do introductory courses disproportionately drive minoritized students out of STEM pathways? *PNAS Nexus*, 1 (4), 1-10, <https://doi.org/10.1093/pnasnexus/pgac167>

Hed, L., Åhag, P., & Norqvist, M. (2025). Changes in mathematical skills among freshman engineering students. *International Journal of Mathematical Education in Science and Technology*, 56(11), 2236–2250. <https://doi.org/10.1080/0020739X.2024.2404427>

Hernández, R., Fernández, C. & Baptista, P. (2014). *Metodología de la investigación*. Mc Graw Hill.

Hewson, S. (2011). *The Mathematical Problems Faced by Advanced Stem Students*. <https://nrich.maths.org/articles/mathematical-problems-faced-advanced-stem-students>

Hyland, D. & O’Shea, A. (2022). The nature and prevalence of diagnostic testing in mathematics at tertiary-level in Ireland, Teaching Mathematics and its Applications. *An International Journal of the IMA*, 41(1), 32-50. <https://doi.org/10.1093/teamat/hrab006>

Keria, R., Md Nor, M., & Noor Mat, S. (2025). The need to investigate student performance through diagnostic tests in higher education. *Education*, 7(24), 589-604. <https://doi.org/10.35631/IJMoe.724041>

- Landgärds-Tarvöll, L. (2024). Understanding the challenges of the secondary-tertiary transition in mathematics for economics in higher education: a literature review, *Teaching Mathematics and its Applications. An International Journal of the IMA*, 43(4). 251–272, <https://doi.org/10.1093/teamat/hrad011>
- Loureiro, S. & Miguez, M. (2016). Favoreciendo la permanencia y el avance en la facultad de ingeniería. *Memorias de las V Jornadas Nacionales y I Latinoamericanas de Ingreso y Permanencia en Carreras Científico-Tecnológicas* (IPECyT), (pp.1-6). Universidad Tecnológica Nacional, Facultad Regional Bahía Blanca, Argentina. https://www.edutecne.utn.edu.ar/ipecyt-2016/10-IPECyT_2016.pdf#page=3
- Marín-Ballon, EM, Romero Gómez, FL, Linares Flores Castro, AE, & Flores Castro Linares, MR (2024). Evaluación de las habilidades de resolución de problemas y procedimentales de estudiantes de primer año en una institución de educación superior peruana. *Eurasia Journal of Mathematics, Science and Technology Education*, 20 (2), em2396. <https://doi.org/10.29333/ejmste/14154>
- Ministerio de Educación Pública. (2012). *Programas de Estudio de Matemáticas: I, II y III Ciclos de la Educación General Básica y Educación Diversificada*. <https://www.mep.go.cr/sites/default/files/media/matematica.pdf>
- Muñiz, J.(2002). *Teoría Clásica de los Test*. Ediciones Pirámide, S.A.
- Nunnally, J., & Bernstein, I. (1995). *Teoría psicométrica*. McGraW-Hill.
- Rojas-Torres, L., & Jiménez-Alfaro, K. (2025). Construcción e implementación de una prueba diagnóstica para cursos de cálculo. *Revista Electrónica Calidad En La Educación Superior*, 16(1), 376–405. <https://doi.org/10.22458/caes.v16i1.5628>
- Rodríguez, P.C. (2017). Creación, desarrollo y resultados de la aplicación de pruebas de evaluación basadas en estándares para diagnosticar

- competencias en matemática y lectura al ingreso a la universidad. *Revista Iberoamericana de Evaluación Educativa*, 10(1), 89-107. <https://doi.org/10.15366/riee2017.10.1.005>
- Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*, 54(2), 189-203. <https://doi.org/10.1016/j.im.2016.06.005>
- Schroeders, U., & Gnams, T. (2025). Sample-size planning in item-response theory: A tutorial. *Advances in Methods and Practices in Psychological Science*, 8(1), 1-36. <https://doi.org/10.1177/2515245925131479>
- Universidad Tecnológica del Uruguay. (s.f.). *Modelo UTEC* [Sitio web]. Recuperado el 8 de junio de 2026. <https://utec.edu.uy/es/sobre-utec/>
- Vergara-Palma, J. S. (2025). Las pruebas de admisión especial a las carreras de música: ¿ Una traba más al acceso igualitario? *Revista enfoques educacionales*, 22(2), 91-113. <http://dx.doi.org/10.5354/2735-7279.2025.76006>
- Zamora-Araya, J.A, (2012). *Validación de la prueba de diagnóstico en Matemática para estudiantes de nuevo ingreso a la universidad nacional de costa rica, por medio de un modelo de Rasch y de un modelo de Ecuaciones estructurales para predecir la nota*. [Tesis de Maestría inédita]. Universidad de Costa Rica, San José, Costa Rica.
- Zamora Araya, J. A., & Moreira Mora, T. E. (2025). Abandono estudiantil en el curso de Matemática General: identificación de variables relevantes para su predicción. *Revista Educación*, 49(1),1-23. <https://doi.org/10.15517/revedu.v49i1.61275>