



Análisis de Rasch en la medición de competencias matemáticas en estudiantes de nuevo ingreso de la PUCMM

Jeremías **Willmore** Metivier

Escuela de Ciencias Naturales y Exactas, Pontificia Universidad Católica Madre y Maestra
República Dominicana

j.willmore@ce.pucmm.edu.do

Carmen Miguelina **Santos** Abreu

Escuela de Ciencias Naturales y Exactas, Pontificia Universidad Católica Madre y Maestra
República Dominicana

cm.santos@ce.pucmm.edu.do

Resumen

El desempeño matemático de los estudiantes de nuevo ingreso en la Pontificia Universidad Católica Madre y Maestra ha sido motivo de preocupación en los últimos años. Este estudio se enfoca en evaluar sus competencias mediante una prueba diagnóstica diseñada bajo el modelo de Rasch, con el propósito de identificar sus fortalezas y debilidades en aspectos fundamentales. La prueba se aplicó a 325 estudiantes y los datos fueron analizados utilizando métodos psicométricos. Los resultados indican que más del 50% presenta dificultades en operaciones básicas y resolución de problemas matemáticos esenciales para el éxito académico. A partir de estos hallazgos, se recomienda la implementación de estrategias de nivelación y la aplicación periódica de pruebas para monitorear avances. Este estudio contribuye al diseño de instrumentos de medición más precisos y a la optimización de los procesos de enseñanza de Matemáticas en educación superior.

Palabras clave: Análisis de Rasch; Prueba diagnóstica; Competencias matemáticas; Estudiantes universitarios; Medición educativa; Pruebas Estandarizadas; Modelo de Rasch.

Definición y relevancia del problema

Los docentes de Matemáticas en la Pontificia Universidad Católica Madre y Maestra han identificado deficiencias significativas en la comprensión matemática de los estudiantes de nuevo ingreso. Estas dificultades afectan la resolución de problemas, la interpretación de gráficos y la ejecución de cálculos básicos, habilidades esenciales para el éxito en asignaturas que dependen de las Matemáticas. Precálculo es la primera asignatura matemática que deben cursar los estudiantes universitarios para fortalecer competencias en el análisis y modelado de funciones algebraicas y trascendentales. No obstante, muchos presentan dificultades al aplicar conocimientos previos fundamentales, lo que indica una preparación insuficiente. Según el currículo dominicano (MINERD, 2022) y el marco de referencia de las pruebas nacionales (MINERD, 2020), estos conocimientos deberían haberse adquirido en la educación preuniversitaria, lo que justifica la necesidad de un diagnóstico preciso.

Ante esta problemática, surge la necesidad de responder a preguntas clave como: ¿cuál es el nivel de desempeño académico en Matemáticas de los estudiantes de nuevo ingreso? y ¿cuáles son las principales deficiencias en relación con las competencias matemáticas requeridas para la educación superior? Para abordar estas interrogantes, esta investigación tiene como objetivo diseñar y validar instrumentos de evaluación confiables mediante la teoría de medición de Rasch.

A través de una prueba estandarizada, se evaluará el desempeño matemático de los estudiantes, lo que facilitará la identificación de áreas de mejora en competencias clave. Los resultados de esta investigación permitirán optimizar los procesos de enseñanza, mediante la implementación de talleres diseñados para fortalecer las habilidades matemáticas esenciales, beneficiando así a la Universidad en términos de medición y evaluación del aprendizaje.

Diversos estudios han destacado la eficacia del modelo Rasch en el diseño de pruebas diagnósticas. Cupani y Cortes (2016) señalan la importancia del análisis psicométrico usando Rasch, lo cual facilita la validación del instrumento y mejora la precisión de la medición. En esta misma línea, Backhoff, Andrade, Balcázar y Téllez (2022) analizan sus ventajas y limitaciones en la calificación de respuestas y en la estimación del nivel de competencia estudiantil. Jiménez y Montero (2022) aplicaron el modelo en una prueba diagnóstica universitaria en Costa Rica, enfocándose en la evaluación de conocimientos matemáticos y en la identificación de factores que influyen en el bajo rendimiento académico. Además, el Instituto Colombiano de Evaluación (Icfes, 2023) ha desarrollado una guía detallada para el análisis, calibración y validación de pruebas basadas en la teoría de respuesta al ítem, destacando así el uso del modelo Rasch en contextos educativos.

El modelo Rasch se ha consolidado como una herramienta clave en la evaluación educativa por su capacidad de generar mediciones precisas y objetivas en pruebas estandarizadas. No obstante, en la República Dominicana, su implementación en el ámbito universitario sigue siendo limitada, con escasa presencia en la investigación académica. Este estudio busca reducir esta brecha, contribuyendo al diseño y validación de instrumentos de evaluación, fortaleciendo el análisis psicométrico y promoviendo el desarrollo de esta metodología en el país.

Referentes Teóricos

Fundamentos del modelo de Rasch dicotómico

La base teórica del análisis de Rasch se enfoca en la medición objetiva de una única habilidad o rasgo latente a través de modelos probabilísticos, los cuales garantizan la invarianza del instrumento y la comparabilidad de las puntuaciones de los examinados. Desde la teoría de la respuesta al ítem (TRI), el rendimiento de los estudiantes puede evaluarse mediante modelos psicométricos que permiten estimar tanto la dificultad de los ítems como la habilidad de los estudiantes (Wright & Stone, 1998). Este enfoque se sustenta en cinco principios fundamentales: aditividad de los puntajes, unidimensionalidad del instrumento, invarianza de la medida, distribución de la dificultad de los ítems, estimación de la habilidad de los examinados.

El modelo de Rasch dicotómico parte de la premisa de que un individuo con mayor habilidad tiene más probabilidades de responder correctamente un ítem en comparación con alguien de menor habilidad. De igual manera, si un ítem es más difícil que otro, cualquier sujeto tendrá una mayor probabilidad de responder correctamente el ítem más sencillo (Rasch, 1960). La objetividad del modelo Rasch radica en su formulación matemática, basada exclusivamente en la dificultad del ítem y la habilidad del individuo, eliminando la influencia de factores externos como la muestra utilizada en la calibración o la estructura del instrumento de evaluación.

En el modelo de Rasch dicotómico la probabilidad de respuesta de un sujeto a un ítem se modela mediante una función logística del nivel de dificultad y la habilidad del sujeto:

$P_i\{X_i = 1\} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$. En esta expresión matemática, P_i representa la probabilidad de que la persona n con habilidad β_n conteste correctamente el ítem con nivel de dificultad δ_i .

La Figura 1 representa gráficamente esta relación para tres ítems de distinta dificultad. En el eje horizontal se mide la habilidad de las personas (sujetos) y en el vertical la probabilidad de respuesta correcta. Para un sujeto con habilidad media ($x = 0$), la probabilidad de responder correctamente el ítem C es 0.08, para el ítem B es 0.50 y para el ítem A es 0.90, lo que confirma que el ítem C es el más difícil, seguido por el ítem B, mientras que el ítem A es el más fácil.

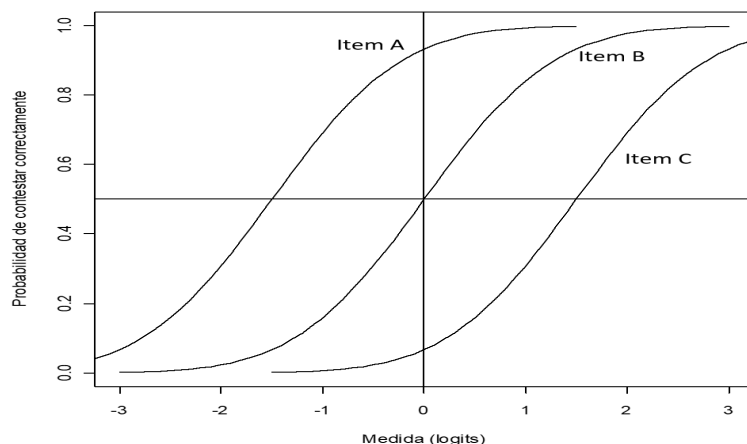


Figura 1. Curvas características de ítems

El resultado del análisis de Rasch es una escala, en unidades logit, que queda definida por la posición de los ítems y en la cual se pueden ubicar las personas (sujetos) de acuerdo con su nivel de habilidad (véase la Figura 2).

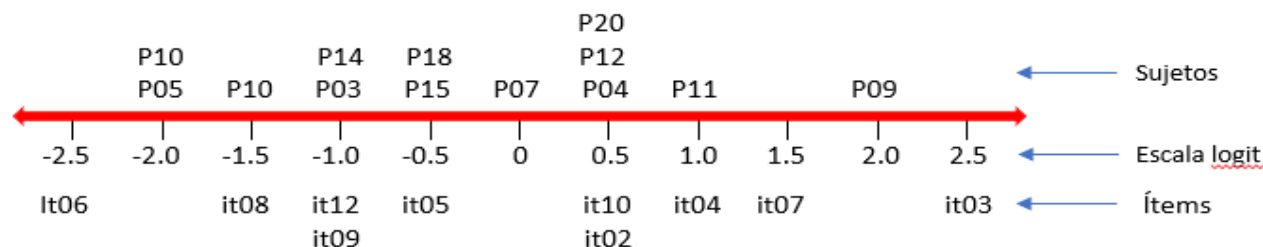


Figura 2. Mapa ítem-personas(sujetos) típico.

En esta escala, tanto la dificultad de los ítems como la habilidad de las personas (sujetos) aumentan de izquierda a derecha. Así, el sujeto P09 tiene la habilidad más alta, mientras que el ítem it03 es el más difícil. Dado que la escala es lineal, la diferencia en habilidad entre P05 y P03 es la misma que entre P11 y P09, lo que indica que la escala es de intervalos.

El mapa de ítems-personas permite construir niveles de desempeño, ya que los ítems se ordenan de menor a mayor dificultad y las personas (sujetos) de menor a mayor habilidad. Generalmente, se identifican tres categorías: personas (sujetos) con baja habilidad y múltiples carencias, personas con habilidad media y con alto desempeño en la prueba.

Elementos teóricos y conceptuales de la medición

Entre los principales conceptos considerados en este estudio se encuentra el análisis de Rasch, entendido como un conjunto de procedimientos que permiten evaluar la validez y confiabilidad de una prueba psicométrica (Bond & Fox, 2015). La validez se refiere al grado en que una prueba mide e informa con precisión sobre el constructo de interés (Messick, 1989), mientras que la invarianza se define como la propiedad de la escala que asegura su independencia respecto a la muestra utilizada y a los individuos evaluados (Wright & Stone, 1979).

Otro concepto fundamental es el de variable latente, que alude a un rasgo o habilidad no observable de forma directa, pero que puede inferirse a partir de las respuestas proporcionadas por los examinados a los ítems del instrumento (Embretson & Reise, 2000). En cuanto a los índices de ajuste, el infit expresa la relación entre los valores observados y esperados dentro del rango de habilidad del examinado, permitiendo valorar la consistencia interna de las respuestas (Linacre, 2002), mientras que el outfit capta esa relación fuera del rango esperado, siendo útil para detectar patrones atípicos o inconsistentes (Smith, 2004).

Finalmente, las correlaciones biserialles puntuales permiten estimar el grado de asociación entre los ítems y el nivel de habilidad de los examinados, y se emplean como criterio para valorar la capacidad discriminativa de cada ítem dentro de una prueba educativa (De Ayala, 2009).

Metodología

Este estudio adopta un enfoque cuantitativo, no experimental y descriptivo, utilizando el modelo de Rasch para diseñar instrumentos válidos y confiables en el diagnóstico de Matemática.

Participantes

La muestra estuvo conformada por 325 estudiantes de nuevo ingreso de la Pontificia Universidad Católica Madre y Maestra (PUCMM), campus Santo Domingo, durante el mes de septiembre de 2024.

Diseño del instrumento

La prueba diagnóstica fue elaborada por el equipo de profesores de Precálculo, considerando los conocimientos previos esenciales para el nivel universitario. Se evaluaron tópicos como conjuntos numéricos, operaciones fundamentales, potencias y raíces, simplificación de radicales, expresiones algebraicas, cálculo de porcentajes, factorización y ecuaciones de primer grado. La prueba incluyó preguntas de dificultad básica, intermedia y avanzada, con un tiempo de resolución de dos horas lectivas. La prueba consta de 33 ítems: 25 de opción múltiple y 8 de verdadero o falso.

Procesamiento de datos

Los datos fueron recolectados a través de una prueba integrada en la plataforma Moodle (PVA PUCMM). Se realizó la tabulación de los datos de las respuestas de los estudiantes en Excel, se estructuraron y se llevaron a cabo análisis descriptivos y univariados, incluyendo cantidad de respuestas correctas, proporción de respuestas correctas por opción, omisiones y multimarcas. Finalmente, se realizó un análisis de Rasch que incluyó análisis de ítems, análisis de personas y medición conjunta.

Resultados

Es fundamental considerar estos resultados para futuras versiones de la prueba, ya que establecen la línea base del diagnóstico en Matemática de los estudiantes de nuevo ingreso a la Universidad.

Fase 1. Preprocesamiento

El análisis se llevó a cabo utilizando el software Jamovi, aplicando el modelo dicotómico de Rasch para estimar la habilidad de los estudiantes y la dificultad de los ítems. En el análisis de confiabilidad, no se eliminaron ítems, ya que todos presentaron un buen comportamiento. Sin embargo, en el análisis de los ítems, se eliminaron R21dic, R24dic y R9.4 debido a baja correlación biserial (< 0.2) y desajustes en *infit* y *outfit*. La versión final de la prueba es de 30 ítems.

En el análisis de las personas, se identificaron dos estudiantes (38 y 108) con habilidades superiores al nivel de la prueba. Además, cinco estudiantes mostraron desajustes en infit y outfit, lo que indica un patrón de respuesta inconsistente, posiblemente aleatoria y no ajuste al modelo.

Análisis de ítems. El análisis de cada ítem incluyó la estimación de la media, el nivel de dificultad, las correlaciones biseriales puntuales, así como los índices de ajuste Infit y Outfit. A continuación, se presenta la Tabla 1 como ejemplo representativo del procedimiento aplicado.

Tabla 1
Estadísticas de ítems

	Media	Dificultad	S.E.	Infit	Outfit	Correlación biserial
R25dic	0.204	1.3757	0.149	1.007	1.062	0.337

Fuente: Elaboración propia. 2024.

A modo de ejemplo, la Figura 3 muestra las respuestas al ítem R25dic, considerado ítem difícil, cuya respuesta correcta es la opción D (línea roja). Este ítem solo es respondido correctamente por estudiantes con alta habilidad. Así mismo, se presenta su curva característica, donde las probabilidades de respuesta en estudiantes con baja habilidad son cercanas a 0, mientras que en los de mayor habilidad son altas.

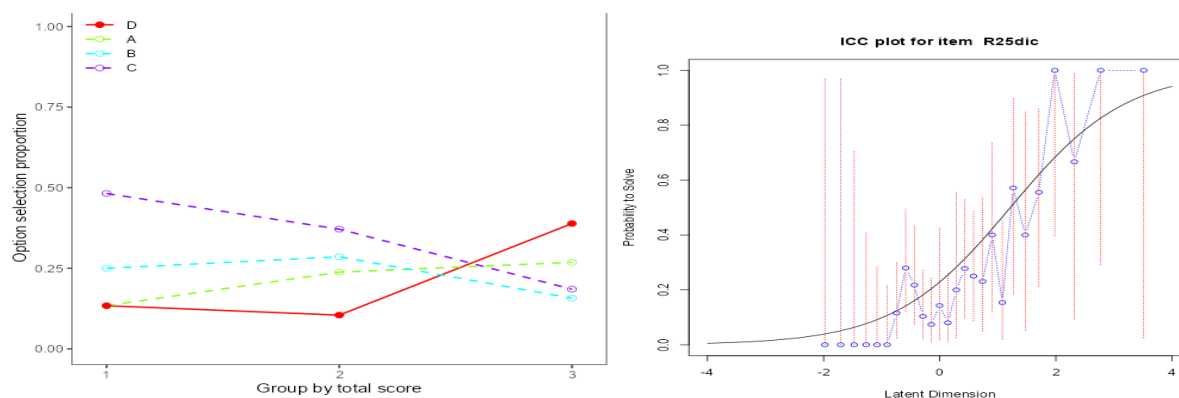


Figura 3. Proporciones de respuesta y curva característica ítem R25dic

Análisis de las personas. La Figura 4 muestra las medidas de ajuste (infit y outfit) de los examinados dentro del intervalo aceptable (0.5, 1.5). Se identificaron cinco estudiantes con desajustes en el modelo, lo que indica patrones de respuesta atípicos.

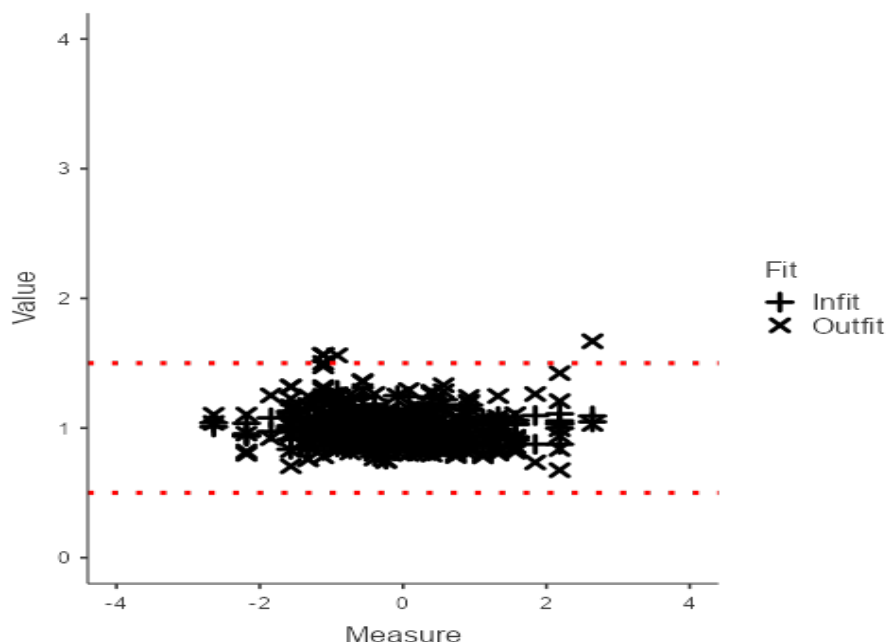


Figura 4. Gráfico de infit y outfit de personas

Fase 2. Procesamiento

Tras los ajustes de la fase 1, se realizó el análisis definitivo. Se generó un gráfico ítems-personas para visualizar la distribución a lo largo del continuo de medición. Se analizaron las correlaciones residuales de los ítems para validar la unidimensionalidad del instrumento. Finalmente, se revisaron las estadísticas de ajuste (infit y outfit) en ítems y personas para identificar patrones de respuesta ordenados. En la tabla 2, se pueden observar índices de confiabilidad y separación cercanas 0.8, además de la chi-cuadrada con p-valor de 0.21 que evidencien buen ajuste de los datos.

Tabla 2
Estadísticas de ajustes

Alpha Cronbach	Chi-Cuadrada	Índice de Separación
0.77	35.84 (30gl,p=0.21)	0.776

Fuente: Elaboración propia. 2024.

La Figura 5 muestra el mapa ítems-personas, que representa la relación entre la dificultad de los ítems y las habilidades de los estudiantes. Se observa que ambas distribuciones están centradas, cercanas a 0 logit con una desviación estándar de 1, y siguen una distribución aproximadamente normal. El histograma representa las personas y la parte de abajo del histograma se puede apreciar los ítems.

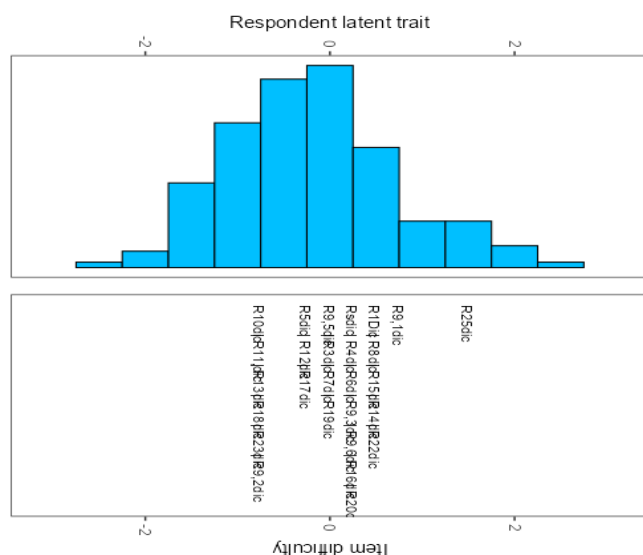


Figura 5. Mapa ítems-persona(sujeto)

Discusión y Conclusiones

Para evaluar las habilidades de los estudiantes con precisión, es fundamental contar con instrumentos válidos, confiables e invariantes (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2018; Wilson, 2005; Wright & Stone, 1998). En este estudio, el modelo de Rasch demostró su eficacia para medir la habilidad matemática de los estudiantes de nuevo ingreso, proporcionando datos ajustados a un modelo robusto de medición (Andrich & Marais, 2019; Wright & Masters, 1982).

El alfa de Cronbach de 0.77 y el índice de separación de 0.776 indican una alta confiabilidad del instrumento, permitiendo diferenciar de manera efectiva los niveles de habilidad de los estudiantes (Icfes, 2023). La eliminación de ítems con baja correlación biserial mejoró la precisión de la prueba, garantizando que cada ítem contribuya significativamente a la medición del constructo (Cupani & Cortez Cipsi, 2016; Liu, 2020).

Los resultados mostraron que los ítems y las habilidades de los estudiantes se distribuyen de manera adecuada en un continuo de medición entre (-3,3) logits, alineándose con los principios del modelo de Rasch (Soriano Rodríguez, 2014). Además, la prueba permitió diferenciar con claridad a los estudiantes de baja y alta habilidad, así como identificar los ítems más fáciles y difíciles dentro del intervalo (-0.83,1.38) logits. Este hallazgo respalda la utilidad del modelo para separar efectivamente los niveles de desempeño en Matemáticas (Wright & Stone, 1998).

Los resultados también revelaron que el 55% de los estudiantes presenta habilidades bajas en Matemáticas, y dentro de este grupo, el 15% no logró responder preguntas sencillas, como resolver operaciones aritméticas combinadas (ítem 11), ecuaciones lineales básicas (ítem 18) o problemas simples de tanteo o sistemas de ecuaciones (ítem 23).

Los hallazgos obtenidos en este estudio confirman que el modelo de Rasch constituye una herramienta eficaz para evaluar las habilidades matemáticas en estudiantes de nuevo ingreso, ya que la alta confiabilidad y precisión del instrumento garantizan datos consistentes, válidos y

relevantes para realizar análisis educativos fundamentados; asimismo, el adecuado ajuste de los ítems y la distribución apropiada de los niveles de habilidad permiten clasificar con precisión a los estudiantes e identificar oportunamente a aquellos con dificultades significativas en Matemáticas, lo que evidencia su validez psicométrica para futuras investigaciones orientadas al monitoreo continuo y a la mejora del desempeño matemático en contextos universitarios, destacando la necesidad de fortalecer la enseñanza y aprendizaje de la Matemática en la educación preuniversitaria.

Referencias y bibliografía

- Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social, and health sciences*. Springer Nature Singapore.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas* (M. Lieve, Trad.). American Educational Research Association. (Obra original publicada en 2014).
- Backhoff, E., Andrade, E., Balcázar, A., & Téllez, R. (2022). Uso del modelo de crédito parcial de Rasch y Masters en la evaluación de competencias matemáticas. REICE. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 20(1), 41-55. <https://doi.org/10.15366/reice2022.20.1.003>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- Cupani, M., & Cortez Cipsi, F. (2016). Análisis psicométrico del subtest de razonamiento numérico utilizando el modelo de Rasch. *Revista Psicología*, 25(2), 1-16. <https://www.revistapsicologia.uchile.cl>
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Icfes. (2023). *Aspectos metodológicos sobre calibración y calificación de las pruebas*. Subdirección de estadísticas. Bogotá, Colombia.
- Jiménez, K., & Montero, E. (2022). Aplicación del modelo de Rasch en el análisis psicométrico de una prueba de diagnóstico en matemática. *Revista Educación y Ciencia*, 27(1), 85–104. <https://doi.org/10.19053/01207105.v27.n1.2022.13267>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Liu, X. (2020). *Use and development of measurement instruments in science education: A Rasch modeling approach* (2nd ed.). Information Age Publishing.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). American Council on Education/Macmillan.
- Ministerio de Educación de la República Dominicana. (2020). *Marco de referencia para las Pruebas Nacionales del Segundo Ciclo de Educación Secundaria*. MINERD.
- Ministerio de Educación de la República Dominicana. (2022). *Adecuación curricular: Nivel Secundario*. Viceministerio de Servicios Técnicos y Pedagógicos, Dirección General de Currículo.
- Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Danish Institute for Educational Research.
- Smith, R. M. (2004). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 5(2), 199-218.
- Soriano Rodríguez, A. (2014). Diseño y validación de instrumentos de medición. *Editorial Universidad Don Bosco*, 8(13), 19-40.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.
- Wright, B. D., & Stone, M. H. (1998). *Diseño de mejores pruebas utilizando la técnica de Rasch*. CENEVA.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.