

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo



MAESTRÍA

“Aplicación de modelos de diagnóstico cognitivo en pruebas de admisión a la universidad”

Tesis
Que para obtener el grado de
Maestra en Ciencias Educativas

Presenta

Edna Gisela Navarro Hernández

Ensenada, Baja California, México

Febrero, 2025



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas



**“Aplicación de modelos de diagnóstico cognitivo en
pruebas de admisión a la universidad”**

Tesis
Que para obtener el grado de
Maestra en Ciencias Educativas

Presenta
Edna Gisela Navarro Hernández

Aprobado por:

Dr. Carlos David Díaz López
Director de tesis

**Dr. Juan Carlos Rodríguez
Macías**
Sinodal

Dr. Joaquín Caso Niebla
Sinodal

Agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnologías (Conahcyt) por el apoyo económico brindado, el cual contribuyó al logro de este proyecto de investigación y a mi formación como investigadora.

Agradecimientos

A mi mamá, por apoyarme en la construcción de la vida que siempre he querido. A mi hermano y cuñada, por su acompañamiento y apoyo al logro de mis metas. A Sammy, por estar a mi lado en mis noches de desvelo desde mis inicios en la psicometría y por adaptarse a los cambios de ciudad mejor que yo. Y a Mila, por llegar en medio de este proceso y complementar nuestra familia.

A todas las personas que me han acompañado en este proceso y han contribuido a la investigadora que aspiro ser:

A Leslie Vázquez, por motivarme a salir de mi zona de confort, por reconocer y celebrar mis logros, y por apoyarme en cada etapa del camino. A Alfonso Medina, por apoyar mi decisión de mudarme a otra ciudad para estudiar a pesar de lo que eso implicaba, pero especialmente por apoyarme cuando ya no quería hacerlo. Gracias por todas tus enseñanzas y por confiar en mí. A Josué Mendoza, por enseñarme estadística y R, por compartir tu pasión por la psicometría y, sobre todo, por presentarme a los modelos de diagnóstico cognitivo e inspirar esta tesis. A Rocío Gutiérrez y su familia, por su apoyo incondicional y cariño a lo largo de estos dos años. Sin ustedes, mi paso por Ensenada hubiera sido completamente diferente.

Al Dr. Ramsés Vázquez y a los miembros del *Computational Psychometrics lab* de la facultad de Psicología de la UNAM, por ayudarme a comprender mejor los modelos de diagnóstico cognitivo y por el apoyo que sigo recibiendo hasta hoy.

A los miembros de mis comités de expertos, por su disposición y valiosas aportaciones al desarrollo de esta investigación. A mis sinodales, el Dr. Juan Carlos Rodríguez, por hacer que mi paso por la maestría valiera totalmente la pena, y el Dr. Joaquín Caso, por contribuir significativamente a esta investigación, incluso a la distancia. Por último, a mi director de tesis, el Dr. Carlos Díaz, por su infinita paciencia, por brindarme la libertad de tomar mis propias decisiones y enseñarme a asumir la responsabilidad de ellas.

Resumen

Los exámenes de conocimientos y habilidades forman parte del proceso de admisión de aspirantes a la universidad. Aunque su propósito principal es la selección de aspirantes, Atkinson y Geiser (2009) y Sánchez (2020) sugieren que la información obtenida por las pruebas de admisión puede utilizarse con fines diagnósticos o formativos. Los modelos de diagnóstico cognitivo (MDC) se originaron en el campo de la medición educativa como una herramienta psicométrica para proporcionar información más específica sobre los aspirantes, y a partir de esta implementar estrategias de apoyo a su aprendizaje (de la Torre y Sorrel, 2023). El proyecto de investigación tuvo como propósito evaluar la aplicación de MDC a una prueba de matemáticas para el ingreso a la Universidad Autónoma de Baja California. Para ello, se propuso un modelo cognitivo subyacente a la prueba mediante comités de expertos, quienes determinaron la relación de siete atributos cognitivos a la prueba. Posteriormente, se aplicaron los modelos DINA, DINO y G-DINA y se evaluó su ajuste y capacidad diagnóstica. Como principal hallazgo, se encontró que los MDC permiten una representación más completa de los conocimientos y habilidades de los estudiantes que los modelos psicométricos tradicionales (la teoría clásica de los tests y la teoría de respuesta al ítem). Como consecuencia, los MDC pueden contribuir a optimizar tanto el proceso de selección como el acompañamiento de los aspirantes a lo largo de su proceso de formación universitaria.

Palabras clave: modelos de diagnóstico cognitivo, teoría clásica de los tests, teoría de respuesta al ítem, pruebas de admisión a la universidad, medición educativa

Índice

CAPÍTULO I. Planteamiento del problema	1
Introducción	1
Antecedentes	3
Evaluación del aprendizaje	3
Evaluaciones diagnósticas cognitivas	5
Modelos de diagnóstico cognitivo	7
Las pruebas de acceso a la universidad	10
Evaluaciones diagnósticas al ingreso a la universidad	12
Planteamiento del problema	14
Objetivos	18
Justificación	19
CAPÍTULO II. Marco teórico	21
Modelos psicométricos	21
Teoría clásica de los tests	21
Confiabilidad en la TCT	23
Análisis de los ítems en la TCT	24
Teoría de respuesta al ítem	25
Modelo de Rasch	27
Modelo logístico de dos parámetros	27
Modelo logístico de tres parámetros	28
Modelos de diagnóstico cognitivo	28
Modelos no compensatorios	30
Modelo DINA	30
Modelo NIDA	32

Modelo NC-RUM	34
Modelos compensatorios	37
Modelo DINO	38
Modelo NIDO	39
Modelo C-RUM	40
Modelos generalizados	41
Modelos de diagnóstico cognitivo en la admisión a la universidad	42
CAPÍTULO III. Método	48
Fase 1. Información sobre la prueba y los aspirantes	49
Objetivo de la fase	49
Instrumento y participantes	49
Análisis de datos	49
Fase 2. Especificación de atributos cognitivos	51
Etapas 1. Identificación de atributos cognitivos	51
Etapas 2. Validación de atributos cognitivos	53
Etapas 3. Determinación de la matriz Q	55
Fase 3. Aplicación de modelos de diagnóstico cognitivo	56
Etapas 1. Ajuste de los modelos	56
Etapas 2. Comparación de los modelos	56
Fase 4. Interpretación y comparación de resultados	58
Objetivos de la fase	58
Procedimiento	58
CAPÍTULO IV. Resultados	60
Evaluación de la calidad técnica de la prueba	60
Resultados bajo el enfoque de la teoría clásica de los tests	60
Resultados bajo en enfoque de la teoría de respuesta al ítem	63

Aplicación de modelos de diagnóstico cognitivo	67
Construcción y validación de la matriz Q	67
Ajuste de los modelos	72
Precisión en la clasificación de aspirantes	74
Estimación de los parámetros de los ítems de la prueba	77
Interpretación y comparación de resultados bajo los tres enfoques psicométricos .	81
Nivel de dominio de los atributos cognitivos en la prueba	81
Desempeño de los aspirantes por enfoque psicométrico	83
CAPÍTULO V. Discusión y conclusiones	91
Discusión de hallazgos	91
Limitaciones del estudio	93
Futuras líneas de investigación	94
Referencias	95
Anexos	108
Anexo A. Proceso de identificación de atributos cognitivos	108
Anexo B. Materiales para la etapa 1 del comité de desarrollo y validación	114
Anexo C. Materiales para la etapa 2 del comité de desarrollo y validación	119
Anexo D. Materiales para el comité de determinación de la matriz Q	121
Anexo E. Análisis de ítems con el enfoque de la teoría clásica de los tests	130
Anexo F. Análisis de ítems con el enfoque de la teoría de respuesta al ítem	131
Anexo G. Tabla de atributos y descriptores	132
Anexo H. Parámetros de adivinación y desliz de los ítems de la prueba	134
Índice de tablas	
Tabla 1 Ejemplo de matriz Q de aritmética básica	29

Tabla 2	Ejemplo de estimaciones de parámetros del modelo DINA	32
Tabla 3	Ejemplo de estimación de parámetros del modelo NIDA	34
Tabla 4	Número de ítems de la prueba de matemáticas del ExIES por sección	49
Tabla 5	Criterios de calificación para la validación de contenido de descriptores	54
Tabla 6	Criterios de los estadísticos de ajuste del modelo	58
Tabla 7	Frecuencia de selección de los atributos cognitivos en los ítems de la prueba de matemáticas del ExIES	69
Tabla 8	Atributos cognitivos subyacentes a la prueba de matemáticas del ExIES	71
Tabla 9	Matriz Q de la prueba de matemáticas del ExIES	71
Tabla 10	Índices de ajuste y comparación de modelos de diagnóstico cognitivo	72
Tabla 11	Comparación del PVAF de los ítems entre modelos de diagnóstico cognitivo	74
Tabla 12	Índices de precisión y consistencia de la prueba y los atributos con los modelos DINA, DINO y G-DINA	75
Tabla 13	Clases latentes identificadas con más de 1 % de aspirantes, por modelo de diagnóstico cognitivo DINA, DINO y G-DINA	77
Tabla 14	Prevalencia de los aspirantes en los atributos cognitivos	81
Tabla 15	Relación de los parámetros de los aspirantes por enfoque	84

Índice de figuras

Figura 1	Proceso de retrofitting de modelos de diagnóstico cognitivo	48
Figura 2	Índices de dificultad y discriminación de los ítems bajo el enfoque de la teoría clásica de los tests	62
Figura 3	Distribución de los puntajes totales de los aspirantes	63
Figura 4	Curvas características de los ítems de la prueba de matemáticas del ExIES	65

Figura 5	Nivel de habilidad matemática de los aspirantes e información de la prueba	66
Figura 6	Parámetros de adivinación de los modelos DINA, DINO y G-DINA .	78
Figura 7	Parámetros de desliz de los modelos DINA, DINO y G-DINA	79
Figura 8	Red de correlaciones y prevalencias de los atributos cognitivos subyacentes a la prueba de matemáticas del ExIES	83
Figura 9	Relación entre los atributos cognitivos dominados por los aspirantes y su nivel de habilidad	85
Figura 10	Relación entre el nivel de habilidad y los patrones de atributos de los aspirantes	86
Figura 11	Relación entre el puntaje y los perfiles de atributos de los aspirantes por nivel de habilidad	88
Figura 12	Aspirantes con el mismo puntaje en la prueba (15 aciertos)	90
Figura 13	Aspirantes con el mismo nivel de habilidad ($\theta = 0.82$)	90

CAPÍTULO I. Planteamiento del problema

Introducción

En el ámbito de la educación superior, el proceso de admisión de aspirantes a la universidad incluye la aplicación de exámenes de conocimientos y habilidades que, bajo los enfoques psicométricos clásicos, permiten realizar inferencias sobre el nivel de habilidad que tienen los estudiantes para dominar los contenidos medidos en las pruebas. Ello permite ubicar y ordenar a los aspirantes en función de su nivel de habilidad o establecer puntos de corte para tomar decisiones de selección. Por su naturaleza, las pruebas de admisión a la universidad no tienen como finalidad principal proporcionar información diagnóstica sobre los aspirantes admitidos. Sin embargo, distintos autores sugieren que los resultados de las pruebas de admisión, además de proporcionar información para seleccionar a los aspirantes, también pueden usarse con finalidades diagnósticas o formativas (Atkinson y Geiser, 2009; Sánchez, 2020).

La identificación de las fortalezas y áreas de oportunidad de los estudiantes de nuevo ingreso a la universidad proporciona información que puede contribuir a la implementación de estrategias remediales para los estudiantes que requieren de atención educativa. Para este propósito, los modelos de diagnóstico cognitivo (MDC) son modelos psicométricos que proporcionan información detallada, acerca del nivel de dominio de los estudiantes en distintos atributos cognitivos, para la implementación de estrategias de apoyo al aprendizaje de los estudiantes (de la Torre y Sorrel, 2023).

Los MDC, desarrollados hacia finales de la década de los 80, son modelos multidimensionales de variables latentes categóricas que se utilizan principalmente en evaluaciones formativas (de la Torre et al., 2016). Su implementación en la evaluación del aprendizaje da cuenta de las fortalezas y debilidades que tienen los estudiantes a nivel cognitivo en áreas específicas del aprendizaje, dado que permiten identificar la presencia o ausencia de múltiples atributos cognitivos necesarios para dominar los contenidos evaluados (Lee et al., 2012). Los MDC se sustentan bajo el enfoque de las evaluaciones diagnósticas

cognitivas (EDC), las cuales reconocen y hacen explícitos los procesos y atributos cognitivos que utiliza un estudiante para responder ítems de una prueba de conocimientos y habilidades (Leighton y Gierl, 2007).

La aplicación de MDC en las pruebas de admisión a la universidad tiene el potencial de proporcionar información complementaria al proceso de admisión, tanto para la toma de decisiones de selección, como para coadyuvar al proceso de aprendizaje de los aspirantes admitidos. El presente proyecto de tesis buscó examinar la aplicación de modelos de diagnóstico cognitivo a la prueba de matemáticas del examen de ingreso a la educación superior (ExIES) aplicado en el proceso de admisión de la Universidad Autónoma de Baja California (UABC) del año 2018. Como resultado, se validó la aplicación de MDC en pruebas de admisión a la universidad y se identificó la capacidad de estos modelos para proporcionar información complementaria a los tomadores de decisión para fines diagnósticos y formativos.

Antecedentes

Evaluación del aprendizaje

La evaluación educativa ha evolucionado desde sus inicios en el siglo XIX hasta la actualidad, por lo que su conceptualización no es una actividad puntual ni estática (Madaus y Stufflebeam, 2002; Phillips, 2018). Ralph W. Tyler, considerado el padre de la evaluación educativa, acuñó el término como el proceso sistemático de medir el grado en el que se han alcanzado los objetivos y metas establecidas (Madaus y Stufflebeam, 2002). Por su parte, Cronbach (1963) definió la evaluación educativa como un proceso de recopilación y uso de información para orientar la toma de decisiones, las cuales pueden dirigirse hacia la mejora del curso, los estudiantes, los docentes o los procesos administrativos. Asimismo, Scriven (1967) definió la evaluación como un proceso que permite determinar el mérito o valor de algo.

Dentro de la evaluación educativa, la evaluación del aprendizaje es un proceso sistemático de recolección de información para promover el aprendizaje en los estudiantes (Sánchez, 2020). Esta evaluación implica emitir juicios sobre el logro de los estudiantes en un área particular, siendo estos un potencial de mejora de su desempeño académico (Clarke y Madaus, 2003). Para reunir evidencias acerca del aprendizaje de los estudiantes, se utilizan diversos instrumentos cuyos resultados permiten tomar decisiones sobre su aprendizaje o sobre la forma en que se lleva a cabo el proceso de enseñanza (Clarke y Madaus, 2003; Instituto Nacional para la Evaluación de la Educación [INEE], 2011).

Las evaluaciones del aprendizaje pueden clasificarse, en función de su objetivo y momento de implementación, en evaluaciones sumativas, evaluaciones formativas y evaluaciones diagnósticas (Jornet et al., 2009; Sánchez, 2020). La evaluación diagnóstica por lo general se realiza al inicio de un curso o actividad, con la intención de describir, predecir o determinar el estado actual de conocimientos, habilidades o actitudes de los estudiantes para así proveer de información al docente que le permita hacer adecuaciones en el contenido o implementación de las actividades académicas programadas (Sánchez,

2020; Vera, 2020). En el contexto de la educación superior, son ejemplo de este tipo de evaluaciones las que lleva a cabo la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia de la Universidad Nacional Autónoma de México (CUAIEED-UNAM), que elabora anualmente una evaluación diagnóstica a los estudiantes de nuevo ingreso de todas sus licenciaturas. La CUAIEED diseña y analiza exámenes de conocimientos específicos por cada área del conocimiento y tiene la finalidad de informar a los profesores y estudiantes acerca de las áreas de oportunidad de los estudiantes en asignaturas o temas específicos (UNAM, 2023).

La evaluación formativa se define como un proceso que puede aplicarse a lo largo del curso, permitiendo realizar valoraciones continuas, brindar retroalimentación y identificar fortalezas y áreas de oportunidad en los conocimientos, habilidades o actitudes de los estudiantes (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2021; Ramos et al., 2009). Dado que la evaluación formativa se caracteriza por proporcionar retroalimentación continua, su principal escenario de aplicación es el aula (Popham, 2013). Black et al. (2003) identificaron cuatro estrategias en las que se puede impulsar el proceso formativo en el aula: hacer preguntas, retroalimentación, autoevaluación, coevaluación y preparación para pruebas sumativas.

La evaluación sumativa, por su parte, tiene como propósito determinar el valor de lo que se evalúa, medir los efectos de lo evaluado, su calidad, y en general, valorar el cumplimiento de los objetivos de la evaluación (Ramos et al., 2009). En evaluaciones dentro del aula, la evaluación sumativa representa el conjunto de valoraciones efectuadas durante el curso, para determinar si los objetivos de la enseñanza se alcanzaron y así asignar una calificación final (Sánchez y Delgado, 2022). También, las evaluaciones sumativas suelen emplearse para tomar decisiones sobre el progreso del estudiante o la eficacia de un programa, institución o sistema educativo (Jornet et al., 2009). Ejemplo de estas evaluaciones son las que se realizan para seleccionar aspirantes a estudiar una especialidad médica en México, a través del Examen Nacional para Aspirantes a Residencias Médicas

(ENARM). A partir de los resultados de esta prueba, los médicos aspirantes a una especialidad médica son seleccionados y, a su vez, estos mismos resultados dan cuenta de la calidad educativa de las instituciones educativas donde se formaron (Roblero Ochoa et al., 2023).

Diversos autores proponen una complementariedad entre las evaluaciones sumativas y las evaluaciones formativas (Bennett, 2011; Black et al., 2003; Taras, 2009). Bennett (2011) planteó que las evaluaciones sumativas y formativas cumplen propósitos primarios y secundarios: una evaluación sumativa debe cumplir su propósito primario que consiste en documentar lo que los estudiantes saben y pueden hacer al término de un periodo determinado. Sin embargo, los resultados de esta evaluación también deben cumplir un propósito secundario como el apoyo a la mejora del aprendizaje. De manera similar, Taras (2009) concibió a las evaluaciones formativas y a las evaluaciones sumativas no solo como funciones y productos de evaluación, sino también como procesos que pueden ser implícitos o explícitos. Por ejemplo, la autora argumentó que una evaluación no puede ser únicamente formativa, sino que la evaluación sumativa puede estar implícita al tomar decisiones sobre los resultados de aprendizaje que se requieren mejorar.

En esta misma línea, las pruebas sumativas pueden generar usos formativos a lo largo del proceso de aprendizaje. Durante la preparación para la prueba, el estudiante puede identificar sus fortalezas y áreas de oportunidad con respecto al contenido y, por otro lado, el docente puede centrarse en las áreas débiles de los estudiantes para proveer retroalimentación y mejorar su aprendizaje (Black et al., 2003; Taras, 2009; Lau, 2016). Esto indica que tanto el estudiante como el docente participan en el proceso de aprendizaje con la finalidad de alcanzar una calificación final.

Evaluaciones diagnósticas cognitivas

Durante la década de los 80, se planteó la necesidad por combinar la ciencia cognitiva y la psicometría con el propósito de utilizar los resultados de las evaluaciones para informar sobre los procesos cognitivos de los estudiantes y mejorar la enseñanza y el

aprendizaje (Huff y Goodman, 2007; Nichols, 1994). Al respecto, Snow (1980) planteó que el logro de los estudiantes tiene una función tanto organizativa como de adquisición,

El rendimiento escolar ya no debe entenderse simplemente como la acumulación de hechos y habilidades específicas de un contenido... [Más bien] una parte significativa de la tarea del alumno es continuamente reunir, reensamblar, estructurar y sintonizar el cuerpo de conocimientos acumulado en sistemas funcionales para su uso en el pensamiento y en el aprendizaje posterior. (pp. 42-43)

Por lo anterior, para dar cuenta de la capacidad de los estudiantes en la resolución de problemas, las evaluaciones pueden integrar los conocimientos y las habilidades de procesamiento de los estudiantes para identificar sus fortalezas y debilidades cognitivas (Gierl et al., 2008). Bajo este enfoque, los objetivos de aprendizaje deben inculcar en los estudiantes un conjunto de estructuras cognitivas que sustenten la comprensión de los conocimientos, habilidades y contenidos educativos que coadyuven a la resolución de problemas en el mundo real (Haertel y Calfee, 1983).

Nichols (1994) definió el término de evaluaciones diagnósticas cognitivas (EDC) al desarrollo de evaluaciones que combinan las aportaciones de la ciencia cognitiva y la psicometría, con la finalidad de proporcionar resultados diagnósticos de los estudiantes. Las EDC reconocen y hacen explícitos los procesos y estructuras cognitivas que usa un examinado para responder una tarea, como una prueba de conocimientos (Leighton y Gierl, 2007; Nichols, 1994).

Para emplear las EDC es necesario un marco interpretativo acerca de los procesos y estructuras cognitivas que usan los estudiantes para resolver los ítems de una prueba. Los modelos cognitivos proporcionan el marco de referencia necesario para vincular las inferencias basadas en las habilidades cognitivas de los examinados con las interpretaciones específicas y detalladas de las puntuaciones de una prueba (Gierl et al., 2008). En el contexto de evaluaciones del aprendizaje, un modelo cognitivo se refiere a una descripción

simplificada del proceso de resolución de tareas o ítems de una prueba de conocimientos o habilidades (Leighton y Gierl, 2007).

Los modelos cognitivos se generan a partir del estudio del conocimiento, las habilidades y las estrategias empleadas por los estudiantes cuando estos responden a los ítems de una prueba, así como por métodos de informe verbal (como los protocolos de pensamiento en voz alta o las entrevistas retrospectivas), los cuales permiten identificar los procesos de pensamiento cognitivo de los estudiantes (Gierl et al., 2008). Sin embargo, se ha señalado la ausencia de teorías cognitivas en Educación que informen sobre los procesos y habilidades cognitivas subyacentes al desempeño de los estudiantes (Leighton y Gierl, 2007). A pesar de esto, diversos estudios han identificado los atributos cognitivos necesarios para responder correctamente los ítems de pruebas en distintas áreas de conocimiento o habilidades como las matemáticas y la comprensión lectora. A este respecto, Gierl et al. (2009) identificaron y validaron cuatro modelos cognitivos que integran los atributos cognitivos que utilizan los estudiantes para resolver ítems de la sección de Álgebra de la prueba SAT (College Board, 2023a) un examen utilizado principalmente en Estados Unidos y aceptado en instituciones de educación superior de varios países, como Canadá y el Reino Unido, como criterio de admisión a la universidad. De manera similar, Chin y Chew (2023) desarrollaron y validaron tres modelos cognitivos que integran los atributos cognitivos que utilizan los estudiantes de quinto grado de primaria al realizar operaciones de sustracción con unidades de tiempo. Como resultado, presentaron recomendaciones a los profesores para mejorar la enseñanza de estas operaciones, particularmente aquellas que implican cálculos con años y siglos.

Modelos de diagnóstico cognitivo

Los modelos de diagnóstico cognitivo (MDC) son modelos multidimensionales que identifican los atributos cognitivos que subyacen a los niveles de competencia o habilidad de los estudiantes. Los MDC se originaron en el campo de la medición educativa como una herramienta psicométrica para proporcionar información más detallada y adecuada para la

evaluación formativa. Haertel (1989) introdujo un modelo restrictivo de clases latentes para clasificar a las personas con respecto a si poseían o no un conjunto de variables latentes, llamados atributos cognitivos (p. e. comprensión de oraciones); y donde las clases latentes fueron tratadas como variables categóricas. Este modelo es considerado el primero de la familia de los MDC (Liu y Shi, 2020). Los MDC también son llamados modelos psicométricos cognitivos, modelos de clasificación diagnóstica o modelos de respuesta latente (Rupp et al., 2010).

En su planteamiento, los MDC incorporan algunas de las características de otros enfoques y teorías tales como la teoría clásica de los tests (TCT), la teoría de respuesta al ítem (TRI), el análisis factorial, los modelos de ecuaciones estructurales, el análisis de datos categóricos, y la estadística bayesiana (Rupp et al., 2010). El propósito de los MDC es clasificar a los examinados en perfiles de dominio de atributos cognitivos para proporcionar una descripción detallada y precisa de las fortalezas y debilidades subyacentes al desempeño de los estudiantes en tareas o evaluaciones específicas (Rupp et al., 2010). Por lo anterior, los MDC han cobrado relevancia en su aplicación a evaluaciones formativas (de la Torre et al., 2016), principalmente en pruebas de matemáticas, lectura y escritura de educación básica y media superior. También se han realizado aplicaciones de estos modelos en exámenes de competencia del inglés a gran escala, como la Batería de evaluación del idioma inglés de Michigan (MELAB) y la Prueba de Inglés como Lengua Extranjera (TOEFL) (Kim, 2014; Lee y Sawaki, 2009a; Ravand, 2016; Wang, 2023; Xie, 2016). Al respecto, You et al. (2019) presentaron los hallazgos de un Sistema de Diagnóstico de Aprendizaje del idioma Chino (CLDS, por sus siglas en inglés) en educación básica, donde implementaron CDA para proporcionar informes de retroalimentación para las escuelas, los maestros, los padres y los estudiantes con respecto al perfil de atributos cognitivos de estos últimos. Con el perfil de atributos cognitivos proporcionado por el CLDS, los estudiantes obtuvieron retroalimentación inmediata sobre su desempeño en clases, lo que les permitió trabajar en aspectos específicos para mejorar su aprendizaje durante el curso. Asimismo,

los maestros redujeron su tiempo de clase grupal entre un tercio y la mitad, y se enfocaron en instrucciones más detalladas e individualizadas a los estudiantes que, de acuerdo a su perfil de atributos cognitivos, presentaban mayores áreas de oportunidad (You et al., 2019).

Aunque con menor frecuencia, los MDC también han sido implementados en la educación superior. Mei y Chen (2022) realizaron un estudio en una universidad de estudios extranjeros en China, cuyo objetivo fue identificar los atributos cognitivos que subyacen a la traducción del idioma Chino al inglés. Por medio de jueces expertos, identificaron los atributos cognitivos principales involucrados en la traducción, tales como la transmisión de información clave, reproducción de estilos, adaptación a los hábitos del lenguaje, entre otros. Como resultado, los autores identificaron que los atributos cognitivos que subyacen a la competencia en traducción difieren entre grupos cuyos resultados en la prueba los clasifican con alta o baja competencia. La integración de la CDA permitió obtener una aproximación a las habilidades a reforzar durante la enseñanza y aprendizaje de la traducción (Mei y Chen, 2022).

Por otro lado, se han realizado aplicaciones de MDC a pruebas educativas a gran escala desarrolladas bajo otros modelos psicométricos, como la TRI o TCT. El propósito de estas aplicaciones es generar resultados para fines diagnósticos o formativos con respecto a los atributos cognitivos que subyacen a dichas pruebas. Las pruebas que se han analizado con los MDC son el informe del Programa Internacional para la Evaluación de Estudiantes (PISA, por sus siglas en inglés) que examina los conocimientos y habilidades de estudiantes de 15 años de edad en las áreas de lectura, matemáticas y ciencias (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2019); el Estudio Internacional para el Progreso de la Comprensión Lectora (PIRLS, por sus siglas en inglés) que evalúa la comprensión lectora de estudiantes de cuarto grado de primaria (Mullis et al., 2023); y el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS), que evalúa los conocimientos de estudiantes de cuarto y octavo grado en ciencias y matemáticas (Mullis et al., 2020). No obstante, el alcance de los estudios ha sido exploratorio con el objetivo de

examinar el potencial de la información obtenida a través de los MDC, pero no se reporta el uso de los resultados para la toma de decisiones o para mejorar el aprendizaje de los estudiantes (Hu et al., 2021; Toprak-Yildiz, 2021; Wu et al., 2021).

También se ha tenido interés en comparar resultados obtenidos bajo los enfoques de la TRI y los MDC para obtener información complementaria sobre el aprendizaje de los estudiantes. Ma et al. (2020) analizaron una prueba de razonamiento proporcional aplicada a estudiantes de secundaria, donde encontraron que estudiantes con un mayor nivel de habilidad medidos desde el enfoque de la TRI tienden a dominar un mayor número de atributos identificados por los MDC. También se identificó que estudiantes con el mismo nivel de habilidad pueden tener distintos patrones de dominio de atributos cognitivos, y viceversa. Estos hallazgos sugieren que los MDC ofrecen resultados más específicos respecto a las diferencias individuales de los estudiantes a nivel cognitivo.

Las pruebas de acceso a la universidad

El ingreso a la educación superior es un proceso que involucra una serie de requisitos que orientan a las instituciones educativas a tomar decisiones de selección. Estos requisitos pueden considerarse de carácter sumativo y de altas consecuencias tanto para los aspirantes como para las propias instituciones de educación superior (IES). En México, cada IES determina sus criterios de selección de acuerdo con su normatividad, criterios técnicos y disponibilidad de recursos (Sánchez et al., 2020). Por lo general, los criterios de selección consideran los logros y aptitudes de los aspirantes, como lo son el promedio de bachillerato, pruebas de personalidad y aptitudes, exámenes de conocimientos y habilidades, entrevistas, ensayos, certificación de dominio del idioma inglés u otros idiomas, entre otros (Sánchez et al., 2020; Trost, 1993). No obstante, durante el proceso de selección se asume que el ingreso de los aspirantes dependerá únicamente de su nivel de conocimientos y habilidades medidos a través de pruebas o exámenes objetivos (Buendía y Rivera, 2010).

Los exámenes de ingreso son el instrumento de medición principal para la valoración de los conocimientos y habilidades de los aspirantes. Jiménez-Moreno (2021) identificó que

las 35 universidades públicas estatales (UPE) mexicanas hicieron uso de al menos un examen de conocimientos y habilidades como herramienta para la selección de aspirantes a nivel licenciatura. Al respecto, se identificó que el 70 % de las UPE utilizó el Examen Nacional de Ingreso a licenciatura (EXANI-II), desarrollado por el Centro Nacional de Evaluación para la Educación Superior (Ceneval). Mientras que el 11 % de las UPE aplicó el Examen de Competencias Básicas (EXCOBA) desarrollado por Métrica Educativa. El 14 % de las UPE utilizó instrumentos desarrollados por la misma institución, y finalmente, el 5 % utilizaron la Prueba de Aptitud Académica (PAA) desarrollada por el College Board (Jiménez-Moreno, 2021).

Los exámenes utilizados por las UPE difieren principalmente en términos de los contenidos, conocimientos y habilidades que evalúan, además del formato de administración (lápiz y papel, a computadora, entre otros). El EXANI-II mide las habilidades básicas que los estudiantes desarrollan a lo largo de su formación previa. El examen se divide en cinco módulos que evalúan las habilidades y conocimientos en comprensión lectora, redacción indirecta y pensamiento matemático, así como los módulos que evalúan conocimientos específicos relacionados con la carrera a la que aspiran ingresar los estudiantes (Ceneval, 2023). Por su parte, el EXCOBA mide las competencias académicas esenciales de español, matemáticas, ciencias naturales y ciencias sociales de educación básica, así como conocimientos específicos de educación media superior afines a la carrera elegida por el aspirante (Métrica Educativa, 2020a). Finalmente, la PAA mide el conocimiento adquirido por los estudiantes a lo largo de su vida académica en tres áreas: razonamiento verbal y lectura crítica; razonamiento lógico-matemático y la redacción en español (College Board, 2020).

Los resultados de los exámenes de admisión tienen principalmente un uso sumativo que permite tomar decisiones respecto a la selección de aspirantes a partir del nivel de conocimientos, habilidades o competencias que cada prueba evalúa. Sin embargo, distintos autores coinciden en que estos resultados también pueden tener una finalidad diagnóstica o

formativa si se comunica la información a los estudiantes y docentes. Esto facilitaría identificar el nivel de conocimientos y habilidades de los estudiantes al ingresar a un nuevo nivel educativo, permitiendo así la implementación de estrategias orientadas a mejorar el aprendizaje (Atkinson y Geiser, 2009; Sánchez, 2020).

Al respecto, Atkinson y Geiser (2009) señalaron la importancia de proporcionar a cada aspirante información diagnóstica acerca de sus fortalezas y áreas de oportunidad identificadas durante el proceso de admisión, además de su puntaje o posición en una escala determinada por la misma prueba. Los autores señalan que brindar información diagnóstica a los aspirantes permite que trabajen en sus áreas de oportunidad y lleguen mejor preparados al siguiente nivel educativo.

Evaluaciones diagnósticas al ingreso a la universidad

Evaluar a los estudiantes al ingresar a la universidad es una práctica clave para identificar sus fortalezas, áreas de oportunidad y nivel de preparación académica. Las evaluaciones diagnósticas no solo orientan a los estudiantes en su transición hacia la educación superior, sino que también permiten a las instituciones educativas fundamentar su toma de decisiones para ajustar los programas y planes de estudio. En distintos contextos, estas prácticas han evolucionado para incluir sistemas y programas diseñados por las propias universidades, así como evaluaciones externas desarrolladas por organizaciones especializadas.

Como ejemplo de estas prácticas, se ha identificado que en México las IES que cuentan con sistemas de diagnóstico de estudiantes de nuevo ingreso incorporan los resultados de las pruebas de ingreso a la universidad. Tal es el caso del programa Corexani-Educación media superior de la Universidad Veracruzana. Este programa informa el promedio de cada plantel de bachillerato en cada módulo del examen de ingreso, así como el porcentaje de estudiantes con derecho a ingreso a la universidad. El programa permite la comparación de resultados de un plantel con el promedio global del examen, o bien, con los resultados de otras escuelas. Otro programa desarrollado por la misma

universidad es Corexani- UV, que permite identificar las fortalezas y áreas de oportunidad de los aspirantes y comparar los resultados de los puntajes del examen de ingreso con respecto a un programa, área académica o del total de estudiantes que ingresaron a la Universidad Veracruzana (Universidad Veracruzana, 2007).

De manera similar, la Universidad Nacional Autónoma de México (UNAM) desde 1995 diseña y aplica exámenes diagnósticos del desempeño de estudiantes de primer ingreso en las asignaturas de matemáticas, física, química, biología, historia universal, historia de México, geografía, literatura y filosofía. Los resultados son utilizados para la mejora constante de los estudiantes, programas académicos y planes de estudio (UNAM, 2023). También se realizan evaluaciones externas a las IES por medio de instituciones privadas que diseñan exámenes diagnósticos de conocimientos y habilidades de los estudiantes de nuevo ingreso a la universidad. Un ejemplo de ello es Métrica Educativa (2020b), una organización mexicana sin fines de lucro que ofrece a las instituciones la aplicación de tres exámenes diagnósticos: inglés, competencias básicas en matemáticas y expresión escrita.

En un marco internacional y en línea con la demanda de proporcionar diagnósticos sobre los atributos cognitivos de los estudiantes (Haertel y Calfee, 1983; Huff y Goodman, 2007; Nichols, 1994), el College Board (2023a) adaptó la Prueba Preliminar de Calificación para la Beca de Mérito Nacional/Scholastic Aptitude Test (PSAT/NMSQT, por sus siglas en inglés). La prueba tiene la finalidad de medir los conocimientos y habilidades que los estudiantes han adquirido durante el bachillerato y que son necesarias para tener éxito en la universidad. Los resultados permiten a los estudiantes obtener oportunidades de beca para la universidad y a su vez, sirven como diagnóstico para prepararse para la prueba SAT que se aplica para el proceso de admisión a la universidad (College Board, 2023b). Los desarrolladores del PSAT/NMSQT, a través de la aplicación de las CDA, identificaron los atributos cognitivos necesarios para resolver cada ítem de la prueba, lo que permitió crear perfiles de atributos cognitivos por examinado y generar reportes para los estudiantes sobre sus áreas de oportunidad en la prueba y sobre cómo mejorarlas (Huff y Goodman, 2007).

Planteamiento del problema

Las pruebas de admisión a la universidad mexicanas evalúan los conocimientos y habilidades básicas necesarias para tener éxito en la vida personal y académica, entre las que se destacan las habilidades de lectura, escritura y matemáticas (Larrondo et al., 2018 y Zorrilla, 2015). Estas habilidades son esenciales, ya que el dominio insuficiente de competencias como la comprensión de textos, la escritura o el razonamiento matemático puede limitar las oportunidades educativas y laborales de los estudiantes, perpetuando su marginación y exclusión social (Zorrilla, 2015). Además, Cortés y Palomar (2008) destacan que los mecanismos de selección de aspirantes, en el nivel superior, son fundamentales para garantizar la calidad académica de los egresados. Por ello, subrayan la importancia de aplicar instrumentos técnicamente sólidos que proporcionen resultados confiables y válidos, permitiendo evaluar de manera precisa el desempeño de los aspirantes y apoyar decisiones de admisión basadas en criterios objetivos.

En el caso de la Universidad Autónoma de Baja California (UABC), cada año oferta dos convocatorias para el ingreso a la licenciatura, en las cuales aplica el Examen de Ingreso a la Educación Superior (ExIES), instrumento desarrollado por la misma institución (Universidad Autónoma de Baja California [UABC], 2023a). El ExIES mide la capacidad de aplicación de conocimientos y habilidades que han adquirido los aspirantes durante su formación básica y media superior, las cuales les permitirán resolver las demandas de una formación a nivel superior (UABC, 2023b). La prueba se estructura en tres componentes: 1) Lectura, que evalúa la capacidad para leer y comprender textos literarios e informativos; 2) Lengua Escrita, que evalúa la capacidad para revisar y editar textos con contenidos académicos, así como medir la capacidad para expresar ideas en apego a las convenciones del español escrito; y 3) Matemáticas, que evalúa la capacidad para la aplicación, manejo y comprensión de conceptos matemáticos, y la habilidad para la resolución de problemas e interpretación de datos, tablas, planos, figuras y gráficos (UABC, 2023b). Los resultados que proporciona el ExIES dan cuenta del estado actual de

conocimientos y habilidades de los aspirantes en las habilidades básicas descritas.

De manera particular, la tendencia del desempeño de los estudiantes mexicanos en matemáticas ha ido a la baja. El Programa para la Evaluación Internacional de Estudiantes (PISA) – que evalúa los conocimientos y habilidades de estudiantes de 15 años, en matemáticas, lectura y ciencias – publicó los resultados de su evaluación en México del año 2022 (OCDE, 2023). Se encontró que los estudiantes obtuvieron puntuaciones inferiores con respecto a la evaluación del 2018 en matemáticas y ciencias, mientras que en lectura las puntuaciones se mantuvieron similares. Por su parte, Osuna y Díaz (2019) realizaron un análisis histórico de la competencia matemática de adolescentes mexicanos desde el año 2003 al 2015 de los resultados obtenidos por las pruebas PISA, y las pruebas nacionales mexicanas Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE) y el Plan Nacional para la Evaluación de los Aprendizajes (PLANEA). Los resultados obtenidos en las evaluaciones de PISA en todos los años mostraron un puntaje menor a la media establecida por la OCDE. En concreto, las autoras apuntan a que solo el 3 % de los estudiantes obtuvo un nivel de competencia aceptable, y un 38 % de los estudiantes no reflejaron las competencias matemáticas más básicas. Respecto a las pruebas ENLACE y PLANEA, los resultados que se obtuvieron del 2003 al 2015 tuvieron una tendencia similar a los de PISA.

De acuerdo con Feng (2023) los conocimientos matemáticos adquiridos en la educación básica y media superior desempeñan un papel fundamental en la comprensión y aprendizaje de los contenidos sobre matemáticas que se imparten en la universidad. Por su parte, McNaught y Hoyne (2011) argumentan que algunas carreras, como las de las áreas de ciencia, tecnología, ingeniería y matemáticas, tienen como requisito un cierto nivel en matemáticas para garantizar el éxito académico de los estudiantes. Incluso, en áreas en donde las matemáticas no se consideran un requisito evidente, como en ciencias sociales y administrativas, se requieren habilidades matemáticas y estadísticas para completar con éxito la universidad (Stenberg et al., 2010).

De manera consistente con las evaluaciones realizadas en bachillerato, evaluaciones realizadas a estudiantes de nuevo ingreso a la universidad también han evidenciado un bajo desempeño en matemáticas. Por ejemplo, Torres y Campos (2015) identificaron bajo desempeño en resolución de tareas que reflejan las competencias de argumentación y modelación en estudiantes de nuevo ingreso en ingeniería, especialmente en ejercicios que combinaban pensamientos algebraico, numérico, geométrico y variacional. Sosa et al. (2014) encontraron un bajo desempeño en áreas como álgebra, geometría y precálculo en estudiantes de nuevo ingreso de distintas carreras, quienes obtuvieron, en promedio, de 28 % a 30 % de respuestas correctas en las áreas evaluadas. Por último, Sánchez-Luján y Moreno (2019) detectaron problemas en la aplicación de algoritmos de fracciones y en la comunicación del razonamiento matemático, especialmente en problemas de interpretación indirecta, aunque los estudiantes tuvieron un mejor desempeño en problemas matemáticos que incluyen contextos familiares y planteamientos directos.

El bajo desempeño de los estudiantes en competencias matemáticas identificadas desde la educación básica hasta el inicio de la educación superior plantea un reto significativo para las universidades. Derivado de ello, surge la necesidad de obtener diagnósticos más específicos de los estudiantes de nuevo ingreso para identificar sus debilidades cognitivas y crear estrategias pedagógicas. Las prácticas institucionales en materia de diagnóstico al ingreso a la educación superior no utilizan los resultados de los exámenes de ingreso a la universidad con fines diagnósticos, sino que consideran otros criterios involucrados en el proceso de admisión, como el promedio de bachillerato, datos sociodemográficos, o bien, se elaboran exámenes diagnósticos para identificar los conocimientos específicos de los estudiantes (Fernández et al., 2015; Organista-Sandoval et al., 2012; Torres-Zapata et al., 2019).

Diversos autores proponen que evaluaciones con fines sumativos – como los exámenes de admisión – tienen la capacidad de proporcionar información que puede emplearse con fines diagnósticos o formativos (Bennett, 2011; Black et al., 2003; Taras,

2019). Para ello, los *Estándares para las Pruebas Educativas y Psicológicas* establecen que "si el puntaje de una prueba se interpreta para un uso determinado de una manera que no ha sido validada, corresponde al usuario justificar la nueva interpretación para ese uso, proporcionando una razón fundamental y reuniendo nueva evidencia, si fuera necesario"(AERA et al., 2018, p. 26).

Los modelos de diagnóstico cognitivo (MDC) surgen como respuesta a la necesidad de determinar estrategias apropiadas para la retroalimentación oportuna de los estudiantes sobre sus fortalezas y debilidades (de la Torre y Sorrel, 2023). Su implementación en evaluaciones formativas, realizadas en educación básica y media superior, ha tenido un auge en las últimas décadas. No obstante, su implementación en contextos de admisión a la educación superior es un campo poco explorado. La implementación de los MDC implica el reconocimiento de los procesos o atributos cognitivos que expliquen el desempeño de los estudiantes. Bajo el enfoque del diagnóstico cognitivo, se pueden obtener diagnósticos del nivel de competencia de los estudiantes de nuevo ingreso, así como diagnósticos más detallados que vinculan los contenidos de las pruebas de conocimientos y habilidades con los atributos cognitivos involucrados.

La identificación oportuna de los atributos cognitivos que poseen los estudiantes tiene el potencial de mejorar no solo las decisiones de admisión, sino proporcionar una comprensión más completa del potencial académico de los aspirantes. A partir de la necesidad de proporcionar información diagnóstica cognitiva sobre los estudiantes de nuevo ingreso a la universidad, surgen las siguientes preguntas: ¿Cuáles son las diferencias en el dominio de las matemáticas de los aspirantes a ingresar a la UABC y cuáles son los atributos cognitivos que lo explican? y ¿En qué medida los modelos de diagnóstico cognitivo pueden proporcionar información complementaria con fines diagnósticos y formativos?

Objetivos

A partir de los planteamientos anteriores se propone el siguiente objetivo general:

Examinar la aplicación de modelos de diagnóstico cognitivo a la prueba de matemáticas del ExIES de la UABC con fines diagnósticos.

Para responder al objetivo general, se proponen los siguientes objetivos específicos:

1. Determinar la calidad técnica de la prueba de matemáticas del ExIES bajo la teoría clásica de los tests y la teoría de respuesta al ítem.
2. Determinar el perfil de atributos cognitivos que subyace al nivel de dominio de las matemáticas de los aspirantes a ingresar a la UABC.
3. Describir el perfil de atributos cognitivos con respecto al nivel de habilidad de los aspirantes.

Justificación

En México, las pruebas de admisión son instrumentos clave para la selección de estudiantes, especialmente en las universidades públicas, donde la oferta educativa enfrenta una alta demanda de aspirantes. En el caso de la UABC, el Examen de Ingreso a la Educación Superior (ExIES) se ha consolidado como una herramienta fundamental para evaluar las competencias académicas de los aspirantes y determinar su ingreso a la institución. Sin embargo, a pesar de su importancia, este tipo de pruebas se han centrado tradicionalmente en enfoques psicométricos que priorizan puntajes totales o niveles de habilidad continua, dejando de lado información más detallada sobre los atributos cognitivos subyacentes al desempeño de los estudiantes.

El presente proyecto de tesis aborda esta limitación mediante la aplicación de modelos de diagnóstico cognitivo (MDC) a la prueba de matemáticas del ExIES. Los MDC, al permitir una descomposición más granular del desempeño de los estudiantes, ofrecen perfiles de atributos cognitivos que identifican fortalezas y áreas de oportunidad específicas.

Aunque este enfoque ha tenido un desarrollo significativo en evaluaciones diagnósticas y formativas en Educación Básica, su aplicación en pruebas de admisión a nivel universitario es aún limitada. Estudios empíricos previos se han centrado en aspectos teóricos y metodológicos de los MDC, pero pocos han explorado su implementación en contextos de admisión. Este estudio contribuye al conocimiento al evaluar la calidad técnica de una prueba existente bajo este enfoque, y propone lineamientos que pueden ser replicados en contextos similares. Esto no solo amplía la comprensión sobre la aplicabilidad de los MDC, sino que también refuerza su valor como una herramienta que combina evaluaciones diagnósticas y sumativas.

Además, este proyecto no solo contribuirá al desarrollo teórico de los MDC en este contexto, sino que también proporcionará evidencia empírica sobre su viabilidad y utilidad. En el caso específico de la UABC, se espera que los resultados permitan robustecer el ExIES mediante la generación de perfiles de atributos cognitivos, lo que beneficiará tanto a

los desarrolladores de la prueba como a los estudiantes y docentes. Por un lado, se ofrecerán lineamientos para mejorar las especificaciones de nuevos ítems y asegurar la calidad técnica del instrumento; por otro, los resultados diagnósticos facilitarán la identificación de necesidades educativas al ingreso, ayudando a estudiantes y docentes a implementar y adaptar estrategias pedagógicas.

En última instancia, este proyecto tiene el potencial de transformar los usos de las evaluaciones de admisión al integrar resultados sumativos, diagnósticos y formativos. Esto contribuirá significativamente a los procesos de admisión de la UABC y podrá servir como modelo para otras instituciones de educación superior en México o para evaluaciones en contextos similares.

CAPÍTULO II. Marco teórico

Modelos psicométricos

El proceso del diseño y desarrollo de pruebas educativas estandarizadas está respaldado por diversas teorías que buscan garantizar su validez, confiabilidad y calidad de los ítems. La teoría clásica de los tests (TCT) y la teoría de respuesta al ítem (TRI) son los dos principales marcos de referencia utilizados en la medición educativa, los cuales se basan en distintos supuestos y enfoques estadísticos. Inicialmente, la TCT fue la teoría predominante en el desarrollo y análisis de pruebas estandarizadas; sin embargo, décadas más tarde, la TRI comenzó a aplicarse como un complemento que permitió una evaluación más detallada del comportamiento de los ítems y de las habilidades de los examinados (Awopeju y Afolabi, 2016). A continuación, se describen las teorías.

Teoría clásica de los tests

A inicios del siglo XX, Spearman (1904) desarrolló un modelo denominado modelo clásico o modelo lineal clásico en donde definió los conceptos de puntuación verdadera y error. A partir de las aportaciones de Spearman y subsecuentes desarrollos a éstas, se formó la teoría clásica de los tests (TCT), la cual se convirtió en el principal modelo de referencia para la construcción y evaluación de tests psicológicos (Abad et al., 2011). De acuerdo con Berg et al. (2018), en el contexto educativo, la TCT se emplea comúnmente durante la etapa inicial del desarrollo de instrumentos y en evaluaciones en el aula.

La TCT tiene como supuesto central que la puntuación observada de un examinado en una prueba corresponde a la suma de su puntuación verdadera, la cual se considera un valor hipotético que refleja la verdadera capacidad del examinado, y un componente de error aleatorio (AERA et al., 2018; Bond y Fox, 2001). El modelo se expresa en la siguiente ecuación general:

$$X = T + \epsilon \quad (1)$$

Donde X hace referencia a la puntuación observada de la prueba, T la puntuación verdadera, y ϵ el error de medición. La TCT parte del supuesto de que los efectos sistemáticos entre las respuestas de los examinados se deben únicamente a la variación en la habilidad medida por la prueba (p. ej. Competencia en matemáticas), mientras que las demás fuentes de variación son constantes en todas las puntuaciones de la muestra (van der Linden y Hambleton, 1997). Las fuentes de variación se deben a factores propios del examinado, de la construcción de la prueba o al contexto de aplicación.

El error estándar de medición, que se asume constante para todas las puntuaciones, puede calcularse para interpretar qué tan variable es la puntuación de un examinado dadas las distintas fuentes de variación (Geisinger et al., 2013). El error estándar de medición es la desviación estándar de los errores, la cual se expresa como sigue:

$$s_e = s_x \sqrt{1 - r_{xx}} \quad (2)$$

Adicionalmente, el modelo clásico tiene los siguientes supuestos (Desjardins y Bulut, 2018):

1. El valor esperado de la puntuación observada es la puntuación verdadera; $E(X) = T$.
2. La puntuación verdadera y el error de medición son independientes; $Cov(T, E) = 0$.
3. Los errores de medición en un test son independientes a los errores de medición de otro test; $Cov(E_1, E_2) = 0$.
4. Los errores de medición en un test son independientes de la puntuación verdadera en otro test; $Cov(E_1, T_2) = 0$.

Aunado a lo anterior, las propiedades de los ítems y de la prueba, como los índices de discriminación, dificultad y consistencia interna, dependen de la muestra de examinados y el número de ítems, lo que puede limitar la generalización de los resultados estimados (Desjardins y Bulut, 2018).

Confiabilidad en la TCT. La confiabilidad es la proporción de la varianza de la puntuación observada que puede atribuirse a la varianza de la puntuación real (Desjardins y Bulut, 2018). Abad et al. (2011) define la confiabilidad como un indicador de precisión en la medición. La ecuación general de la confiabilidad se expresa como sigue:

$$confiabilidad = \frac{\sigma_T^2}{\sigma_X^2} \quad (3)$$

Existen cuatro métodos para estimar la confiabilidad: test-retest (coeficiente de estabilidad), formas paralelas (coeficiente de equivalencia), formas alternativas (confiabilidad de formas alternativas) y consistencia interna (Desjardins y Bulut, 2018). A continuación se presentará el método de estimación de consistencia interna al ser el comúnmente utilizado para estimar la confiabilidad de pruebas.

La consistencia interna puede calcularse a través de diferentes métodos que examinan la similitud de las respuestas de los examinados en distintas partes de la misma prueba (Geisinger et al., 2013), de tal modo que la confiabilidad se define como “el grado en que diferentes subconjuntos de ítems covarían, correlacionan o son consistentes entre sí” (Abad et al., 2011, p. 97). Existen dos métodos para calcular la consistencia interna:

- **Método dos mitades (Spearman-Brown).** El método de dos mitades de Spearman-Brown expresa que la consistencia entre dos mitades de la prueba, es mayor que la correlación de Pearson entre ambas mitades. La fórmula de Spearman-Brown se expresa como sigue:

$$SB_{XX}^r = \frac{2r_{X_1X_p}}{1 + r_{X_1X_p}} \quad (4)$$

Los valores de la fórmula oscilan en un rango de 0 y 1 y se interpreta como la proporción de varianza de las puntuaciones de la prueba que es debida a las puntuaciones verdaderas (Abad et al., 2011).

-
- **Coefficiente alfa de Cronbach.** El coeficiente calcula la consistencia interna a través de la homogeneidad entre las puntuaciones de los ítems (Abad et al., 2011). Los ítems que se correlacionan altamente entre sí contribuyen más a la varianza de las puntuaciones de la prueba que los ítems que no se correlacionan altamente entre sí (Geisinger et al., 2013). La fórmula del alfa de Cronbach se calcula como sigue:

$$r_{xx} = \frac{k}{k-1} \left[1 - \frac{\sum s_i^2}{s_x^2} \right] \quad (5)$$

con k el número de puntuaciones de la prueba, $\sum s_i^2$ la suma de la varianza de los ítems y s_x^2 la varianza total de la prueba.

Análisis de los ítems en la TCT. El análisis de ítems tiene por objetivo el mejoramiento de una prueba a través de la revisión de la calidad de los ítems (Krishnan, 2013). A continuación, en los siguientes apartados se presenta el índice de dificultad y discriminación de los ítems.

Índice de dificultad. La dificultad del ítem en la TCT se define como la proporción de examinados que responden correctamente al ítem (Finch y French, 2019). El índice de dificultad se calcula como sigue:

$$p = \frac{n_c}{N} \quad (6)$$

Con n_c como el número de respuestas correctas y N el total de examinados. El índice de dificultad oscila entre 0 y 1, donde los valores cercanos a 0 indican ítems más difíciles y los valores cercanos a 1 indican ítems más fáciles (Finch y French, 2019).

La varianza del índice de dificultad se calcula como:

$$S^2 = p(1-p) \quad (7)$$

Y el error estándar de la dificultad se calcula de la siguiente manera:

$$SE_p = \sqrt{\frac{S^2}{N}} \quad (8)$$

Índice de discriminación. La discriminación hace referencia al grado en que un ítem distingue a los examinados con alto nivel de habilidad de aquellos con bajo nivel (French y Finch, 2019). La estimación de la discriminación puede ser a través del cálculo de la correlación biserial o punto biserial entre la respuesta al ítem y la puntuación total en la prueba (Finch y French, 2019).

El cálculo de la correlación punto biserial se calcula como sigue:

$$\rho_{PBIS} = \frac{(\bar{x}_c - \bar{x}_n)}{s} \sqrt{\frac{\frac{n_c}{N}}{1 - \frac{n_c}{N}}} \quad (9)$$

Con $\bar{x}_c$ como la media de las puntuaciones de los examinados que respondieron correctamente los ítems de la prueba, $\bar{x}_n$ como la media de las puntuaciones de todos los examinados y, s como la desviación estándar de las puntuaciones de todos los examinados.

Teoría de respuesta al ítem

Los cimientos de la teoría de respuesta al ítem (TRI) fueron expuestos a inicios del siglo XX, pero fue hasta la década de los cincuenta que tuvieron su mayor auge (van der Linden, 2016). La TRI constituye una familia muy extensa de modelos psicométricos, cada uno con sus características específicas. Estos modelos tienen en común que relacionan las características latentes de los ítems de una prueba (como la dificultad y la discriminación de los ítems) y las características de los examinados (su nivel de habilidad), con el fin de llegar a afirmaciones probabilísticas acerca de la respuesta de cada examinado en cada ítem (Leenen, 2013).

A diferencia de la TCT, la TRI plantea la hipótesis de que las respuestas a los ítems son manifestaciones observables de habilidades, rasgos o atributos humanos no observables, y además, se enfoca en el comportamiento de cada ítem de manera individual en lugar de la prueba en su totalidad (Bond y Fox, 2001). Cada modelo de la TRI tiene como principio

la estimación de parámetros para los efectos del nivel de habilidad (rasgo o aptitud) del examinado, así como para las propiedades de los ítems, lo que hace posible realizar inferencias tanto de los examinados como de los ítems de la prueba (van der Linden y Hambleton, 1997; van der Linden, 2016).

Los modelos de la TRI, de acuerdo con Embretson y Reise (2000), tienen los siguientes supuestos básicos: i) las curvas características del ítem (CCI) tienen una forma específica y ii) la independencia local de los ítems. La forma de una CCI indica el cambio en el nivel de habilidad y su relación con los cambios en la probabilidad de respuesta específica. Para ítems dicotómicos en donde la respuesta específica se considera “correcta”, la CCI indica la probabilidad de responder correctamente en un nivel de rasgo. Por otro lado, para ítems politómicos, la CCI hace una regresión de la probabilidad de respuestas en cada categoría a nivel del rasgo (Embretson y Reise, 2000).

La independencia local de los ítems hace referencia a que la probabilidad de responder correctamente un ítem no cambiará a la luz de información adicional sobre las respuestas en otros ítems (Leenen, 2013). Por ejemplo, si un examinado responde correctamente tres ítems consecutivos, ello no implica que acertará el cuarto ítem. Por último, otro supuesto tiene que ver con la cantidad de rasgos latentes necesarios para responder a un ítem (Attoresi et al., 2009; Embretson y Reise, 2000). Si se asume que la respuesta al ítem está en función de una única habilidad latente, los modelos son denominados unidimensionales (Attoresi et al., 2009). Por otro lado, existen modelos multidimensionales que asumen más de una habilidad latente involucrada en la respuesta a un ítem (Reckase, 1997).

La TRI tiene sus aplicaciones principalmente en exámenes a gran escala, dado que sus parámetros, a diferencia de la TCT, no dependen de la muestra, lo que permite la generalización de resultados. Los modelos de la TRI son comúnmente utilizados y han sido la base de muchas pruebas estandarizadas en medición educativa; por ejemplo, el Estudio Internacional de Matemáticas y Ciencias Naturales (TIMSS, por sus siglas en inglés) ha

utilizado modelos de la TRI para obtener estimaciones sobre el rendimiento de los estudiantes (Osterlind y Wang, 2018).

Los primeros modelos desarrollados fueron los modelos dicotómicos unidimensionales que son: el modelo de Rasch que estima los parámetros de dificultad de los ítems y el nivel de habilidad de los examinados (van der Linden, 2016); el modelo logístico de dos parámetros, que además de estimar los parámetros de dificultad de los ítems y el nivel de rasgo de los examinados, considera el parámetro de discriminación entre los ítems (Leenen, 2013); y el modelo logístico de tres parámetros, que añade un parámetro de pseudo-adivinación para tomar en cuenta el rendimiento distinto de cero de los examinados con bajo nivel de rasgo (van der Linden y Hambleton, 1997).

Modelo de Rasch. El modelo de Rasch utiliza las respuestas dicotómicas de n examinados y un conjunto fijo de k ítems que miden la misma habilidad latente θ (Fisher y Molenaar, 1995). El modelo asume que cada examinado tiene un parámetro θ_j que refleja su nivel de habilidad que es medida en los ítems; asimismo, cada ítem tiene un parámetro β_i que denota la dificultad de ese ítem (Fisher y Molenaar, 1995). La dificultad del ítem describe la cantidad de habilidad requerida para responder correctamente el ítem, por lo tanto, es una medida de posición del ítem en la escala de medida de la habilidad (Martínez-Arias et al., 2014).

El modelo de Rasch se expresa de la siguiente manera:

$$P(X = x|\theta, \beta) = \frac{\exp[x(\theta - \beta)]}{1 + \exp(\theta - \beta)} \quad (10)$$

Con β como un vector de k parámetros de los ítems y θ como un vector con n parámetros de la habilidad de los examinados y x como valores de 1 si el examinado responde correctamente el ítem y 0 si responde incorrectamente (Fisher y Molenaar, 1995).

Modelo logístico de dos parámetros. El modelo logístico de dos parámetros (2PL), al igual que el modelo de Rasch, considera el parámetro de habilidad de los examinados (θ) y el parámetro de dificultad de los ítems (β). El modelo 2PL incorpora un

parámetro de discriminación para cada ítem representado por a (Desjardins y Bulut, 2018). La discriminación del ítem representa la tasa de cambio en la probabilidad de éxito a medida que aumenta el nivel de habilidad (Martínez-Arias et al., 2014).

El modelo 2PL se expresa de la siguiente manera:

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - \beta)}} \quad (11)$$

Modelo logístico de tres parámetros. El modelo logístico de tres parámetros (3PL) es una extensión del modelo 2PL, añadiendo un parámetro de adivinación c que representa la probabilidad de éxito por azar. El valor del parámetro c se interpreta directamente como una probabilidad de responder correctamente el ítem (Martínez-Arias et al., 2014).

El modelo 3PL se expresa de la siguiente manera:

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - \beta)}} \quad (12)$$

Modelos de diagnóstico cognitivo

A diferencia de la TRI, que se centra en estimar una habilidad general de los examinados, los modelos de diagnóstico cognitivo (MDC) son modelos multidimensionales de variables latentes categóricas que tienen como objetivo principal la clasificación de los examinados en clases latentes dicotómicas o politómicas que reflejan la presencia o ausencia de habilidades o atributos cognitivos latentes en cada ítem (Luna, 2017; Ma et al., 2020; Rupp y Templin, 2008; Rupp et al., 2010). Los MDC tuvieron sus inicios en la década de los ochenta a partir del desarrollo de modelos psicométricos de clasificación con variables categóricas en lugar de variables continuas (Haertel, 1989).

Un componente esencial para la implementación de los MDC es el establecimiento *a priori* de la relación hipotética de cómo un conjunto de atributos cognitivos se vincula con el desempeño esperado de un examinado en una tarea (von Davier y Lee, 2019). En el

contexto de las evaluaciones diagnósticas cognitivas, se establece la relación de los atributos cognitivos con los ítems de una prueba. Esta relación se expresa en una matriz Q de especificaciones (Tatsuoka, 1985).

De manera formal, la matriz Q es una matriz de ceros y unos de tamaño $J \times K$, en donde el elemento de la j -ésima fila y la k -ésima columna, q_{jk} , indica si se requiere el atributo k para responder correctamente al ítem j (de la Torre, 2009). De acuerdo con Rupp et al. (2010) "la matriz Q es el componente por excelencia en cualquier MDC porque representa la operacionalización de la teoría sustantiva que ha dado lugar al diseño de la evaluación diagnóstica"(p. 49). En la tabla 1 se presenta un ejemplo de matriz Q con ítems de una prueba de aritmética básica. En la primera columna se presentan los ítems de la prueba, mientras que los atributos cognitivos se presentan en las columnas subsecuentes. Si se requiere el atributo para responder correctamente al ítem, se indica con un 1, y si no se requiere el atributo se indica con un 0. Se puede observar que el primer ítem de la prueba requiere de los atributos Suma y Resta, el segundo ítem requiere División, el tercer ítem requiere de Resta y Multiplicación y, finalmente, el último ítem requiere de Suma.

Tabla 1
Ejemplo de matriz Q de aritmética básica

Ítems	Suma	Resta	Multiplicación	División
$2 + 3 - 1 = ?$	1	1	0	0
$\frac{4}{2} = ?$	0	0	0	1
$5 \times 3 - 4 = ?$	0	1	1	0
$8 + 12 = ?$	1	0	0	0

Nota. Adaptado de Diagnostic Measurement (p. 116), por A. A. Rupp et al., 2010, The Guilford Press.

Los conjuntos de MDC pueden clasificarse en función del tipo de interacción que los atributos cognitivos tienen entre sí, en modelos compensatorios y modelos no compensatorios. Los modelos compensatorios suponen que es posible compensar el dominio en un atributo con el dominio en un atributo diferente (Rupp et al., 2010). Los modelos

compensatorios frecuentemente utilizados son: DINO (The deterministic input, noisy or gate model), NIDO (The noisy input deterministic or gate model) y C-RUM (The compensatory reparameterized unified model)(Rupp et al., 2010).

Por su parte, los modelos no compensatorios suponen que el dominio en un atributo no puede compensarse por el dominio en otro diferente (Rupp et al., 2010). Son ejemplo de estos modelos: DINA (The deterministic input, noisy and gate model), NIDA (The noisy input deterministic and gate model) y NC-RUM (The noncompensatory reparameterized unified model)(Rupp et al., 2010).

Modelos no compensatorios

Los modelos de diagnóstico cognitivo no compensatorios, también denominados conjuntivos, asumen que el bajo dominio en un atributo es independiente al alto dominio en otro (Rupp et al., 2010). Por lo tanto, el estudiante requiere dominar todos los atributos necesarios para resolver correctamente un ítem (von davier y Lee, 2019). Un ejemplo de estos modelos son DINA, NIDA y NC-RUM que se describen a continuación.

Modelo DINA. El modelo DINA (The deterministic-input, noisy-and-gate model) es un modelo con una regla de condensación conjuntiva, la cual indica que un examinado requiere dominar todos los atributos presentes en un ítem para responderlo correctamente (Rupp et al., 2010). Al ser un modelo conjuntivo, la habilidad de los examinados se reduce a dos clases de dominio por ítem: una clase con los examinados que dominan todos los atributos presentes en el ítem y una clase con los examinados que no dominan al menos uno de los atributos presentes en el ítem (von Davier y Lee, 2019). Ello implica que se trata de un modelo restrictivo que no distingue entre examinados que carecen de uno o más atributos presentes en el ítem (Rupp et al., 2010).

El modelo DINA considera X_{ij} la respuesta del examinado i al ítem j , con $i = 1, \dots, I$, $j = 1, \dots, J$, y $\alpha_i = \{\alpha_{ik}\}$ el vector de atributos binarios del examinado, $k = 1, \dots, K$, donde un 1 en el k -ésimo elemento denota la presencia o dominio del atributo k y 0, la ausencia o el no dominio del atributo (de la Torre, 2009).

El vector de atributos binarios del examinado $\alpha_i = \{\alpha_{ik}\}$ y la matriz Q producen un vector de respuesta latente $\eta_i = \{\eta_{ij}\}$, donde

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \quad (13)$$

El vector de respuesta latente es igual a 1 si el examinado posee todos los atributos requeridos para responder correctamente el ítem, y 0 si el examinado no domina al menos uno de los atributos requeridos (de la Torre, 2009).

El modelo DINA tiene dos parámetros para cada ítem, el de adivinación g_j y el de desliz s_j . El parámetro de adivinación es la probabilidad de que los examinados que no dominan al menos uno de los atributos presentes en el ítem adivinen la respuesta correcta. La adivinación en este contexto involucra no solo a obtener la respuesta correcta a pesar de no poseer los atributos necesarios, sino a utilizar otro tipo de atributos no considerados en la matriz Q (de la Torre y Douglas, 2004). Por su parte, el parámetro de desliz es la probabilidad de que los examinados que dominan todos los atributos presentes en el ítem se equivoquen y lo respondan incorrectamente (Rupp et al., 2010). Las probabilidades se expresan como sigue:

$$s_j = P(X_{ij} = 0 | \eta_{ij} = 1) \quad (14)$$

$$g_j = P(X_{ij} = 1 | \eta_{ij} = 0) \quad (15)$$

Por tanto, la probabilidad de que el examinado i con un vector de atributos α_i responda correctamente al ítem j está dada por

$$P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = g_j^{1-\eta_{ij}} (1 - s_j)^{\eta_{ij}} \quad (16)$$

En la tabla 2 se presenta un ejemplo de estimaciones de parámetros del modelo DINA para ítems de aritmética. Puede observarse que el ítem más informativo es el ítem

($8 + 12 = ?$) dado que las estimaciones de sus parámetros son las más bajas. En concreto, para un examinado que domina el atributo Suma hay 5 % de probabilidad de responder incorrectamente el ítem ($s_4 = 0.05$). Por su parte, para un examinado que no domina el atributo Suma hay un 2 % de probabilidad de que adivine la respuesta correcta ($g_4 = 0.02$).

Tabla 2

Ejemplo de estimaciones de parámetros del modelo DINA

Ítems	Suma	Resta	Multipliación	División	s_j	g_j
$2 + 3 - 1 = ?$	1	1	0	0	0.08 (0.01)	0.30 (0.02)
$\frac{4}{2} = ?$	0	0	0	1	0.05 (0.01)	0.12 (0.01)
$5 \times 3 - 4 = ?$	0	1	1	0	0.15 (0.02)	0.02 (0.01)
$8 + 12 = ?$	1	0	0	0	0.05 (0.01)	0.02 (0.01)

Nota. Error estándar entre paréntesis. Adaptado de Diagnostic Measurement (p. 119), por A. A. Rupp et al., 2010, The Guilford Press.

El modelo DINA proporciona un parámetro de desliz y uno de adivinación por ítem con restricciones de igualdad entre atributos, por lo que el número de parámetros del modelo está influenciado únicamente por el número de ítems (Rupp et al., 2010).

Asimismo, las probabilidades de los ítems en el modelo DINA tienen una restricción tal que $(1 - s_j) > g_j$, lo que implica que un examinado que domina todos los atributos presentes en el ítem tenga una mayor probabilidad de responder correctamente que un examinado que no domina al menos uno de los atributos presentes en el ítem (Rupp et al., 2010).

Modelo NIDA. El modelo NIDA (The noisy-input, deterministic-and-gate model), al igual que el modelo DINA, es un modelo no compensatorio con una regla de condensación conjuntiva. Sin embargo, contrario al modelo DINA, el modelo NIDA proporciona un parámetro de desliz y un parámetro de adivinación por cada atributo, por lo que el número de parámetros está influenciado únicamente por el número de atributos cognitivos presentes en la matriz Q (Rupp et al., 2010). El modelo supone una diferenciación entre las probabilidades de dominio de cada atributo presente en la matriz Q, por lo que representa una distinción más precisa entre examinados que no dominan al menos un atributo presente (Rupp et al., 2010).

Dado que los parámetros del modelo se encuentran a nivel del atributo, el parámetro de desliz indica la probabilidad de que el examinado aplique incorrectamente el atributo a pesar de dominarlo. Asimismo, el parámetro de adivinación indica la probabilidad de que el examinado aplique correctamente el atributo aunque no lo domine. Las probabilidades se expresan como sigue:

$$s_k = P(\eta_{ijk} = 0 \mid \alpha_i = 1) \quad (17)$$

$$g_k = P(\eta_{ijk} = 1 \mid \alpha_i = 0) \quad (18)$$

Por tanto, la probabilidad de que el examinado i con un vector de atributos α_i responda correctamente el ítem j está dada por

$$P(X_{ij} = 1 \mid \alpha_i) = \prod_{k=1}^K \left[(1 - s_k)^{\alpha_i} g_k^{1-\alpha_i} \right]^{q_{ik}} \quad (19)$$

En la tabla 3 se presenta un ejemplo de estimaciones de parámetros de atributos de aritmética. Puede observarse que el atributo de Multiplicación tiene los valores más altos en los parámetros de desliz y adivinación $s_3 = 0.29$; $g_3 = 0.26$. Estos valores indican que un examinado que domina el atributo de Multiplicación tiene un 29 % de tener un desliz y aplicar incorrectamente el atributo en el ítem. Asimismo, un examinado que no domina el atributo de Multiplicación tiene un 26 % de probabilidad de aplicar correctamente el atributo en el ítem. Estos valores pueden indicar una incorrecta especificación del atributo o bien, una incorrecta indexación del atributo entre los ítems de la matriz Q (Rupp et al., 2010).

Tabla 3*Ejemplo de estimación de parámetros del modelo NIDA*

Ítems	Suma	Resta	Multiplicación	División
$2 + 3 - 1 = ?$	1	1	0	0
$4/2 = ?$	0	0	0	1
$5 * 3 - 4 = ?$	0	1	1	0
$8 + 12 = ?$	1	0	0	0
<i>Parámetro de desliz</i>	$s_1 = 0.04 (0.02)$	$s_2 = 0.09 (0.01)$	$s_3 = 0.29 (0.02)$	$s_4 = 0.12 (0.04)$
<i>Parámetro de adivinación</i>	$g_1 = 0.15 (0.01)$	$g_2 = 0.36 (0.03)$	$g_3 = 0.26 (0.05)$	$g_4 = 0.09 (0.01)$

Nota. Error estándar entre paréntesis. Adaptado de Diagnostic Measurement (p. 123), por A. A. Rupp et al., 2010, The Guilford Press.

Modelo NC-RUM. El modelo NC-RUM (noncompensatory reparameterized unified model), es un modelo no compensatorio que estima los parámetros de desliz y adivinación a nivel de ítem como el modelo DINA, y también estima los parámetros de desliz y adivinación a nivel de atributo, como el modelo NIDA. El modelo NC-RUM asume que es posible que un atributo tenga distintos niveles de complejidad entre un ítem y otro y, por lo tanto, distintas probabilidades de responder correctamente el ítem (Rupp et al., 2010). Existen dos variantes del modelo NC-RUM, el modelo completo y el modelo reducido (Rupp et al., 2010).

El modelo NC-RUM reducido tiene, como primer componente, el parámetro base π_j^* el cual representa la probabilidad de responder correctamente al ítem dado que el examinado domina todos los atributos requeridos (Rupp et al., 2010). Dado que la probabilidad se define a nivel de ítem, el modelo estima el número de parámetros base según el número de ítems en la evaluación diagnóstica cognitiva. El segundo componente en el modelo es el parámetro de penalización r_{jk}^* para el ítem j y el atributo k , el cual es la penalización para un examinado que no ha dominado el atributo k y reduce la probabilidad de una respuesta correcta para el ítem j (Rupp et al., 2010).

Para definir las probabilidades de los parámetros base y penalización se definen, en primer lugar, las siguientes probabilidades:

$$\pi_{ij} = P(\eta_{ijk} = 1 \mid \alpha_k = 1) = (1 - s_{jk}) \quad (20)$$

$$r_{jk} = P(\eta_{ijk} = 1 \mid \alpha_k = 0) = g_{jk} \quad (21)$$

π_{ij} es la probabilidad de dominar el atributo k y aplicarlo correctamente al ítem j , la cual representa la probabilidad de no tener un desliz a nivel de atributo (Rupp et al., 2010). Por su parte, r_{jk} es la probabilidad de aplicar correctamente el atributo k al ítem j a pesar de no dominar el atributo, la cual representa la probabilidad de adivinar a nivel de atributo (Rupp et al., 2010).

De acuerdo con las probabilidades anteriores, se define el parámetro base como sigue:

$$\pi_j^* = \prod_{k=1}^K \pi_{jk}^{q_{jk}} \quad (22)$$

El parámetro base representa el producto de las probabilidades de aplicar correctamente cada atributo presente en la evaluación diagnóstica cognitiva (Rupp et al., 2010).

Por su parte, el parámetro de penalización se define como sigue:

$$r_{jk}^* = \frac{r_{jk}}{\pi_{jk}} \quad (23)$$

La penalización es la proporción entre adivinar y no tener un desliz en aplicar el atributo. r_{jk}^* toma valores entre 0 y 1, en donde valores bajos indican una disminución en la probabilidad de obtener una respuesta correcta en el ítem si no se domina el atributo, por lo que resulta ideal obtener valores cercanos a cero (Rupp et al., 2010). Se considera el parámetro de penalización como un parámetro de discriminación diagnóstica para el atributo k y el ítem j (Rupp et al., 2010).

Considerando todas las probabilidades definidas, la probabilidad de responder correctamente el ítem en el modelo NC-RUM reducido es:

$$\pi_{ij} = P(X_{ij} = 1 \mid \alpha_i) = \pi_j^* \prod_{k=1}^K r_{jk}^* (1 - \alpha_{ik})^{q_{ik}} \quad (24)$$

La fórmula anterior indica que cada examinado tiene la probabilidad inicial de π_j^* de responder correctamente al ítem j , asimismo, la fórmula indica la influencia de los atributos presentes en la evaluación diagnóstica cognitiva que no han sido dominados por los examinados (Rupp et al., 2010).

Para ejemplificar el modelo, Rupp et al. (2010) toma como referencia el siguiente ítem: $(2 + 3 - 1 = ?)$, y considera la probabilidad de responder correctamente al ítem dado que el examinado domina los atributos de Suma y Resta que es $\pi_j^* = 0.90$, la probabilidad de penalización por no dominar la Suma es de $r_{j1}^* = 0.20$ y la probabilidad de penalización por no dominar la Resta es de $r_{j2}^* = 0.40$. Por tanto, dado que el examinado que domina Suma y Resta tiene una probabilidad de 0.90 de responder correctamente el ítem, si domina únicamente la Suma, entonces tiene una probabilidad de $\pi_j^* * r_{j1}^* = 0.90 * 0.40 = 0.36$; Si domina únicamente la Resta, entonces tiene una probabilidad de $\pi_j^* * r_{j2}^* = 0.90 * 0.20 = 0.18$; y si no domina ninguno de los dos atributos (Suma y Resta), entonces tiene una probabilidad de $\pi_j^* * r_{j1}^* * r_{j2}^* = 0.90 * 0.40 * 0.20 = 0.07$.

Por su parte, el modelo NC-RUM completo añade un término de interacción latente que da cuenta de un desajuste del modelo proveniente de la falta de completitud de la matriz Q, y mejora las posibilidades de ajuste del modelo (Rupp et al., 2010). La matriz Q está incompleta si no se incluyen todos los atributos presentes en un ítem. El término de interacción latente se considera como un parámetro de penalización adicional a nivel de ítem, como se expresa como sigue:

$$\pi_{ij} = P(X_{ij} = 1 \mid \alpha_i, \eta_c) = \left[\pi_j^* \prod_{k=1}^K r_{jk}^* (1 - \alpha_{ik})^{q_{ik}} \right] P_{c_i}(\eta_c) \quad (25)$$

Con π_{ij} como la probabilidad del examinado de responder correctamente el ítem, X_{ij} como la respuesta observada del examinado en el ítem, q_{jk} como el indicador de la matriz Q que indica si el atributo k está presente en el ítem j , α_{ik} como el indicador de dominio del atributo, π_j^* como la probabilidad de responder correctamente al ítem j si se dominan todos los atributos, r_{jk}^* es la penalización a la probabilidad de responder correctamente al ítem j por no haber dominado el atributo k , η_c es la variable latente para todos los atributos no indicados en la matriz Q, c_i es el parámetro de facilidad para el ítem j para todos los atributos no indicados en la matriz Q y α_i es el vector de atributos dominados del examinado (Rupp et al., 2010).

La fórmula anterior indica que, para cada ítem, además de tener una probabilidad base π_j^* , también tiene un parámetro de penalización base $P_{c_i}(\eta_c)$.

$$P_{c_i}(\eta_c) = \frac{\exp(\eta_c + c_i)}{1 + \exp(\eta_c + c_i)} \quad (26)$$

η_c mide la habilidad del examinado asociada con los atributos que no están presentes en la matriz Q. De acuerdo con Rupp et al. (2010) la habilidad medida solo es relevante si se tienen valores pequeños de c_i , de lo contrario, si el valor de c_i es grande, la probabilidad de la penalización base es muy pequeña para valores bajos de η_c . En concreto, el mismo autor señaló que un valor grande de c_i indica que los examinados con baja habilidad emplean atributos no especificados por la matriz Q, mientras que valores pequeños de c_i indican que tanto los examinados con mayor o menor habilidad emplean atributos no especificados.

Modelos compensatorios

Los modelos de diagnóstico cognitivo compensatorios asumen que la baja probabilidad de dominio de un atributo puede compensarse con la alta probabilidad de dominio de otros atributos (von Davier y Lee, 2019). Esto implica que el estudiante requiere dominar al menos un atributo dentro de un conjunto de atributos para resolver

correctamente un ítem. Los modelos compensatorios más frecuentes son DINO, NIDO y C-RUM que se describen a continuación.

Modelo DINO. El modelo DINO (The deterministic-input noisy-or-gate model) posee una regla de condensación disyuntiva, por lo que se requiere que el examinado domine al menos un atributo para responder correctamente un ítem (Rupp et al., 2010). Al igual que en el modelo DINA y NIDA, para modelar la probabilidad de responder correctamente un ítem en el modelo DINO, se define un vector de respuesta latente ω_{ij} que se expresa como sigue (Rupp et al., 2010):

$$\omega_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}} \quad (27)$$

Con $\omega_{ij} = 1$ si el examinado posee al menos un atributo presente en el ítem, y $\omega_{ij} = 0$ si el examinado no posee ningún atributo. Por tanto, para el modelo DINO no tiene relevancia cuales son los atributos dominados por el examinado, ya que cualquier combinación de atributos tendría una alta probabilidad de responder correctamente el ítem (Rupp et al., 2010).

El modelo DINO estima los parámetros de desliz s_j y adivinación g_j a nivel de ítem, como en el modelo DINA. El parámetro de desliz es la probabilidad de que los examinados respondan incorrectamente al ítem a pesar de dominar al menos un atributo presente en el ítem, mientras que el parámetro de adivinación representa la probabilidad de que el examinado que no domina ningún atributo presente en el ítem lo responda correctamente (Rupp et al., 2010). Las probabilidades de los parámetros se expresan como sigue:

$$s_j = P(X_{ij} = 0 | \omega_{ij} = 1) \quad (28)$$

$$g_j = P(X_{ij} = 1 | \omega_{ij} = 0) \quad (29)$$

Considerando los parámetros de los ítems del modelo y el vector de respuesta latente, la probabilidad de que el examinado i con un vector de atributos α_i responda

correctamente al ítem j está dada por

$$\pi_{ij} = P(X_{ij} = 1 | \omega_i) = g_j^{1-\omega_{ij}} (1 - s_j)^{\omega_{ij}} \quad (30)$$

Con π_{ij} como la probabilidad del examinado de responder correctamente al ítem j , X_{ij} como la respuesta observada del examinado en el ítem j y ω_i como el vector de respuesta latente.

Modelo NIDO. El modelo NIDO (The noisy input, deterministic-or-gate model) es la versión compensatoria del modelo NIDA, el cual modela los parámetros a nivel de atributo con igualdad entre ítems (Rupp et al., 2010). El modelo estima dos parámetros, el intercepto $(\lambda_{.,0,(k)})$ y la pendiente $(\lambda_{.,1,(k)})$, los cuales se integran en un núcleo de valores continuos de la siguiente manera (Rupp et al., 2010):

$$\text{kernel}_j = \sum_{k=1}^K \left(\lambda_{.,0,(k)} + \lambda_{.,1,(k)} \alpha_{jk} \right) q_{jk} \quad (31)$$

El primer subíndice de los parámetros representa el ítem al que corresponde el parámetro, y es un punto (.) debido a que todos los parámetros para los atributos del modelo son los mismos para todos los ítems. El segundo subíndice representa el nivel del parámetro, con 0 como el intercepto para cada ítem y 1 como el conjunto de efectos principales para cada ítem. Por último, los términos entre paréntesis corresponden a los atributos a los que se refiere el parámetro. Por lo tanto, el parámetro $(\lambda_{.,1,(k)})$ se considera el parámetro de efectos principales o pendiente del modelo NIDO para el atributo k .

La probabilidad del examinado de responder correctamente al ítem en el modelo NIDO se expresa de manera similar a la fórmula básica de un modelo multidimensional de la TRI (Rupp et al., 2010):

$$\pi_{ij} = P(X_{ij} = 1 | \alpha_i) = \frac{\exp \left(\sum_{k=1}^K \left(\lambda_{.,0,(k)} + \lambda_{.,1,(k)} \alpha_{jk} \right) q_{jk} \right)}{1 + \exp \left(\sum_{k=1}^K \left(\lambda_{.,0,(k)} + \lambda_{.,1,(k)} \alpha_{jk} \right) q_{jk} \right)} \quad (32)$$

La fórmula anterior es equivalente a la siguiente transformación logística, la cual

mantiene la probabilidad de respuesta en un rango entre 0 y 1 (Rupp, et al., 2010):

$$\ln \left(\frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \sum_{k=1}^K \left(\lambda_{.,0,(k)} + \lambda_{.,1,(k)} \alpha_{jk} \right) q_{jk} \quad (33)$$

De acuerdo con Rupp et al. (2010) obtener un logit de 0 representa una probabilidad de 0.5 de responder correctamente al ítem; valores positivos representan probabilidades de respuesta mayores a 0.5 y valores negativos tienen probabilidades de respuesta menores a 0.5.

Modelo C-RUM. El modelo C-RUM (The compensatory reparameterized unified model) es la versión compensatoria del modelo NC-RUM, el cual estima la probabilidad de respuesta correcta a nivel tanto de ítem como de atributo. A diferencia del modelo NC-RUM, el modelo C-RUM no especifica un parámetro que de cuenta de la falta de completitud de la matriz Q (Rupp et al., 2010).

Para modelar la probabilidad de respuesta de los examinados, el modelo contiene el parámetro de intercepto $\lambda_{j,0}$ definido a nivel de los ítems y el parámetro de la pendiente $\lambda_{j,1,(k)}$ definido a nivel de los atributos para cada ítem (Rupp et al., 2010). Por lo tanto, el modelo C-RUM estima un parámetro de intercepto para cada ítem y parámetros de la pendiente de acuerdo al número de atributos presentes en la evaluación diagnóstica. Similar al modelo NIDO, los parámetros se integran en un núcleo de la siguiente manera (Rupp et al., 2010):

$$\text{kernel}_j = \lambda_{j,0} + \sum_{k=1}^K \lambda_{j,1,(k)} \alpha_{jk} q_{jk} \quad (34)$$

La probabilidad del examinado de responder correctamente el ítem en el modelo C-RUM está dada por la siguiente fórmula (Rupp et al., 2010):

$$\pi_{ij} = P(X_{ij} = 1 \mid \alpha_j) = \frac{\exp \left(\lambda_{j,0} + \sum_{k=1}^K \lambda_{j,1,(k)} \alpha_{jk} q_{jk} \right)}{1 + \exp \left(\lambda_{j,0} + \sum_{k=1}^K \lambda_{j,1,(k)} \alpha_{jk} q_{jk} \right)} \quad (35)$$

Con π_{ij} como la probabilidad de responder correctamente el ítem j , X_{ij} como la

respuesta observada del examinado en el ítem j , q_{jk} como el indicador de la matriz Q que señala si un atributo k se encuentra presente en el ítem j , α_{jk} como el indicador de dominio del atributo, $\lambda_{j,0}$ como el parámetro de intercepto para el ítem j y $\lambda_{j,0}$ como el parámetro de la pendiente para el ítem j y atributo k .

Modelos generalizados

El modelo G-DINA (Generalized Deterministic Inputs, Noisy and gate) es una generalización del modelo DINA con supuestos más flexibles. Al aplicarse las restricciones adecuadas, distintos modelos de diagnóstico cognitivo son casos especiales de los modelos generales, tales como los modelos DINA, DINO, R-RUM, entre otros (de la Torre, 2011).

El modelo G-DINA divide las clases latentes en $2^{K_j^*}$ grupos latentes, donde

$$K_j^* = \sum_{k=1}^K q_{jk}$$

representa el número de atributos requeridos para responder correctamente el ítem j . Para describir el modelo, se define α_{lj}^* como el vector reducido de atributos cuyos elementos corresponden a los atributos requeridos para el ítem j . Por lo que, si solo se requieren los dos primeros atributos para el ítem j , entonces el vector de atributos α_{lj} se reduce a

$$\alpha_{lj}^* = (\alpha_{l1}, \alpha_{l2})'$$

La probabilidad de que un examinado con patrón de atributos α_{lj}^* responda correctamente el ítem, está dada por $P(X_j = 1 | \alpha_{lj}^*) = P(\alpha_{lj}^*)$.

Para la especificación de modelos generales de diagnóstico cognitivo, es posible emplear diversas funciones de enlace, como la función de identidad, logit y log (de la Torre, 2011). La fórmula original del modelo G-DINA, que representa un modelo saturado, utiliza la función de enlace de identidad. Esta fórmula se descompone en la suma de los efectos asociados a la presencia de atributos específicos y sus interacciones, de la siguiente manera (de la Torre, 2011):

$$P(\alpha_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jk'k} \alpha_{lk} \alpha_{lk'} + \cdots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk} \quad (36)$$

Con δ_{j0} como la intersección para el ítem j ; δ_{jk} como el efecto principal debido a α_k ; $\delta_{jk'k'}$ como el efecto de interacción debido a α_k y $\alpha_{k'}$; y $\delta_{j12\dots K_j^*}$ como el efecto de interacción debido a $\alpha_1, \dots, \alpha_{K_j^*}$.

δ_0 representa la probabilidad base, es decir, la probabilidad de responder correctamente el ítem a pesar de no estar presente ningún atributo requerido; δ_k es el cambio en la probabilidad de una respuesta correcta como resultado de dominar un solo atributo (es decir, α_k); $\delta_{kk'}$ es un efecto de interacción de primer orden, el cual representa el cambio en la probabilidad de una respuesta correcta debido al dominio de ambos α_k y $\alpha_{k'}$, por lo que está por encima de una relación aditiva del dominio de esos dos atributos; y $\delta_{12\dots K_j^*}$ representa el cambio en la probabilidad de una respuesta correcta debido al dominio de todos los atributos requeridos, que está por encima de la relación aditiva de los efectos principales y de interacción de menor orden (de la Torre, 2011).

Modelos de diagnóstico cognitivo en la admisión a la universidad

Los modelos de diagnóstico cognitivo (MDC) han mostrado un gran potencial en la evaluación de habilidades y conocimientos en contextos educativos. Sin embargo, su aplicación en el ámbito de la admisión universitaria sigue siendo un campo en desarrollo, con escasos estudios empíricos que exploren su utilidad en este contexto. En la práctica, los MDC se han implementado principalmente para obtener información diagnóstica de evaluaciones existentes diseñadas con otros enfoques psicométricos. Este proceso, conocido como *retrofitting*, supera en número a las evaluaciones diseñadas inicialmente con un enfoque diagnóstico cognitivo (Williamson, 2023). Este apartado analiza las investigaciones que han implementado MDC, a través de *retrofitting*, en pruebas de ingreso universitario, centrándose en sus contribuciones al diagnóstico cognitivo y en las características de las

pruebas utilizadas. La revisión incluye estudios recientes que aplican MDC a pruebas de comprensión lectora, destacando los enfoques metodológicos utilizados, los modelos aplicados y los hallazgos más relevantes, con el fin de ilustrar el impacto de los MDC en la interpretación de los resultados y en la toma de decisiones educativas.

Ravand (2016) aplicó modelos de diagnóstico cognitivo a la sección de comprensión lectora del Examen de Ingreso a la Universidad Nacional Iraní (INUEE, por sus siglas en inglés). El INUEE se aplica anualmente para la admisión de aspirantes a programas de maestría en estudios del idioma inglés. La prueba consistió en 20 ítems de opción múltiple con cuatro opciones de respuesta. Los análisis se realizaron de manera exploratoria a una muestra de 10,000 aspirantes del año 2012. Para la construcción de la matriz Q, el autor convocó a un comité de jueces expertos en el área que identificaron cinco atributos cognitivos de comprensión lectora: Lectura para detalles, Lectura para hacer inferencias, Lectura para identificar la idea principal, Sintaxis y Vocabulario. Posteriormente, un segundo comité de jueces expertos identificó los atributos subyacentes a cada uno de los 20 ítems de la sección de comprensión lectora del INUEE. El autor ajustó el modelo G-DINA que clasificó a los aspirantes en 32 clases latentes. Se encontró que las clases que tuvieron mayor frecuencia de aspirantes fueron la clase latente = [11111] con aproximadamente 37 % de los aspirantes, y la clase latente = [00000] con aproximadamente 15 % de aspirantes. Estos resultados indican que el 15 % de los aspirantes que aplicaron el INUEE no dominaron ningún atributo de comprensión lectora. Por el contrario, el 37 % de los aspirantes dominaron todos los atributos. Los perfiles de atributos en donde el examinado domina todos los atributos o ninguno de los atributos medidos en la prueba se denominan “perfiles planos”, los cuales pueden deberse a que la prueba utilizada es unidimensional, o bien, existe una alta correlación entre atributos que solo se pueden explicar en conjunto (Lee y Wasaki, 2009b). De acuerdo con Lee y Wasaki (2009b) cuando se tiene una cantidad considerable de examinados clasificados en los perfiles planos, se proporciona poca información adicional más allá de la que puede ofrecer una puntuación total o un nivel de

habilidad.

En un estudio posterior, Ravand y Robitzsch (2018) compararon distintos modelos de diagnóstico cognitivo en términos del ajuste del modelo, la proporción de perfiles de dominio de atributos, las distribuciones univariadas de probabilidades de dominio y la precisión y consistencia de las clases latentes. Los autores utilizaron la matriz Q desarrollada anteriormente por Ravand (2016) para la prueba de comprensión lectora del INUEE. Los modelos comparados fueron DINA, DINO, G-DINA, C-RUM, NC-RUM y ACDM. De manera general, todos los modelos se desempeñaron adecuadamente, siendo el G-DINA el que tuvo un mejor desempeño en casi todas las áreas evaluadas. Los autores señalaron que resulta plausible ajustar distintos modelos dentro de una misma evaluación. Los ítems de una prueba pueden asumir distintas relaciones entre atributos, ya sea compensatoria, no compensatoria, jerárquica o no jerárquica.

En un estudio similar, Hemati y Baghaei (2020) aplicaron los modelos de diagnóstico cognitivo G-DINA, DINA, DINO y ACDM para el módulo de comprensión lectora de la sección del idioma inglés del Examen de Ingreso a la Universidad Iraní (IUEE, por sus siglas en inglés). Los autores analizaron de manera exploratoria las respuestas de una muestra de 10,000 aspirantes del año 2012. Dado que no se contó con información de un modelo cognitivo sobre la comprensión lectora de los examinados, se utilizaron datos de una prueba previamente desarrollada para extraer información diagnóstica acerca de la habilidad de comprensión lectora. Asimismo, se consultaron modelos teóricos y estudios previos en donde se identificaran atributos cognitivos en el campo de la comprensión lectora. Para la construcción de la matriz Q , los autores reunieron a un conjunto de profesores de inglés que especificaron los atributos cognitivos subyacentes a los ítems de la prueba de comprensión lectora. Se especificaron cinco atributos cognitivos que fueron: Realizar inferencias, Extraer información explícita, Identificar el significado de las palabras a partir del contexto, Identificar referencias pronominales y Evaluar las opciones de respuesta. Los autores reportaron que el modelo G-DINA se ajustó mejor a los datos en

comparación de los modelos DINA, DINO y ACDM. Como resultado, se obtuvieron 32 clases latentes de las cuales la clase [00000] obtuvo el 57 % de aspirantes clasificados, mientras que la segunda clase con mayor porcentaje de clasificados fue la clase [11111] con el 10 % de los aspirantes. Esto indica que el 57 % de los aspirantes no dominó ninguno de los cinco atributos cognitivos en comprensión lectora. Por su parte, solamente el 10 % de los aspirantes dominó todos los atributos cognitivos. Finalmente, los autores ilustraron que dos aspirantes que obtuvieron las mismas puntuaciones totales en la prueba pertenecen a diferentes clases latentes y por tanto, podrían tener diferentes perfiles de dominio de atributos. Estos hallazgos ilustran que los estudiantes pueden tener distintas fortalezas y áreas de oportunidad a pesar de tener el mismo puntaje en una prueba (Hemati y Bagahei, 2020).

Ravand et al. (2013) exploraron la capacidad diagnóstica de la prueba de comprensión lectora que se aplicó a los aspirantes a ingresar a doctorado de la Universidad de Isfahán, Irán a través de la aplicación del modelo DINA. Los autores analizaron las respuestas de 1,500 aspirantes a doctorado del año 2009 a la prueba de comprensión lectora la cual estuvo compuesta por 34 ítems de opción múltiple. Un conjunto de expertos especificaron cinco atributos cognitivos subyacentes a la prueba: Vocabulario, Sintaxis, Extracción de información explícita, Conexión y síntesis y Realización de inferencias. Para explorar la capacidad diagnóstica de la prueba, los autores estimaron las clases latentes de los aspirantes así como los parámetros de adivinación, desliz e índice de discriminación de los ítems de la prueba. De acuerdo con Rupp et al. (2010) los ítems que proporcionan información diagnóstica son aquellos con bajos valores de desliz y adivinación, lo que da como resultados altos índices de discriminación. Como resultado del estudio, se obtuvieron bajos índices de discriminación y altos valores de adivinación y desliz de los ítems, lo que implica que los ítems no proporcionan suficiente información diagnóstica para diferenciar a los aspirantes que dominan más atributos de los que dominan menos atributos (Ravand, et al., 2013).

Los modelos no compensatorios, como el DINA, asumen que los estudiantes deben poseer todos los atributos necesarios para resolver un ítem correctamente, mientras que los modelos compensatorios, como el DINO, permiten que la fortaleza en ciertos atributos pueda compensar la ausencia de otros. Por su parte, el modelo G-DINA ofrece un marco más flexible al incorporar tanto interacciones compensatorias como no compensatorias entre atributos.

La elección del modelo depende de las características de los ítems y de las interpretaciones que se deseen obtener. Los modelos no compensatorios son ideales cuando el éxito en un ítem requiere dominar todos los atributos específicos, como en problemas de matemáticas avanzadas donde cada paso depende del dominio de conocimientos previos. En contraste, los modelos compensatorios son más adecuados en contextos donde las fortalezas en ciertos atributos pueden equilibrar las debilidades en otros, como en evaluaciones de habilidades generales de comprensión lectora. Por último, el modelo G-DINA es una opción versátil que permite capturar una amplia gama de interacciones entre atributos, siendo útil en contextos donde no se tiene certeza sobre la relación exacta entre los atributos y el desempeño.

El uso de modelos no compensatorios genera perfiles cognitivos más estrictos, donde un estudiante solo es clasificado como competente si domina todos los atributos necesarios, lo que puede resultar en diagnósticos más conservadores. Por otro lado, los modelos compensatorios producen interpretaciones más flexibles, al permitir que un estudiante con fortalezas en algunos atributos sea clasificado como competente. El modelo G-DINA, al integrar ambas perspectivas, puede ofrecer diagnósticos más precisos y ajustados a la naturaleza del ítem, pero también demanda una mayor complejidad en su estimación e interpretación. En el caso de la prueba de matemáticas del ExIES, la selección del modelo debe considerar la naturaleza de los ítems y los atributos cognitivos evaluados. Por ejemplo, si los ítems requieren dominar múltiples habilidades interdependientes, como el razonamiento algebraico y el manejo de ecuaciones, un modelo no compensatorio como el

DINA sería apropiado. Sin embargo, si los ítems permiten que ciertas habilidades compensen deficiencias en otras, un modelo compensatorio como el DINO sería más adecuado. Alternativamente, el modelo G-DINA podría proporcionar una perspectiva más integral al capturar las relaciones complejas entre los atributos evaluados en la prueba.

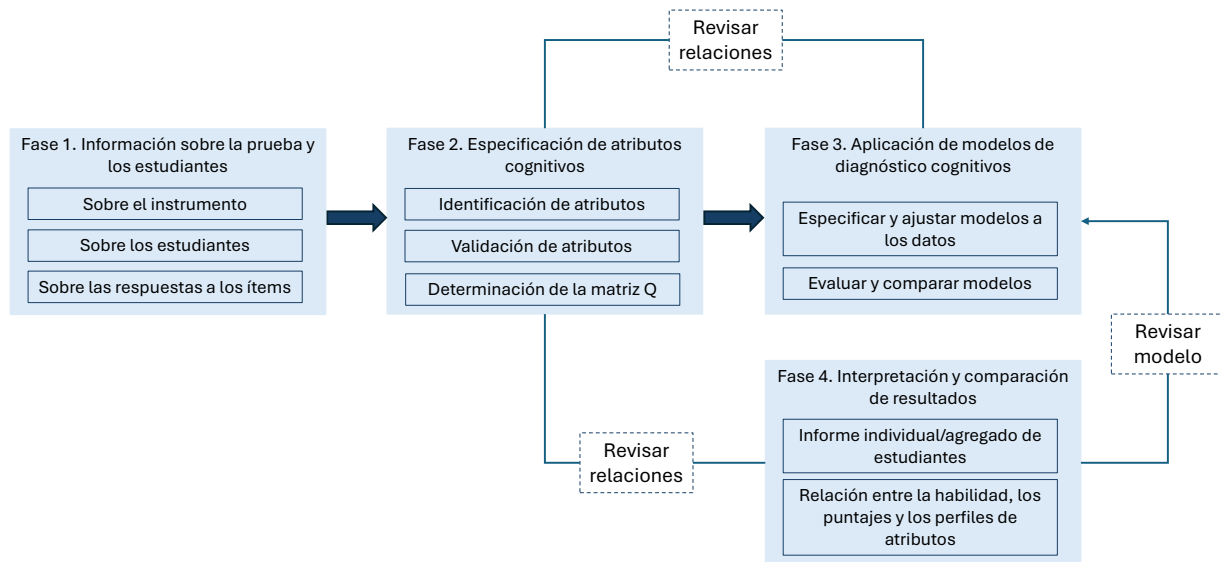
CAPÍTULO III. Método

La aplicación de modelos de diagnóstico cognitivo a evaluaciones existentes, diseñadas bajo otros enfoques psicométricos como la TCT o la TRI, es una aproximación plausible para obtener perfiles de atributos cognitivos de los aspirantes. Esta aproximación, conocida como *retrofitting*, incluye cuatro etapas propuestas por Liu et al. (2018): recopilación de información, especificación de atributos y relaciones atributo-ítem, modelado de respuestas a los ítems e interpretación de los resultados.

Tomando como referencia la metodología propuesta por Liu et al. (2018) se definieron cuatro fases de trabajo para el desarrollo de la presente investigación, como se muestra en la figura 1: Información sobre la prueba y los aspirantes, Especificación de atributos cognitivos, Aplicación de modelos de diagnóstico cognitivo, e Interpretación y comparación de resultados. Dentro de cada fase se describe el objetivo de la misma, así como los instrumentos, participantes y procedimientos realizados.

Figura 1

Proceso de retrofitting de modelos de diagnóstico cognitivo



Fase 1. Información sobre la prueba y los aspirantes

En la primera fase de la investigación se utilizaron los datos obtenidos de la aplicación del Examen de Ingreso a la Educación Superior ([ExIES]; UABC 2018) que se utilizó en el proceso de admisión a la Universidad Autónoma de Baja California (UABC) en mayo de 2018. La aplicación y corrección del examen fueron coordinadas y realizadas por la misma universidad.

Objetivo de la fase

Evaluar la calidad técnica de la prueba y de las puntuaciones de los aspirantes bajo los enfoques de la teoría clásica de los tests (TCT) y la teoría de respuesta al ítem (TRI).

Instrumento y participantes

Se analizaron las respuestas de 13,953 aspirantes, 7,693 hombres (55.13 %) y 6,260 mujeres (44.86 %). La prueba estuvo compuesta por 50 ítems de opción múltiple con una sola opción de respuesta correcta. Los ítems de la prueba se organizaron en cuatro secciones: Herramientas algebraicas, Problemas, probabilidad y análisis de datos, Matemáticas avanzadas y Temas adicionales en matemáticas. La tabla 4 muestra el número de ítems que conformaron cada sección.

Tabla 4

Número de ítems de la prueba de matemáticas del ExIES por sección

Sección	Número de ítems
Herramientas algebraicas	10
Problemas, probabilidad y análisis de datos	15
Matemáticas avanzadas	15
Temas adicionales en matemáticas	10
Total	50

Análisis de datos

Se estimaron las propiedades psicométricas de la prueba bajo los enfoques de la TCT y la TRI a través del software R Studio versión 1.4.1725 con los paquetes CTT (Willse, 2022) y mirt (Chalmers, 2012). El código del análisis se encuentra disponible en el

repositorio de GitHub, el cual puede ser accedido a través del siguiente enlace:

<https://github.com/egiselanavarro/MDC-admision-universidad>.

Bajo el enfoque de la TCT, se estimaron los siguientes índices y puntuaciones sobre la prueba de matemáticas y los aspirantes:

- El índice de dificultad de los ítems (p), el cual es la proporción de aspirantes que respondieron correctamente al ítem, cuyos valores oscilan entre 0 y 1 (Crocker y Algina, 2006).
- El índice de discriminación de los ítems ($rpbis$), el cual es el grado en que un ítem distingue a los aspirantes con un alto nivel de habilidad de aspirantes con un bajo nivel. Se tomaron en cuenta los siguientes criterios (Abad et al., 2011; Crocker y Algina, 2006):
 - Si $rpbis > 0.40$, el ítem discrimina satisfactoriamente
 - Si $0.30 < rpbis < 0.39$, se requiere poca o ninguna revisión del ítem
 - Si $0.20 < rpbis < 0.29$, se necesita revisar el ítem
 - Si $rpbis < 0.19$, el ítem se debe eliminar o revisar por completo.
- El índice de confiabilidad alfa de Cronbach que es una medida de consistencia interna para estimar la precisión en la medición de la prueba. Valores cercanos a 1 indican que el instrumento tiene una consistencia interna elevada (Abad et al., 2011).
- El error estándar de medición, el cual refleja el grado en que la medición se desvía de la puntuación verdadera (Abad et al., 2011).
- Puntuación total observada por el aspirante, la cual representa el número total de aciertos en la prueba de matemáticas.

Bajo el enfoque de la TRI, se ajustó el modelo unidimensional logístico de tres parámetros (3PL) para ítems dicotómicos y se obtuvieron las siguientes estimaciones:

-
- Parámetro de dificultad (β) de los ítems de la prueba.
 - Parámetro de discriminación (α) de los ítems de la prueba. Se tomaron en cuenta los siguientes criterios (Baker, 2001):
 - Muy baja discriminación: 0.01 – 0.34
 - Baja discriminación: 0.35 – 0.64
 - Discriminación moderada: 0.65 – 1.34
 - Discriminación elevada: 1.35 – 1.69
 - Discriminación muy elevada: > 1.70
 - Parámetro de adivinación (c) de los ítems de la prueba, el cual es la probabilidad de adivinar la respuesta correcta del ítem, cuyos valores oscilan entre 0 y 0.35 (Martínez-Arias, 2014).
 - Parámetro de la habilidad latente (θ) de los aspirantes.
 - Curva de información del test, la cual refleja el nivel de información proporcionado por la prueba en distintos niveles de habilidad latente (Baker, 2001).
 - Curvas características de los ítems, las cuales representan la relación entre la probabilidad de responder correctamente un ítem y el nivel de habilidad latente de los examinados (Baker, 2001).

Fase 2. Especificación de atributos cognitivos

La fase 2 se dividió en tres etapas que comprendieron la identificación de atributos cognitivos, validación de los atributos cognitivos y sus descriptores y la construcción de la matriz Q.

Etapas 1. Identificación de atributos cognitivos

- Objetivos de la etapa. Identificar los atributos cognitivos que subyacen al dominio de las matemáticas en aspirantes a ingresar a la universidad.

-
- **Participantes.** Se conformó un comité de desarrollo y validación de atributos cognitivos conformado por dos investigadores en la docencia de la matemática y una maestra en educación con licenciatura en matemáticas aplicadas.
 - **Procedimiento.** Previo a la reunión con el comité de expertos, se realizó una revisión de la literatura y se identificaron 21 atributos cognitivos de matemáticas propuestos por Gierl et al. (2009). La descripción del proceso de identificación de atributos se encuentra en el anexo A. Posteriormente, se convocó al comité de desarrollo y validación para operacionalizar los 21 atributos cognitivos de matemáticas. Los miembros del comité elaboraron de cinco a seis descriptores que representaran a cada atributo cognitivo. Los descriptores se categorizaron en conocimientos y destrezas con las siguientes definiciones que se tomaron como referencia de Haladyna y Rodriguez (2013):

- **Destreza.** Capacidad de realizar una tarea. Algunas destrezas consisten en una sola acción, mientras que otras destrezas son procedimientos que involucran dos o más pasos.
- **Conocimiento.** El conocimiento puede clasificarse como un hecho, un concepto, un principio o un procedimiento.
 - Un hecho es una verdad conocida por experiencia u observación.
 - Un concepto es una idea de algo que se forma combinando mentalmente todas sus características o detalles.
 - Un principio es una regla aceptada de acción o conducta, una ley o verdad fundamental, primaria o general de la que se derivan otras.
 - Un procedimiento es un curso de acción, físico o mental, que tiene un resultado previsto.

Se convocó una sesión en línea con los tres miembros del comité. Durante la sesión se les capacitó a los expertos con respecto a la actividad a realizar y se les compartió

una carpeta de Google Drive con los materiales para la elaboración de descriptores. Dentro de la carpeta se adjuntó el instructivo de elaboración de descriptores en formato pdf y un archivo de google sheets en donde realizaron la tarea. Los expertos tuvieron 10 días hábiles para realizar la tarea de manera individual. Los materiales para el comité de desarrollo y validación se encuentran en el anexo B.

Etapas 2. Validación de atributos cognitivos

- **Objetivos de la etapa.** Validar la operacionalización de los descriptores de los atributos cognitivos desarrollados en la etapa 1.
- **Participantes.** Comité de desarrollo y validación de la etapa 1, conformado por dos investigadores en la docencia de la matemática y una maestra en educación con licenciatura en matemáticas aplicadas.
- **Procedimiento.** El comité de desarrollo y validación valoró el contenido de los descriptores de cada atributo cognitivo de acuerdo con los siguientes criterios:
 - **Claridad.** El descriptor se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.
 - **Coherencia.** El descriptor tiene relación lógica con el atributo que representa.
 - **Relevancia.** El descriptor es esencial o importante, es decir, debe incluirse.
 - **Suficiencia.** Los descriptores que pertenecen a un mismo atributo bastan para representarlo.

Cada criterio se valoró con una escala de calificación del uno al cuatro de acuerdo con los criterios presentados en la tabla 5.

Tabla 5*Criterios de calificación para la validación de contenido de descriptores*

Criterio	Calificación
Suficiencia. Los descriptores que pertenecen a un mismo atributo bastan para representarlo.	1 = Los descriptores no son suficientes para medir el atributo. 2 = Los descriptores solo representan algún aspecto del atributo. 3 = Debe incrementarse el número de descriptores para representar el atributo. 4 = Los descriptores son suficientes y representan el atributo.
Claridad. El descriptor se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1 = El descriptor no es claro. 2 = El descriptor requiere bastantes modificaciones o una modificación sustantiva. 3 = Se requiere una modificación muy específica de algunos términos del descriptor. 4 = El descriptor es claro.
Coherencia. El descriptor tiene relación lógica con el atributo que representa.	1 = El descriptor no tiene relación lógica con el atributo. 2 = El descriptor tiene una relación tangencial con el atributo. 3 = El descriptor tiene una relación moderada con el atributo. 4 = El descriptor está completamente relacionado con el atributo.
Relevancia. El descriptor es esencial o importante, es decir, debe incluirse.	1 = El descriptor puede eliminarse sin que se vea afectada la representación del atributo. 2 = El descriptor tiene alguna relevancia, pero otro descriptor puede estar incluyendo lo que éste representa. 3 = El descriptor es relativamente importante para representar el atributo. 4 = El descriptor es muy relevante y debe incluirse para representar el atributo.

Se realizó una sesión en línea con los tres miembros del comité y se les indicó la tarea a realizar. Se les proporcionó un archivo en Google Sheets para realizar la tarea de manera individual durante la sesión. Posteriormente, se calculó la proporción de acuerdo de los jueces por cada descriptor. El material para el comité de desarrollo y validación para esta etapa se encuentra en el anexo C.

Etapas 3. Determinación de la matriz Q

- **Objetivos.** Determinar la relación entre los atributos cognitivos y los ítems de la prueba de matemáticas del ExIES y construir la matriz Q.
- **Participantes.** Comité de determinación de matriz Q conformado por cuatro profesores de bachillerato con experiencia en el desarrollo de ítems.
- **Materiales:** Un cuadernillo con la descripción del proyecto de tesis, los objetivos del comité y las indicaciones de la actividad a realizar, el listado de atributos cognitivos con sus descriptores, cuatro hojas de trabajo para establecer la relación entre los atributos y los 50 ítems de la prueba (dos hojas de trabajo con 10 ítems y dos hojas con 15 ítems), un cuestionario de datos de identificación y un aviso de confidencialidad (Anexo D).
- **Procedimiento.** Para construir la matriz Q se convocó a los profesores a una reunión presencial que tuvo una duración de 3 horas. La sesión se organizó de la siguiente manera:
 - Se presentó el proyecto de tesis, los objetivos del comité y las actividades a realizar durante la sesión. Posteriormente, se destinó un tiempo para que los expertos se familiarizaran con los materiales.
 - Se les pidió a los jueces seleccionar, para cada uno de los 50 ítems de la prueba, de uno a tres atributos que consideraran necesarios para responder correctamente cada ítem.
 - Los expertos llenaron de manera individual cada hoja de trabajo. Una vez terminada cada hoja, se concentró la información en un archivo de Excel y se presentaron los resultados preliminares a los expertos con la finalidad de discutir modificaciones, recomendaciones y llegar a un consenso entre los expertos. El proceso se repitió para todas las hojas de trabajo.

Al finalizar la sesión, se analizó y concentró la información para construir la matriz Q.

Fase 3. Aplicación de modelos de diagnóstico cognitivo

La fase 3 del estudio consistió en la aplicación de MDC dentro de la clasificación de modelos compensatorios y no compensatorios. Los modelos que se ajustaron fueron DINA, DINO y G-DINA. Esta fase consistió en el ajuste y comparación de los modelos.

Etapas 1. Ajuste de los modelos

- **Objetivo de la etapa.** Ajustar modelos de diagnóstico cognitivo para la estimación del dominio de atributos cognitivos de los aspirantes.
- **Instrumentos.** Respuestas de los aspirantes en los ítems de la prueba de matemáticas del ExIES en un formato de respuesta binaria de 0 y 1 y matriz Q.
- **Procedimiento.** Se estimaron los modelos DINA, DINO y G-DINA con el software libre R Studio versión 1.4.1725 con el paquete GDINA versión 2.9.4 (Ma y de la Torre, 2020) y el paquete CDM (Robitzsch et al., 2016).

Etapas 2. Comparación de los modelos

- **Objetivo de la etapa.** Determinar los MDC que presentan un mejor ajuste a los datos observados.
- **Procedimiento.** Para realizar la comparación de los modelos, se analizaron los siguientes criterios:
 - Estadísticos de ajuste relativo y ajuste absoluto, para cada modelo estimado. Se seleccionaron los modelos con mejor ajuste de acuerdo a los criterios que se presentan en la tabla 6.
 - Proporción de varianza explicada (PVAE), la cual se define como el GDI (índice de discriminación general, por sus siglas en inglés) expresado como una

proporción respecto al GDI máximo posible para el ítem. El GDI corresponde a la varianza de las probabilidades de responder correctamente a un ítem, dado un vector q de atributos cognitivos (Williamson, 2023). Se tomó en cuenta un punto de corte de $PVAF = 0.95$ (de la Torre y Chiu, 2016).

- Índices de precisión P_a y consistencia P_c de los atributos y las clases latentes (Cui et al., 2012).
- Parámetros de adivinación g y desliz s de los ítems de la prueba con los siguientes criterios para determinar la calidad de los ítems (Iaconangelo, 2017):
 - Calidad baja: $g = s = 0.3$
 - Calidad media: $g = s = 0.2$
 - Calidad alta: $g = s = 0.1$
- Producto. Se seleccionó el modelo de diagnóstico cognitivo con mejores índices de ajuste, mejor precisión y consistencia en la clasificación, mejores parámetros de adivinación y desliz, y que proporcionó información interpretable para usos diagnósticos y formativos.

Tabla 6*Criterios de los estadísticos de ajuste del modelo*

Estadístico	Descripción	Criterio
RMSEA	Evalúa la discrepancia por grado de libertad en el modelo.	< 0.06
SRMR	Evalúa la diferencia entre los valores observados y los valores predichos por el modelo.	< 0.08
CFI	Evalúa el ajuste del modelo comparando el modelo propuesto con un modelo nulo.	> 0.95
TLI/NNFI	Evalúa el ajuste del modelo comparando el modelo propuesto con un modelo nulo, pero penaliza por la complejidad del modelo.	> 0.95
AIC	Evalúa la cantidad de información perdida al aproximar la realidad con un modelo dado.	En comparación con otro modelo, valores más bajos indican mejor ajuste
BIC	Evalúa la calidad de un modelo en función de su ajuste a los datos y su complejidad.	En comparación con otro modelo, valores más bajos indican mejor ajuste

Nota. Burnham y Anderson, 2002; Hu y Bentler, 1999; Schwarz, 1978.

Fase 4. Interpretación y comparación de resultados

La fase 4 del estudio consistió en la interpretación del modelo de diagnóstico cognitivo que tuvo el mejor ajuste a los datos. Posteriormente, se describió el desempeño de los aspirantes según el enfoque psicométrico.

Objetivos de la fase

Examinar el nivel de dominio de los aspirantes en los atributos cognitivos evaluados por la prueba de matemáticas del ExIES, y establecer comparaciones con sus puntuaciones obtenidas en la prueba y su nivel de habilidad.

Procedimiento

Para examinar el nivel de dominio de los aspirantes en los atributos cognitivos, se obtuvieron las siguientes medidas:

- Prevalencia de los atributos cognitivos, definida como la proporción de aspirantes que dominan cada atributo cognitivo (Shi et al., 2021).

-
- Correlación entre los atributos cognitivos.

El desempeño de los aspirantes se comparó tomando en consideración las siguientes medidas:

- Estadísticos descriptivos de los atributos cognitivos dominados, el puntaje total en la prueba y el nivel de habilidad de los aspirantes.
- Comparación de los perfiles de atributos cognitivos, así como del puntaje total en la prueba y el nivel de habilidad de los aspirantes.

CAPÍTULO IV. Resultados

El presente capítulo se encuentra organizado en tres apartados que responden a los objetivos de investigación. En el primer apartado se presentan las evidencias de la calidad técnica de la prueba de matemáticas del ExIES bajo los enfoques de la teoría clásica de los tests (TCT) y la teoría de respuesta al ítem (TRI). En el segundo apartado se presenta el proceso de desarrollo y especificación de la matriz Q por juicio de expertos, así como la aplicación y comparación de los modelos de diagnóstico cognitivo DINA, DINO y G-DINA. En el tercer apartado se comparan los resultados obtenidos por los MDC con el nivel de habilidad y el puntaje total de los aspirantes.

Evaluación de la calidad técnica de la prueba

A continuación, se presenta el análisis de la calidad técnica de la prueba de matemáticas del ExIES bajo los enfoques de la TCT y la TRI, tomando en consideración las recomendaciones de los *Estándares para Pruebas Educativas y Psicológicas* (AERA et al., 2018). Se analizaron los datos obtenidos por la prueba de matemáticas del ExIES, compuesta por 50 ítems de opción múltiple con una sola opción de respuesta correcta, la cual fue respondida por 13,953 aspirantes a ingresar a la UABC en el año 2018.

Resultados bajo el enfoque de la teoría clásica de los tests

El análisis de los ítems de la prueba de matemáticas del ExIES bajo el enfoque de la TCT arrojó una dificultad media de 0.46. Esto indica que, en promedio, cada ítem de la prueba fue respondido correctamente por el 46 % de los aspirantes. Del análisis de la dificultad por ítem, se encontró que 38 ítems obtuvieron una dificultad moderada entre 0.3 y 0.8. Mientras que tres ítems de la prueba obtuvieron una dificultad mayor a 0.80, los cuales se consideran ítems fáciles, y nueve ítems obtuvieron una dificultad menor a 0.30, por lo que se consideran ítems difíciles. De manera particular, el ítem 47 fue el más difícil para la muestra de aspirantes con un índice de dificultad de 0.12.

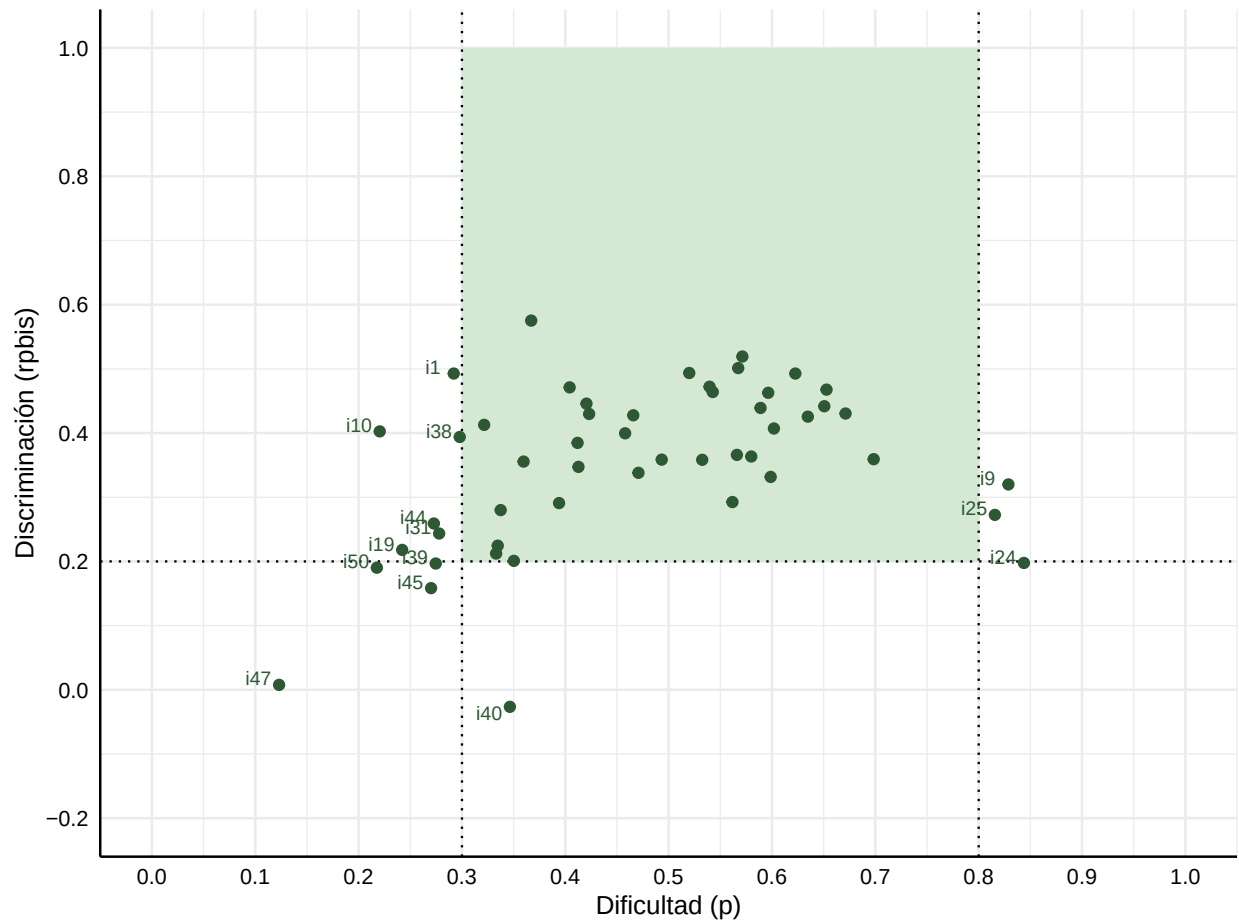
En relación con la discriminación de los ítems, se obtuvo una discriminación media

de 0.35. Se identificaron siete ítems con baja discriminación ($rpbis < 0.20$), lo que indica que tienen problemas para distinguir entre los aspirantes con un alto desempeño en la prueba y los que tienen un bajo desempeño (Abad et al., 2011; Crocker y Algina, 2006). En particular, el ítem 40 fue el único que presentó una discriminación negativa ($rpbis = -0.03$), lo que indica que una mayor proporción de aspirantes con bajo desempeño en la prueba respondió correctamente el ítem con respecto a los aspirantes con un alto desempeño. La figura 2 muestra el gráfico de dispersión con los índices de dificultad y discriminación de los ítems. En el área sombreada se ubican los ítems con índices de dificultad y discriminación adecuados ($rpbis > 0.2; 0.3 < p < 0.08$). Puede observarse que la mayoría de los ítems se encuentran dentro de esta área. Los índices de dificultad, discriminación y las variaciones en el alfa de Cronbach se encuentran en el anexo E.

La confiabilidad de la prueba de matemáticas se estimó mediante el coeficiente alfa de Cronbach, obteniéndose un valor de .892 (IC del 95 %: .889, .894). De acuerdo con Abad et al. (2011), valores cercanos a 1 en el alfa de Cronbach reflejan una alta consistencia interna, lo que indica que los ítems de la prueba están midiendo de manera coherente los conocimientos y habilidades evaluados. Además, el error estándar de medida de 3.08, indica que las puntuaciones observadas en la prueba presentan una variabilidad aproximada de tres puntos, con respecto a la puntuación verdadera de los aspirantes, reflejando la precisión del instrumento. Con respecto al análisis de las variaciones en el alfa de Cronbach si se eliminara un ítem de la prueba, se identificó que la eliminación del ítem 40 aumentaría marginalmente el índice de confiabilidad de la prueba de 0.89 a 0.90.

Figura 2

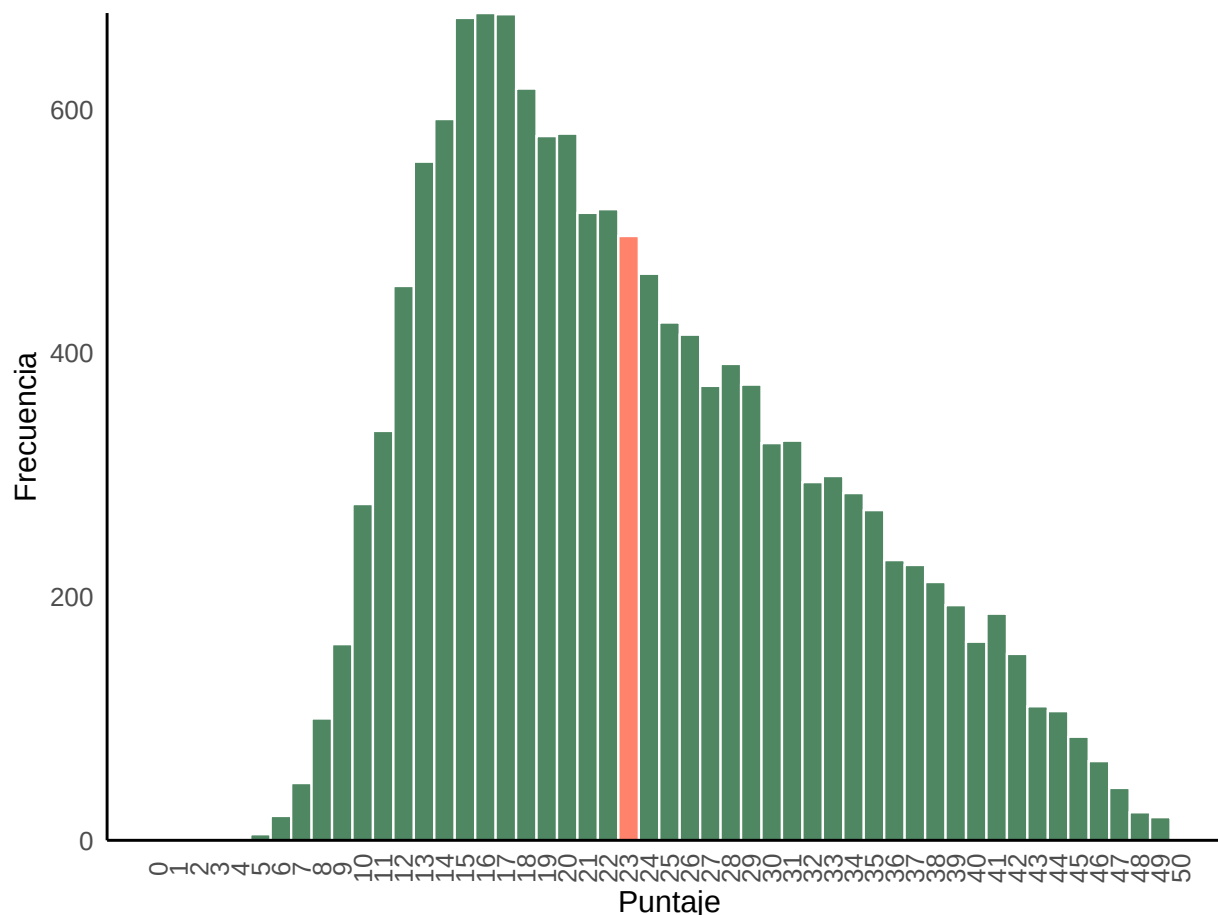
Índices de dificultad y discriminación de los ítems bajo el enfoque de la teoría clásica de los tests



Nota. El gráfico muestra la relación entre el índice de discriminación y el índice de dificultad de los ítems. Las líneas punteadas indican los valores deseables para el índice de dificultad (.3 y .8) y para el índice de discriminación (≥ 0.2).

En cuanto a las puntuaciones de los aspirantes, se obtuvo un puntaje de aciertos mínimo de 4 y un máximo de 50. El puntaje medio de los aspirantes fue de 23 aciertos, lo cual es ligeramente por debajo del promedio teórico de la distribución. En la figura 3 se muestra la distribución de las puntuaciones de los aspirantes. En el histograma se observa la tendencia hacia valores por debajo de la media de la muestra de aspirantes.

Figura 3
Distribución de los puntajes totales de los aspirantes



Nota. La barra roja corresponde a la media de los puntajes ($\bar{x} = 23$).

Resultados bajo en enfoque de la teoría de respuesta al ítem

Se ajustó el modelo logístico de tres parámetros (3PL) el cual, de acuerdo con Crocker y Algina (2006) es un modelo ideal para pruebas con ítems de opción múltiple en donde, a pesar de que los aspirantes tengan valores muy bajos de habilidad, pueden adivinar la respuesta correcta. Se obtuvieron los estadísticos de ajuste absoluto y ajuste relativo del modelo, los cuales indican un buen ajuste a los datos (RMSEA=0.016, SRMSR=0.018, TLI=0.989, CFI=0.990, $p=0.000$) de acuerdo con los criterios de interpretación (Burnham y Anderson, 2002 y Hu y Bentler, 1999).

El ajuste del modelo permitió identificar los parámetros de dificultad (b),

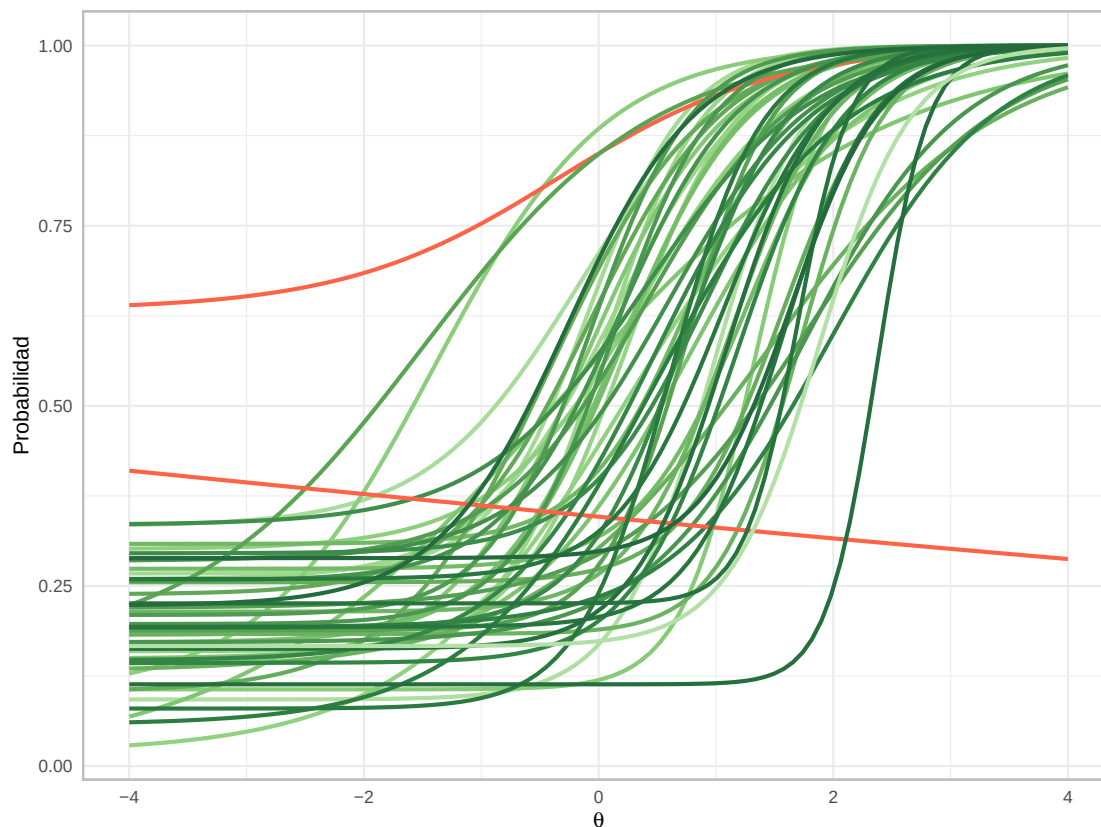
discriminación (a) y adivinación (c) de los ítems de la prueba de matemáticas (anexo F). Se encontró que más de la mitad de los ítems de la prueba (38 ítems) obtuvieron una dificultad moderada con valores de b entre -1.5 y 1.5 . Por su parte, 10 ítems se consideraron difíciles con valores de $b > 1.5$, y dos ítems se consideraron fáciles, con valores de $b < -1.5$. En particular, el ítem 40 obtuvo una dificultad inusualmente baja de ($b = -9.35$).

En relación con la discriminación de los ítems, se obtuvieron 31 ítems con discriminación muy alta, 11 ítems con discriminación alta, siete ítems con discriminación moderada y, en particular, un ítem (el ítem 40) que no cumplió con los criterios de interpretación al presentar una discriminación negativa de $d = -0.07$ (Baker, 2001).

El parámetro de adivinación c , representa la probabilidad de que los aspirantes respondan correctamente el ítem por adivinación, sin importar su nivel de habilidad. Por lo general, los valores de c oscilan entre 0.0 y 0.35 (Martínez-Arias, 2014). Los resultados arrojaron que todos los ítems a excepción de uno obtuvieron valores de $c < 0.35$, lo que indica que hay poca probabilidad de que los aspirantes respondan correctamente a los ítems por adivinación. En particular, el ítem 24 obtuvo un valor de $c = 0.63$, la cual es una probabilidad de adivinar la respuesta correcta inusualmente alta. En la figura 4 se presentan las curvas características de los ítems de la prueba, las cuales reflejan los parámetros de dificultad, discriminación y adivinación.

Figura 4

Curvas características de los ítems de la prueba de matemáticas del ExIES



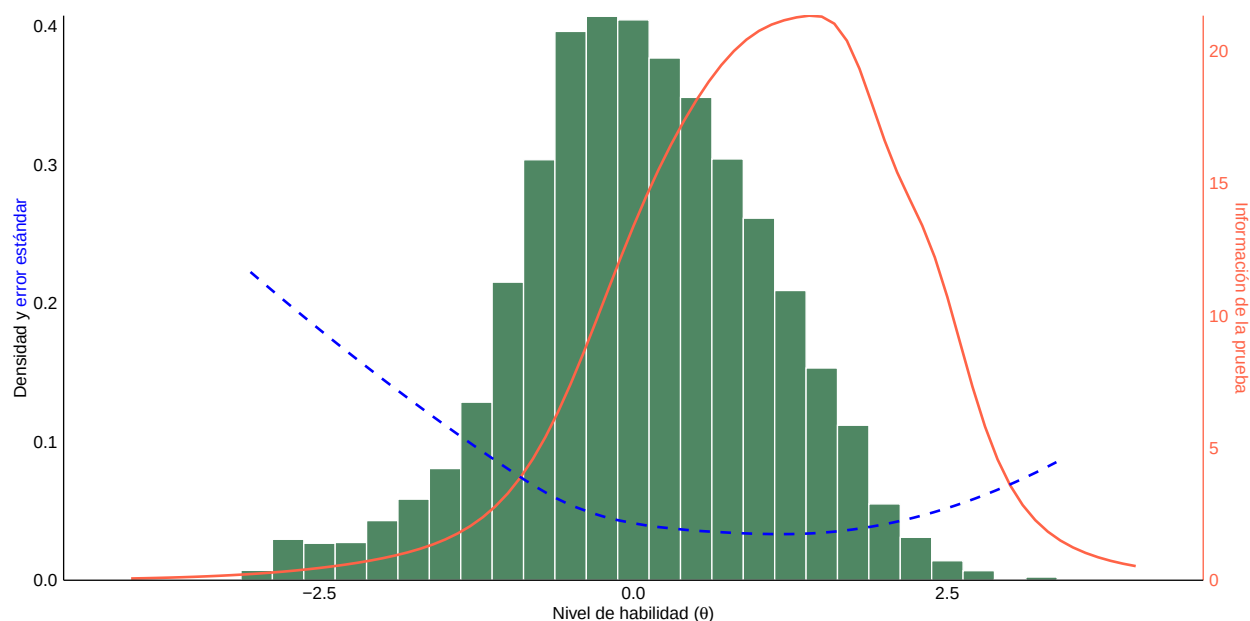
Nota. Las curvas de color rojo corresponden a los ítems 24 (arriba) y 40 (abajo).

Para evaluar la precisión en la medición desde el enfoque de la TRI, se calculó la curva de información de la prueba (figura 5) en conjunto con la distribución del nivel de habilidad de los aspirantes. La curva muestra la cantidad de información que proporciona la prueba sobre los aspirantes en función de sus diferentes niveles de habilidad. Puede observarse que la curva alcanza un punto máximo alrededor de $\theta = (1.5, 2.0)$, lo que sugiere que la prueba es más precisa para medir la habilidad matemática de los aspirantes que tienen una habilidad en ese rango. Después del punto máximo, la información comienza a disminuir a medida que la habilidad de los aspirantes aumenta. De manera general, la prueba es más informativa y precisa para discriminar a los aspirantes con un nivel de habilidad medio y alto. Este comportamiento resulta ideal para pruebas de admisión, dado

que se espera seleccionar a los aspirantes con altos puntajes en la prueba. No obstante, los niveles de habilidad de los aspirantes se encuentran ligeramente por debajo de la media teórica de $\theta = 0$. En la figura 5, la curva punteada de color azul representa los errores de estimación de θ , la cual indica la precisión de la estimación de los diferentes niveles de habilidad θ . Puede observarse una mayor precisión de estimación en niveles de habilidad en donde se concentra la mayoría de los aspirantes, mientras que la precisión disminuye en los niveles de habilidad muy bajos o muy altos.

Figura 5

Nivel de habilidad matemática de los aspirantes e información de la prueba



Nota. El error estándar está normalizado para ser comparable visualmente con la densidad.

En concordancia con los objetivos de investigación, los análisis realizados demuestran que la prueba de matemáticas del ExIES cumple con los estándares de calidad técnica tanto desde el enfoque de la teoría clásica de los tests (TCT) como de la teoría de respuesta al ítem (TRI). Respecto a los ítems y a la prueba en general, bajo el enfoque de la TCT se obtuvo un coeficiente de confiabilidad elevado, mientras que los valores de discriminación y confiabilidad reflejaron que los ítems son adecuados para medir el desempeño en matemáticas de los aspirantes. Bajo el enfoque de la TRI, se encontró que la

prueba se alinea a los usos propuestos de seleccionar a los aspirantes con mejor desempeño, y la mayoría de los ítems obtuvieron valores adecuados de dificultad, discriminación y adivinación.

En particular, se observó que el ítem 40 obtuvo discriminación negativa en ambos enfoques ($rpbis = -0.03$; $d = -0.07$), indicando que conforme aumenta el nivel de desempeño o habilidad de los aspirantes, su probabilidad de responder correctamente al ítem disminuye, lo cual es un comportamiento incongruente con lo esperado en una prueba de desempeño. Además, derivado del análisis de confiabilidad en la TCT, se encontró que si se eliminara el ítem 40 el coeficiente de confiabilidad de la prueba mejoraría ligeramente. Por último, este ítem presentó una dificultad inusualmente baja en el análisis de ítems en la TRI. Por su parte, el ítem 24 presentó valores adecuados de dificultad y discriminación en ambos enfoques, pero obtuvo una probabilidad de adivinación considerablemente alta ($c = 0.63$).

Dado los resultados anteriores, se sugiere la revisión del contenido de los ítems 24 y 40 tanto en redacción como en las opciones de respuesta para considerar su modificación o eliminación. Para los propósitos de esta investigación, se utilizará la prueba de matemáticas en su totalidad, lo que permitirá analizar el desempeño de los ítems con funcionamiento deficiente según los enfoques de la TCT y la TRI.

Aplicación de modelos de diagnóstico cognitivo

En este apartado se presenta el proceso de determinación de la matriz Q a través del juicio de expertos, seguido de la aplicación de los modelos de diagnóstico cognitivo DINA, DINO y G-DINA.

Construcción y validación de la matriz Q

A continuación, se describe el resultado del trabajo de los dos comités de jueces expertos para la identificación y validación de los atributos cognitivos que subyacen a la prueba de matemáticas del ExIES. El comité de desarrollo y validación elaboró y desarrolló

los descriptores de 21 atributos cognitivos (tabla 7) y, posteriormente, se realizó una reunión con el comité para la revisión de los mismos. En el anexo G se encuentran los atributos cognitivos con sus descriptores. Durante la revisión con los jueces, se realizaron las siguientes recomendaciones y modificaciones:

- Se realizaron 52 recomendaciones de redacción para mejorar la claridad en los descriptores. Por ejemplo: *“Quizá reformular la expresión como: Identifica las partes de la expresión (variables, exponentes, coeficientes, constantes, valor a sustituir).”*
- Se identificaron siete descriptores que carecían de relevancia para el atributo cognitivo, o bien, eran redundantes con otro descriptor, por lo que se recomendó su eliminación. Un ejemplo de modificación fue la siguiente: *“El H2A2D10 y H2A2D9 pueden ir juntos ya que hay varios que dejan ver que sabe aplicar las propiedades de los exponentes”.*
- Se realizaron 11 recomendaciones respecto al número de descriptores adecuados para representar al atributo cognitivo. Una recomendación fue la siguiente: *“Se debe añadir uno [descriptor] donde se mencione que Reconoce si el resultado de una expresión es indefinido, comenzará simplificar antes de sustituir para validar el resultado en la expresión más simplificada.”*

Posteriormente, el comité de determinación de la matriz Q estableció la relación de los atributos cognitivos subyacentes a cada uno de los 50 ítems de la prueba de matemáticas del ExIES. Se analizó la frecuencia con la que fueron relacionados los atributos cognitivos con los ítems (tabla 7). Como criterio para la selección de atributos cognitivos en la construcción de la matriz Q, se excluyeron aquellos con una frecuencia inferior a 2, con el fin de evitar estimaciones imprecisas para dichos atributos cognitivos. En consecuencia, se excluyeron los siguientes 10 atributos cognitivos:

- 6. Ordenamiento de series geométricas o cálculo numérico consecutivo.

-
- 7. Realización de operaciones exponenciales o suma de potencias.
 - 11. Aplicación de reglas de factorización.
 - 15. Resolución de series geométricas en un patrón abstracto.
 - 16. Representación o ejecución de múltiples soluciones algebraicas básicas.
 - 17. Representación o ejecución de múltiples soluciones algebraicas avanzadas.
 - 18. Multiplicación de potencias o aplicación de reglas en operaciones de exponenciales.
 - 19. Establecimiento o resolución de ecuaciones cuadráticas.
 - 21. Resolución de una desigualdad cuadrática con dos variables.

Tabla 7

Frecuencia de selección de los atributos cognitivos en los ítems de la prueba de matemáticas del ExIES

Atributo cognitivo	Frecuencia
1. Establecimiento de una única proporción comparando dos cantidades.	7
2. Resolución de una ecuación lineal.	11
3. Aplicación de operaciones aritméticas básicas.	31
4. Sustitución de valores en expresiones algebraicas.	7
5. Comprensión de problemas matemáticos contextualizados.	24
6. Ordenamiento de series geométricas o cálculo numérico consecutivo.	0
7. Realización de operaciones exponenciales o suma de potencias.	2
8. Graficación de una función.	5
9. Aplicación de las propiedades de factorización.	2
10. Sustitución de números en expresiones algebraicas.	6
11. Aplicación de reglas de factorización.	2
12. Sustitución de expresiones algebraicas por otras expresiones algebraicas.	6
13. Comprensión de fórmulas matemáticas.	22
14. Resolución de sistemas de ecuaciones lineales.	4
15. Resolución de series geométricas en un patrón abstracto.	0
16. Representación o ejecución de múltiples soluciones algebraicas básicas.	1
17. Representación o ejecución de múltiples soluciones algebraicas avanzadas.	2
18. Multiplicación de potencias o aplicación de reglas en operaciones de exponenciales.	0
19. Establecimiento o resolución de ecuaciones cuadráticas.	1
20. Reconocimiento de la relación entre variables independientes y dependientes o su representación gráfica.	7
21. Resolución de una desigualdad cuadrática con dos variables.	1

Durante la sesión, los expertos emitieron las siguientes recomendaciones respecto a

la similitud de contenido en los atributos cognitivos:

- Los descriptores de los atributos cognitivos “4. Sustitución de valores en expresiones algebraicas”, “10. Sustitución de números en expresiones algebraicas” y “12. Sustitución de expresiones algebraicas por otras expresiones algebraicas” son muy similares y difíciles de distinguir, por lo que se recomendó la agrupación de los atributos cognitivos a uno nuevo denominado “Manipulación de expresiones algebraicas”.
- Los descriptores de los atributos cognitivos “8. Graficación de una función” y “20. Reconocimiento de la relación entre variables dependientes e independientes y su representación gráfica” son muy similares, por lo que se recomendó la agrupación de los atributos cognitivos a uno nuevo denominado “Relación entre variables y su representación gráfica”.
- Los descriptores de los atributos cognitivos “2. Resolución de una ecuación lineal” y “14. Resolución de sistemas de ecuaciones lineales” son muy similares, por lo que se recomendó agrupar los atributos cognitivos a uno nuevo denominado “Resolución de ecuaciones lineales”.

Por lo anterior, y considerando tanto la exclusión de atributos cognitivos por tener una baja frecuencia como la agrupación de atributos cognitivos por similitud en contenido, se construyó la matriz Q con los siete atributos cognitivos resultantes (tabla 8). De los 50 ítems de la prueba, se identificaron 24 ítems uni-atributo, es decir, ítems en los cuales se requiere únicamente un atributo cognitivo para responderlos correctamente, y 26 ítems multi-atributo, que son ítems en los que se requiere de dos a tres atributos cognitivos para obtener una respuesta correcta.

La matriz Q de la prueba de matemáticas del ExIES se presenta en la tabla 9. En el encabezado se especifica el código de cada atributo cognitivo, mientras que cada fila representa uno de los 50 ítems de la prueba de matemáticas. La matriz utiliza el valor de

“0” para indicar que el atributo cognitivo no se requiere para responder correctamente el ítem, y “1” para indicar que sí se requiere el atributo cognitivo.

Tabla 8

Atributos cognitivos subyacentes a la prueba de matemáticas del ExIES

Código	Atributo cognitivo
A1	Establecimiento de una única proporción comparando dos cantidades.
A2	Resolución de ecuaciones lineales.
A3	Aplicación de operaciones aritméticas básicas.
A4	Manipulación de expresiones algebraicas.
A5	Relación entre variables y su representación gráfica.
A6	Comprensión de fórmulas matemáticas.
A7	Comprensión de problemas matemáticos contextualizados.

Tabla 9

Matriz Q de la prueba de matemáticas del ExIES

Ítem	A1	A2	A3	A4	A5	A6	A7	Ítem	A1	A2	A3	A4	A5	A6	A7
i1	0	0	0	1	0	0	1	i26	0	0	0	1	0	0	0
i2	0	0	1	1	0	0	0	i27	0	0	0	0	0	1	0
i3	0	0	0	1	1	0	0	i28	0	0	0	0	1	0	0
i4	0	1	0	0	0	0	0	i29	0	1	0	0	0	0	0
i5	0	1	0	0	0	0	1	i30	0	0	1	1	0	0	0
i6	0	1	0	1	0	0	1	i31	0	0	0	0	1	0	1
i7	0	1	0	0	0	0	0	i32	0	0	0	0	0	0	1
i8	0	1	0	0	0	0	0	i33	0	0	0	1	0	1	0
i9	0	0	0	1	0	0	1	i34	0	0	0	1	0	0	0
i10	0	0	0	0	1	0	0	i35	0	0	0	1	0	0	0
i11	1	0	1	0	0	0	1	i36	0	0	0	0	0	0	1
i12	1	0	1	0	0	0	0	i37	0	0	0	1	0	0	0
i13	1	0	1	0	0	0	0	i38	0	1	0	0	0	0	0
i14	0	0	1	0	0	0	0	i39	0	0	0	0	1	0	0
i15	1	0	1	0	0	0	0	i40	0	0	0	0	0	1	0
i16	0	0	1	0	0	0	0	i41	0	0	1	0	0	0	0
i17	0	0	1	1	0	1	0	i42	0	0	1	0	0	1	0
i18	0	0	0	0	1	0	0	i43	1	0	1	0	0	0	0
i19	0	0	0	0	1	0	0	i44	0	0	1	0	0	1	1
i20	0	0	1	0	0	1	1	i45	0	0	1	0	0	1	0
i21	0	0	1	0	0	1	0	i46	0	0	1	0	0	1	0
i22	0	0	1	0	0	0	1	i47	0	0	1	0	0	1	0
i23	0	0	1	0	0	1	0	i48	0	0	1	0	0	1	0
i24	0	0	0	0	0	1	1	i49	0	0	0	0	0	1	0
i25	0	0	1	0	0	0	0	i50	0	0	0	1	0	0	0

Nota. A1= Establecimiento de una única proporción comparando dos cantidades, A2= Resolución de ecuaciones lineales, A3= Aplicación de operaciones aritméticas básicas, A4= Manipulación de expresiones algebraicas, A5= Relación entre variables y su representación gráfica, A6= Comprensión de fórmulas matemáticas y A7= Comprensión de problemas matemáticos contextualizados.

Ajuste de los modelos

En esta sección se presentan los resultados del ajuste y desempeño de los modelos DINA, DINO y G-DINA, así como su capacidad para validar la matriz Q y clasificar a los aspirantes en clases latentes. Estos hallazgos permiten identificar las fortalezas y limitaciones de cada modelo en el análisis de la prueba de matemáticas, aportando criterios para seleccionar el modelo más adecuado.

Para identificar el modelo de diagnóstico cognitivo que ofrece un mejor ajuste a los datos, se utilizaron los modelos DINA (no compensatorio), DINO (compensatorio) y G-DINA (generalizado). Se obtuvieron los índices de ajuste absoluto de los modelos (tabla 10). Como resultado, los tres modelos presentaron un buen ajuste a los datos, siendo el modelo G-DINA el que presentó un mejor ajuste dados los índices RMSEA y SRMSR. Asimismo, se obtuvieron los índices de ajuste relativo a través de un ANOVA para realizar la comparación de los tres modelos. En congruencia con los resultados de ajuste absoluto, el modelo G-DINA obtuvo los valores más bajos de devianza, AIC y BIC, por lo que este modelo tuvo un mejor ajuste a los datos con respecto a los modelos DINA y DINO (tabla 10).

Tabla 10

Índices de ajuste y comparación de modelos de diagnóstico cognitivo

Modelo	Parámetros	RMSEA	SRMSR	AIC	BIC	Chi-cuadrada	gl	p-value
DINA	227	0.024	0.042	803804.12	805516.49	6519.96	72	<0.001
DINO	227	0.024	0.041	803206.22	804918.58	5922.05	72	<0.001
G-DINA	299	0.019	0.037	797428.17	799683.66	—	—	—

Nota. Los criterios de interpretación de los estadísticos de ajuste son $RMSEA < .06$, $SRMSR < .08$, AIC y BIC, donde los valores más bajos indican un mejor ajuste.

La proporción de varianza explicada (PVAF) es una métrica que mide qué tan bien un MDC puede explicar la varianza de las respuestas a un ítem en función de los atributos cognitivos necesarios para resolverlo correctamente. Como criterio, se empleó un punto de

corte de $PVAF = 0.95$, recomendado por de la Torre y Chiu (2016), para evaluar la adecuación del modelo a cada ítem. De manera general, los ítems de la prueba obtuvieron valores deseables de $PVAF$ en los tres modelos DINA, DINO y G-DINA (tabla 11). Aunque 26 ítems presentaron valores ligeramente inferiores al punto de corte, estos siguen siendo proporciones de varianza explicada adecuadas. Sin embargo, el ítem 40 presentó valores significativamente bajos de $PVAF$ en los tres modelos. Por su parte, el ítem 47 mostró valores bajos en los modelos DINA y DINO, pero un valor cercano al punto de corte en el modelo G-DINA. Estos resultados indican que la mayoría de los ítems de la prueba tuvieron una correcta especificación de atributos cognitivos. Los bajos valores de $PVAF$ en los ítems 40 y 47 sugieren que los atributos cognitivos involucrados no son adecuados para explicar la varianza de las respuestas de los aspirantes.

Tabla 11*Comparación del PVAF de los ítems entre modelos de diagnóstico cognitivo*

Ítem	DINA	DINO	G-DINA	Ítem	DINA	DINO	G-DINA
i1	0.964	0.963	0.974	i26	.957	.979	.963
i2	0.970	0.937	0.970	i27	0.968	0.964	0.887
i3	0.937	0.924	0.896	i28	0.960	0.937	0.861
i4	0.993	0.989	0.985	i29	0.987	0.990	0.983
i5	0.992	0.996	0.995	i30	0.993	0.984	0.989
i6	0.996	0.996	0.997	i31	0.974	0.965	0.967
i7	0.981	0.974	0.958	i32	0.959	0.971	0.914
i8	0.984	0.982	0.970	i33	0.988	0.992	0.973
i9	0.931	0.908	0.896	i34	0.985	0.976	0.969
i10	0.958	0.989	0.974	i35	0.955	0.965	0.931
i11	0.993	0.994	0.998	i36	0.954	0.970	0.920
i12	0.985	0.972	0.997	i37	0.978	0.968	0.958
i13	0.987	0.979	0.995	i38	0.901	0.866	0.795
i14	0.981	0.937	0.974	i39	0.962	0.965	0.940
i15	0.996	0.994	0.996	i40	0.357	0.054	0.008
i16	0.837	0.918	0.723	i41	0.971	0.964	0.936
i17	0.985	0.992	0.993	i42	0.992	0.982	0.985
i18	0.993	0.990	0.981	i43	0.913	0.921	0.987
i19	0.962	0.976	0.938	i44	0.948	0.969	0.991
i20	0.990	0.989	0.991	i45	0.861	0.864	0.980
i21	0.905	0.946	0.984	i46	0.970	0.965	0.994
i22	0.990	0.965	0.982	i47	0.341	0.616	0.943
i23	0.991	0.962	0.991	i48	0.979	0.942	0.990
i24	0.959	0.967	0.933	i49	0.938	0.967	0.957
i25	0.981	0.891	0.937	i50	0.762	0.828	0.660

Nota. Las celdas resaltadas en color verde corresponden a valores de $PVAF > 0.95$, las celdas resaltadas en color naranja corresponden a valores de $PVAF < 0.95$, y las celdas resaltadas en rojo corresponden a valores de $PVAF < 0.5$.

Precisión en la clasificación de aspirantes

Para determinar la confiabilidad en la prueba, en los atributos y en las clases latentes que estimó cada MDC, se obtuvieron los índices de precisión y consistencia. La precisión es la probabilidad de clasificar con exactitud a un aspirante en función de sus respuestas en los ítems de una prueba, mientras que la consistencia hace referencia a la probabilidad de clasificar a un aspirante de manera estable y congruente en cada medición (Ciu et al., 2012).

Los resultados para la prueba en general fueron moderados (tabla 12). En el modelo DINA se obtuvo un índice de precisión de 0.532 y un índice de consistencia de 0.795. En el modelo DINO se obtuvo un índice de precisión de 0.674 y un índice de consistencia de 0.787. Y en el modelo G-DINA se obtuvo un índice de precisión de 0.670 y un índice de consistencia de 0.621. En general, la prueba es más consistente con el modelo DINA, pero es más precisa con el modelo DINO o G-DINA. En relación con la consistencia y precisión de los atributos cognitivos, el atributo $A1 = \text{Establecimiento de una única proporción comparando dos cantidades}$ obtuvo los índices más bajos de precisión y consistencia en los tres modelos de diagnóstico cognitivo. El resto de los atributos cognitivos obtuvieron índices de precisión y consistencia altos, siendo el modelo DINO el que estimó los valores más altos con respecto a los modelos DINA y G-DINA.

Tabla 12

Índices de precisión y consistencia de la prueba y los atributos con los modelos DINA, DINO y G-DINA

	DINA		DINO		G-DINA	
	P_a	P_c	P_a	P_c	P_a	P_c
Prueba	0.532	0.795	0.674	0.787	0.670	0.621
A1	0.675	0.904	0.790	0.940	0.808	0.794
A2	0.942	0.917	0.945	0.917	0.939	0.902
A3	0.940	0.910	0.957	0.931	0.937	0.900
A4	0.945	0.924	0.953	0.924	0.940	0.904
A5	0.942	0.918	0.941	0.918	0.949	0.926
A6	0.882	0.919	0.900	0.906	0.924	0.898
A7	0.942	0.919	0.944	0.916	0.932	0.894

Nota. $P_a = \text{Índice de precisión en la clasificación}$, $P_c = \text{Índice de consistencia en la clasificación}$.

Respecto a la clasificación de aspirantes en los perfiles o clases latentes, teóricamente los modelos clasifican a los aspirantes en 2^k clases latentes posibles, siendo k el número de atributos cognitivos presentes en la evaluación. Por lo tanto, cada uno de los modelos puede identificar hasta $2^7 = 128$ clases latentes para clasificar a los aspirantes. Se tomó como criterio la identificación de las clases latentes con un índice de precisión mayor

a cero y con más del 1 % de aspirantes clasificados.

De manera inicial, el modelo DINA clasificó a los aspirantes en 77 clases latentes, pero solo cuatro clases latentes clasificaron con precisión a más del 1 % de los aspirantes. El modelo DINO clasificó a los aspirantes en 87 clases latentes, pero solo siete clases latentes clasificaron con precisión a más del 1 % de los aspirantes. Y el modelo G-DINA clasificó a los aspirantes en 69 clases latentes, pero 13 clases latentes clasificaron con precisión a más del 1 % de los aspirantes. En total, los tres modelos clasificaron a la mayoría de los aspirantes en 15 clases latentes (tabla 13).

Un gran porcentaje de aspirantes no dominó ninguno de los atributos cognitivos subyacentes a la prueba, los cuales fueron clasificados en la clase [0000000], mientras que alrededor de una cuarta parte o menos de los aspirantes dominaron todos los atributos cognitivos, los cuales fueron clasificados en la clase [1111111]. Respecto a la clase latente [0000000], los modelos DINA, DINO y G-DINA clasificaron al 44.6 %, 54.1 % y 40.7 % de los aspirantes, respectivamente. En un menor porcentaje, los modelos DINA, DINO y G-DINA clasificaron al 26 %, 16.6 % y 16.4 % de los aspirantes en la clase latente [1111111], respectivamente. Ambas clases latentes obtuvieron índices de precisión elevados en los tres modelos. Por su parte, el resto de las clases latentes obtuvieron índices de precisión de moderados a bajos.

Tabla 13

Clases latentes identificadas con más de 1 % de aspirantes, por modelo de diagnóstico cognitivo DINA, DINO y G-DINA

Clase latente	DINA		DINO		G-DINA	
	Porcentaje	P_a	Porcentaje	P_a	Porcentaje	P_a
0000000	44.6	0.945	54.1	0.981	40.7	0.945
0010000	-	-	-	-	2.4	0.445
0010001	-	-	-	-	2.1	0.509
0100000	-	-	1.5	0.288	-	-
0110000	1.1	0.394	-	-	2.3	0.394
0110001	-	-	1.5	0.483	4.3	0.644
0111000	-	-	-	-	2.5	0.510
0111001	-	-	-	-	3.4	0.489
1000000	-	-	1.6	0.135	7.1	0.224
1000001	-	-	-	-	1.7	0.393
1100000	-	-	1.1	0.380	-	-
1111101	1.8	0.261	1.2	0.166	1.7	0.261
1111111	26.0	0.951	16.6	0.957	16.4	0.951

Nota. P_a = Índice de precisión en la clasificación. El guión (-) en las celdas indica que la clase latente no fue identificada por el modelo.

Estimación de los parámetros de los ítems de la prueba

Se obtuvieron las estimaciones de los parámetros de adivinación (g) y desliz (s) de cada ítem de la prueba, utilizando los modelos DINA, DINO y G-DINA. Según Rupp et al. (2010), un ítem ofrece información diagnóstica cuando presenta valores bajos tanto en adivinación como en desliz. Para determinar la calidad de los ítems, se adoptaron los criterios propuestos por Iaconangelo (2017): los ítems de alta calidad presentan valores de $g = s = 0.1$, los de calidad media tienen valores de $g = s = 0.2$, y los de calidad baja presentan valores de $g = s = 0.3$. La figura 6 y figura 7 muestran la dispersión de los parámetros de cada modelo, con un punto de corte en 0.3. Puede observarse que la distribución de los parámetros es similar en los tres modelos, identificándose ítems de calidad alta, media y baja. En particular, se presentaron valores del parámetro de desliz

más bajos con respecto al parámetro de adivinación en los tres MDC. Los índices de adivinación y desliz de cada MDC se encuentran en el anexo H.

Figura 6

Parámetros de adivinación de los modelos DINA, DINO y G-DINA

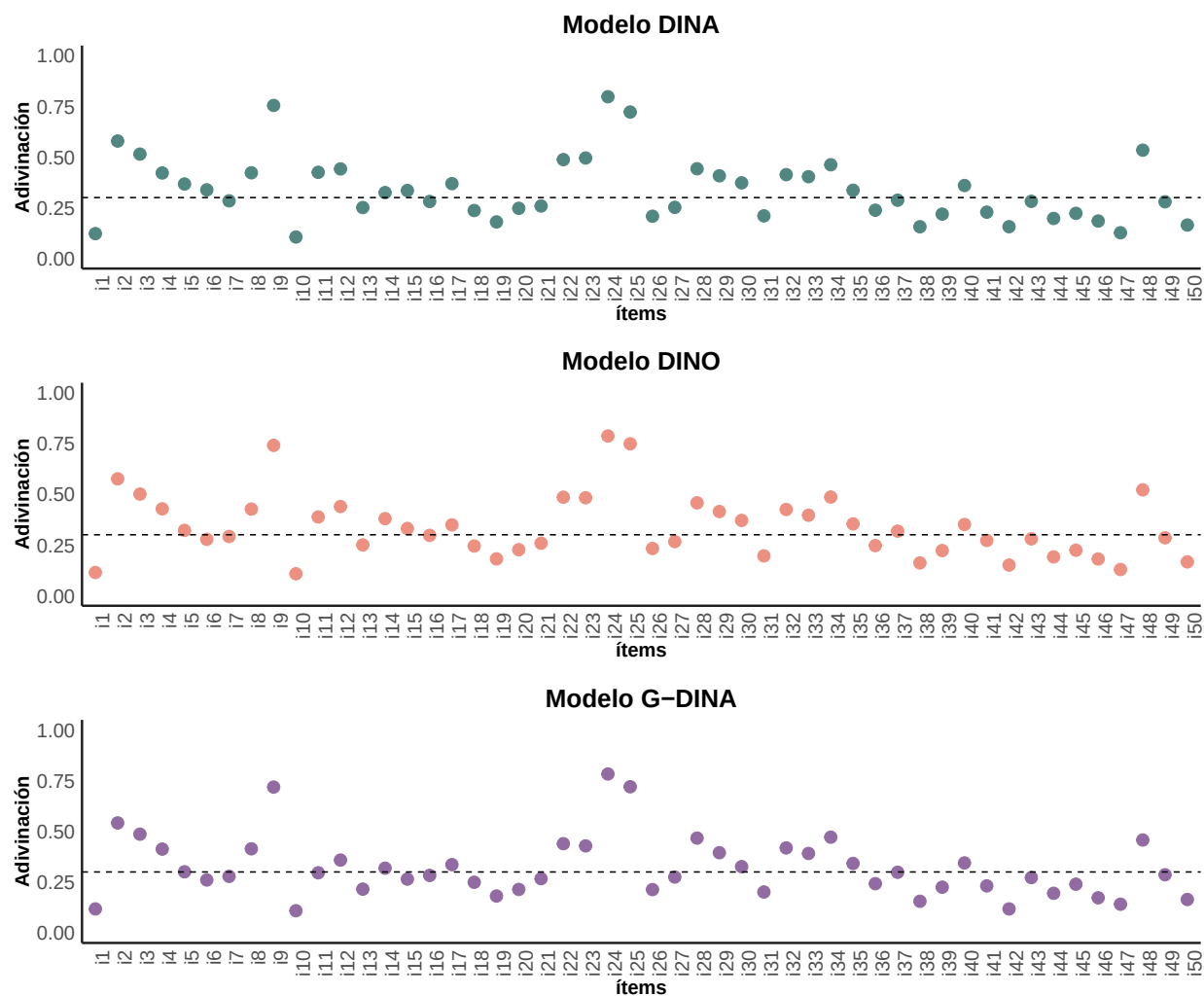
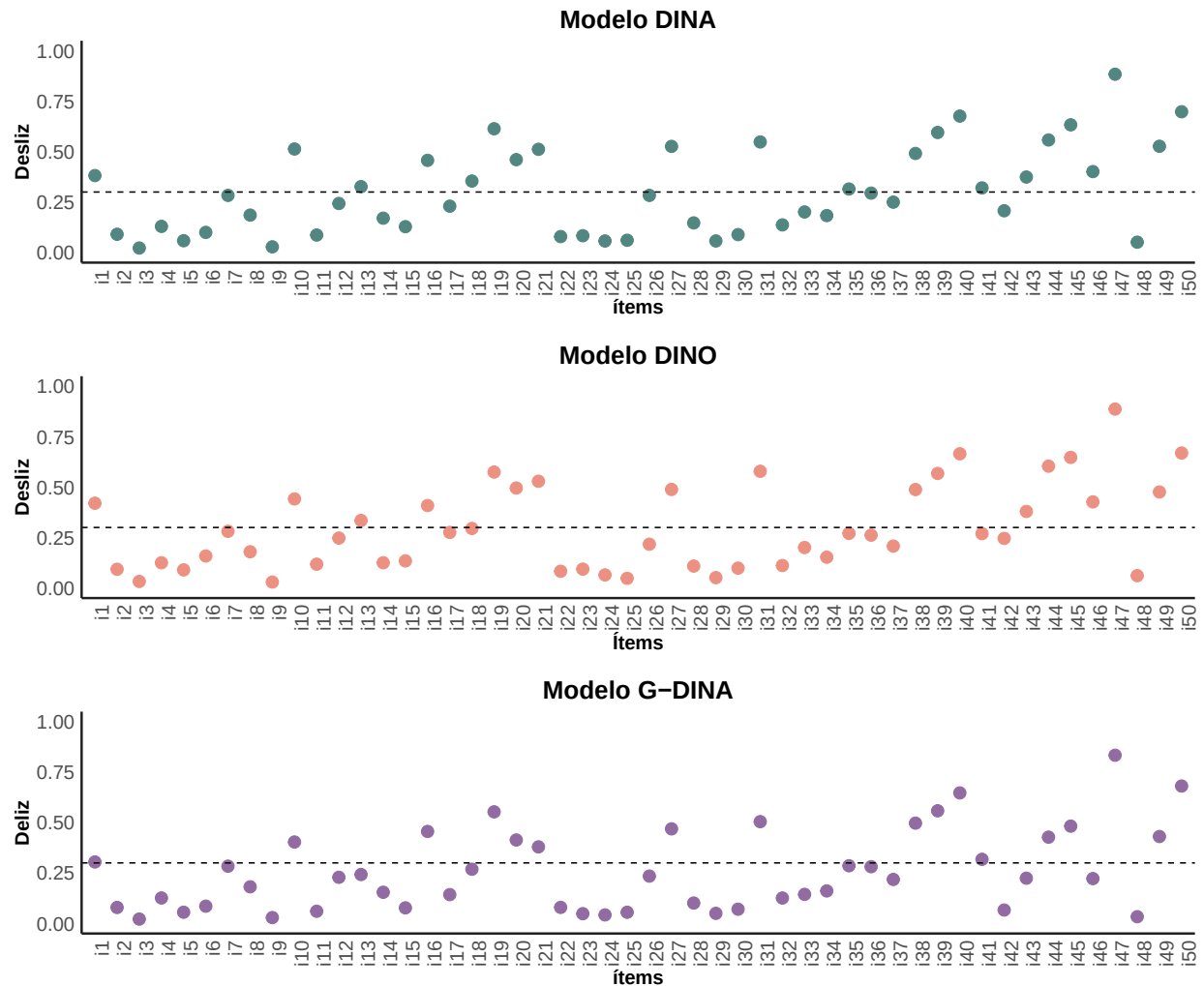


Figura 7
Parámetros de desliz de los modelos DINA, DINO y G-DINA



Los resultados obtenidos del ajuste de los modelos DINA, DINO y G-DINA muestran un buen desempeño de los tres modelos. Respecto a la validación de la matriz Q , los tres modelos presentaron valores de PVAF similares en los ítems de la prueba, destacando que el ítem 40 obtuvo valores muy bajos en los tres modelos. Este ítem en particular presentó problemas desde el análisis de los ítems realizado con los enfoques de la TCT y la TRI, lo que indica congruencia en los resultados obtenidos por los tres enfoques psicométricos.

El modelo DINA obtuvo los índices de ajuste relativo y absoluto más bajos, con un bajo índice de precisión en la prueba ($P_a = 0.532$), clasificando en un 53.2 % a los aspirantes de manera precisa; sin embargo, obtuvo un índice de consistencia moderado ($P_c = 0.795$), siendo este el más alto con respecto a los modelos DINO y G-DINA. En relación con la clasificación de aspirantes en las clases latentes, el modelo DINA clasificó a la mayoría de los aspirantes en cuatro clases latentes.

Por su parte, el modelo DINO obtuvo índices de ajuste relativo y absoluto ligeramente menores que el modelo DINA, índices de precisión y consistencia moderados en la prueba ($P_a = 0.674$, $P_c = 0.787$), y una clasificación de la mayoría de los aspirantes en 7 clases latentes. Por último, el modelo G-DINA obtuvo los valores más bajos de ajuste absoluto y relativo, índices moderados de precisión y consistencia en la prueba ($P_a = 0.670$, $P_c = 0.621$), ligeramente menores que el modelo DINO, y una clasificación de la mayoría de los aspirantes en 13 clases latentes, siendo el modelo que clasificó en menor porcentaje a los aspirantes en las clases latentes [000000] y [111111].

El modelo DINA (no compensatorio) asume que los atributos cognitivos interactúan de forma conjunta, por lo que se requiere del dominio completo de todos los atributos cognitivos requeridos en cada ítem para tener una alta probabilidad de responderlo correctamente. Por el contrario, el modelo DINO es menos restrictivo ya que asume que el bajo dominio de un atributo requerido en el ítem puede ser compensado por el alto dominio en otro. Por su parte, el modelo G-DINA, al ser generalizado, permite capturar interacciones de atributos cognitivos tanto compensatorias como no compensatorias, lo que lo hace un modelo más flexible. En la práctica, los siete atributos cognitivos evaluados en matemáticas no seguirán necesariamente la misma relación compensatoria o no compensatoria en todos los ítems de la prueba, lo que hace que el modelo G-DINA sea más adecuado para representar la naturaleza compleja de estos atributos. Por este motivo, el modelo G-DINA fue seleccionado para llevar a cabo el análisis.

Interpretación y comparación de resultados bajo los tres enfoques psicométricos

En este apartado se presenta la interpretación y comparación de los resultados obtenidos por la teoría clásica de los tests, la teoría de respuesta al ítem y el modelo de diagnóstico cognitivo G-DINA. En primer lugar, se presenta el dominio de los atributos cognitivos de manera general, y posteriormente, la comparación del desempeño de los estudiantes con los tres enfoques psicométricos.

Nivel de dominio de los atributos cognitivos en la prueba

El nivel de dominio de los atributos cognitivos en la prueba se determinó mediante la prevalencia, entendida como la proporción de aspirantes que dominan cada atributo cognitivo (Shi et al., 2021). En general, se presentó un bajo dominio en los siete atributos cognitivos en matemáticas (tabla 14). El atributo cognitivo *A5. Relación entre variables y su representación gráfica* fue el atributo con menor prevalencia, siendo dominado por el 22.6 % de los aspirantes. En contraste, el atributo cognitivo con una mayor prevalencia fue el *A3. Aplicación de operaciones aritméticas básicas* siendo dominado por el 42.5 % de los aspirantes.

Tabla 14
Prevalencia de los aspirantes en los atributos cognitivos

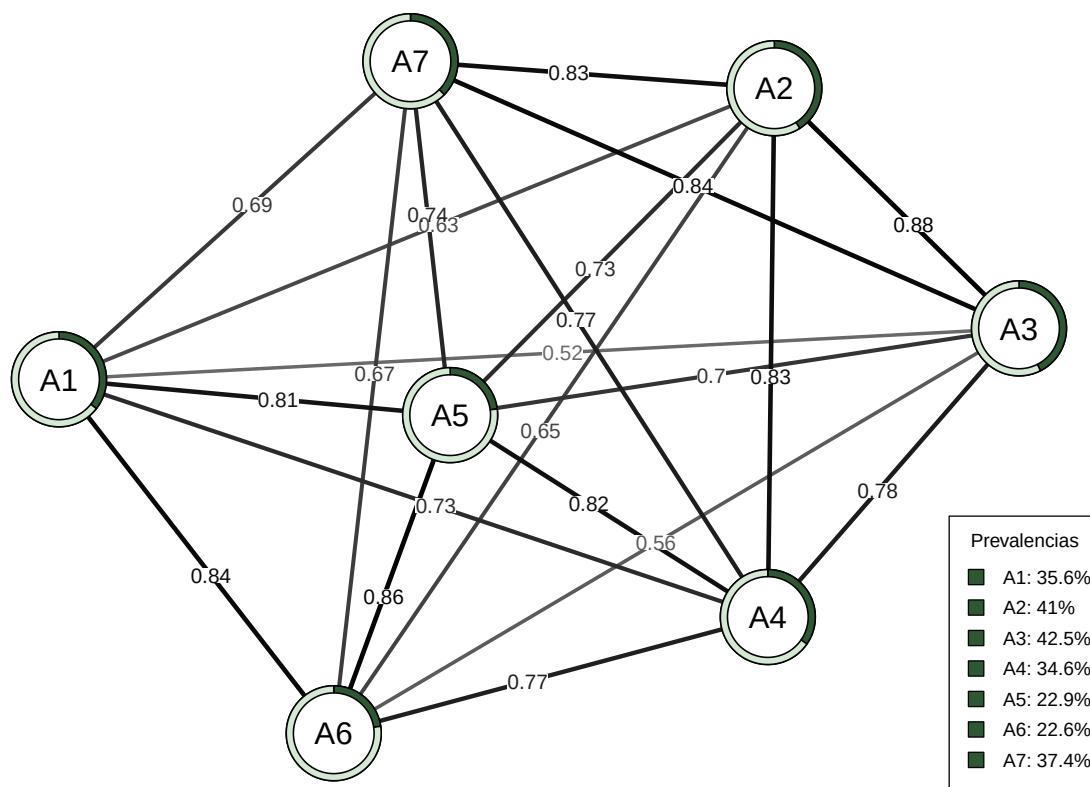
Atributo cognitivo	Prevalencia
A1. Establecimiento de una única proporción comparando dos cantidades	0.356
A2. Resolución de ecuaciones lineales	0.410
A3. Aplicación de operaciones aritméticas básicas	0.425
A4. Manipulación de expresiones algebraicas	0.346
A5. Relación entre variables y su representación gráfica	0.229
A6. Comprensión de fórmulas matemáticas	0.226
A7. Comprensión de problemas matemáticos contextualizados	0.374

De acuerdo con Liu et al. (2018), cuando se aplican modelos de diagnóstico cognitivo a pruebas diseñadas originalmente con un enfoque unidimensional, es esperado

que las correlaciones entre los atributos cognitivos sean altas o cercanas a 1, reflejando así dicha unidimensionalidad. Por este motivo, y para examinar las relaciones entre los atributos cognitivos, se calcularon las correlaciones entre los siete atributos cognitivos de matemáticas (figura 8). La correlación más alta fue entre los atributos *A2. Resolución de ecuaciones lineales* y *A3. Aplicación de operaciones aritméticas básicas* con un coeficiente de correlación de 0.876. Mientras que la correlación más baja fue entre los atributos *A1. Establecimiento de una única proporción comparando dos cantidades* y *A3. Aplicación de operaciones aritméticas básicas* obteniendo un coeficiente de 0.52. Puede observarse que, aunque los atributos cognitivos están correlacionados entre sí, ninguno de los coeficientes alcanza valores cercanos a 1. Este resultado sugiere la posibilidad de identificar más de una dimensión en la prueba de matemáticas, lo que implica que los atributos cognitivos pueden explicarse de manera independiente. Cabe destacar que las correlaciones altas entre atributos cognitivos podrían explicarse por relaciones jerárquicas, en las que un atributo cognitivo actúa como prerrequisito para dominar otro (Rupp et al., 2010; de la Torre y Douglas, 2004). Además, según Ma et al. (2020), dado que no todos los atributos cognitivos están fuertemente correlacionados entre sí, resulta útil reportar los perfiles de atributos para ofrecer información detallada.

Figura 8

Red de correlaciones y prevalencias de los atributos cognitivos subyacentes a la prueba de matemáticas del ExIES



Nota. A1= Establecimiento de una única proporción comparando dos cantidades, A2= Resolución de ecuaciones lineales, A3= Aplicación de operaciones aritméticas básicas, A4= Manipulación de expresiones algebraicas, A5= Relación entre variables y su representación gráfica, A6= Comprensión de fórmulas matemáticas y A7= Comprensión de problemas matemáticos contextualizados.

Desempeño de los aspirantes por enfoque psicométrico

Se obtuvieron los estadísticos descriptivos de los atributos cognitivos dominados por cada aspirante, los puntajes totales en la prueba y sus niveles de habilidad estimados (tabla 15). El 46 % de los aspirantes no dominó ningún atributo cognitivo y obtuvo una puntuación promedio en la prueba de 16 aciertos, con un nivel de habilidad promedio de -0.752 . Por su parte, el 17 % de los aspirantes dominó todos los atributos cognitivos y obtuvo una puntuación promedio en la prueba de 39 aciertos, con un nivel de habilidad

promedio de 1.492. Puede observarse que, a medida que aumenta el dominio de atributos dominados, los aspirantes tienden a obtener puntuaciones más altas en la prueba, lo que también refleja un mayor nivel de habilidad.

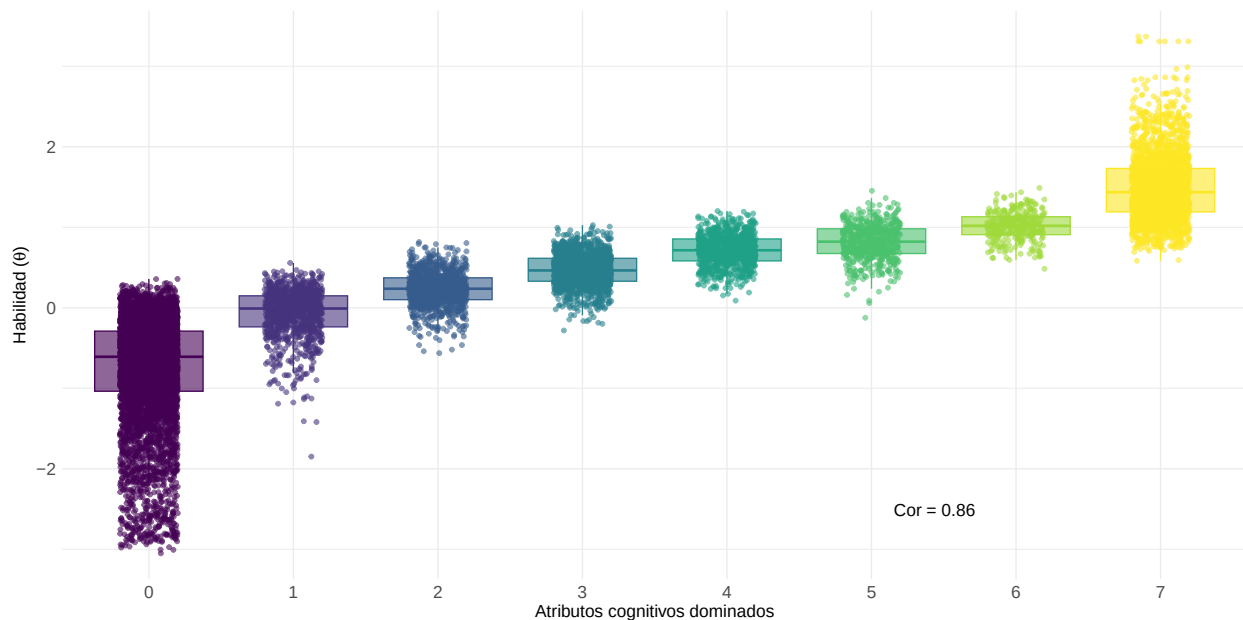
Tabla 15
Relación de los parámetros de los aspirantes por enfoque

Atributos dominados	n aspirantes	Enfoque	Mínimo	Q1	Mediana	Media	Q3	D.E.	Máximo
0	6,447 (46 %)	TCT	4	13	16	16	18	3.795	27
		TRI	-3.050	-1.036	-0.608	-0.752	-0.290	0.643	0.358
1	980 (7 %)	TCT	10	19	21	21	23	3.292	28
		TRI	-1.848	-0.237	-0.009	-0.069	0.148	0.309	0.557
2	1,102 (8 %)	TCT	15	22	24	24	26	2.822	32
		TRI	-0.563	0.101	0.236	0.234	0.372	0.212	0.816
3	1,378 (10 %)	TCT	18	24	26	26	28	2.898	35
		TRI	-0.282	0.329	0.464	0.467	0.614	0.207	1.026
4	836 (6 %)	TCT	21	28	29	29	31	2.698	37
		TRI	0.088	0.582	0.714	0.712	0.855	0.197	1.202
5	548 (4 %)	TCT	22	29	31	31	33	2.954	39
		TRI	-0.124	0.674	0.821	0.812	0.981	0.226	1.453
6	306 (2 %)	TCT	27	32	33	33	35	2.359	39
		TRI	0.484	0.908	1.018	1.013	1.131	0.176	1.489
7	2,356 (17 %)	TCT	29	35	38	39	42	4.269	50
		TRI	0.583	1.190	1.436	1.492	1.731	0.418	3.368

La figura 9 muestra la relación entre los atributos cognitivos dominados por los aspirantes y su nivel de habilidad. En el gráfico de cajas y bigotes puede observarse un ligero solapamiento en el número de atributos cognitivos dominados consecutivos, lo que indica que el nivel de habilidad de los grupos no está claramente diferenciado. Este patrón sugiere que aspirantes con niveles de habilidad similares pueden presentar combinaciones diferentes de atributos dominados. Por lo tanto, de acuerdo con Ma et al. (2020), la habilidad estimada por la TRI puede no identificar las fortalezas y áreas de oportunidad de los aspirantes tan bien como los perfiles de atributos cognitivos.

Figura 9

Relación entre los atributos cognitivos dominados por los aspirantes y su nivel de habilidad

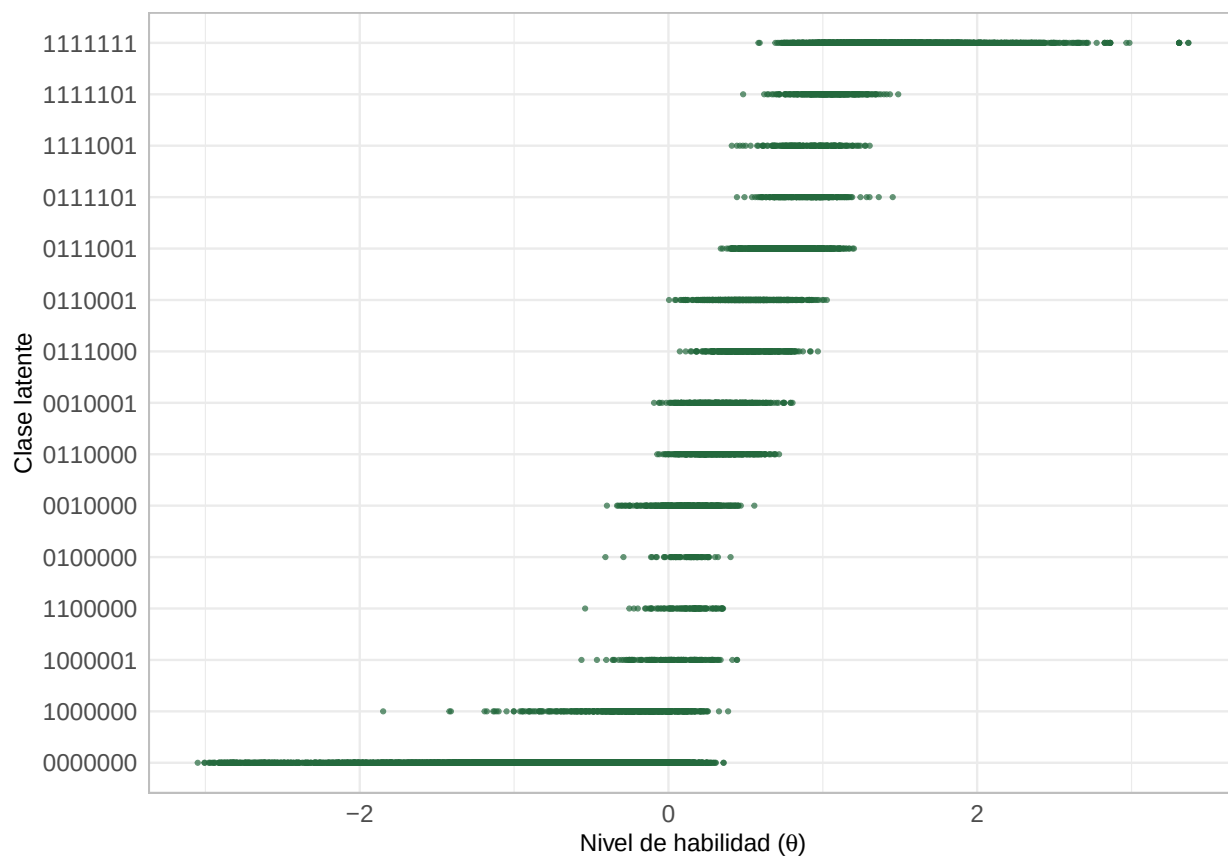


Nota. La habilidad (θ) fue estimada por el modelo 3PL y el dominio de los atributos cognitivos por el modelo G-DINA.

Se obtuvo la relación entre el nivel de habilidad y los patrones de atributos de los aspirantes (figura 10). Puede observarse que aspirantes con niveles de habilidad similares, especialmente alrededor de $\theta = 0$, dominan distintos conjuntos de atributos cognitivos. Este hallazgo refuerza la idea de que los aspirantes pueden tener diferentes fortalezas y áreas de oportunidad (Ma et al., 2020). En concreto, la figura muestra que los aspirantes que dominan únicamente el atributo *A1. Establecimiento de una única proporción comparando dos cantidades* (clase latente 1000000) tienen un menor nivel de habilidad en comparación con los aspirantes que dominan únicamente el atributo *A2. Resolución de ecuaciones lineales* (clase latente 0100000) o el atributo *A3. Aplicación de operaciones aritméticas básicas* (clase latente 0010000). Estos resultados sugieren que el dominio de atributos específicos conlleva a niveles de habilidad más altos.

Figura 10

Relación entre el nivel de habilidad y los patrones de atributos de los aspirantes



En la figura 11 se muestra la relación entre el puntaje obtenido y los perfiles de atributos de los aspirantes de acuerdo con el nivel de habilidad estimado, el cual se dividió en tres niveles de acuerdo con los siguientes rangos de θ :

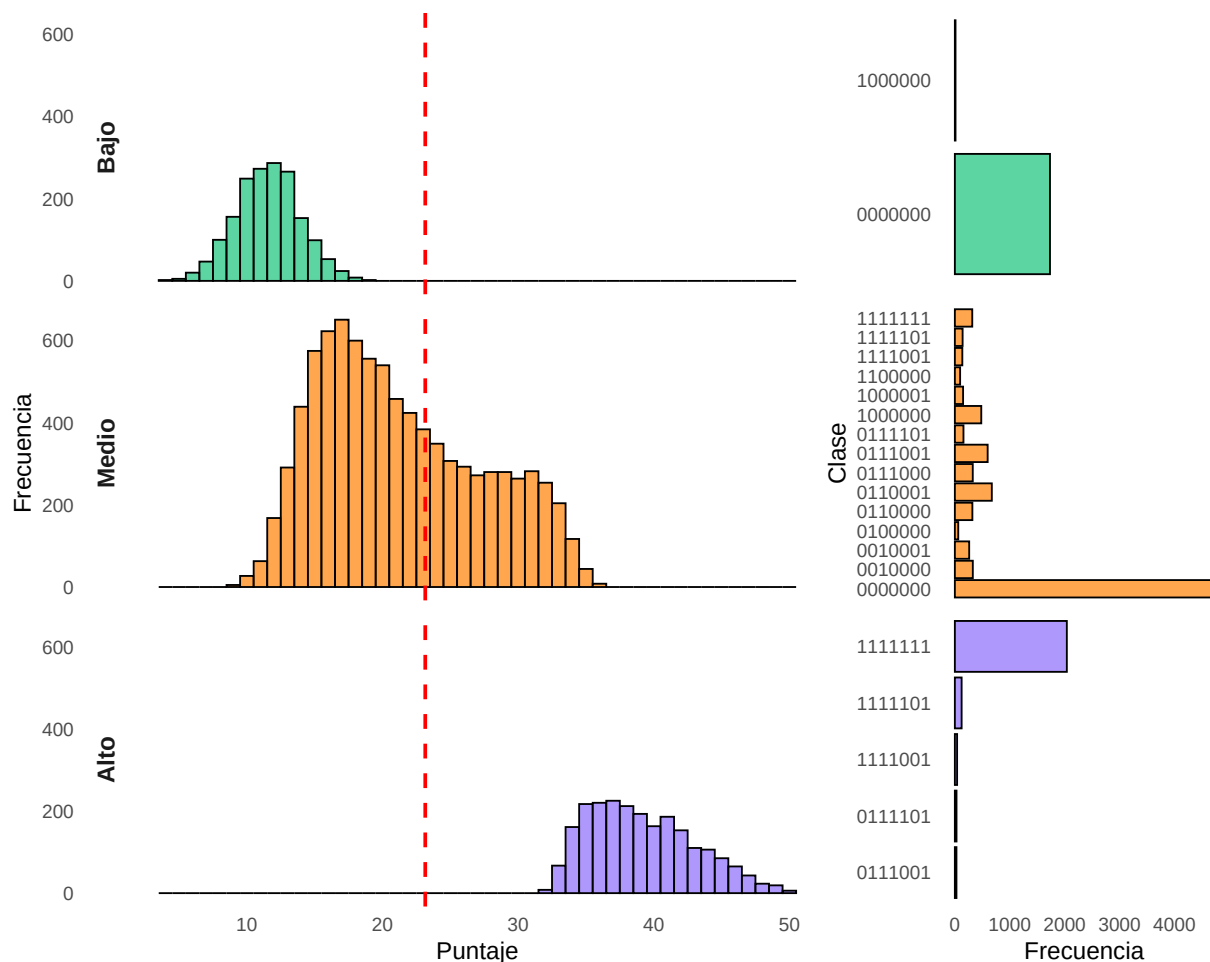
- Bajo = $\theta < -1$
- Medio = $-1 \leq \theta \leq 1$
- Alto = $\theta \geq 1$

Respecto a los puntajes en la prueba, se observa que conforme aumenta el nivel de habilidad, los aspirantes tienden a obtener puntajes más altos en la prueba. Aquellos con baja habilidad fueron clasificados en dos clases latentes, [0000000] y [1000000], siendo la

clase que no dominó ningún atributo la de mayor frecuencia. En el nivel de habilidad medio, se identificó una mayor variedad de clases latentes, predominando de igual manera la clase [0000000]. Por último, los aspirantes con una alta habilidad fueron clasificados hasta en cinco clases latentes, siendo la clase [1111111] la de mayor frecuencia. Cabe destacar que en el nivel más alto predominan los perfiles con mayor dominio de atributos. La figura muestra que los modelos de diagnóstico cognitivo proporcionan información diferente enfocada en los perfiles de clasificación, lo que representa información potencialmente útil para el establecimiento de estrategias remediales (Kunina-Habenicht et al., 2009).

Figura 11

Relación entre el puntaje y los perfiles de atributos de los aspirantes por nivel de habilidad



Nota. Se estimó el nivel de habilidad (θ) por el modelo 3PL y el dominio de los atributos cognitivos por el modelo G-DINA. El desempeño bajo corresponde a $\theta < -1$, el desempeño medio corresponde a $-1 < \theta < 1$ y el desempeño alto corresponde a $\theta > 1$. La línea punteada roja corresponde a la media de los puntajes ($\bar{x} = 23$).

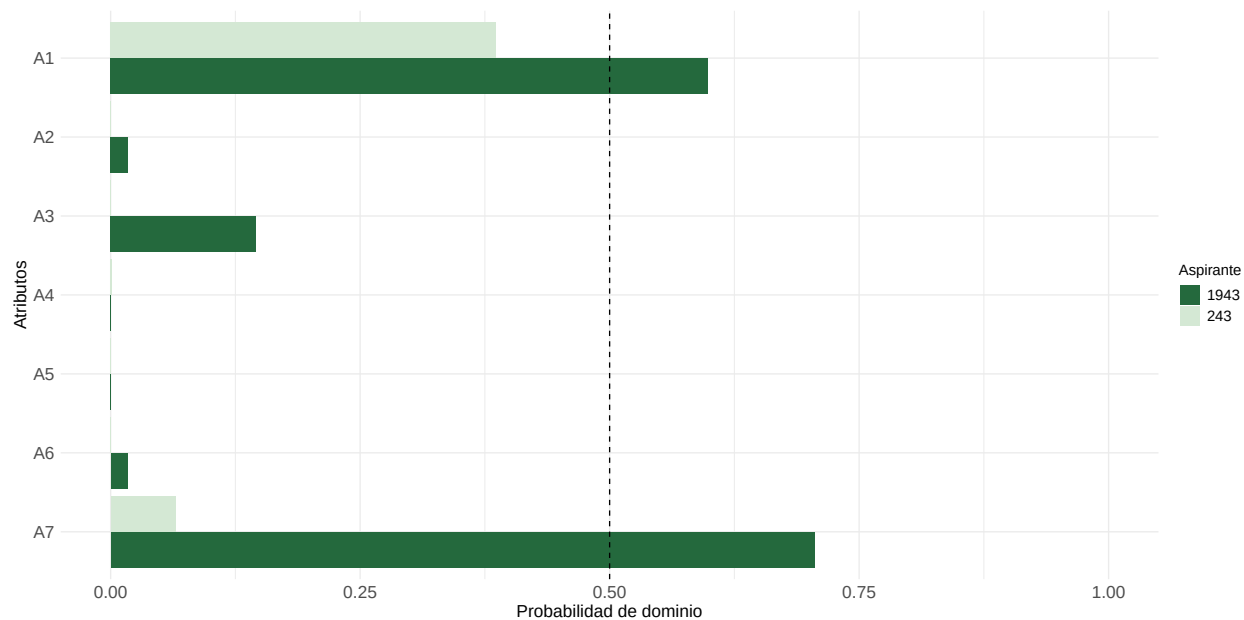
Por último, se realizaron comparaciones individuales entre los perfiles de atributos cognitivos, las puntuaciones en la prueba y los niveles de habilidad estimados de los aspirantes. La figura 12 muestra los perfiles de dos aspirantes, con ID 243 y 1943. Ambos aspirantes obtuvieron 15 aciertos en la prueba de matemáticas, pero presentaron distintos niveles de habilidad ($\theta_{243} = -0.79$; $\theta_{1943} = -0.23$) y perfiles de dominio de atributos cognitivos diferentes. El aspirante 243 no dominó ningún atributo cognitivo, mientras que

el aspirante 1943 dominó el establecimiento de una única proporción comparando dos cantidades (A1) y la comprensión de problemas matemáticos contextualizados (A7). Estos resultados son similares a los hallazgos reportados por Hemati y Bagahei (2020), los cuales sugieren que los aspirantes pueden tener diferentes fortalezas y áreas de oportunidad a pesar de tener puntajes similares en una prueba.

La figura 13 muestra los perfiles de dos aspirantes con el mismo nivel de habilidad ($\theta = 0.82$). El aspirante con ID 5110 obtuvo 32 aciertos en la prueba, mientras que el aspirante 9560 obtuvo 31 aciertos. A pesar de que ambos aspirantes obtuvieron puntajes similares en la prueba y poseen un mismo nivel de habilidad, su perfil de atributos cognitivos es diferente. El aspirante 5110 se clasificó en la clase latente [0111011], lo que indica que dominó la mayoría de los atributos cognitivos a excepción del establecimiento de una única proporción comparando dos cantidades (A1) y la relación entre variables y su representación gráfica (A5). Por su parte, el aspirante 9560 se clasificó en la clase latente [1101101], lo que indica que dominó la mayoría de los atributos a excepción de la aplicación de operaciones aritméticas básicas (A3) y la comprensión de fórmulas matemáticas (A6). Estos resultados pueden dar cuenta de las diferencias individuales de los aspirantes y sus necesidades educativas.

Figura 12

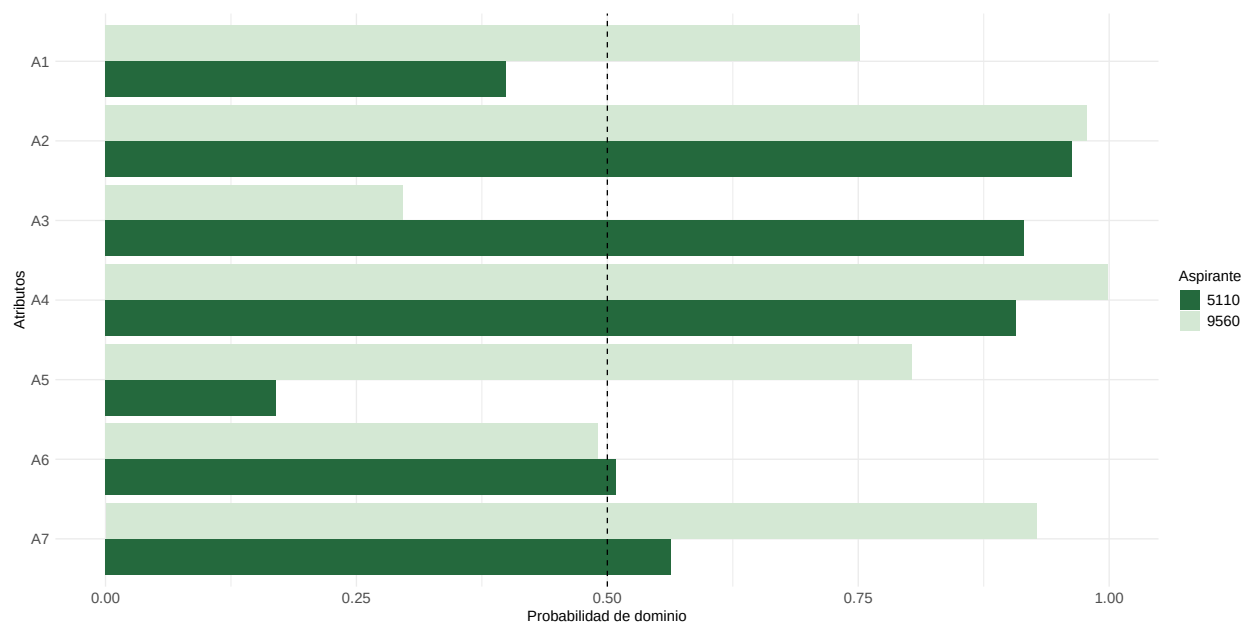
Aspirantes con el mismo puntaje en la prueba (15 aciertos)



Nota. El aspirante 243 obtuvo un nivel $\theta = -0.79$ y se clasificó en la clase latente [0000000]. El aspirante 1943 obtuvo un nivel $\theta = -0.23$ y se agrupó en la clase latente [1000001].

Figura 13

Aspirantes con el mismo nivel de habilidad ($\theta = 0.82$)



Nota. El aspirante 5110 obtuvo 32 aciertos en la prueba y se clasificó en la clase latente [0111011]. El aspirante 9560 obtuvo 31 aciertos en la prueba y se clasificó en la clase latente [1101101].

CAPÍTULO V. Discusión y conclusiones

El presente capítulo está dividido en tres apartados principales. En primer lugar, se discuten los hallazgos obtenidos a lo largo del proyecto de tesis, enfocándose en su relación con los objetivos planteados y con la literatura previa. En segundo lugar, se describen las limitaciones encontradas durante el desarrollo del proyecto. Por último, se presentan las futuras líneas de investigación que surgen a partir de este trabajo, resaltando su relevancia para el campo de estudio.

Discusión de hallazgos

La aplicación de modelos de diagnóstico cognitivo (MDC) a pruebas desarrolladas bajo enfoques psicométricos tradicionales, conocida como *retrofitting*, es una práctica cada vez más explorada para generar información diagnóstica sin necesidad de crear nuevos instrumentos (Liu et al., 2018). En este proyecto, el *retrofitting* fue implementado para responder a la pregunta de investigación: ¿En qué medida los modelos de diagnóstico cognitivo pueden proporcionar información complementaria con fines diagnósticos y formativos para la prueba de matemáticas del Examen de Ingreso a la Educación Superior (ExIES)?

El primer objetivo consistió en determinar la calidad técnica de la prueba de mediante los enfoques de la teoría clásica de los tests (TCT) y la teoría de respuesta al ítem (TRI). Los resultados indicaron que la prueba mide adecuadamente el desempeño de los aspirantes, especialmente aquellos con niveles medios y altos de habilidad. No obstante, se observó que los puntajes y niveles de habilidad promedio se encuentran ligeramente por debajo de la media, lo que es congruente con desempeños reportados en matemáticas en evaluaciones nacionales e internacionales como PISA, PLANEA y ENLACE (OCDE, 2023; Osuna y Díaz, 2019; Sánchez-Luján y Moreno, 2019; Sosa et al., 2014; Torres y Campos, 2015). El análisis también identificó dos ítems con mal funcionamiento según los criterios psicométricos: los ítems 24 y 40. En el caso del ítem 40, su baja discriminación y problemas

detectados en el índice de dificultad fueron consistentes entre los enfoques de la TCT, la TRI y los MDC. Esto destaca la importancia de revisar y ajustar los ítems problemáticos para mejorar la calidad de la prueba, en línea con lo sugerido por Baker (2001) sobre la necesidad de eliminar o modificar ítems que comprometan la confiabilidad y validez del instrumento.

El segundo objetivo se centró en determinar el perfil de atributos cognitivos que subyacen al dominio de matemáticas de los aspirantes. Para ello, se aplicaron los modelos DINA, DINO y G-DINA. Aunque los tres modelos presentaron un ajuste aceptable, el modelo G-DINA demostró ser el más adecuado debido a su capacidad para capturar relaciones tanto compensatorias como no compensatorias entre los atributos. Esta flexibilidad es crucial, ya que los atributos evaluados en matemáticas no necesariamente siguen una relación estrictamente jerárquica o independiente, como lo destacan estudios previos (de la Torre y Douglas, 2004). Cabe destacar que los tres modelos obtuvieron un mayor porcentaje de aspirantes clasificados en lo que Lee y Wasaki (2009b) denominan “perfiles planos”. Es decir, la clase latente en donde no se domina ningún atributo cognitivo [0000000] y la clase latente en donde se dominan todos los atributos cognitivos [1111111]. Investigaciones previas que aplicaron modelos de diagnóstico cognitivo a pruebas de ingreso a la universidad también reportaron resultados consistentes con los perfiles planos al ajustar el modelo G-DINA (Hemati y Baghaei, 2020; Ravand, 2016; Ravand y Robitzsch, 2018).

El tercer objetivo fue describir el perfil de atributos cognitivos con respecto al nivel de habilidad de los aspirantes. Los resultados evidenciaron que aspirantes con un mismo puntaje total y nivel de habilidad pueden tener diferentes perfiles de dominio de atributos. Este hallazgo resalta las diferencias individuales y subraya el potencial de los MDC para identificar fortalezas y áreas de oportunidad específicas en cada estudiante. En línea con lo mencionado por Huff y Goodman (2007), esta información puede ser útil para diseñar estrategias pedagógicas personalizadas que atiendan las necesidades específicas de los aspirantes.

Los MDC son útiles para respaldar la toma de decisiones de manera informada y personalizada, ya que proporcionan información detallada y precisa de las fortalezas y áreas de oportunidad de los estudiantes. Junto con la información que puede obtenerse de los enfoques tradicionales como la TCT y la TRI, los MDC permiten una representación más completa de los conocimientos y habilidades de los estudiantes. Los resultados obtenidos a través de estos modelos también ofrecen información que puede utilizarse para optimizar el proceso de selección de aspirantes, asegurando que los aspirantes admitidos cuenten con las habilidades necesarias para tener éxito en su vida académica.

La aplicación de los MDC en los procesos de admisión a la universidad requiere de la aceptación gradual por parte de las instituciones de educación superior. La investigación sobre MDC se ha centrado más en su evaluación metodológica que en el uso de sus resultados (Williamson, 2023). Existe poca evidencia empírica que respalde su uso, incluso para modificar las prácticas del docente y de los estudiantes en el aula, lo que genera incertidumbre entre los actores educativos y los desarrolladores de pruebas para tomar decisiones de alto impacto basadas en estos modelos (Sessoms y Henson, 2018). Lo anterior resalta la relevancia de investigaciones empíricas que evalúen la utilidad de los MDC en contextos universitarios, asegurando la confiabilidad de sus resultados y su contribución a la mejora de los procesos de admisión y de acompañamiento a estudiantes.

Limitaciones del estudio

En el proceso de determinación de la matriz Q, se identificaron las siguientes limitaciones:

1. Validación de la matriz Q. Los atributos cognitivos fueron especificados originalmente por Gierl et al. (2009) para la prueba SAT y no fueron validados mediante protocolos verbales en una población mexicana. Las diferencias culturales y educativas entre Estados Unidos y México pueden influir en la aplicabilidad de dichos atributos.
2. Aunque se recibieron recomendaciones para mejorar la matriz Q, no todos los

descriptores de los atributos cognitivos fueron revisados por al menos dos jueces, debido a la falta de respuesta de uno de los participantes del comité. Esto impidió calcular una proporción de acuerdo para los descriptores.

3. Sobrecarga cognitiva. El comité elaboró 129 descriptores para 21 atributos cognitivos, lo que pudo generar una sobrecarga cognitiva para los expertos que utilizaron estos descriptores en el comité de determinación de la Matriz Q. Es necesario simplificar y definir los atributos de manera más clara para facilitar su uso y aplicación.

Futuras líneas de investigación

A partir de este estudio, se proponen las siguientes líneas de investigación:

- a) Validación de la matriz Q mediante protocolos verbales adaptados al contexto mexicano para analizar cómo los aspirantes resuelven problemas matemáticos. Esta validación permitirá la construcción de un modelo cognitivo. Así como para obtener interpretaciones más precisas de los perfiles de dominio de atributos y generar información para el desarrollo de nuevos ítems.
- b) Generación de evidencias de validez mediante la comparación de los resultados de la prueba de matemáticas del ExIES con otras medidas de desempeño en matemáticas. Este estudio proporcionaría información adicional para asegurar la calidad de la prueba, así como para reforzar la utilidad e interpretación de sus resultados.
- c) Acompañar los procesos de selección y admisión de estudiantes con estudios comparativos del desempeño en matemáticas de aspirantes admitidos y no admitidos por los diferentes programas educativos, para identificar fortalezas y áreas de oportunidad específicas.

Referencias

- Abad, F., Olea, J., Ponsoda, V., y García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Síntesis.
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. (2018). *Estándares para Pruebas Educativas y Psicológicas*. American Educational Research Association.
<https://doi.org/10.2307/j.ctvr43hg2>
- Atkinson, R. C., y Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665–676. <https://doi.org/10.3102/0013189X09351981>
- Attorresi, H. F., Lozzia, G. S., Abal, F. J., y Galibert, M. S. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos. *Revista Argentina de Clínica Psicológica*, 13(1), 179-188.
- Awopeju, O. A., y Afolabi, E. R. I. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal*, 12 (28), 263-284.
<https://doi.org/10.19044/esj.2016.v12n28p263>
- Bennett, R. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5-25.
<https://doi.org/10.1080/0969594X.2010.513678>
- Berg, D., Schaughency, E., van der Meer, J., y Smith, J. (2018). Using Classical Test Theory in higher education. (pp. 252-269). En C. Secolsky y D. Denison. *Handbook on Measurement, Assessment, and Evaluation in Higher Education*. Routledge.
- Black, P., Harrison, C., Lee, C., Marshall, B., y Wiliam, D. (2003). *Assessment for learning. Putting it into practice*. Open University Press.
- Bond, T. G., y Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

-
- Buendía, M. A., y Rivera, R. (2010). Modelo de selección para el ingreso a la Educación Superior: El caso de la UACH. *Revista de la Educación Superior*, 39(156), 55-72.
- Burnham, K. P., y Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2da ed.). Springer.
- Crocker, L., y Algina, J. (2006). Introduction to classical and modern test theory (2da ed.). Cengage Learning.
- Centro Nacional de Evaluación para la Educación Superior (2023). *Guía para el sustentante EXANI-II*. <https://online.flippingbook.com/view/278435445/6/>
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1–29.
<https://doi.org/10.18637/jss.v048.i06>
- Chin, H., y Chew, C. M. (2023). Validation of cognitive models for subtraction of time involving years and centuries. *International Journal of Assessment Tools in Education*, 10(2), 175-196. <https://doi.org/10.21449/ijate.1160120>
- College Board (2020). *Guía de estudio. Prueba de aptitud académica, Puerto Rico, College Board*. https://latam.collegeboard.org/wp-content/uploads/2020/04/Guia_de_estudios_PAA_WEB_.pdf
- College Board (2023a). *What's on the SAT?*.
<https://satsuite.collegeboard.org/sat/whats-on-the-test>
- College Board (2023b). *The PSAT/NMSQT*. <https://satsuite.collegeboard.org/psat-nmsqt>
- Cortés, A. y Palomar, J. (2008). El proceso de admisión como predictor del rendimiento académico en la educación superior. *Universitas Psychologica*, 7(1), 197-213.
- Cronbach, L. J. (1963). Course Improvement through Evaluation. *Teachers College Record*, 64(8), 1–13. <https://doi.org/10.1177/016146816306400802>
- Cui, Y., Gierl, M.J. y Chang, H.-H. (2012), Estimating Classification Consistency and Accuracy for Cognitive Diagnostic Assessment. *Journal of Educational Measurement*, 49, 19-38. <https://doi.org/10.1111/j.1745-3984.2011.00158.x>
-

-
- de la Torre, J. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115–130.
<http://www.jstor.org/stable/40263519>
- de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika* 76, 179–199. <https://doi.org/10.1007/s11336-011-9207-7>
- de la Torre, J., y Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333–353. doi:10.1007/bf02295640
- de la Torre, J. y Chiu, C-Y. (2016). A General Method of Empirical Q-matrix Validation. *Psychometrika*, 81, 253-273.
- de la Torre, J., Carmona, G., Kieftenbeld, V., Tjoe, H., y Lima, C. (2016). Diagnostic Classification Models and Mathematics Education Research: Opportunities and Challenges. *Journal for Research in Mathematics Education. Monograph*, 15, 53-72.
- de la Torre, J., y Sorrel, M. (2023). Cognitive Diagnosis Models. En F. Ashby, H. Colonius, y E. Dzhaferov (Eds.), *New Handbook of Mathematical Psychology* (Cambridge Handbooks in Psychology, pp. 385-420). Cambridge: Cambridge University Press.
doi:10.1017/9781108902724.010
- Desjardins, C., y Bulut, O. (2018). *Handbook of Educational Measurement and Psychometrics Using R*. Taylor & Francis Group.
- Embretson, S. E., y Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum.
- Fernández, M., Mijares, B., Álvarez, J. L., y León, A. (2015). Habilidades académicas en estudiantes de nuevo ingreso a la Universidad del Zulia. *Revista de Ciencias Sociales*, 21(2), 361-372.
- Feng, X. (2023). Important Evaluation Model of University Mathematics Curriculum. *Highlights in Science, Engineering and Technology*, 49, 310-314.
<https://doi.org/10.54097/hset.v49i.8523>.
- Finch, W. H., y French, B. F. (2019). *Educational and psychological measurement*.
-

Routledge.

Fischer, G. H., y Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. Springer-Verlag.

Geisinger, K. F., Bracken, B. A., Carlson, J. F., Hansen, J.-I. C., Kuncel, N. R., Reise, S. P., y Rodríguez, M. C. (Eds.). (2013). *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology*. American Psychological Association.

<https://doi.org/10.1037/14047-000>.

Gierl, M. J., Tan, X., y Wang, Ch. (2005). Identifying content and cognitive dimensions on the SAT. Research Report n° 2005-11. College Board.

Gierl, M. J., Zheng, Y., y Cui, Y. (2008). Using the Attribute Hierarchy Method to Identify and Interpret Cognitive Skills that Produce Group Differences. *Journal of Educational Measurement*, 45, 65-89. <https://doi.org/10.1111/j.1745-3984.2007.00052.x>

Gierl, M. J., Leighton, J. P., Wang, C., Zhou, J., Gokiert, R., y Tan, A. (2009). Validating cognitive models of task performance in algebra on the SAT (College Board Research 2009-3). The College Board.

Haertel, E. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 301-321.

Haertel, E., y Calfee, R. (1983). School Achievement: Thinking about What to Test. *Journal of Educational Measurement*, 20(2), 119-132.

<http://www.jstor.org/stable/1434660>

Haladyna, T. M. y Rodríguez, M. C. (2013). *Developing and Validating Test Items*. Routledge.

Hemati, S. J., y Baghaei, P. (2020). A Cognitive Diagnostic Modeling Analysis of the English Reading Comprehension Section of the Iranian National University Entrance Examination. *International Journal of Language Testing*, 10(1), 11-32.

Hu, T., Yang, J., Wu, R. y Wu, X. (2021). An International Comparative Study of

-
- Students' Scientific Explanation Based on Cognitive Diagnostic Assessment. *Frontiers in Psychology*, 12. doi:10.3389/fpsyg.2021.795497.
- Hu, L.-T., y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Huff, K., y Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. En J. P. Leighton y M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 19-60). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511611186.002>
- Iaconangelo, C. (2017). *Uses of Classification Error Probabilities in the Three-Step Approach to Estimating Cognitive Diagnosis Models*. (Unpublished doctoral dissertation). New Brunswick, Rutgers University.
- Instituto Nacional para la Evaluación de la Educación (2011). Evaluación de los aprendizajes en el aula. Opiniones y prácticas de docentes de primaria en México.
- Jimenez-Moreno, J. A. (2021). Mecanismos de ingreso de las universidades públicas estatales en México: un análisis bajo el concepto de equidad. *Revista Iberoamericana de Educación Superior*, 35(12), <https://orcid.org/0000-0003-0704-7883>.
- Jornet, J. M., Leyva, B. y Sánchez, P. (2009). Dimensiones de clasificación de los procesos de evaluación educativa (pp. 65-75). En J. M. Jornet y R. Leyva. *Conceptos, metodología y profesionalización en la evaluación educativa*. INETE.
- Kim, A. Y. (2014). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, 32(2), 227-258.
- Krishnan, V. (2013). *The Early Child Development Instrument (EDI): An item analysis using Classical Test Theory (CTT) on Alberta's data*.
<https://www.ualberta.ca/en/community-university-partnership/media-library/community-university-partnership/research/ecmap->
-

reports/ediitemanalysisctt.pdf

- Kunina-Habenicht, O., Rupp, A. A., Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies in Educational Evaluation*, 35(2-3), 64-70. <https://doi.org/10.1016/j.stueduc.2009.10.003>.
- Larrondo, T., Lara, M., Figueroa, C., Rojas, M., y Caro, A. (2018). Desarrollo de habilidades básicas en lenguaje y matemáticas en egresados de Pedagogía. Un estudio comparativo. *Calidad en la Educación*, (27), 150-176.
- Lau, A. (2016). Formative good, summative bad? - A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40(4), 509-525, DOI: 10.1080/0309877X.2014.984600
- Lee, YS., de la Torre, J. y Park, Y.S. (2012). Relationships between cognitive diagnosis, CTT, and IRT indices: an empirical investigation. *Asia Pacific Education Review*, 13, 333-345. <https://doi.org/10.1007/s12564-011-9196-3>
- Lee Y.-W., Sawaki Y. (2009a). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, 6(3), 239-263.
- Lee, Y.-W., y Sawaki, Y. (2009b). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, 6, 172-189.
doi:10.1080/15434300902985108
- Leenen, I. (2013). Virtudes y limitaciones de la teoría de respuesta al ítem para la evaluación educativa en las ciencias médicas. *Investigación en Educación Médica*, 3(9), 40-55.
- Leighton, J., y Gierl, M. (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge: Cambridge University Press.
- Liu, R., Huggins-Manley, A. C., y Bulut, O. (2018). Retrofitting Diagnostic Classification Models to Responses From IRT-Based Assessment Forms. *Educational and Psychological Measurement*, 78(3), 357-383.

-
- <https://doi.org/10.1177/0013164416685599>
- Liu, R., y Shi, D. (2020). Using Diagnostic Classification Models in Psychological Rating Scales. *The Quantitative Methods for Psychology*, 16(5), 442-456.
- Luna, D. A., (2017). Modelos de diagnóstico cognitivo: fundamentos, didáctica y aplicaciones en ciencias del comportamiento. *Salud y Sociedad*, 8(1), 68-80.
- Ma, W., Minchen, N. y de la Torre, J. (2020). Choosing between CDM and Unidimensional IRT: The Proportional Reasoning Test Case. *Measurement: Interdisciplinary Research and Perspectives*, 18(2), 87-96. <https://doi.org/10.1080/15366367.2019.1697122>
- Ma, W., y de la Torre, J. (2020). GDINA: An R Package for Cognitive Diagnosis Modeling. *Journal of Statistical Software*, 93(14), 1-26. <https://doi.org/10.18637/jss.v093.i14>
- Martínez-Arias, R. (2014). *Psicometría: Teoría de los tests psicológicos y educativos*. Síntesis.
- Madaus, G. F. y Stufflebeam, D. L. (2002). Program Evaluation: A Historical Overview (pp. 3-18). En D. L., Stufflebeam, G. F., Madaus y T., Kellaghan. *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*. Kluwer Academic Publishers.
- McNaught, K. y Hoyne, G. (2011). Mathematics for first-year success. https://www.researchgate.net/publication/254724640_Mathematics_for_first-year_success
- Mei, H. y Chen, H. (2022). Assessing Students' Translation Competence: Integrating China's Standards of English With Cognitive Diagnostic Assessment Approaches. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.872025>
- Métrica Educativa (2020a). Examen de Competencias Básicas (Excoba). <https://metrica.edu.mx/wp-content/uploads/2020/05/Excoba-Examen-de-Competencias-Basicas.pdf>
- Métrica Educativa (2020b). Exámenes diagnósticos. <https://metrica.edu.mx/estudiantes/examenes-diagnosticos/>
-

-
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., y Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*.
<https://timssandpirls.bc.edu/timss2019/international-results/>
- Mullis, I.V.S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K.A., y Wry, E. (2023). *PIRLS 2021 International Results in Reading*. Boston College, TIMSS y PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- Nichols, P. D. (1994). A Framework for Developing Cognitively Diagnostic Assessments. *Review of Educational Research*, 64(4), 575-603.
- Organista-Sandoval, J., McAnally-Salas, L., y Henríquez-Ritchie, P. (2012). Clasificación de estudiantes de nuevo ingreso a una universidad pública, con base a variables de desempeño académico, uso de tecnología digital y escolaridad de los padres. *Revista Electrónica de Investigación Educativa*, 14, 34-55.
- Organización para la Cooperación y el Desarrollo Económicos (2019). *PISA 2018 Results. What students know and can do. Volume I*.
<https://www.oecd.org/pisa/publications/pisa-2018-results-volume-i-5f07c754-en.htm>
- Organización para la Cooperación y el Desarrollo Económicos (2023). PISA 2022 Country Notes: México. PISA, OECD.
https://www.oecd.org/pisa/publications/Countrynote_MEX_Spanish.pdf
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2021). *Evaluación formativa: Una oportunidad para transformar la educación en tiempos de pandemia*. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE).
- Osterlind, S. J. y Wang, Z. (2018). Item Response Theory in measurement, assessment, and evaluation for Higher Education. (pp. 270-283). En C. Secolsky y D. Denison. *Handbook on Measurement, Assessment, and Evaluation in Higher Education*. Routledge
- Osuna, C. y Díaz, K. (2019). El logro de los aprendizajes en matemáticas en PISA, ENLACE y PLANEA en adolescentes mexicanos. Un análisis retrospectivo. *Archivos*
-

-
- analíticos de políticas educativas*, 28(28).
- <https://repositorio.cetys.mx/handle/60000/856>
- Phillips, D. C. (2018). The many functions of evaluation in education. *Education Policy Analysis Archives*, 26(46). <http://dx.doi.org/10.14507/epaa.26.3811>
- Popham, J. W. (2013). *Evaluación transformativa. El poder transformador de la evaluación educativa*. Narcea.
- Ramos, G., Perales, M. J. y Pérez Carbonell, M. A. (2009). El concepto de evaluación educativa. (pp. 57-64). En J. M. Jornet y R. Leyva. *Conceptos, metodología y profesionalización en la evaluación educativa*. INETE.
- Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, 34(8), 782–799.
- <https://doi.org/10.1177/0734282915623053>
- Ravand, H., y Robitzsch, A. (2018). Cognitive diagnostic model of best choice: a study of reading comprehension. *Educational Psychology*, 38(10), 1255–1277.
- <https://doi.org/10.1080/01443410.2018.1489524>
- Reckase, M. D. (1997). The Past and Future of Multidimensional Item Response Theory. *Applied Psychological Measurement*, 21(1), 25–36.
- <https://doi.org/10.1177/0146621697211002>
- Robitzsch, A., Kiefer, T., Groß, J., Ünlü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. *Journal of Statistical Software*, 74(2), 1–24.
- [doi:10.18637/jss.v074.i02](https://doi.org/10.18637/jss.v074.i02).
- Roblero Ochoa, S. R., Jiménez Pirrón, T. de los Ángeles, Mandujano Trujillo, Z. P., Velasco Martínez, R. M., y Castro Padilla, I. (2023). Evaluación de egresados de una facultad de Medicina Humana a través de resultados en el EGEL y puntaje ENARM. *Espacio I+D, Innovación más Desarrollo*, 12(34).
- <https://doi.org/10.31644/IMASD.34.2023.a04>
- Rupp, A. A., y Templin, J. L. (2008). Unique characteristics of diagnostic classification
-

-
- models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219–262.
<https://doi.org/10.1080/15366360802490866>
- Rupp, A. A., Templin, J. L., y Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. The Guilford Press.
- Sánchez, M. (2020). Evaluación del aprendizaje (pp. 17-40). En M. Mendiola y A. Martínez. *Evaluación del y para el aprendizaje: instrumentos y estrategias*. Coordinación de Desarrollo Educativo e Innovación Curricular, UNAM.
- Sánchez-Luján, B. I., y Moreno, R. A. (2019). Competencias matemáticas en fracciones en alumnos de nuevo ingreso a nivel universitario. *In Crescendo*, 9(3), 525-539.
- Sánchez, M. y Delgado, L. (2022). Evaluación sumativa y exámenes de alto impacto. (pp. 81-97). En M. Mendiola y A. Martínez. *Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos*. CUAIEED-UNAM.
- Sánchez, M., García, M., Martínez, A. y Buzo, E. (2020). El examen de ingreso a la Universidad Nacional Autónoma de México: Evidencias de validez de una prueba de alto impacto y gran escala. *Revista Iberoamericana de Evaluación Educativa*, 13(2), 107-128. <https://doi.org/10.15366/riee2020.13.2.006>
- Scriven, M. (1967). The Methodology of Evaluation. En R. W. Tyler, R. M. Gagne, y M. Scriven (Eds.), *Perspectives of Curriculum Evaluation Chicago* (pp. 39-83). Rand McNally.
- Schwarz, G. E. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464. doi:10.1214/aos/1176344136.
- Shi, Q., Ma, W., Robitzsch, A., Sorrel, M.A., Man, K. (2021). Cognitively Diagnostic Analysis Using the G-DINA Model in R. *Psych*, 3, 812–835.
<https://doi.org/10.3390/psych3040052>
- Snow, R. E. (1980). Aptitude and achievement. En W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade*. New Directions for Testing and Measurement.
-

-
- Jossey-Bass.
- Sosa, L., Tuyub, I., y Aparicio, E. (2014). Diagnóstico en estudiantes de nuevo ingreso a nivel superior: competencias y dificultades matemáticas. Comité Latinoamericano de Matemática Educativa.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Stenberg, L. C., Varua, M., y Yong, J. (2010). Mathematics aptitude, attitude, secondary schools and student success in quantitative methods for business subject in an Australian Catholic university experience. Proceedings Of The 39Th Australian Conference Of Economists (Ace10), Sydney, Australia, 27-29 Septiembre 2010.
- Taras, M. (2009). Summative assessment: the missing link for formative assessment. *Journal of further and higher education*, 33(1), 57-69.
<https://doi.org/10.1080/03098770802638671>
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, 12, 55-73.
<https://doi.org/10.3102/107699860100010>
- Toprak-Yildiz, T. (2021). An international comparison using Cognitive Diagnostic Assessment: Fourth graders' diagnostic profile of reading skills on PIRLS 2016. *Studies in Educational Evaluation*, 70. <https://doi.org/10.1016/j.stueduc.2021.101057>
- Torres, A. y Campos, M. (6-8 de mayo de 2015). Análisis de las competencias matemáticas de estudiantes de nuevo ingreso en ingeniería. Congreso Internacional sobre la Enseñanza y Aplicación de las Matemáticas, UNAM.
- Torres-Zapata, A. E., Acuña-Lara, J., Acevedo-Olvera, G., Villanueva, J. R. (2019). Caracterización del perfil de ingreso a la universidad. Consideraciones para la toma de decisiones. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 9(18). <https://doi.org/10.23913/ride.v9i18.435>
- Trost, G. (1993). Principios y prácticas en la selección para la admisión a la educación
-

-
- superior. *Revista de la Educación Superior*, 22(85), 1-10.
- Universidad Autónoma de Baja California (2023a). Cuenta UABC con nuevo examen para selección de aspirantes a ingresar. https://gaceta.uabc.mx/notas/academia/cuenta-uabc-con-nuevo-examen-para-seleccion-de-aspirantes-ingresar?fbclid=IwAR2bCljbaqtRPR2lkWRl4m1PdFkp839XftD0ssRKxbf6ACwEaI2gr7l_b0Y
- Universidad Autónoma de Baja California (2023b). Guía de estudios examen de ingreso a la Educación Superior. <https://admisiones.uabc.mx/>
- Universidad Nacional Autónoma de México (2023). Examen para el diagnóstico de conocimientos. Resultados de los alumnos que ingresan al nivel licenciatura de la generación 2023.
- https://cuaieed.unam.mx/descargas/evaluacion/Publicacion_licenciatura_2023.pdf
- Universidad Veracruzana (2007). Maestría en Investigación Educativa (MIE).
- https://www.uv.mx/mie/files/2012/10/Replica_MIE_2007.pdf
- van der Linden, W. J. (2016). *Handbook of Item Response Theory (Vol. 1)*. Taylor & Francis Group.
- van der Linden, W. J. y Hambleton, R. K. (1997). *Handbook of modern item response theory*. Springer-Verlag.
- Vera, F. (2020). La importancia del proceso de enseñanza-aprendizaje y la evaluación diagnóstica. *Revista Atlante Cuadernos de Educación y Desarrollo*.
- von Davier, M. y Lee, Y-S. (2019). *Handbook of diagnostic classification models. Models and model extensions, applications, software packages*. Springer.
- Wang, M. (2023). Bridging assessment and learning: A cognitive diagnostic analysis of a large-scale Spanish proficiency test. *Porta Linguarum: An International Journal of Foreign Language Teaching and Learning*, 40, 9–24.
- <https://doi.org/10.30827/portalin.vi40.15930>
- Williamson, J. (2023). *Cognitive Diagnostic Models and how they can be useful*. Cambridge University Press & Assessment.
-

-
- Willse, J. (2022). Package 'CTT'. <https://cran.r-project.org/web/packages/CTT/CTT.pdf>
- Wright, B. D., y Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370. <https://www.rasch.org/rmt/rmt83b.htm>
- Wu, X., Zhang, Y., Wu, R. y Chang, H. (2021). A comparative study on cognitive diagnostic assessment of mathematical key competencies and learning trajectories. *Current Psychology*, 41, 7854-7866. <https://doi.org/10.1007/s12144-020-01230-0>.
- Xie, Q. (2016). Diagnosing university students' academic writing in English: Is cognitive diagnostic modelling the way forward? *Educational Psychology*, 37(1), 26-47. doi:10.1080/01443410.2016.1202900
- You, X., Li, M., Xiao, Y. y Liu, H. (2019). The Feedback of the Chinese Learning Diagnosis System for Personalized Learning in Classrooms. *Frontiers in Psychology*, 10(1751). doi:10.3389/fpsyg.2019.01751
- Zorrilla, J. (2015). Preguntas y sugerencias sobre el Sistema Educativo Nacional. *Perfiles Educativos*, 37, 7-16.

Anexos

Anexo A. Proceso de identificación de atributos cognitivos

Para identificar los atributos cognitivos involucrados en el dominio de las matemáticas de aspirantes a ingresar a la universidad se realizaron dos revisiones de la literatura. La primera revisión se realizó en las bases de datos Scopus y Web of Science con la siguiente cadena de búsqueda que incluye las palabras “cognitive model”, “cognitive attributes” y “mathematics” en el título, resumen y palabras clave y excluye áreas temáticas que no fueran de medición educativa:

- (TITLE-ABS-KEY (“cognitive model”) OR TITLE-ABS-KEY (“cognitive attributes”) AND TITLE-ABS-KEY (“mathematics”)) AND (EXCLUDE (SUBJAREA , “NEUR”) OR EXCLUDE (SUBJAREA , “MEDI”) OR EXCLUDE (SUBJAREA , “PHYS”) OR EXCLUDE (SUBJAREA , “ENVI”) OR EXCLUDE (SUBJAREA , “BUSI”) OR EXCLUDE (SUBJAREA , “BIOC”) OR EXCLUDE (SUBJAREA , “MATE”) OR EXCLUDE (SUBJAREA , “AGRI”) OR EXCLUDE (SUBJAREA , “PHAR”) OR EXCLUDE (SUBJAREA , “NURS”) OR EXCLUDE (SUBJAREA , “IMMU”) OR EXCLUDE (SUBJAREA , “ENER”) OR EXCLUDE (SUBJAREA , “CHEM”))

Se obtuvieron 455 resultados en las bases de datos y se revisaron con los siguientes criterios de inclusión:

- Cualquier tipo de documento: artículo, revisión de libro, libro, capítulo de libro, conferencia.
- Resultados que en el título o resumen incluyeran algunas de las siguientes palabras: Cognitive diagnostic, Cognitive diagnostic assessment, Cognitive diagnostic models, Cognitive attributes, Cognitive models, Attribute hierarchy method, Rule space method, Educational assessment.

-
- Resultados que en el título o resumen incluyeran algún modelo de diagnóstico cognitivo (p. ej. DINA, NIDA, DINO, G-DINA).
 - Resultados que en el título o resumen incluyeran el uso de un instrumento de medición en matemáticas.

Posterior a la revisión se obtuvieron 43 resultados que se revisaron considerando un segundo filtro de criterios de inclusión:

- Artículos con aplicaciones en educación media o superior.
- Artículos basados en el marco de referencia de Evaluación Diagnóstica Cognitiva (Cognitive diagnostic assessment).
- Artículos con temática de competencia matemática o algún área de las matemáticas.

Posterior a la revisión se obtuvieron 6 resultados los cuales se revisaron considerando un tercer filtro con los siguientes criterios de inclusión:

- Artículos que describan el proceso de construcción y definición de atributos cognitivos/modelo cognitivo.
- Artículos que describan atributos cognitivos/modelos cognitivos que sean aplicables a contextos de admisión a la universidad.
- Artículos que utilicen modelos de diagnóstico cognitivo en el análisis psicométrico de los datos.

Al finalizar la revisión, ningún artículo proporcionó información sobre atributos cognitivos en matemáticas en el nivel superior.

Se realizó una segunda revisión de la literatura en google académico. Se modificó la cadena de búsqueda para excluir las áreas temáticas que no fueran de medición educativa:

“cognitive models”, “cognitive attributes”, “mathematics”. La búsqueda arrojó 183 resultados de los cuales se revisaron con base en los siguientes criterios de inclusión:

-
- Artículos con aplicaciones en el área de matemáticas.
 - Artículos que tengan como marco de referencia la evaluación diagnóstica cognitiva.
 - Artículos que describan el proceso de construcción y definición de atributos cognitivos o modelos cognitivos que sean aplicables a contextos de admisión a la universidad.

Se obtuvieron 4 resultados finales de los cuales se revisaron y finalmente se tomaron como referencia los atributos cognitivos de matemáticas que fueron validados en Gierl, et al. (2009) a través de la aplicación de protocolos de respuesta en voz alta a estudiantes que presentaron el examen SAT el cual es utilizado para la admisión a la universidad en Estados Unidos y otros países.

Como resultado de la identificación de atributos cognitivos, se adaptaron y tradujeron del inglés al español 21 atributos cognitivos del área de álgebra.

Gierl et al. (2009) desarrollaron los atributos cognitivos a partir de cuatro habilidades primarias descritas en Gierl et al. (2005):

- **Aplicación de conocimientos matemáticos básicos.** El estudiante puede aplicar contenidos y procedimientos matemáticos que generalmente se aprenden en cursos de álgebra o cursos previos. Este contenido incluye conceptos aritméticos básicos, operaciones, procedimientos, traducción entre expresiones verbales y ecuaciones, geometría básica, gráficos, lectura de gráficos, definiciones y terminología.
- **Aplicación de conocimientos matemáticos avanzados.** El estudiante puede aplicar el contenido y los procedimientos matemáticos enseñados en cursos de geometría y álgebra que se imparten en bachillerato de manera avanzada. Esta habilidad incluye procedimientos, traducción de expresiones verbales a ecuaciones, gráficos, definiciones y terminología más allá de los conocimientos matemáticos básicos.

-
- **Gestión de la complejidad.** El estudiante puede administrar una gran cantidad de información o seguir varios pasos para resolver un problema con un mayor nivel de complejidad. Responder correctamente a un problema de varios pasos requiere que el estudiante tenga la iniciativa de continuar en cada paso de la solución. En algunos casos, esa iniciativa puede ser necesaria para dar el primer paso e intentar la solución de un problema no rutinario.
 - **Modelamiento matemático.** El estudiante puede utilizar los conocimientos adquiridos y plantear ejemplos para responder una pregunta con un mayor nivel de dificultad. Un estudiante que puede modelar, o crear una representación, tiene la capacidad de formular ecuaciones para problemas planteados cuando eso implica algo más que un proceso memorístico. El estudiante también puede crear o agregar gráficos o figuras para resolver problemas. Parte de la habilidad de crear representaciones es decidir qué tipo de representación sería útil para resolver un problema.

Los 21 atributos cognitivos seleccionados se presentan en la siguiente tabla:

Tabla A1

Atributos cognitivos seleccionados para el estudio

H1. Aplicación de conocimientos matemáticos básicos

H1A1. Establecer una única proporción comparando dos cantidades.

H1A2. Resolver una ecuación lineal (por ejemplo, resta o división en ambos lados).

H1A3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).

H1A4. Sustituir valores en expresiones algebraicas.

H1A5. Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.

H2. Aplicación de conocimientos matemáticos avanzados

H2A1. Ordenar series geométricas o cálculo numérico consecutivo.

Este atributo implica el conocimiento sobre series geométricas (por ejemplo, la naturaleza de la relación entre términos) o el cálculo numérico consecutivo (por ejemplo, multiplicación y división).

H2A2. Realizar operaciones exponenciales o suma de potencias.

H2A3. Graficar una función.

Este atributo implica el conocimiento de la gráfica de una función o la sustitución de puntos en la gráfica.

H2A4. Aplicar las propiedades de factorización.

H2A5. Sustituir números en expresiones algebraicas.

H3. Gestión de la complejidad

H3A1. Aplicar reglas de factorización.

H3A2. Sustituir expresiones algebraicas por otras expresiones algebraicas.

H3A3. Comprender o aplicar reglas.

H3A4. Resolver sistemas de ecuaciones lineales.

H4. Modelamiento matemático

H4A1. Resolver series geométricas en un patrón abstracto.

H4A2. Representar o ejecutar múltiples soluciones algebraicas básicas.

H4A3. Representar o ejecutar múltiples soluciones algebraicas avanzadas.

Este atributo implica establecer una proporción única, habilidades para transformar fracciones y habilidades de análisis cómo, cuándo, dónde y/o cómo realizar la transformación.

H4A4. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.

H4A5. Establecer o resolver ecuaciones cuadráticas.

H4A6. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.

H4A7. Resolver una desigualdad cuadrática con dos variables.

Anexo B. Materiales para la etapa 1 del comité de desarrollo y validación

Instructivo de elaboración de descriptores de atributos cognitivos para una prueba de Matemáticas de admisión a la universidad

Estimado miembro del comité de desarrollo y validación:

Como parte del desarrollo del proyecto de tesis “Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad”, el objetivo del presente comité es elaborar y validar los descriptores de un conjunto de atributos cognitivos que subyacen a una prueba matemáticas.

Materiales

Para el desarrollo de la actividad de elaboración, usted contará con los siguientes materiales:

Formato de elaboración de descriptores. Documento en formato Google sheets en el cual realizará la actividad de elaboración de descriptores de un conjunto de atributos cognitivos asignados.

El documento contiene las siguientes hojas:

- **Presentación.** En esta hoja se describe el contenido del documento.
- **Datos de identificación.** En esta hoja deberá llenar sus datos de identificación.
- **Instrucciones.** En esta hoja se presentan las indicaciones para realizar la actividad.
- **Hoja de trabajo 1.** En esta hoja se elaboran los descriptores para los atributos asociados con una habilidad.
- **Hoja de trabajo 2.** En esta hoja se elaboran los descriptores para los atributos asociados con una habilidad.
- **Ejemplos.** En esta hoja se presentan tres ejemplos de la actividad que debe realizar.

Presentación

Estimado juez,

Este archivo contiene las siguientes hojas:

Datos de identificación. En esta hoja deberá llenar sus datos de identificación.

Instrucciones. En esta hoja se presentan las indicaciones para realizar la actividad.

Hoja de trabajo 1. En esta hoja se elaboran los descriptores para los atributos asociados con la habilidad de *aplicación de conocimientos matemáticos básicos*.

Hoja de trabajo 2. En esta se elaboran los descriptores para los atributos asociados con la habilidad de *aplicación de conocimientos matemáticos avanzados*.

Ejemplos. En esta hoja se presentan tres ejemplos de la actividad que debe realizar.

Al finalizar la elaboración de descriptores, le pedimos que notifique a través del correo electrónico que ha concluido con la actividad. Si tiene alguna duda durante el proceso, puede contactarnos al correo gnavarro17@uabc.edu.mx.

Datos de identificación

Nombre:

Edad:

Sexo:

Escriba aquí una semblanza profesional:

La información que proporciona tiene un carácter confidencial y anónimo, y se encuentra protegida por el marco normativo en materia de protección de datos personales que rige a nuestra a la Universidad Autónoma de Baja California. Puedes consultar el Aviso de Privacidad Académico en la siguiente liga. http://transparencia.uabc.mx/Aviso_Privacidad/Avisos/ACADEMICO.pdf

Instrucciones

Este archivo contiene dos hojas de trabajo, 'Hoja de trabajo 1' y 'Hoja de trabajo 2'. Ambas hojas se llenarán de la misma manera. Las hojas de trabajo contienen una tabla con la siguiente información:

Atributo. Esta columna contiene una descripción de los atributos cognitivos que subyacen al dominio de las matemáticas en estudiantes próximos a ingresar a la universidad.

Descriptor. En esta columna deberá escribir los conocimientos y destrezas que mejor representen al atributo.

Categoría. En esta columna deberá seleccionar si el descriptor que escribió corresponde a una destreza o un conocimiento.

Como apoyo, a continuación se presentan las definiciones de conocimientos y destrezas que se tomarán como referencia para esta actividad:

Conocimiento. El conocimiento puede clasificarse como un hecho, un concepto, un principio o un procedimiento.

Un hecho es una verdad conocida por experiencia u observación

Un concepto es una idea de algo que se forma combinando mentalmente todas sus características o detalles.

Un principio es una regla aceptada de acción o conducta, una ley o verdad fundamental, primaria o general de la que se derivan otras.

Un procedimiento es un curso de acción, físico o mental, que tiene un resultado previsto.

Destreza. Es la capacidad de realizar una tarea. Algunas destrezas consisten en una sola acción, mientras que otras destrezas son procedimientos que involucran dos o más pasos.

[Si tiene alguna duda de cómo realizar la actividad, haga clic aquí para consultar el instructivo en extenso.](#)

Hoja de trabajo: Aplicación de conocimientos matemáticos básicos

El estudiante puede aplicar contenidos y procedimientos matemáticos que generalmente se aprenden en cursos de álgebra o cursos previos. Este contenido incluye conceptos aritméticos básicos, operaciones, procedimientos, traducción entre expresiones verbales y ecuaciones, geometría básica, gráficos, lectura de gráficos, definiciones y terminología.

Estimado juez: Escriba en la columna "Descriptor" los conocimientos y destrezas que usted considere que describen al atributo correspondiente.

Atributo	Descriptor	Categoría
H1A1. Establecer una única proporción comparando dos cantidades.		
H1A2. Resolver una ecuación lineal (por ejemplo, resta o división en ambos lados).		

Hoja de trabajo: Aplicación de conocimientos matemáticos avanzados

El estudiante puede aplicar el contenido y los procedimientos matemáticos enseñados en cursos de geometría y álgebra que se imparten en bachillerato de manera avanzada. Esta habilidad incluye procedimientos, traducción de expresiones verbales a ecuaciones, gráficos, definiciones y terminología más allá de los conocimientos matemáticos básicos.

Estimado juez: Escriba en la columna "Descriptor" los conocimientos y destrezas que usted considere que describen al atributo correspondiente.

Atributo	Descriptor	Categoría
H2A1. Ordenar series geométricas o cálculo numérico consecutivo. *Este atributo implica el conocimiento sobre series geométricas (por ejemplo, la naturaleza de la relación entre términos) o el cálculo numérico consecutivo (por ejemplo, multiplicación y división).		
H2A2. Realizar operaciones exponenciales o suma de potencias.		
H2A3. Graficar una función. * Este atributo implica el conocimiento de la gráfica de una función o la sustitución de puntos en la gráfica.		
H2A4. Aplicar las propiedades de factorización.		
H2A5. Sustituir números en expresiones algebraicas.		
H3A3. Comprender o aplicar reglas.		

Ejemplo de descriptores de atributos cognitivos

Estimado juez: La siguiente tabla presenta ejemplos de atributos de matemáticas junto con sus correspondientes descriptores y categorías.

Atributo	Descriptor	Categoría
Conocer y aplicar una función lineal	Conoce la fórmula general de la función lineal que se puede escribir como $f(x) = a + bx$	Conocimiento
	Sabe que "a" es la constante y "bx" es b multiplicado por x	Conocimiento
	Sabe que "x" es la variable independiente y "y" la variable dependiente	Conocimiento
	Puede resolver más de una incógnita en una expresión simple cuando se especifica la relación entre las incógnitas (por ejemplo, $a+b = 21$ y $b = 2a$; puede considerarse el conjunto más simple de ecuaciones simultáneas)	Destreza
	Puede resolver incógnitas en expresión simple (por ejemplo, $3a = 21$)	Destreza
Conocer y aplicar una solución cuadrática	Conoce la fórmula general cuadrática: $ax^2 + bx + c = 0$	Conocimiento
	Reconoce que puede haber más de una solución	Conocimiento
	Puede reorganizar una expresión en forma cuadrática.	Destreza
	Resuelve las incógnitas en una cuadrática usando factorización.	Destreza
	Resuelve fórmula cuadrática usando la ecuación $(-b \pm \sqrt{b^2 - 4ac})/2a$	Destreza
	Puede hacer una referencia cruzada de los términos a, b y c con la expresión en cuestión.	Destreza
Comprender textos	Conoce el significado de los términos producto, multiplicado, igual, resultado, cuadrado, diferencia, suma y total en un contexto matemático.	Conocimiento
	Identifica el símbolo para el cual se requiere solución.	Conocimiento
	Identifica relaciones entre incógnitas (p. ej., "b es 3 veces mayor que a" implica que $b = 3a$, $b/a = 3$ o $a = b/3$)	Conocimiento
	Puede derivar de un texto una expresión matemática cuando se identifican símbolos que representan una, dos o tres cantidades desconocidas.	Destreza
	Puede derivar de un texto una expresión matemática cuando no se identifican los símbolos desconocidos; El examinado asigna símbolos para identificar las incógnitas.	Destreza

Fuente: Gierl, M.J., Leighton, J.P., y Hunka, S. (2007). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills. En J. P. Leighton y M. J. Gierl (Eds.), Cognitive diagnostic assessment for education: Theory and applications. (pp. 242–274). Cambridge University Press.

Anexo C. Materiales para la etapa 2 del comité de desarrollo y validación

INSTRUCCIONES		Criterio	Escala
● Estimado juez, por cada descriptor que se presenta en la tabla de abajo otorgue una valoración de acuerdo con los criterios que se presentan a lado derecho de estas instrucciones. ● La escala de valoración va del 1 al 4, donde 1 indica que el descriptor no cumple con el criterio, y 4 indica que el descriptor cumple con el criterio en su totalidad.		Claridad. El descriptor se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1= El descriptor no es claro. 2= El descriptor requiere bastantes modificaciones o una modificación sustantiva. 3= Se requiere una modificación muy específica de algunos términos del descriptor. 4= El descriptor es claro.
● Los criterios a evaluar son claridad, coherencia, relevancia y suficiencia. Los primeras tres deben valorarse por cada descriptor, mientras que el criterio de suficiencia corresponde a la valoración de los descriptors de todo un atributo. ● En el caso de que otorgue una calificación menor a 4, por favor indique en la sección de Observaciones la justificación de tu valoración y recomendaciones o modificaciones al descriptor.		Coherencia. El descriptor tiene relación lógica con el atributo que representa.	1 = El descriptor no tiene relación lógica con el atributo. 2 = El descriptor tiene una relación tangencial con el atributo. 3 = El descriptor tiene una relación moderada con el atributo. 4 = El descriptor está completamente relacionado con el atributo.
		Relevancia. El descriptor es esencial o importante, es decir, debe incluirse.	1= El descriptor puede eliminarse sin que se vea afectada la representación del atributo. 2= El descriptor tiene alguna relevancia, pero otro descriptor puede estar incluyendo lo que éste representa. 3= El descriptor es relativamente importante para representar el atributo. 4= El descriptor es muy relevante y debe incluirse para representar el atributo.
		Suficiencia. Los descriptors que pertenecen a un mismo atributo bastan representarlo.	1= Los descriptors no son suficientes para medir el atributo. 2= Los descriptors solo representan algún aspecto del atributo. 3= Debe incrementarse el número de descriptors para representar el atributo. 4= Los descriptors son suficientes y representan el atributo.

Instrucciones: Otorgue una valoración del 1 al 4, 1 indica que el descriptor no cumple con el criterio, y 4 indica que el descriptor cumple con el criterio en su totalidad.							
Atributo	Descriptor	Relevancia	Coherencia	Claridad	Suficiencia	Observaciones	Comentarios generales
H4A1. Resolver series geométricas en un patrón abstracto.	H4A1D1. Comprende la definición y las propiedades de las series geométricas						
	H4A1D2. Conoce la fórmula para la suma de n términos						
	H4A1D3. Reconoce patrones abstractos en configuraciones geométricas y numéricas						
	H4A1D4. Conoce los principios de convergencia y divergencia de series						
	H4A1D5. Ejecuta manipulaciones algebraicas para simplificar y resolver series geométricas						
	H4A1D6. Aplica correctamente la fórmula de la suma de una serie geométrica al problema dado						
H4A2. Representar o ejecutar múltiples soluciones algebraicas básicas.	H4A2D1. Reconoce las variables y constantes en expresiones y ecuaciones algebraicas						
	H4A2D2. Conoce los métodos de factorización de expresiones algebraicas						
	H4A2D3. Sabe las propiedades de las operaciones algebraicas (asociativa, conmutativa, distributiva) y cómo aplicarlas para simplificar y resolver ecuaciones						
	H4A2D4. Ejecuta manipulaciones simbólicas para simplificar o reescribir expresiones algebraicas en formas equivalentes						
	H4A2D5. Comprueba la validez de las soluciones obtenidas mediante la sustitución de valores en las ecuaciones originales						
	H4A2D6. Resuelve sistema de ecuaciones utilizando diferentes métodos						
H4A3. Representar o ejecutar múltiples soluciones algebraicas avanzadas. * Este atributo implica establecer una proporción única, habilidades para transformar fracciones y habilidades de análisis cómo, cuándo, dónde y/o cómo realizar la transformación.	H4A3D1. Conoce las propiedades y métodos de simplificación y transformación de fracciones algebraicas (incluyendo fracciones complejas y parciales)						
	H4A3D2. Comprende el concepto de función algebraica						
	H4A3D3. Utiliza métodos adecuados para resolver ecuaciones plinómicas de grado superior, ecuaciones racionales y ecuaciones algorítmicas						
	H4A3D4. Realiza manipulación algebraica avanzada para reescribir o simplificar expresiones complejas						
	H4A3D5. Comprueba la validez de las soluciones obtenidas						
H4A4. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.	H4A4D1. Conoce las reglas de las potencias (multiplicación y división de potencias con la misma base, potencia a otra potencia, potencias negativas y potencias fraccionarias)						
	H4A4D2. Conoce las leyes de los logaritmos y cómo se relacionan con las operaciones exponenciales, incluyendo la propiedad de logaritmo de un producto, cociente y potencia						
	H4A4D3. Comprende las propiedades de las funciones exponenciales (incluyendo su comportamiento y gráficas) y su relación inversa con las funciones logarítmicas						
	H4A4D4. Aplica correctamente las reglas de exponentes en la simplificación y resolución de expresiones algebraicas						
	H4A4D5. Transforma expresiones que involucran exponentes negativos, fraccionarios y raíces						
	H4A4D6. Resuelve ecuaciones exponenciales utilizando técnicas algebraicas y logarítmicas						
H4A5. Establecer o resolver ecuaciones cuadráticas.	H4A5D1. Comprende la forma general de una ecuación cuadrática ($ax^2+bx+c=0$, donde a, b y c son coeficientes reales)						
	H4A5D2. Conoce la fórmula general para resolver una ecuación cuadrática y cómo se aplica						
	H4A5D3. Comprende que el discriminante (b^2-4ac) le permite determinar la naturaleza de las soluciones de una ecuación cuadrática (reales distintas o iguales, o complejas)						
	H4A5D4. Aplica técnicas de factorización para resolver ecuaciones cuadráticas						
	H4A5D5. Usa el discriminante para determinar la naturaleza de las soluciones de una ecuación cuadrática						
	H4A5D6. Resuelve una ecuación cuadrática mediante el método de completar un trinomio cuadrado perfecto						
	H4A5D7. Formula una ecuación cuadrática a partir de un problema verbal.						
H4A6. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.	H4A6D1. Identifica variables dependientes e independientes.						
	H4A6D2. Comprende el concepto de función donde la variable dependiente se expresa en términos de la independiente.						
	H4A6D3. Conoce las propiedades y características de las gráficas de funciones: pendiente, intersección con los ejes, máximos y mínimos, comportamiento asintótico						
	H4A6D4. Grafica funciones representando adecuadamente la relación entre variables en un plano cartesiano						
	H4A6D5. Interpreta la relación entre las variables a través de analizar las gráficas						
	H4A6D6. Resuelve problema contextualizados mediante el análisis de las gráficas						
H4A7. Resolver una desigualdad cuadrática con dos variables.	H4A7D1. Conoce las propiedades y características de las curvas cuadráticas (circunferencia, parábola, elipse, hipérbola)						
	H4A7D2. Entiende cómo las desigualdades cuadráticas dividen al plano						
	H4A7D3. Transforma las desigualdades en igualdades para facilitar la resolución						
	H4A7D4. Usa puntos de prueba para determinar las regiones del plano que satisfacen la desigualdad						
	H4A7D5. Verifica las soluciones obtenidas mediante la sustitución de puntos						
H1A3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).	H1A3D1. Conoce los números reales y su clasificación						
	H1A3D2. Conoce las propiedades de los números reales						
	H1A3D3. Comprende la relación de los signos al momento de realizar operaciones básicas						
	H1A3D4. Resuelve ejercicios donde hay operaciones de suma, resta, multiplicación y división						
	H1A3D5. Ejecuta manipulación aritmética para simplificar y resolver problemas rutinarios o de contexto real						
	H1A3D6. Identifica el símbolo para el cual se requiere solución						
H1A4. Sustituir valores en expresiones algebraicas.	H1A4D1. Identifica la operación que está realizando la variable						
	H1A4D2. Identifica el símbolo para el cual se requiere solución						
	H1A4D3. Reconoce el coeficiente y la variable						
	H1A4D4. Aplica correctamente las operaciones básicas (suma, resta, multiplicación y división) en la simplificación y resolución de expresiones algebraicas						
H1A5. Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.	H1A5D1. Reconoce cuando sumar o restar y el signo que domina el resultado						
	H1A5D2. Reconoce el como trabajar las operaciones básicas (suma, resta, multiplicación y división) tanto en procesos aritméticos como algebraicos						
	H1A5D3. Resuelve ejercicios donde identifica que operación realizar dependiendo los signos y los términos semejantes						

Anexo D. Materiales para el comité de determinación de la matriz Q



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo



Maestría en Ciencias Educativas

Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad

Estimado miembro del comité:

Como parte del desarrollo del proyecto de tesis “Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad”, el objetivo del presente comité es determinar la relación entre un conjunto de atributos cognitivos con los ítems de la sección de matemáticas del Examen de Ingreso a la Educación Superior (ExIES).

Para el desarrollo de la actividad, usted cuenta con los siguientes materiales:

1. Listado de atributos con sus descriptores.
2. 50 ítems de la prueba de matemáticas del ExIES.
3. 4 hojas de trabajo.

El presente cuadernillo contiene una sección de datos de identificación, la descripción del proyecto de tesis y las indicaciones para la actividad a realizar.

Datos de identificación

- Nombre: _____
- Edad: _____
- Nivel de estudios: _____
- Años de experiencia en docencia de las matemáticas: _____
- Años de experiencia en elaboración y revisión de ítems: _____

Descripción del proyecto

La admisión de estudiantes a la educación superior en México involucra la aplicación de exámenes de conocimientos y habilidades que, bajo los enfoques psicométricos clásicos como la Teoría clásica de los tests o la Teoría de respuesta al ítem, permiten realizar inferencias sobre el nivel de habilidad que tienen los estudiantes para dominar los contenidos medidos en las pruebas. Estas inferencias permiten ubicar y comparar el nivel de habilidad de los aspirantes para orientar la toma de decisiones de selección.

Las pruebas de admisión a las universidades mexicanas evalúan los conocimientos y las habilidades básicas necesarias para tener éxito en la vida personal y académica, entre las que se destacan las habilidades de lectura, escritura y matemáticas (Larrondo et al., 2018 y Zorrilla, 2015). Al respecto, Zorrilla (2015) argumenta que los estudiantes que no dominan las habilidades de escribir, comprender un texto o aplicar las matemáticas en determinados contextos tienen implicaciones de marginación y exclusión social.

La Universidad Autónoma de Baja California (UABC) aplica el Examen de Ingreso a la Educación Superior (ExIES) para seleccionar a sus aspirantes de licenciatura (Universidad Autónoma de Baja California [UABC], 2023^a). El ExIES mide la capacidad de aplicación de conocimientos y habilidades que han adquirido los aspirantes durante su formación básica y media superior, las cuales les permitirán resolver las demandas de una formación a nivel superior (UABC, 2023b). El ExIES mide el dominio de tres componentes: 1) Lectura, que evalúa la capacidad para leer y comprender textos literarios e informativos; 2) Lengua escrita, que evalúa la capacidad para revisar y editar textos con contenidos académicos, así como medir la capacidad para expresar ideas en apego a las convenciones del español escrito y; 3) Matemáticas, que evalúa la capacidad para la aplicación, manejo y comprensión de conceptos matemáticos, y la habilidad para la resolución de problemas e interpretación de datos, tablas, planos, figuras y gráficos (UABC, 2023b).

En su planteamiento, las pruebas de admisión a la universidad no tienen como finalidad proporcionar información diagnóstica o formativa sobre los aspirantes admitidos. Sin embargo, distintos autores sugieren que los resultados de las pruebas de admisión, además de brindar resultados para la selección de aspirantes, también pueden usarse con finalidades diagnósticas o formativas (Atkinson y Geiser, 2009; Sánchez, 2020). En esa misma línea, distintos autores argumentan que evaluaciones con fines sumativos, - como lo es el ExIES de la UABC-, tienen la capacidad de proporcionar información para mejorar la enseñanza y el aprendizaje de los estudiantes (Bennett, 2011; Black et al., 2003; Taras, 2019).

En las últimas décadas se ha incrementado la necesidad de proporcionar resultados de evaluaciones que informen sobre los procesos cognitivos de los estudiantes para mejorar la enseñanza y el aprendizaje (Huff y Goodman, 2007; Nichols, 1994). Por ello, resulta necesario para las universidades no solo obtener diagnósticos del nivel de competencia en los conocimientos y habilidades básicas de los estudiantes de nuevo ingreso, sino también se requieren diagnósticos que permitan vincular los contenidos de

las pruebas de conocimientos y habilidades con los mecanismos cognitivos involucrados en el dominio de esos contenidos.

Nichols (1994) definió el término de evaluaciones diagnósticas cognitivas (EDC) a evaluaciones que combinan las aportaciones de la ciencia cognitiva y la psicometría con la finalidad de proporcionar resultados diagnósticos de los estudiantes. Las EDC son evaluaciones que reconocen y hacen explícitos los procesos y estructuras cognitivas que usa un estudiante para responder una tarea, como una prueba de conocimientos (Leighton y Gierl, 2007; Nichols, 1994). De acuerdo con Gierl et al. (2009), identificar los atributos cognitivos que subyacen al desempeño de una prueba conlleva a realizar inferencias sobre las fortalezas y debilidades cognitivas que tienen los estudiantes.

Para obtener inferencias sobre los atributos cognitivos de los estudiantes, se originaron los modelos de diagnóstico cognitivo (MDC) en el campo de la psicometría y la medición, los cuales surgen como respuesta a la necesidad de determinar estrategias apropiadas para la retroalimentación oportuna de los estudiantes sobre sus fortalezas y debilidades cognitivas (de la Torre y Sorrel, 2023). Los MDC son modelos multidimensionales de variables latentes categóricas que tienen como objetivo principal la clasificación de los examinados en clases latentes dicotómicas o politómicas que reflejan la presencia o ausencia de habilidades o atributos cognitivos latentes en cada ítem (Luna, 2017; Rupp y Templin, 2008; Rupp et al., 2010).

Un componente esencial para la implementación de los MDC es el establecimiento a priori de la relación hipotética de cómo un conjunto de atributos cognitivos se vincula con el desempeño esperado de un examinado en una tarea (von Davier y Lee, 2019). En el contexto de las evaluaciones diagnósticas cognitivas, se establece la relación de los atributos cognitivos con los ítems de una prueba. Esta relación se expresa en una matriz Q de especificaciones (Tatsuoka, 1985).

De manera formal, la matriz Q es una matriz de ceros y unos de tamaño $J \times K$, en donde el elemento de la j-ésima fila y la k-ésima columna, q^{jk} , indica si se requiere el atributo k para responder correctamente al ítem j (de la Torre, 2009). De acuerdo con Rupp et al. (2010) “la matriz Q es el componente por excelencia en cualquier MDC porque representa la operacionalización de la teoría sustantiva que ha dado lugar al diseño de la evaluación diagnóstica”(p. 49).

La siguiente tabla presenta un ejemplo de matriz Q con ítems de una prueba de aritmética básica. En la primera columna se presentan los ítems de la prueba, mientras que los atributos cognitivos se presentan en las columnas subsecuentes. Si se requiere el atributo para responder correctamente al ítem, se indica con un 1, y si no se requiere el atributo se indica con un 0. Se puede observar que el primer ítem de la prueba requiere de los atributos Suma y Resta, el segundo ítem requiere División, el tercer ítem requiere de Resta y Multiplicación y finalmente, el último ítem requiere de Suma.

Tabla 1

Ejemplo de matriz Q de aritmética básica

Ítems	Suma	Resta	Multiplicación	División
$2 + 3 - 1 = ?$	1	1	0	0
$4/2 = ?$	0	0	0	1
$5 * 3 - 4 = ?$	0	1	1	0
$8 + 12 = ?$	1	0	0	0

Nota. Adaptado de Diagnostic Measurement (p. 116), por A. A. Rupp et al., 2010, The Guilford Press.

La implementación de los MDC en evaluaciones formativas realizadas en educación básica y media superior han tenido un gran desarrollo en las últimas décadas. Sin embargo, su implementación en contextos de admisión a la educación superior es un campo poco explorado.

La identificación oportuna de las habilidades cognitivas que poseen los estudiantes tiene el potencial de mejorar no solo las decisiones de admisión, sino proporcionar una comprensión más completa del potencial académico del aspirante. A partir de la necesidad de proporcionar información diagnóstica cognitiva sobre los estudiantes de nuevo ingreso a la universidad, surgen las siguientes preguntas: ¿Cuáles son las diferencias en el dominio de las matemáticas de los aspirantes a ingresar a la UABC y cuáles son los atributos cognitivos que lo explican? Y ¿En qué medida los modelos de diagnóstico cognitivo pueden proporcionar información complementaria con fines diagnósticos y formativos?

Instrucciones

1. Identifique las 4 hojas de trabajo. Cada hoja contiene una tabla con 21 atributos en las filas y los ítems enumerados en las columnas.




Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas
Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad

Nombre: _____ Fecha: _____

Atributos	Ítems									
	1	2	3	4	5	6	7	8	9	10
1. Establecer una única proporción comparando dos cantidades.										
2. Resolver una ecuación lineal.										
3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).										
4. Sustituir valores en expresiones algebraicas.										
5. Ordenar series geométricas o cálculo numérico consecutivo.										
6. Realizar operaciones exponenciales o suma de potencias.										
7. Graficar una función.										
8. Aplicar las propiedades de factorización.										
9. Sustituir números en expresiones algebraicas.										
10. Aplicar reglas de factorización.										
11. Sustituir expresiones algebraicas por otras expresiones algebraicas.										
12. Comprender o aplicar reglas.										
13. Resolver sistemas de ecuaciones lineales.										
14. Resolver series geométricas en un patrón abstracto.										
15. Representar o ejecutar múltiples soluciones algebraicas básicas.										
16. Representar o ejecutar múltiples soluciones algebraicas avanzadas.										
17. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.										
18. Establecer o resolver ecuaciones cuadráticas.										
19. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.										
20. Resolver una desigualdad cuadrática con dos variables.										
* Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.										

Hoja 1 de 4

2. Lea con atención el listado de atributos con sus descriptores.
3. Lean con atención cada uno de los ítems de la prueba de matemáticas y marque con una $\checkmark$ de 1 a 3 atributos que considere que se requieren para responder correctamente el ítem.
4. Una vez que finalice cada hoja de trabajo, notifíquelo a los organizadores.



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas

Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad

Nombre: _____

Fecha: _____

Atributos	Ítems									
	1	2	3	4	5	6	7	8	9	10
1. Establecer una única proporción comparando dos cantidades.										
2. Resolver una ecuación lineal.										
3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).										
4. Sustituir valores en expresiones algebraicas.										
5. Ordenar series geométricas o cálculo numérico consecutivo.										
6. Realizar operaciones exponenciales o suma de potencias.										
7. Graficar una función.										
8. Aplicar las propiedades de factorización.										
9. Sustituir números en expresiones algebraicas.										
10. Aplicar reglas de factorización.										
11. Sustituir expresiones algebraicas por otras expresiones algebraicas.										
12. Comprender o aplicar reglas.										
13. Resolver sistemas de ecuaciones lineales.										
14. Resolver series geométricas en un patrón abstracto.										
15. Representar o ejecutar múltiples soluciones algebraicas básicas.										
16. Representar o ejecutar múltiples soluciones algebraicas avanzadas.										
17. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.										
18. Establecer o resolver ecuaciones cuadráticas.										
19. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.										
20. Resolver una desigualdad cuadrática con dos variables.										
* Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.										



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas
Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad

Nombre: _____ Fecha: _____

Atributos	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Establecer una única proporción comparando dos cantidades.															
2. Resolver una ecuación lineal.															
3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).															
4. Sustituir valores en expresiones algebraicas.															
5. Ordenar series geométricas o cálculo numérico consecutivo.															
6. Realizar operaciones exponenciales o suma de potencias.															
7. Graficar una función.															
8. Aplicar las propiedades de factorización.															
9. Sustituir números en expresiones algebraicas.															
10. Aplicar reglas de factorización.															
11. Sustituir expresiones algebraicas por otras expresiones algebraicas.															
12. Comprender o aplicar reglas.															
13. Resolver sistemas de ecuaciones lineales.															
14. Resolver series geométricas en un patrón abstracto.															
15. Representar o ejecutar múltiples soluciones algebraicas básicas.															
16. Representar o ejecutar múltiples soluciones algebraicas avanzadas.															
17. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.															
18. Establecer o resolver ecuaciones cuadráticas.															
19. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.															
20. Resolver una desigualdad cuadrática con dos variables.															
* Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.															



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas
Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad

Nombre: _____ Fecha: _____

Atributos	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1. Establecer una única proporción comparando dos cantidades.															
2. Resolver una ecuación lineal.															
3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).															
4. Sustituir valores en expresiones algebraicas.															
5. Ordenar series geométricas o cálculo numérico consecutivo.															
6. Realizar operaciones exponenciales o suma de potencias.															
7. Graficar una función.															
8. Aplicar las propiedades de factorización.															
9. Sustituir números en expresiones algebraicas.															
10. Aplicar reglas de factorización.															
11. Sustituir expresiones algebraicas por otras expresiones algebraicas.															
12. Comprender o aplicar reglas.															
13. Resolver sistemas de ecuaciones lineales.															
14. Resolver series geométricas en un patrón abstracto.															
15. Representar o ejecutar múltiples soluciones algebraicas básicas.															
16. Representar o ejecutar múltiples soluciones algebraicas avanzadas.															
17. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.															
18. Establecer o resolver ecuaciones cuadráticas.															
19. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.															
20. Resolver una desigualdad cuadrática con dos variables.															
* Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.															



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas

Aplicación de modelos de diagnóstico cognitivo a pruebas de admisión a la universidad

Nombre: _____

Fecha: _____

Atributos	Ítems									
	41	42	43	44	45	46	47	48	49	50
1. Establecer una única proporción comparando dos cantidades.										
2. Resolver una ecuación lineal.										
3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).										
4. Sustituir valores en expresiones algebraicas.										
5. Ordenar series geométricas o cálculo numérico consecutivo.										
6. Realizar operaciones exponenciales o suma de potencias.										
7. Graficar una función.										
8. Aplicar las propiedades de factorización.										
9. Sustituir números en expresiones algebraicas.										
10. Aplicar reglas de factorización.										
11. Sustituir expresiones algebraicas por otras expresiones algebraicas.										
12. Comprender o aplicar reglas.										
13. Resolver sistemas de ecuaciones lineales.										
14. Resolver series geométricas en un patrón abstracto.										
15. Representar o ejecutar múltiples soluciones algebraicas básicas.										
16. Representar o ejecutar múltiples soluciones algebraicas avanzadas.										
17. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.										
18. Establecer o resolver ecuaciones cuadráticas.										
19. Reconocer la relación entre variables independientes y dependientes o su representación gráfica.										
20. Resolver una desigualdad cuadrática con dos variables.										
* Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.										

Anexo E. Análisis de ítems con el enfoque de la teoría clásica de los tests

Tabla E1

Análisis de ítems bajo el enfoque de la teoría clásica de los tests

Item	D.E	p	d	α si se elimina	Item	D.E	p	d	α si se elimina
i1	0.45	0.29	0.49	0.89	i26	0.49	0.40	0.47	0.89
i2	0.46	0.70	0.36	0.89	i27	0.47	0.33	0.21	0.89
i3	0.48	0.65	0.47	0.89	i28	0.50	0.57	0.37	0.89
i4	0.49	0.60	0.41	0.89	i29	0.48	0.62	0.49	0.89
i5	0.49	0.57	0.52	0.89	i30	0.50	0.57	0.50	0.89
i6	0.50	0.52	0.49	0.89	i31	0.45	0.28	0.24	0.89
i7	0.50	0.46	0.40	0.89	i32	0.49	0.59	0.44	0.89
i8	0.49	0.58	0.36	0.89	i33	0.50	0.53	0.36	0.89
i9	0.38	0.83	0.32	0.89	i34	0.49	0.60	0.33	0.89
i10	0.41	0.22	0.40	0.89	i35	0.50	0.47	0.34	0.89
i11	0.49	0.60	0.46	0.89	i36	0.49	0.42	0.45	0.89
i12	0.50	0.56	0.29	0.89	i37	0.50	0.47	0.43	0.89
i13	0.49	0.41	0.38	0.89	i38	0.46	0.30	0.39	0.89
i14	0.50	0.54	0.46	0.89	i39	0.45	0.27	0.20	0.89
i15	0.50	0.54	0.47	0.89	i40	0.48	0.35	-0.03	0.90
i16	0.49	0.39	0.29	0.89	i41	0.49	0.42	0.43	0.89
i17	0.50	0.49	0.36	0.89	i42	0.48	0.37	0.57	0.89
i18	0.48	0.36	0.36	0.89	i43	0.49	0.41	0.35	0.89
i19	0.43	0.24	0.22	0.89	i44	0.45	0.27	0.26	0.89
i20	0.47	0.34	0.28	0.89	i45	0.44	0.27	0.16	0.89
i21	0.47	0.33	0.22	0.89	i46	0.47	0.32	0.41	0.89
i22	0.48	0.65	0.44	0.89	i47	0.33	0.12	0.01	0.89
i23	0.48	0.63	0.43	0.89	i48	0.47	0.67	0.43	0.89
i24	0.36	0.84	0.20	0.89	i49	0.48	0.35	0.20	0.89
i25	0.39	0.82	0.27	0.89	i50	0.41	0.22	0.19	0.89

Anexo F. Análisis de ítems con el enfoque de la teoría de respuesta al ítem

Tabla F1

Análisis de ítems bajo el enfoque de la teoría de respuesta al ítem con el modelo 3PL

Item	b	a	c	Item	b	a	c
i1	0.96	2.49	0.09	i26	0.76	2.74	0.19
i2	-0.18	1.56	0.33	i27	1.80	1.25	0.22
i3	-0.08	2.36	0.27	i28	0.40	1.58	0.29
i4	0.14	1.72	0.26	i29	-0.07	2.31	0.21
i5	0.04	2.35	0.16	i30	0.13	2.32	0.20
i6	0.24	2.11	0.16	i31	1.74	1.50	0.17
i7	0.25	1.07	0.02	i32	0.18	1.96	0.26
i8	0.38	1.58	0.30	i33	0.59	1.61	0.29
i9	-1.41	1.36	0.10	i34	0.41	1.43	0.33
i10	1.32	3.21	0.11	i35	0.92	1.91	0.30
i11	-0.10	1.70	0.14	i36	0.74	2.25	0.19
i12	-0.37	0.73	0.00	i37	0.48	1.55	0.15
i13	0.79	1.41	0.15	i38	1.17	2.24	0.14
i14	0.19	1.83	0.17	i39	1.99	1.45	0.19
i15	0.22	1.92	0.18	i40	-9.35	-0.07	0.00
i16	1.26	2.16	0.27	i41	0.47	1.28	0.06
i17	0.81	2.17	0.31	i42	0.61	2.57	0.08
i18	1.14	2.33	0.21	i43	1.06	2.20	0.26
i19	1.72	2.84	0.18	i44	1.57	2.59	0.19
i20	1.39	1.01	0.13	i45	1.69	4.44	0.23
i21	1.56	2.36	0.26	i46	1.09	2.46	0.16
i22	-0.42	1.54	0.10	i47	2.37	4.65	0.11
i23	-0.07	1.75	0.24	i48	-0.27	1.81	0.22
i24	-0.34	1.09	0.63	i49	1.65	2.61	0.29
i25	-1.50	1.01	0.16	i50	1.88	2.55	0.17

Anexo G. Tabla de atributos y descriptores

Tabla de descriptores de los atributos cognitivos

Habilidad	Atributos cognitivos	Descriptores	Categoría del descriptor
H1. Aplicación de conocimientos matemáticos básicos	H1A1. Establecer una única proporción comparando dos cantidades.	H1A1D1. Sabe que una razón es la comparación entre dos magnitudes que suele expresarse como una fracción (a/b).	Conocimiento
		H1A1D2. Sabe que una proporción es la relación de igualdad entre dos razones. Es decir, si a/b es una razón, entonces la igualdad a/b=c/d es una proporción, si y sólo si ad=bc, con b y d diferentes de cero.	Conocimiento
		H1A1D3. Identifica las cantidades que deben compararse para formar una proporción.	Conocimiento
		H1A1D4. Expresa en forma de fracción las dos razones que deben compararse para formar una proporción.	Destreza
		H1A1D5. Utiliza la proporción entre dos cantidades conocidas para encontrar una cantidad desconocida, utilizando la fórmula: a/b=c/d (procedimiento mejor conocido como regla de tres).	Destreza
		H1A1D6. Verifica la exactitud de una proporción, usando la ecuación de proporción cruzada: ad=bc.	Destreza
	H1A2. Resolver una ecuación lineal (por ejemplo, resta o división en ambos lados).	H1A2D1. Conoce la forma general de una ecuación lineal (ax+b=c, tal que a≠0, x es la incógnita, b y c son números reales).	Conocimiento
		H1A2D2. Sabe que resolver una ecuación implica encontrar el valor de incógnita que satisface la ecuación.	Conocimiento
		H1A2D3. Entiende el principio de la igualdad y que se debe mantener el equilibrio de la ecuación al realizar diferentes operaciones.	Conocimiento
		H1A2D4. Realiza operaciones de suma y resta en ambos lados de la ecuación para despejar la variable.	Destreza
		H1A2D5. Realiza operaciones de multiplicación y división en ambos lados de la ecuación para despejar la variable.	Destreza
		H1A2D6. Sabe verificar la solución de una ecuación al sustituir el valor encontrado en la ecuación.	Destreza
		H1A2D7. Puede resolver ecuaciones lineales de la forma ax+b=c.	Destreza
		H1A2D8. Puede resolver ecuaciones lineales de la forma ax+b=cx+d.	Destreza
	H1A3. Aplicar operaciones aritméticas básicas (por ejemplo, suma, resta, multiplicación y división).	H1A3D1. Conoce los números reales y su clasificación	Conocimiento
		H1A3D2. Conoce las propiedades de los números reales	Conocimiento
		H1A3D3. Comprende la relación de los signos al momento de realizar operaciones básicas	Conocimiento
		H1A3D4. Resuelve ejercicios donde hay operaciones de suma, resta, multiplicación y división	Destreza
		H1A3D5. Ejecuta manipulación aritmética para simplificar y resolver problemas rutinarios o de contexto real	Destreza
	H1A4. Sustituir valores en expresiones algebraicas.	H1A3D6. Identifica el símbolo para el cual se requiere solución	Conocimiento
		H1A4D1. Identifica la operación que está realizando la variable	Conocimiento
		H1A4D2. Identifica el símbolo para el cual se requiere solución	Conocimiento
		H1A4D3. Reconoce el coeficiente y la variable	Conocimiento
		H1A4D4. Aplica correctamente las operaciones básicas (suma, resta, multiplicación y división) en la simplificación y resolución de expresiones algebraicas	Destreza
	H1A5. Conocer el lenguaje básico que permite la comprensión del contenido o demanda del ítem.	H1A5D1. Reconoce cuando sumar o restar y el signo que domina el resultado	Conocimiento
		H1A5D2. Reconoce el como trabajar las operaciones básica (suma, resta, multiplicación y división) tanto en procesos aritméticos como algebraicos	Conocimiento
		H1A5D3. Resuelve ejercicios donde identifica que operación realizar dependiendo los signos y los términos semejantes	Destreza
		H1A5D4. Conoce el significado de los términos suma, resultado, resta, diferencia, producto, multiplicado, igual, cuadrado, doble, triple, etc en un contexto matemático.	Conocimiento
H2. Aplicación de conocimientos matemáticos avanzados.	H2A1. Ordenar series geométricas o cálculo numérico consecutivo. *Este atributo implica el conocimiento sobre series geométricas (por ejemplo, la naturaleza de la relación entre términos) o el cálculo numérico consecutivo (por ejemplo, multiplicación y división).	H2A1D1. Sabe que una serie geométrica es la suma de términos que tienen una razón constante entre sus términos sucesivos. Es decir, a+ar+a(r ²)+a(r ³)+..., tal que a es el primer término y r la razón común.	Conocimiento
		H2A1D2. Puede identificar la razón común r en una serie geométrica dada.	Destreza
		H2A1D3. Sabe que para calcular el n-ésimo término de una serie geométrica se utiliza la fórmula: a(n)=ar ⁿ -(n-1).	Conocimiento
		H2A1D4. Puede calcular el n-ésimo término de una serie geométrica.	Destreza
		H2A1D5. Puede ordenar los términos de una serie geométrica de forma creciente o decreciente.	Destreza
	H2A2. Realizar operaciones exponenciales o suma de potencias.	orden de los términos en una serie geométrica.	Destreza
		H2A2D1. Sabe que una potencia es una forma abreviada de representar una multiplicación repetida un n número de veces. De forma general a ⁿ , tal que a es la base y n el número de veces que se multiplica la base por sí misma.	Conocimiento
		H2A2D2. Identifica el exponente en una potencia.	Conocimiento
		H2A2D3. Puede calcular potencias de números reales.	Destreza
		H2A2D4. Sabe que cuando se multiplican potencias con la misma base, éstas se pueden simplificar manteniendo la base y sumando los diferentes exponentes. Es decir, (a ⁿ) (a ^m)=a ⁿ (m+n).	Conocimiento
		H2A2D5. Conoce la propiedad de potencia de una potencia: es decir, ((a ⁿ) ^m)=a ⁿ (nm).	Conocimiento
		H2A2D6. Sabe que la división de potencias con la misma base se puede simplificar manteniendo la base y restando los exponentes. Es decir, (a ⁿ)/(a ^m)=a ⁿ (n-m).	Conocimiento
		H2A2D7. Conoce la propiedad de potencia con exponente cero, sabiendo que a ⁰ =1 para cualquier número a distinto de cero.	Conocimiento
		H2A2D8. Conoce la propiedad de exponente negativo. Es decir, a ⁻ⁿ =1/a ⁿ .	Conocimiento
		H2A2D9. Puede simplificar expresiones exponenciales que requieran el uso de una propiedad de los exponentes.	Destreza
		H2A2D10. Puede simplificar expresiones exponenciales que requieran el uso de más de una propiedad de los exponentes.	Destreza
	H2A3. Graficar una función. * Este atributo implica el conocimiento de la gráfica de una función o la sustitución de puntos en la gráfica.	H2A3D1. Conoce los elementos del plano cartesiano: eje x, eje y, cuadrantes).	Conocimiento
		H2A3D2. Sabe que la fórmula general de una función es y=f(x).	Conocimiento
		H2A3D3. Puede llevar la función a la forma y=f(x) en caso de que no se encuentre en dicha forma.	Destreza
		H2A3D4. Sabe identificar el dominio y el rango de la función.	Conocimiento
		H2A3D5. Puede sustituir valores específicos del dominio (x) en la función para calcular los valores de la variable dependiente (y) correspondientes a un punto.	Destreza
		H2A3D6. Identifica los puntos obtenidos en la sustitución de los valores específicos de x en la función y los representa en coordenadas de la forma (x,y).	Conocimiento
		H2A3D7. Puede trazar con precisión los puntos (x,y) en el plano cartesiano.	Destreza
		H2A3D8. Puede conectar los puntos trazados para generar la gráfica de la función.	Destreza
	H2A4. Aplicar las propiedades de factorización.	H2A4D1. Sabe que la factorización es el proceso de descomponer una expresión algebraica como el producto de factores más simples.	Conocimiento
		H2A4D2. Identifica los términos comunes en la expresión algebraica.	Conocimiento
		H2A4D3. Puede aplicar la factorización por factor común en una expresión algebraica.	Destreza
		H2A4D4. Identifica como trinomio cuadrado, una expresión de la forma ax ² +bx+c.	Conocimiento
		H2A4D5. Puede factorizar trinomios como el producto de dos binomios mediante técnicas específicas según la estructura del trinomio (cuadrado perfecto, agrupación, fórmula cuadrática, etc).	Destreza
		H2A4D6. Reconoce expresiones que se pueden factorizar con diferencia de cuadrados.	Conocimiento
		H2A4D7. Puede factorizar expresiones con diferencia de cuadrados a ² -b ² =(a+b)(a-b).	Destreza
		H2A4D8. Identifica expresiones algebraicas que representan polinomios de grado superior a dos.	Conocimiento
		H2A4D9. Puede factorizar polinomios de grado superior a dos.	Destreza
		H2A4D10. Puede verificar que la factorización realizada es correcta multiplicando los factores para obtener la expresión original.	Destreza
	H2A5. Sustituir números en expresiones	H2A5D1. Identifica las variables en una expresión algebraica.	Conocimiento

algebraicas.	H2A5D2. Sabe que en las variables se debes sustituir los números.	Conocimiento
	H2A5D3. Puede sustituir los valores numéricos dados en el lugar de las variables en una expresión algebraica.	Destreza
	H2A5D4. Sabe utilizar jerarquía de operaciones para la simplificación posterior a la sustitución.	Conocimiento
	H2A5D5. Puede realizar operaciones aritméticas posterior a la sustitución para simplificar la expresión.	Destreza
H3A1. Aplicar reglas de factorización.	H3A1D1. Conoce las reglas, métodos de factorización	Conocimiento
	H3A1D2. Identifica que método, regla utilizar para factorizar	Conocimiento
	H3A1D3. Aplica la factorización al momento de obtener resultados indefinidos para validar la solución	Conocimiento
	H3A1D4. Aplica la factorización para encontrar con mayor facilidad la solución	Destreza
	H3A1D5. Reconoce los métodos para factorizar valores, expresiones algebraicas en una fracción	Destreza
H3A2. Sustituir expresiones algebraicas por otras expresiones algebraicas.	H3A2D1. Conoce lo que es una expresión algebraica	Conocimiento
	H3A2D2. Conoce el proceso de sustitución	Conocimiento
	H3A2D3. Identifica la relación de una expresión algebraica con otra	Conocimiento
	H3A2D4. Convierte expresiones algebraicas equivalentes para poder sustituirlas	Destreza
H3A3. Comprender o aplicar reglas.	H3A3D1. Comprende las reglas matemáticas generales (como la de signos, exponentes, potencias, raíces, jerarquía, etc) aplicables en distintos contextos.	Conocimiento
	H3A3D2. Identifica la necesidad de aplicar una regla específica en la resolución de un problema específico.	Destreza
H3. Gestión de la complejidad	H3A3D3. Aplica las reglas matemáticas específicas para la resolución de problemas.	Destreza
	H3A3D4. Verifica si la aplicación de las reglas matemáticas específicas produce el resultado esperado.	Destreza
	H3A3D5. Adapta la aplicación de reglas matemáticas básicas a situaciones o problemas con mayor complejidad.	Destreza
	H3A3D6. Explica la lógica detrás de la aplicación de una regla matemática en contextos específicos.	Conocimiento
	H3A4D1. Conoce los diferentes métodos para resolver sistemas de ecuaciones lineales	Conocimiento
	H3A4D2. Identifica que método utilizar analizando las ecuaciones lineales	Destreza
	H3A4D3. Ecuación lineal	Conocimiento
	H3A4D4. Sustituye el valor de una variable en otra	Conocimiento
	H3A4D5. Comprueba los resultados llegando a que el MI es igual al MD	Conocimiento
	H3A4D6. Redacta el resultado como parte de una coordenada que indica el punto de intersección S (x, y) o como el valor de las incógnitas o variables: x = , y =	Conocimiento
H4A1. Resolver series geométricas en un patrón abstracto.	H4A1D1. Comprende la definición y las propiedades de las series geométricas	Conocimiento
	H4A1D2. Conoce la fórmula para la suma de n términos	Conocimiento
	H4A1D3. Reconoce patrones abstractos en configuraciones geométricas y numéricas	Destreza
	H4A1D4. Conoce los principios de convergencia y divergencia de series	Conocimiento
	H4A1D5. Ejecuta manipulaciones algebraica para simplificar y resolver series geométricas	Destreza
	H4A1D6. Aplica correctamente la fórmula de la suma de una serie geométrica al problema dado	Destreza
H4A2. Representar o ejecutar múltiples soluciones algebraicas básicas.	H4A2D1. Reconoce las variables y constantes en expresiones y ecuaciones algebraicas	Conocimiento
	H4A2D2. Conoce los métodos de factorización de expresiones algebraicas	Conocimiento
	H4A2D3. Sabe las propiedades de las operaciones algebraicas (asociativa, conmutativa, distributiva) y cómo aplicarlas para simplificar y resolver ecuaciones	Conocimiento
	H4A2D4. Ejecuta manipulaciones simbólicas para simplificar o reescribir expresiones algebraicas en formas equivalentes	Destreza
	H4A2D5. Comprueba la validez de las soluciones obtenidas mediante la sustitución de valores en las ecuaciones originales	Destreza
	H4A2D6. Resuelve sistema de ecuaciones utilizando diferentes métodos	Destreza
H4A3. Representar o ejecutar múltiples soluciones algebraicas avanzadas. * Este atributo implica establecer una proporción única, habilidades para transformar fracciones y habilidades de análisis cómo, cuándo, dónde y/o cómo realizar la transformación.	H4A3D1. Conoce las propiedades y métodos de simplificación y transformación de fracciones algebraicas (incluyendo fracciones complejas y parciales)	Conocimiento
	H4A3D2. Comprende el concepto de función algebraica	Conocimiento
	H4A3D3. Utiliza métodos adecuados para resolver ecuaciones plinómicas de grado superior, ecuaciones racionales y ecuaciones algoritmicas	Destreza
	H4A3D4. Realiza manipulación algebraica avnzada para reescribir o simplificar expresiones complejas	Destreza
	H4A3D5. Comprueba la validez de las soluciones obtenidas	Destreza
	H4A4D1. Conoce las reglas de las potencias (multiplicación y división de potencias con la misma base, potencia a otra potencia, potencias negativas y potencias fraccionarias)	Conocimiento
H4A4. Multiplicar potencias o aplicar reglas en operaciones de exponenciales.	H4A4D2. Conoce las leyes de los logaritmos y cómo se relacionan con las operaciones exponenciales, incluyendo la propiedad de logaritmo de un producto, cociente y potencia	Conocimiento
	H4A4D3. Comprende las propiedades de las funciones exponenciales (incluyendo su comportamiento y gráficas) y su relación inversa con las funciones logaritmicas	Conocimiento
	H4A4D4. Aplica correctamente las reglas de exponentes en la simplificación y resolución de expresiones algebraicas	Destreza
	H4A4D5. Transforma expresiones que involucran exponentes negativos, fraccionarios y raíces	Destreza
	H4A4D6. Resuelve ecuaciones exponenciales utilizando técnicas algebraicas y logaritmicas	Destreza
H4A5. Establecer o resolver ecuaciones cuadráticas.	H4A5D1. Comprende la forma general de una ecuación cuadrática ($ax^2+bx+c=0$, donde a, b y c son coeficientes reales)	Conocimiento
	H4A5D2. Conoce la fórmula general para resolver una ecuación cuadrática y cómo se aplica	Conocimiento
	H4A5D3. Comprende que el discriminante (b^2-4ac) le permite determinar la naturaleza de las soluciones de una ecuación cuadrática (reales distintas o iguales, o complejas)	Conocimiento
	H4A5D4. Aplica técnicas de factorización para resolver ecuaciones cuadráticas	Destreza
	H4A5D5. Usa el discriminante para determinar la naturaleza de las soluciones de una ecuación cuadrática	Destreza
	H4A5D6. Resuelve una ecuación cuadrática mediante el método de completar un trinomio cuadrado perfecto	Destreza
	H4A5D7. Formula una ecuación cuadrática a partir de un problema verbal.	Destreza
	H4A6D1. Identifica variables dependientes e independientes.	Conocimiento
	H4A6D2. Comprende el concepto de función donde la variable dependiente se expresa en términos de la independiente.	Conocimiento
	H4A6D3. Conoce las propiedades y características de las gráficas de funciones: pendiente, intersección con los ejes, máximos y mínimos, comportamiento asintótico	Conocimiento
H4. Modelamiento matemático	H4A6D4. Grafica funciones representando adecuadamente la relación entre variables en un plano cartesiano	Destreza
	H4A6D5. Interpreta la relación entre las variables a través de analizar las gráficas	Destreza
	H4A6D6. Resuelve problema contextualizados mediante el análisis de las gráficas	Destreza
	H4A7D1. Conoce las propiedades y características de las curvas cuadráticas (circunferencia, parábola, elipse, hipérbola)	Conocimiento
	H4A7D2. Entiende cómo las desigualdades cuadráticas dividen al plano	Conocimiento
	H4A7D3. Transforma las desigualdades en igualdades para facilitar la resolución	Destreza
	H4A7D4. Usa puntos de prueba para determinar las regiones del plano que satisfacen la desigualdad	Destreza
	H4A7D5. Verifica las soluciones obtenidas mediante la sustitución de puntos	Destreza
H4A7. Resolver una desigualdad cuadrática con dos variables.		

Anexo H. Parámetros de adivinación y desliz de los ítems de la prueba

Tabla H1

Parámetros de adivinación y desliz de los ítems por MDC

Item	DINA		DINO		G-DINA	
	<i>g</i>	<i>s</i>	<i>g</i>	<i>s</i>	<i>g</i>	<i>s</i>
i1	0.122	0.382	0.114	0.420	0.118	0.305
i2	0.579	0.091	0.576	0.093	0.542	0.079
i3	0.514	0.022	0.501	0.033	0.486	0.022
i4	0.421	0.130	0.428	0.125	0.413	0.126
i5	0.367	0.059	0.322	0.090	0.301	0.056
i6	0.338	0.100	0.278	0.159	0.261	0.086
i7	0.284	0.283	0.292	0.281	0.278	0.284
i8	0.422	0.186	0.427	0.180	0.414	0.182
i9	0.754	0.028	0.741	0.029	0.719	0.029
i10	0.106	0.513	0.112	0.517	0.115	0.500
i11	0.074	0.687	0.078	0.682	0.067	0.706
i12	0.365	0.137	0.366	0.132	0.362	0.139
i13	0.043	0.802	0.052	0.796	0.055	0.789
i14	0.555	0.060	0.550	0.061	0.532	0.065
i15	0.221	0.604	0.222	0.593	0.218	0.617
i16	0.306	0.393	0.304	0.408	0.298	0.385
i17	0.390	0.167	0.384	0.170	0.371	0.164
i18	0.533	0.095	0.540	0.085	0.519	0.091
i19	0.109	0.642	0.104	0.654	0.103	0.635
i20	0.159	0.532	0.165	0.510	0.154	0.540
i21	0.278	0.290	0.275	0.295	0.267	0.275
i22	0.302	0.395	0.301	0.400	0.292	0.385
i23	0.418	0.110	0.405	0.125	0.413	0.114
i24	0.574	0.076	0.577	0.070	0.569	0.072
i25	0.081	0.625	0.077	0.631	0.074	0.615
i26	0.408	0.204	0.406	0.218	0.396	0.211
i27	0.077	0.753	0.070	0.757	0.062	0.749
i28	0.131	0.809	0.132	0.819	0.123	0.817
i29	0.380	0.108	0.371	0.110	0.366	0.111
i30	0.286	0.487	0.283	0.513	0.270	0.483
i31	0.162	0.626	0.164	0.614	0.159	0.618
i32	0.144	0.675	0.148	0.661	0.131	0.668
i33	0.466	0.031	0.476	0.024	0.463	0.030
i34	0.186	0.625	0.196	0.588	0.188	0.615
i35	0.413	0.220	0.409	0.214	0.400	0.217
i36	0.205	0.456	0.206	0.461	0.200	0.476
i37	0.279	0.227	0.274	0.225	0.256	0.222
i38	0.476	0.044	0.476	0.043	0.462	0.043
i39	0.295	0.458	0.297	0.458	0.279	0.474
i40	0.350	0.183	0.340	0.194	0.348	0.163
i41	0.492	0.018	0.491	0.018	0.476	0.019
i42	0.463	0.033	0.459	0.033	0.438	0.034
i43	0.169	0.676	0.171	0.672	0.164	0.686
i44	0.134	0.649	0.133	0.645	0.121	0.656
i45	0.425	0.100	0.418	0.095	0.411	0.103
i46	0.315	0.191	0.318	0.185	0.306	0.197
i47	0.322	0.142	0.319	0.138	0.304	0.130
i48	0.208	0.476	0.202	0.486	0.197	0.480
i49	0.029	0.895	0.030	0.892	0.029	0.886
i50	0.498	0.042	0.499	0.044	0.482	0.045
