**Attribute based identification in mathematic processes in a large scale assessment: The Mexican case**

GDRR, JCPM, RVL, AFCDP

**Resumen**

Los modelos de diagnóstico cognitivo (CDM) son modelos psicométricos desarrollados para identificar con detalle, a partir de la aplicación de una prueba y a lo largo del conjunto de conocimientos y habilidades cognitivas que conforman el dominio evaluado, las fortalezas y debilidades de cada sustentante y permiten trazar perfiles detallados que orienten el desarrollo de acciones de mejora focalizadas. En este trabajo se presenta la aplicación de una técnica de retrofitting usando el modelo DINA (de la Torre, 2009) sobre los datos obtenidos en una evaluación estandarizada de Matemáticas aplicada a gran escala a estudiantes de educación básica en México: PLANEA (INEE, 2017). La matriz Q fue construida a partir de una revisión curricular y entrevistas cognitivas dirigidas a expertos y alumnos. El ajuste del modelo DINA a los datos mostró ser el mejor dentro de los CDM. A partir de este trabajo, se pudo elaborar un diagnóstico cognitivo nacional, cuyas implicaciones educativas, empíricas y de apoyo pedagógico para la mejora en el aula se discuten a detalle.

Keywords: attribute mastery, cognitive diagnosis modeling, DINA model, mathematics, large scale assessment

**Introducción**

Una reciente y robusta línea de investigación en psicometría se ha enfocado en medir múltiples dimensiones o componentes de un constructo latente para proporcionar información específica y detallada acerca del desempeño de los sustentantes, a diagnóstico cognitivo. De manera general, el objetivo principal de los distintos modelos de diagnóstico cognitivo es poder identificar el estado o grado de dominio que los sustentantes evaluados presentan a lo largo de un conjunto de habilidades, conocimientos o atributos latentes para lograr la construcción de un perfil diagnóstico específico, que permita a todo agente involucrado conocer las áreas de fortaleza y mejora identificadas para orientar el desarrollo de estrategias de mejora. En el ámbito educativo, la aplicación de estos modelos permite a padres de familia, docentes, directores y otros agentes educativos, tomar mejores decisiones con respecto al diseño, uso, manejo e implementación de recursos y materiales didácticos de apoyo hacia los estudiantes. Cabe señalar que existe una amplia gama de modelos de diagnóstico cognitivo (CDM), para una revisión más extensa consultar se sugiere consultar a Cohen, (2019), Ferrara, Lai, Reilly, Nichols, Rupp y Leighton, (2017), o a van der Linden, (2016).

Los CDM pueden entenderse como modelos de clases latentes restringidos que modelan las respuestas observadas como una función de variables latentes discretas que capturan las operaciones cognitivas subyacentes (Templin y Henson, 2006). En otras palabras, los CDM interpretan el desempeño de los participantes a través de una serie de atributos cognitivos binarios, a diferencia de los modelos derivados de la Teoría Clásica de los Test (TCT) o la Teoría de Respuesta al Ítem (TRI) que interpretan el desempeño de los sustentantes como reflejo de su posición a lo largo de un solo continuo que representa el dominio de una única variable o habilidad latente evaluada.

Para que una evaluación pueda ser considerada un diagnóstico cognitivo, el diseño de la misma debe considerar la integración de diversas teorías de aprendizaje, cognición y pedagogía con teorías de medición que permitan el desarrollo de evaluaciones que no se limiten a medir y evaluar, sino que sirvan como insumo para impulsar la mejora del aprendizaje de los estudiantes (Chudowsky y Pellegrino, 2003; Shepard, 2000).

La gran mayoría de los modelos de diagnóstico cognitivo requieren de la construcción de una matriz donde se identifique para cada ítem, cuáles son las habilidades o variables latentes asociadas al dominio general evaluado que se consideran “necesarias” para obtener un acierto. El arreglo resultante es conocido en la literatura como matriz Q, y su construcción requiere del trabajo conjunto de expertos en el dominio evaluado, sustentantes que tras responder cada reactivo den cuenta de los procedimientos seguidos y de expertos en psicometría que estén constantemente revisando que los atributos identificados como “requeridos” por cada ítem, hagan sentido a la luz de las respuestas observadas, (referencias).

**Acerca de los Modelos de Diagnóstico Cognitivo**

Cada uno de los CDMs hace supuestos específicos sobre cómo el dominio de los distintos atributos es ponderado para producir una respuesta correcta o incorrecta. Una distinción importante tiene que ver con si el modelo es conjuntivo o disyuntivo (Rupp, Templin y Henson, 2010). Los modelos conjuntivos asumen que se requiere el dominio de todos los atributos asociados a cada ítem para poder responder de manera exitosa. En contraste, los modelos disyuntivos asumen que la falta de dominio de un atributo puede ser compensada por el dominio de otros atributos, (por ejemplo, algunos CDM asumen que el dominio de los atributos tiene un efecto aditivo referencias).

Algunos de los CDMs más desarrollados y utilizados en la literatura son el modelo conjuntivo DINA (entrada determinística, ruidosa "y" puerta; Junker & Sijtsma, 2001; de la Torre, 2009) y su variante disyuntiva, el modelo DINO (entrada determinística, ruidosa "o" puerta; Templin y Henson, 2006), y el A-CDM (CDM aditivo; de la Torre, 2011). Según Rupp, Templin y Henson, (2010), otros CDM bien conocidos son el modelo NIDA (determinista de entrada ruidosa “y”; Junker y Sijtsma, 2001, Maris, 1999), el NIDO (determinista de entrada ruidosa o, Douglas, de la Torre, Chang, Henson & Templin, 2006), y el R-RUM (modelo unificado de reparación reducida; Hartz, 2002). Además, se han propuesto CDM generales que respetan los supuestos de modelos específicos (véase, por ejemplo, Henson, Templin & Willse, 2009; von Davier, 2005), o bien, que funcionan como modelos generales, como es el caso del modelo G-DINA (DINA generalizada; de la Torre, 2011), el modelo de diagnóstico cognitivo log-lineal (LCDM; Henson, Templin y Willse, 2009) y el modelo de diagnóstico general (GDM; von Davier , 2005). Este último grupo de modelos, describe la probabilidad de éxito en términos de la suma de los efectos debidos a la presencia de atributos específicos y sus interacciones.

**Acerca del Modelo DINA**

El modelo DINA constituye uno de los modelos más sencillos dentro de la familia de los CDMs (Junker y Sijtsma, 2001; de la Torre, 2009), pues considera únicamente dos parámetros libres por ítem, con el objetivo de describir la probabilidad de que los aciertos o errores registrados no estén relacionados con el grado de dominio que los sustentantes tienen en las habilidades requeridas. Estos parámetros, conocidos como parámetro de adivinación y parámetro de desliz, refieren a la probabilidad de obtener un acierto aún sin dominar las habilidades necesarias, (es decir, de “atinarle por por puro azar” a la respuesta correcta), y la probabilidad de errar el ítem aún dominando las habilidades necesarias, (en otras palabras, de cometer un “desliz” al momento de seleccionar una respuesta).

El modelo DINA se expresa a partir de la siguiente ecuación,

De acuerdo con la Ecuación 1, el modelo DINA está compuesto por los siguientes elementos:

1.- Una variable binaria para indicar si la persona obtuvo un acierto (1) o un error (0) en el ítem . Esto implica la existencia de un vector Y compuesto por filas e columnas, donde se concentran los aciertos y errores obtenidos por cada persona en cada ítem.

2.- Un vector que contiene a la variable binaria que señala si la persona domina (1) o no (0) el subdominio evaluado en la prueba. Al conjunto de dominios evaluados suele referírseles, en el marco de los CDM, como habilidades, conocimientos o atributos contenidos en el dominio general evaluado por la prueba.

3.- Un vector que por cada ítem , contiene a la variable binaria que señala la respuesta “idealmente” registrada (en términos de acierto o error) por el sustentante en el ítem , dado lo que la matriz Q, el modelo cognitivo detrás de la prueba, ha establecido acerca de qué habilidades se requieren para obtener un acierto en cada ítem ( y lo que el vector nos dice sobre las habilidades dominadas por el participante . De esta forma, tenemos que el vector queda plasmado como el elemento determinista del modelo, definido matemáticamente como:

, donde

4.- Un parámetro de desliz , que nos dice que aunque la respuesta idealmente esperada por el participante al ítem , sea 1 (porque domina todas las habilidades requeridas por el ítem, ), el participante cometa un “desliz” y se observe aun así que . Es decir,

)

5.- Un parámetro de adivinación , que representa la probabilidad de que un examinado que no posee todas las habilidades requeridas por el ítem , (), “le atine” a la respuesta correcta y respondan correctamente el ítem (). Es decir,

El modelo DINA nos dice que para cada ítem , solo los examinados que dominen todos los atributos requeridos tendrán una probabilidad de éxito igual a , mientras que todos los demás examinados tendrán una probabilidad de éxito igual a . Es decir, que como habíamos mencionado previamente, el modelo DINA es un modelo conjuntivo que asume que es necesario el dominio de todos y cada uno de los atributos relacionados con cada ítem.

**Acerca de la Prueba de Matemáticas para Primaria (06) del PLANEA ELCE 2015**

La Prueba de Matemáticas para Primaria (06) con la que se trabajó en el presente proyecto, forma parte del Plan Nacional para las Evaluaciones de los Aprendizajes (PLANEA) diseñado y aplicado en 2015 en México por el entonces Instituto Nacional para la Evaluación de la Educación (INEE), con la intención de contar con una evaluación a gran escala de la eficacia con la que el Sistema Educativo Nacional fomenta el dominio de los aprendizajes esperados de acuerdo al currículo nacional.

En especial, el PLANEA-ELCE toma los centros escolares como unidad de análisis (Evaluación del Logro de los Centros Escolares) y contempla la evaluación de campos formativos específicos (Matemáticas y Lenguaje y comunicación) tomando como variable de agrupación distintos factores (por ejemplo, región, estado, municipio, zona escolar, modalidad, tipo de organización, entre otros), (INEE, 2015). El propósito específico tras su desarrollo fue el ofrecer a cada escuela elementos de retroalimentación para su mejora, identificando el logro alcanzado en el centro escolar tras los seis años que dura la educación primaria.

**Método:**

La presente investigación tuvo como objetivo realizar, mediante la aplicación de técnicas de retrofitting, un diagnóstico cognitivo nacional para identificar el grado de dominio que los estudiantes de sexto de primaria en México tienen en Matemáticas. Para ello se trabajó con la Prueba de Matemáticas para primaria (06) del PLANEA ELCE 2015 y los datos levantados tras su aplicación a gran escala, a un total de no. estudiantes de sexto de primaria. Una ventaja colateral de trabajar con la prueba de Matemáticas de primaria (06) del PLANEA ELCE 2015 es que se trata de una disciplina ampliamente abordada desde el enfoque propuesto por los CDM (Brown y Burton, 1978; Chen y Macdonald, 2011; Gierl, Leighton, Changjiang, Jiawen, Rebecca & Tan, 2009; Ma, Çetin y Green, 2009; Pérez-Morán, 2014; Pérez-Morán, Contreras-Roldan, Hernández, Olivares, Chan, y Díaz, 2014; Pérez-Morán, Larrazolo, Backhoff, y Guaner, 2015; Revuelta y Ponsoda, 1998; Romero, Ponsoda y Ximénez, 2008).

Se realizaron estudios cognitivos para identificar las habilidades cognitivas requeridas por los distintos ítems que conforman la prueba, mediante la aplicación de técnicas de retrofitting, para poder trazar la matriz Q que guiaría la aplicación del modelo DINA a los datos. Para ello se elaboró un análisis cognitivo-reticular detallado con apoyo de un panel de expertos con experiencia en la enseñanza de las Matemáticas a nivel primaria, que además participaron en la aplicación de técnicas de pensamiento en voz alta concurrentes y retrospectivas con los expertos, para identificar los procesos de respuesta asociados a cada ítem que guiaran el modelado matemático de sub-tareas.

Los 50 ítems contenidos en la prueba fueron piloteados con un grupo de 16 estudiantes. A partir de este piloteo, se obtuvieron los reportes verbales de los procesos de respuesta empleados por los participantes a partir del uso de técnicas concurrentes y retrospectivas (Ericsson y Simon, 1984, 1993; Leighton, 2009; Leighton y Gierl, 2007); se realizaron análisis del sendero de la vista (*eye-tracking*; Snow y Lohman, 1989; Sternberg, 1977) y se analizaron las latencias de respuesta (Fredericksen, 1980; Posner, 1978; Posner y Rogers, 1978). Los resultados obtenidos tras el piloteo fueron revisados por el grupo de expertos, previa capacitación, para verificar la correspondencia entre los modelos de respuesta identificados por expertos y estudiantes.

Como resultado de este estudio cognitivo, se desarrolló un modelo cognitivo que diera cuenta de las habilidades, conocimientos y procesos cognitivos evaluados o requeridos por cada ítem contenido la prueba (la matriz Q). Para ello, se buscó determinar la cantidad y el tipo de relaciones existentes entre los ítems y los atributos u operaciones cognitivas sustantivas determinadas por los expertos y con ello, comenzar a elaborar la matriz Q de la prueba, (Rupp, Templin, y Henson, 2010).

En cuanto a la revisión de la calidad técnica de los ítems de Matemáticas del PLANEA ELCE 2015, se obtuvieron indicadores propios de la Teoría Clásica de los Tests (TCT) y se realizó un análisis de estructura interna mediante la aplicación de Análisis Factoriales Exploratorios (AFE) con el método de estimación de mínimos cuadrados ponderados para variables categóricas. Los análisis para la revisión de la calidad técnica y validez de la prueba se realizaron seleccionando una submuestra aleatoria de 5000 estudiantes dentro de la base de datos disponible.

Los indicadores de la TCT fueron obtenidos con la paquetería CTT del programa libre R 2.15.1. (Ihaka, R. y Gentleman, R., 1996; Willse, J.T. & Shu, Z., 2014) para ser evaluados a la luz de los estándares clásicos de calidad técnica. Los indicadores psicométricos revisados fueron el índice de dificultad, el índice de discriminación, el coeficiente de correlación punto-biserial y el coeficiente de consistencia interna para la prueba y si se elimina un ítem (**α** de Cronbach).

Para el análisis de la estructura interna se aplicó un modelo de Análisis Factorial Exploratorio usando los paquetes *nFactors* y *psych* de R (Ihaka, R. y Gentleman, R., 1996; Revelle, W., 20111, Raiche, G, Magis, D., Raiche, M. G., 2013). Los indicadores de ajuste absoluto que se emplearon fueron la raíz de los residuos cuadráticos promedios estandarizados (SRMR) y la raíz del error cuadrático promedio de aproximación (RMSEA), mismos que de acuerdo a la literatura deben ser inferiores a 0,05 para poder asumir un buen ajuste de los modelos factoriales.

Para el estudio de la estructura interna de la prueba se procedió a agrupar los ítems según su pertenencia a los ejes y contenidos curriculares. En cada eje de la prueba se implementó un análisis factorial exploratorio con el fin de comprobar la dimensión dominante. Como ya se mencionó, se utilizó el método de estimación de mínimos cuadrados ponderados dado que se analizaron variables categóricas.

Finalmente, se realizaron las estimaciones correspondientes a la aplicación del modelo DINA, utilizando la librería *CDM* en R, (Ihaka, R. y Gentleman, R., 1996; George, A., Robitzcsh, A., Kiefer, T., Gross, J. & Ünlü, A., 2016). Se computaron los parámetros de desliz y de adivinación de cada ítem y se estimó el grado de dominio que cada uno de los participantes que conforman la muestra tiene en cada una de las 35 habilidades evaluadas.

El ajuste del modelo DINA a los datos recolectados en el PLANEA ELCE 2015 fue evaluado a partir de dos indicadores: el Criterio de Información Bayesiana (BIC) y el Criterio de Información Akaike (AIC). De acuerdo con estos indicadores, el modelo DINA demostró el modelo con el mejor ajuste dentro de los modelos CDM evaluados (DINO, evaluar más modelos).

Como resultado de dicha estimación, se obtuvieron perfiles diagnósticos individualizados por alumno, por centro escolar y por estado, mismos que fungieron como insumo principal para la realización de un primer Diagnóstico Nacional de las habilidades básicas de Matemáticas en alumnos de sexto de primaria.

Resultados

Análisis de la precisión métrica de la prueba

En cuanto al cómputo de los indicadores de la TCT, destaca en general la obtención de valores que cumplen con los estándares de calidad convencionales de manera suficiente.

Los índices de consistencia interna computados para cada eje temático mostraron un valores entre 0.65 y 0.79 de de Cronbach (ver **Tabla 2**). Los ejes *Forma, Espacio y Medida* y *Manejo de la Información*, mostraron índices de confiabilidad satisfactorios al presentar valores alfa de 0.71 y 0.79 respectivamente. Sin embargo, el eje *Sentido numérico* y *Pensamiento algebraico* presentó un índice de confiabilidad bajo (alfa = 0.65), que sugiere una mayor variabilidad en las respuestas observadas en los reactivos que lo integran y que, por lo tanto, señalan la necesidad de revisar a detalle dichos reactivos en tanto que se considera altamente probable exista un efecto ocasionado por errores aleatorios (no sistemáticos) asociados al proceso de medición. Cabe señalar también que el análisis no identificó ningún ítem cuya eliminación incrementara de manera significativa el valor de las alfas en cada eje.

**Tabla 2**. Alfa de Cronbach global para cada eje

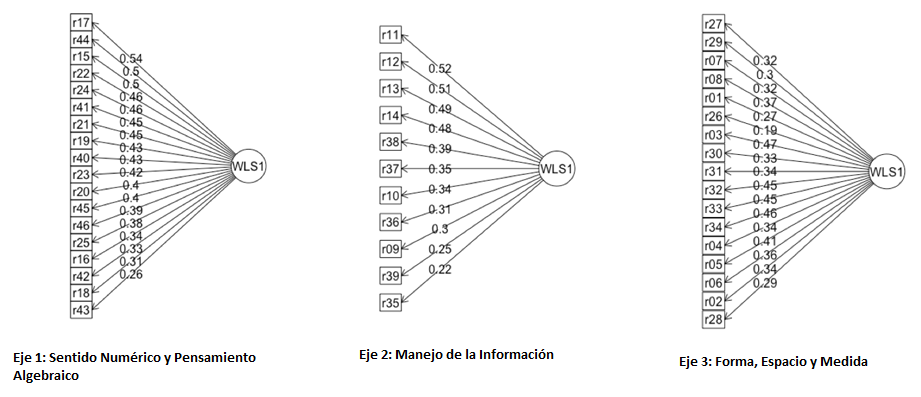
|  |  |
| --- | --- |
| Eje | Alfa de Cronbach (α) |
| Espacio, forma y medida | α=0.71 |
| Sentido numérico y pensamiento algebraico | α=0.65 |
| Manejo de información | α=0.79 |

Los índices de Discriminación **rbis** calculados para la totalidad de los ítems que conforman la prueba superan el umbral establecido de 0.3, con excepción de los ítems PMB01, PMB10 y PMB14. En cuanto a los ejes temáticos, el eje *Espacio, forma y medida* presenta un valor promedio de discriminación superior a 0.3; el eje *Manejo de la Información*, un promedio de .38 y el eje *Sentido Numérico y Pensamiento Alegbraico*, 0.47.

Por su parte, en cuanto a los índices de dificultad computados para la prueba, se encuentran valores entre 0.12 y 0.71, siendo que la mayoría de los ítems presentan niveles intermedios de dificultad, cumpliendo así con los estándares de calidad requeridos para los propósitos del estudio. El eje *Espacio, forma y medida* presenta un índice de dificultad promedio de 0.48; el eje *Manejo de la Información,* 0.46 y el eje de *Sentido Numérico y Pensamiento Algebraico*, un promedio de 0.42.

**Análisis de la estructura interna**

Al realizar un análisis factorial para evaluar la estructura de la dimensión latente evaluada en la prueba, se detectó que todos los ítems mostraron una carga factorial mayor a 0.3 (un umbral usualmente considerado aceptable) en un factor predominante, mismo que estaría asociado al constructo general (“habilidad matemática”) evaluado en la prueba.



*Figura 3.* Se presentan los diagramas correspondientes al análisis factorial realizado para cada uno de los ejes temáticos identificados como parte del modelo cognitivo de la Prueba de Matemáticas para Primaria (06) del PLANEA ELCE.

Adicionalmente, para evaluar el ajuste de la estructura factorial propuesta por el modelo cognitivo elaborado se trabajó con dos estimadores, el Standarized Root Mean Square Residual (SRMR) y el Root Mean Squared Error of Aproximation (RMSEA), que de acuerdo a lo esperado, presentaron valores inferiores a 0.05 para todos los ítems contenidos en cada eje temático. Esto sugiere que en cada eje existe un sólo factor dominante que podría asociarse con el contenido teórico trazado por los desarrolladores de la prueba, por lo que se valida la calidad técnica del diseño y funcionamiento de la prueba. La Figura 3 presenta para cada eje temático un diagrama que señala las cargas factoriales de cada uno de los ítems que lo componen.

**Modelamiento matemático de sub-tareas de respuesta**

Tras aplicar el método de *modelado matemático de sub-tareas de respuesta*, se obtuvieron diferentes los modelos del proceso de respuesta de cada uno de los ítems de la prueba. Por ejemplo, como se observó con el ítem PMA01, los expertos definieron que la secuencia de pasos a seguir para contestarlo correctamente es la siguiente: (a) leer detalladamente las indicaciones del ítem, (b) comprender el contexto del problema, (c) comprender el objetivo de la tarea evaluativa, (d) deducir de elementos del contexto el cálculo del área del triángulo, (e) comprender el objetivo de la tarea evaluativa, (f) recordar y representar la fórmula para calcular el área del triángulo, (g) seleccionar y sustituir los valores de la fórmula, (h) aplicar operaciones aritméticas básicas, (i) reconocer el resultado dentro de las opciones de respuesta y (j) seleccionar la opción de respuesta (ver **Tabla 4**).

**Tabla 4.** Esquema del modelo cognitivo de procesos de respuesta del ítem PMA01

|  |  |
| --- | --- |
| **Ítem PMA01** | **Proceso de respuesta** |
|  | Leer detalladamente las indicaciones del ítem |
| Comprender el contexto del problema |
| Comprender el objetivo de la tarea evaluativa |
| Deducir de elementos del contexto el cálculo del área del triángulo |
| Comprender el objetivo de la tarea evaluativa |
| Recordar y representar la fórmula para calcular el área del triángulo |
| Seleccionar y sustituir los valores de la fórmula |
| Aplicar operaciones aritméticas básicas |
| Reconocer el resultado dentro de las opciones de respuesta |
| Seleccionar la opción de respuesta |

**Análisis de los protocolos verbales**

De acuerdo a los reportes verbales, se confirmaron algunos de los problemas detectados durante el análisis inicial de los ítems de la prueba Matemáticas 06 PLANEA ELCE 2015, pues los sustentantes manifestaron problemas afines al tratar de resolver las tareas evaluativas. Además, se identificaron problemas de diseño que no habían sido considerados previamente; algunos de ellos referentes a problemas de sesgo y/o de diseño. Por ejemplo, en el caso del ítem PMA01 (ya presentado en la Tabla 5), el uso de la palabra “banderola” resultó confuso, tratándose de una palabra desconocida para la mayoría de los estudiantes que participaron en los protocolos verbales, incluso para aquellos que presentaron un alto rendimiento académico.

El análisis de los protocolos verbales permitió, según fue el caso, la validación y la adecuación los modelos de los procesos de respuesta trazados para cada ítem, y se perfeccionó el esquema cognitivo diseñado para dar cuenta de la varianza irrelevante que pudiera estar determinando el comportamiento del ítem.

**Análisis de las operaciones cognitivas que explican la dificultad de cada ítem**

Después del trabajo realizado para analizar y revisar de manera exhaustiva los componentes sustantivos de la Prueba de Matemáticas para Primaria (06) del PLANEA ELCE, se identificaron 35 habilidades cognitivas que son evaluadas a través de tres grandes Ejes Temáticos, descritas en la **Tabla 5**.

**Tabla 5**. Habilidades cognitivas evaluadas en la Prueba de Matemáticas para primaria (06) del PLANEA ELCE por eje temático

|  |  |
| --- | --- |
| Eje | Habilidades |
| 1 Sentido Numérico y Pensamiento Algebraico | **SNPA01:** Comprensión de problemas matemáticos contextualizados  **SNPA02:** Comprensión del Sistema Internacional de Unidades (SIU)  **SNPA03:** Aplicación de operaciones aritméticas básicas  **SNPA04:** Representación del modelo aditivo de números fraccionarios  **SNPA05:** Amplificación de fracciones (Equivalencia de fracciones por amplificación)  **SNPA06:** Representación del modelo aritmético de la división  **SNPA07:** Representación de números fraccionarios  **SNPA08:** Inferencia del patrón que rige una secuencia de números naturales  **SNPA09:** Conversión de texto cardinal a números naturales y viceversa  **SNPA10:** Operación de valores posicionales con números naturales o decimales  **SNPA11:** Representación del modelo multiplicativo de números fraccionarios por naturales.  **SNPA12:** Conversión de una regla verbal de progresión geométrica de ascendente a una sucesión numérica  **SNPA13:** Deducción del patrón de una sucesión con progresión especial |
| 2 Manejo de la Información | **MI01:** Comprensión de problemas matemáticos contextualizados  **MI02:** Comparación de la proporcionalidad de razones  **MI03:** Representación de modelos aritméticos de la media (promedio)  **MI04:** Representación de modelos aritméticos de la mediana  **MI05:** Aplicación de operaciones aritméticas básicas  **MI06:** Representación de datos numéricos en gráficas de barras  **MI07:** Representación del modelo de regla de tres simple  **MI08:** Comprensión de la relación entre porcentajes y fracciones  **MI09:** Comparación de razones con cantidades discretas  **MI10:** Representación de un número fraccionario |
| 3 Forma, Espacio y Medida | **FEM01:** Comprensión de problemas matemáticos contextualizados  **FEM02:** Comprensión del Sistema Internacional de Unidades (SIU)  **FEM03:** Operación de valores posicionales con números naturales y decimales  **FEM04:** Ubicación de una coordenada en el primer cuadrante del plano cartesiano  **FEM05:** Aplicación de operaciones aritméticas básicas  **FEM06:** Definición de tecnicismos del lenguaje formal de la geometría  **FEM07:** Representación viso-espacial de figuras geométricas  **FEM08:** Identificación de las características geométricas de los cuadriláteros  **FEM09:** Identificación gráfica de tipos de líneas rectas (paralelas, perpendiculares y secantes)  **FEM10:** Representación del modelo aritmético para calcular el perímetro de una figura geométrica (triángulo o cuadrilátero)  **FEM11:** Representación del modelo aritmético para calcular el área de cuadriláteros o triángulos  **FEM12:** Deducción de fórmulas para calcular el área mediante descomposición de figuras geométricas |

Por ejemplo, si una vez más tomamos como ejemplo el ítem PMA01, las operaciones que explican su dificultad son: Comprensión de problemas matemáticos contextualizados, Comprensión del Sistema Internacional de Unidades (SIU), Representación del modelo aritmético para calcular el área de cuadriláteros o triángulos y Aplicación de operaciones aritméticas básicas (ver Figura 3).

 *Figura 3*. Diagrama del modelo de un proceso de respuesta erróneo subyacente al ítem PMA01 de la prueba de Matemáticas (06) del PLANEA ELCE 2015

A partir de la Tabla 5, se diseñó la Matriz Q que sintetiza la relación entre las operaciones cognitivas identificadas y los procesos de respuesta identificados en cada uno de los ítems. Como su nombre sugiere, se trata de una Matriz compuesta por tantas columnas como operaciones cognitivas fueron identificadas y de tantas filas como ítems hay en la prueba, que contiene valores dicotómicos para señalar qué operaciones son requeridas por cada ítem (1) y cuáles no (09).  La Matriz Q construida puede consultarse en la **Tabla 6**.

**Tabla 6.** Matriz Q (Extracto de la tabla original)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Código de ítem | **O1** | **O2** | **O3** | **O4** | **O5** | **O6** | **O7** | **O8** | **O9** | **O10** | **O11** | **O12** | **O13** |
| PMA05 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMB09 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMB07 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMA06 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMA04 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMA03 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMB06 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMB08 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| PMB05 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| PMA01 | **1** | **1** | 0 | 0 | **1** | 0 | 0 | 0 | 0 | 0 | 0 | **1** | 0 |
| PMB01 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| PMB04 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| PMA07 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMB02 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| PMA08 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMB03 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PMA02 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Aplicación del Modelo DINA

Con base en la Matriz Q antes presentada, las respuestas registradas por los estudiantes que participaron en la aplicación del PLANEA ELCE fueron r analizadas con el modelo DINA para obtener estimaciones acerca de la probabilidad de dominio que cada participante presenta para cada operación cognitiva evaluada, así como para estimar el valor de los parámetros de desliz y adivinación que caracterizan cada ítem.

En la **Tabla 7** se muestra nuestra primera aproximación a lo que sería un Diagnóstico Cognitivo nacional de las habilidades básicas en matemáticas para estudiantes de sexto año de primaria. Es decir, se presentan las probabilidades de dominio para cada una de las operaciones cognitivas estimadas para la muestra total. Como se puede observar, las probabilidades oscilan entre 0.49 y 0.68, valores que podrían considerarse “moderados” y que podrían estar sugiriendo la necesidad de elaborar e implementar planes de mejora orientados a promover el logro del aprendizaje.

**Tabla 7.** Probabilidades de dominio de las operaciones cognitivas por parte de los examinados en los tres ejes temáticos del PLANEA ELCE (06) 2015 de Matemáticas

|  |  |  |
| --- | --- | --- |
| **Eje** | **Operación cognitiva** | **Probabilidad de dominio** |
| I. Sentido Numérico y Pensamiento Algebraico | **SNPA01:** Comprensión de problemas matemáticos contextualizados | 0.58 |
| **SNPA02:** Comprensión del Sistema Internacional de Unidades (SIU) | 0.59 |
| **SNPA03:** Aplicación de operaciones aritméticas básicas | 0.65 |
| **SNPA04:** Representación del modelo aditivo de números fraccionarios | 0.52 |
| **SNPA05:** Amplificación de fracciones (Equivalencia de fracciones por amplificación) | 0.57 |
| **SNPA06:** Representación del modelo aritmético de la división | 0.59 |
| **SNPA07:** Representación de números fraccionarios | 0.65 |
| **SNPA08:** Inferencia del patrón que rige una secuencia de números naturales | 0.57 |
| **SNPA09:** Conversión de texto cardinal a números naturales y viceversa | 0.51 |
| **SNPA10:** Operación de valores posicionales con números naturales o decimales | 0.61 |
| **SNPA11:** Representación del modelo multiplicativo de números fraccionarios por naturales. | 0.63 |
| **SNPA12:** Conversión de una regla verbal de progresión geométrica de ascendente a una sucesión numérica | 0.57 |
| **SNPA13:** Deducción del patrón de una sucesión con progresión especial | 0.50 |
| II. Manejo de la Información | **MI01:** Comprensión de problemas matemáticos contextualizados | 0.68 |
| **MI02:** Comparación de la proporcionalidad de razones | 0.59 |
| **MI03:** Representación de modelos aritméticos de la media (promedio) | 0.65 |
| **MI04:** Representación de modelos aritméticos de la mediana | 0.59 |
| **MI05:** Aplicación de operaciones aritméticas básicas | 0.68 |
| **MI06:** Representación de datos numéricos en gráficas de barras | 0.63 |
| **MI07:** Representación del modelo de regla de tres simple | 0.63 |
| **MI08:** Comprensión de la relación entre porcentajes y fracciones | 0.66 |
| **MI09:** Comparación de razones con cantidades discretas | 0.49 |
| **MI10:** Representación de un número fraccionario | 0.68 |
| III. Forma, Espacio y Medida | **FEM01:** Comprensión de problemas matemáticos contextualizados | 0.60 |
| **FEM02:** Comprensión del Sistema Internacional de Unidades (SIU) | 0.63 |
| **FEM03:** Operación de valores posicionales con números naturales y decimales | 0.63 |
| **FEM04:** Ubicación de una coordenada en el primer cuadrante del plano cartesiano | 0.51 |
| **FEM05:** Aplicación de operaciones aritméticas básicas | 0.67 |
| **FEM06:** Definición de tecnicismos del lenguaje formal de la geometría | 0.50 |
| **FEM07:** Representación viso-espacial de figuras geométricas | 0.57 |
| **FEM08:** Identificación de las características geométricas de los cuadriláteros | 0.57 |
| **FEM09:** Identificación gráfica de tipos de líneas rectas (paralelas, perpendiculares y secantes) | 0.59 |
| **FEM10:** Representación del modelo aritmético para calcular el perímetro de una figura geométrica (triángulo o cuadrilátero) | 0.61 |
| **FEM11:** Representación del modelo aritmético para calcular el área de cuadriláteros o triángulos | 0.68 |
| **FEM12:** Deducción de fórmulas para calcular el área mediante descomposición de figuras geométricas | 0.57 |

En cuanto a la estimación de los parámetros de adivinación y desliz, el eje *Espacio, forma y medida*, presenta valores de adivinación de entre 0.08 a 0.61, con un promedio de 0.33. Tres de los N ítems que componen este eje presentan valores muy altos de adivinación, indicando que aún los examinados que no dominan las operaciones requeridas tienen una probabilidad mayor a 0.50 de “atinarle” a la respuesta correcta. Por su parte, el parámetro desliz presenta valores entre 0.01 y 0.79 con un valor promedio de 0.25. Nuevamente, tres ítems presentaron valores muy altos de desliz, indicando que aún los estudiantes que dominan las operaciones requeridas, tienen una baja probabilidad de responder correctamente el ítem.

En cuanto al eje *Manejo de información*, los parámetros de adivinación computados presentan valores en el rango de 0.18 a 0.45 con un valor promedio de 0.30. Los parámetros de desliz computado presentan valores entre cero y 0.48 con un valor promedio de 0.22.

Finalmente, el eje *Sentido numérico y pensamiento algebraico*, presenta valores de adivinación que oscilan entre 0.13 y 0.56 con un promedio de 0.28 y valores Del parámetro desliz entre cero y 0.53 con un promedio de 0.25.

**Devolución de resultados**

**Diseño de estrategias de mejora**

**Discusión**

**Conclusiones**

**Referencias**

* Bejar, I. (2002). Item generation: From conception to implementation. In S. H. Irvine and P. C. Kyllonen (Eds.), *Item heneration for test development*, pp. 199-218. Mahwah: Lawrence Erbaum Associates.
* Bejar, I. (2010). Item Generation. Implications for a Validity Argument. In Gierl, Mark J.; Haladyna, Thomas M. (eds.) *Automatic Item Generation: Theory and Practice*, pp. 40-56. New York: Routledge.
* Brown, J. & Burton, R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-192.
* Chen, Y. & Macdonald, G. (2011). Validating Cognitive Sources of Mathematics Item Difficulty: Application of the LLTM to Fraction Conceptual Items. *Psychological Assessment*, *7*, 74–93.

Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What will it take?. *Theory into practice*, *42*(1), 75-83.

Cohen, Y. (2019). The Handbook of Cognition and Assessment; Frameworks, Methodologies, and Applications.

* De La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, *34*(1), 115-130

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179-199.

Douglas, J., de la Torre, J., Chang, H., Henson, R., & Templin, J. (2006, April). Skills diagnosis with latent variable models. In *annual meeting of the National Council on Measurement in Education, San Francisco, CA*.

Embretson, S. (1994). Applications of cognitive design systems to test development. In *Cognitive assessment* (pp. 107-135). Springer, Boston, MA.

* Ericsson, K. & Simon, H. (1984). *Protocol analisys: verbal reports as data*. Cambridge: MIT Press.
* Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT.

Ferrara, S., Lai, E., Reilly, A., Nichols, P. D., Rupp, A. A., & Leighton, J. P. (2017). Principled approaches to assessment design, development, and implementation. *The Handbook of Cognition and Assessment*, 41-74.

* Fredericksen, J. (1980). Component skills in Reading: measurements of individual diferences thought chronometric analisys. In R. E. Snow, P-A. Federico & W. E. Montage (Eds.), *Aptitude, learning, and instructions: Cognitive process analyses of aptitude*, Vol. 1, (pp. 105-138). Hillsdale: Lawrence Erlbaum.
* Gierl, M., Leighton, J., Changjiang, W., Jiawen, Z., Rebecca, G. & Tan, A. (2009). *Validating Cognitive Models of Task Performance in Algebra on the SAT. Research Report 2009-3*. College Board, Research Report, 2009(3). New York.

George, A. C., 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.

Gorin, J. & Embretson, S. (2013). Using Cognitive Psychology to generate Items and Predict Item Characteristics. In Gierl, M. y Haladyna, T. (edit.). *Automatic Item Generation: Theory and Practice,* pp. 40-56. New York: Taylor and Francis Group.

* Haladyna, T. Downing, S. M. & Rodríguez, M. C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309–334.

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*(Doctoral dissertation, ProQuest Information & Learning).

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191.

* INEE. (2015). Plan Nacional para la Evaluación de los aprendizajes (PLANEA). Resultados nacionales
* 2015. Recuperado de http://www.inee.edu.mx/images/stories/2015/planea/inal/fascículosinales/

resultadosPlanea-3011.pdf

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, *5*(3), 299-314.

INEE (2017). Informe de resultados PLANEA 2015. El aprendizaje de los alumnos de sexto de primaria y tercero de secundaria en México. Lenguaje y Comunicación y Matemáticas. México: autor.

* Johnstone, C. (2003). Improving validity of large-scale tests: Universal design and student performance (Technical Report 37). Minneapolis: National Center on Educational Outcomes.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272.

* Leighton, J. & Gierl, M. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees’ thinking processes. *Educational Measurement: Issues and Practice, 26*(2), 3-16.
* Leighton, J. (2009). Two Types of Think Aloud Interviews for Educational Measurement: Protocol and Verbal Analysis Paper presented for symposium How to Build a Cognitive Model for Educational Assessments at the 2009 annual meeting of the National Council on Measurement in Education (NCME), April, 14-16.
* Ma, L. Çetin, E. y Green, K. (2009, April). *Cognitive assessment in Mathematics with the Least Squares Distance Method.* Artículo presentado en el Congreso anual de la AERA 2009. San Diego.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187-212.

Martínez Rizo, F. (2015). Las pruebas ENLACE y Excale. Un estudio de validación. *Cuadernos de Investigación, Ciudad de México: Instituto Nacional para la Evaluación de la Educación*.

* Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3a. ed.), pp. 13-103. New York: Macmillan Publishing Co.
* Pérez-Morán, J. C. (2014). *Análisis del aspecto sustantivo de la validez de constructo de una prueba de habilidades cuantitativas* (tesis doctoral). Universidad Autónoma de Baja California, Baja California, México.
* Pérez-Morán, J. C., Contreras, S., Hernández, E. M., Olivares, C., Chan, P., y Díaz, K. M. (2014). Análisis de las evidencias de validez basadas en el proceso de respuesta de las pruebas de ENLACE MS de Habilidad lectora y Matemáticas. Reporte técnico. México: INEE
* Pérez-Morán, J. C.; Larrazolo, N.; Backhoff, E.; y Guaner, R. (2015). Análisis de la estructura cognitiva del área de habilidades cuantitativas del EXHCOBA mediante el modelo LLTM de Fisher. Revista Internacional de Educación y Aprendizaje, 3(1), 25-38. <http://coleccionderevistasdeeducacionyaprendizaje.cgpublisher.com/product/pub.329/prod.5> ISSN 2255-453X
* Posner, M. I. (1978). *Chronometric exploration of mind.* New York: Jhon Wiley.

Posner, M. I., & Rogers, M. G. K. (1978). Chronometric analysis of abstraction and recognition. In W. K. Estes (Ed.) (1978). *Handbook of learning and cognitive processes* (vol. 6). Hillsdale, N. J.: Lawrecence Erlbaum Associates.

Raiche, G., Magis, D., & Raiche, M. G. (2010). Package ‘nFactors’. *Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test. Available at CRAN repository: http://cran. r-project. org/web/packages/n Factors/n Factors. pdf*.

Revelle, W. (2011). An overview of the psych package. *Department of Psychology Northwestern University. Accessed on March*, *3*, 2012.

Revuelta, J. y Ponsoda, V. (1998). Un test adaptativo informatizado de análisis lógico basado en la generación automática de ítems. *Psicothema, 10*, 753-760.

Romero, S., Ponsoda, V., y Ximenez, C. (2008). Análisis de un test de aritmética mediante el modelo logístico lineal de rasgo latente 1. *Revista Latinoamericana de Psicología, 40*, 85–95.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic assessment: Theory, methods, and applications. *New York: Guilford*.

Secretaría de Educación Pública (2011). Programas de estudios 2011. Guía para el maestro. Educación Básica. Primaria. Sexto grado. México. SEP. URL: <http://edu.jalisco.gob.mx/cepse/sites/edu.jalisco.gob.mx.cepse/files/sep_2011_programas_de_estudio_2011.guia_para_el_maestrosexto_grado.pdf>

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, *29*(7), 4-14.

* Snow, R. & Lohman, D. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3a. ed.), pp. 263-331. New York: Macmillan Publishing Co.
* Sternberg, R. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Oxford: Lawrence Erlbaum.
* TechSmith. (2008). CAMTASIA STUDIO V. 5.0.1. Michigan: Author. Descargado en http://www.techsmith.com/download/camtasia/

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, *11*(3), 287.

Templin, J., & Henson, R. A. (2006). A Bayesian method for incorporating uncertainty into Q-matrix estimation in skills assessment. In *Symposium conducted at the meeting of the American Educational Research Association, San Diego, CA*.

* Thompson, S., Johnstone, C. & Thurlow, M. (2002). *Universal design applied to large scale assessments (Synthesis Report 44).* Minneapolis, MN: National Center on Educational Outcomes.
* Van der Linden, W. J. (Ed.). (2017). *Handbook of Item Response Theory, Volume Three: Applications*. CRC Press.

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, *2005*(2), i-35.

Willse, J. T., & Shu, Z. (2014). CTT: Classical test theory functions. *R package version*, *2*.

* Yang, X. & Embretson, S. (2007). Construct Validity and Cognitivy Diagnostic Assesment. In Leighton, J. y Griel, M. (Edit.). *Cognitive diagnostic assessment for education: Theory and applications*, pp. 85-118. Cambridge: Cambrige University Press.