**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 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).

**Descripción de los modelos**

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, Templin, Henson, y Douglas, 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), 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 , sólo 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 vigente Instituto Nacional para la Evaluación de la Educación (INEE), con la intención de contar con una evaluación a gran escala que permitiera valorar la eficacia del Sistema Educativo Nacional a partir del promedio de los puntajes logrados por los estudiantes evaluados a lo largo de distintas competencias o asignaturas curriculares.

Como antecedente directo del desarrollo del PLANEA, destaca la revisión realizada en 2013 por el mismo INEE acerca de la validez y confiabilidad de los Exámenes Nacionales del Logro Académico en Centros Escolares (ENLACE) y los Exámenes de la Calidad y el Logro Educativo (EXCALE), administrados a nivel nacional con un propósito similar (Martínez-Rizo, 2015). Como resultado de dicha revisión, el PLANEA cuenta con tres distintas modalidades para la evaluación de los aprendizajes de los estudiantes: (1) Evaluación del Logro referida al Sistema Educativo Nacional (ELSEN), (2) Evaluación del Logro referida a los Centros Escolares (ELCE) y (3) Evaluación Diagnóstica Censal (EDC).

En especial, el PLANEA-ELCE tiene como objetivo reunir información acerca de los centros escolares como unidad de análisis, a partir de la evaluación de campos formativos específicos (Matemáticas y Lenguaje y comunicación) y a lo largo de distintos niveles de desagregación (por ejemplo, región, estado, municipio, zona escolar, modalidad, tipo de organización, entre otros), (INEE, 2015).

El propósito general que para entonces se le asignó al PLANEA consistió en conocer la medida en que los estudiantes lograrán el dominio de un conjunto de aprendizajes esenciales alineados con el currículo nacional. En la actualidad no ha cambiado dicho propósito y se espera que con sus resultados se pueda lograr la mejora educativa mediante: (1) Informar a la sociedad sobre el estado que guarda la educación en términos del logro de aprendizaje de los estudiantes y de la equidad (o inequidad) que existe en los resultados educativos; (2) aportar a las autoridades educativas información relevante para el monitoreo, la planeación, programación y operación del sistema educativo y sus centros escolares; (3) ofrecer información pertinente, oportuna y contextualizada a las escuelas y a los docentes, que ayude a mejorar sus prácticas de enseñanza y el aprendizaje de todos sus estudiantes y (4) contribuir al desarrollo de directrices para la mejora educativa con información relevante sobre los resultados educativos y los contextos en que se dan.

Ahora bien, en lo referente a la modalidad ELCE del PLANEA, se puede describir como una modalidad de evaluación de forma única y contextualizada, al ser una versión derivada de uno de los bloques del instrumento matricial de ELSEN. Su propósito específico es ofrecer a cada escuela elementos de retroalimentación para su mejora. Se espera que los resultados de dicha modalidad de PLANEA ofrezcan información de lo que logró el centro escolar con respecto al aprendizaje de sus estudiantes en seis años (para el caso de primaria); indiquen lo que los alumnos logran y no logran aprender; ayuden a identificar las líneas curriculares que la comunidad escolar debe fortalecer desde el primer grado de primaria, y faciliten elementos para identificar acciones que puedan implementarse a fin de lograr los aprendizajes esperados, esta modalidad de PLANEA, corre a cargo de la SEP, tanto para su levantamiento en campo, procesamiento de los datos e informe de resultados, es importante resaltar que las métricas ofrecidas entre PLANEA ELCE y ELSEN, no permiten comparaciones directas debido a los métodos de estimación del logro que cada evaluación emplea, por lo que se debe de ser muy cauto en lo que refiere a las inferencias y decisiones derivadas de cada informe de resultados (Braun y von Davier, 2017; von Davier, Gonzalez y Mislevy, 2009; Wu, 2005).

De forma especial, se espera que ELCE aporte elementos para el fortalecimiento del currículo en materia de habilidades Matemáticas de primaria y secundaria, así como en habilidades de lenguaje y comunicación. Además, se espera que propicie a que el Consejo Técnico Escolar y los supervisores centren su atención en el aprendizaje de los estudiantes; faciliten el monitoreo de la eficacia de cada escuela a través del tiempo; permitan que un centro escolar se compare con los resultados agregados de otros similares; y promuevan el interés y colaboración del colegiado para lograr mejores resultados en los siguientes años.

Por su parte, la modalidad EDC de PLANEA, se considera una evaluación diagnóstica y censal. Consiste en realizar un diagnóstico del conocimiento y habilidades con las que cuentan los estudiantes obtenidos durante los cursos previos. Es por ello, que el formato EDC se describe como un proceso de transición a la segunda etapa de la educación primaria; el objetivo de proveer pruebas estandarizadas a los docentes en esta etapa se sustenta con la necesidad de obtener bases para la construcción de una guía de monitoreo de los estudiantes y que además, contribuya a la mejora en la planificación de trabajo en el aula. Su carácter censal se desprende de la intención de realizar la aplicación de la prueba a prácticamente la totalidad de los estudiantes de cuarto grado de primaria de todos los centros escolares. En síntesis, con PLANEA EDC se dará información a docentes para que mejoren su práctica de enseñanza-aprendizaje.

**Método:**

A partir del Enfoque Sistémico de Diseño Cognitivo (ESDC) propuesto por Embretson (1994), y de la perspectiva top-down para el diseño y validación de pruebas cognitivas (Bejar, 2002, 2010, Gorin y Embretson, 2013 y Messick, 1989b), se desarrolló el modelo metodológico del presente estudio, cuyo foco principal fue la obtención de evidencia que sustentara la validez sustantiva del constructo y corroborara la fidelidad estructural del modelo cognitivo subyacente a la prueba de Matemáticas de primaria (06) del PLANEA ELCE 2015. La Tabla 1 presenta el modelo metodológico propuesto para alcanzar los objetivos del presente estudio.

**Tabla 1.** Modelo metodológico del estudio

|  |  |  |
| --- | --- | --- |
| **Fase** | **Etapa** | **Actividades** |
| **Fase I**  **Diseño de los estudios cognitivos** | 1.1. Selección de técnicas para el análisis cognitivo | -Determinar el conjunto de ítems para su análisis, verificar sus áreas de membresía o dominio y analizar sus características particulares de diseño.  -Definir los métodos y técnicas específicas para el análisis cognitivo con base en las características de los ítems y de la población objetivo. |
| 1.2. Diseño de los estudios cognitivos | -Adaptar el conjunto de ítems y tareas evaluativas para el piloteo de las técnicas cognitivas.  -Definir los procedimientos de operación de los estudios cognitivos.  -Determinar el tipo de herramientas, materiales e instrumentos tecnológicos requeridos para la captura de los datos del análisis cognitivo.  -Adaptar un laboratorio acorde a las necesidades específicas del estudio cognitivo. |
| 1.3. Análisis cognitivo-reticular de los ítems | -Capacitar a expertos en el análisis cognitivo-reticular de los ítems.  **-Elaborar mediante técnicas de análisis cognitivo-reticular la genealogía curricular de los ítems.\***  **-Evaluar mediante expertos la congruencia y alineación de los ítems de la prueba de Matemáticas 06 del PLANEA ELCE con sus especificaciones, la retícula, los libros de texto y los planes de estudio.\*** |
| 1.4. Análisis por expertos de los procesos subyacentes a los ítems | -Capacitar a expertos en el método de modelado matemático de sub-tareas de respuesta  -Aplicar a expertos técnicas de pensamiento en voz alta concurrentes y retrospectivas.  -**Modelar desde la perspectiva de expertos en Matemáticas los procesos posibles de respuesta subyacentes a los ítems**.\*  -Estructurar los sistemas de categorías de los protocolos verbales y modelos cognitivos para su aplicación. |
| **Fase II**  **Piloteo y aplicación de protocolos verbales** | 2.1. Piloteo de protocolos verbales | -Establecer los criterios de selección de los estudiantes participantes del estudio piloto y del grupo de participantes del estudio cognitivo.  -Seleccionar y capacitar a los participantes del estudio piloto.  -Pilotear, probar y ajustar los protocolos, sistemas de categorías y materiales del laboratorio cognitivo. |
| 2.2. Selección del grupo de participantes | -Seleccionar al grupo de participantes del estudio cognitivo.  -Confirmar el consentimiento informado y recabar datos de identificación.  -Establecer cronograma de actividades y citas con los participantes del estudio cognitivo. |
| 2.3. Aplicación en forma de los protocolos verbales | -Entrenar a los participantes del estudio cognitivo.  -Aplicar el estudio cognitivo a los participantes.  -Recopilar la información obtenida durante el estudio cognitivo.  **-Obtener reportes verbales de los procesos de respuesta de estudiantes ante los ítems de las pruebas de Matemáticas 06 del PLANEA ELCE.\*** |
| **Fase III**  **Desarrollo y definición del modelo cognitivo** | 3.1. Análisis de los datos obtenidos durante los estudios cognitivos | -Capacitar a expertos en la verificación del proceso de respuesta de los examinados en contraste con los modelos del proceso de respuesta subyacente a los ítems definidos por docentes y especialistas en Matemáticas.  -Verificar los reportes verbales de los estudiantes participantes con base en los modelos del proceso de respuesta subyacente a los ítems definidos por docentes y especialistas en Matemáticas.  **-Elaborar los modelos sintetizados de los procesos de respuesta subyacentes a los ítems.**\* |
| 3.2. Desarrollo y definición del modelo cognitivo de la prueba | -Determinar la cantidad y el tipo de relaciones entre los ítems de la prueba y los atributos u operaciones cognitivas sustantivas determinadas por los investigadores con base en la síntesis de los resultados de los estudios reticulares y cognitivos.  **-Elaborar la matriz Q de la prueba con base en las operaciones cognitivas sustantivas determinadas *a-priori*.\***  -**Analizar y evaluar el diseño de los ítems** **para identificar posible varianza irrelevante o sesgo**.\* |
| **Fase IV**  **Aplicación del análisis componencial** | 4.1 Revisión de la estructura interna bajo el modelo de *redes nomológicas* | **-Calibrar con la aplicación del modelo de la Teoría Clásica de los Test (TCT) los ítems de la prueba.\***  **-Analizar las características psicométricas de los distractores de los ítems.\***  **-Analizar la estructura interna bajo el *modelo de redes nomológicas* (aplicación del análisis factorial confirmatorio).\*** |
| 4.2 Revisión de la estructura del *modelo cognitivo* de la prueba | -Elegir y aplicar los modelos psicométricos componenciales (MPC) acordes a las características del modelo cognitivo estructurado.  **-Analizar la validez de la matriz Q planteada.\***  -Considerar posibles mejoras y reconfiguraciones en el diseño de la matriz Q. |
| 4.3 Interpretación de los resultados del diagnóstico cognitivo de los examinados | **-Asignar puntuaciones a los examinados con base en el diagnóstico cogitivo de las operaciones definidas en la matriz Q.\***  **-Explorar resultados del diagnóstico cognitivo a nivel nacional, por estado, sexo, modalidad, tipo de servicio u otra agrupación que se considere relevante para su uso a nivel escuela.\*** |
| \*Actividades fundamentales, relacionadas con los objetivos del estudio. | | |

**Fase I**

**Diseño de los estudios cognitivos**

Las primeras fases de nuestro modelo metodológico contemplan la aplicación de estudios cognitivos que permitan identificar los modelos, estrategias y procesos de respuesta adheridos a cada ítem contenido en la prueba de Matemáticas de primaria (06) del PLANEA ELCE 2015. Esto obedece a la necesidad señalada por autores como Yang y Embretson, (2007) de que toda prueba de diagnóstico cognitivo diseñada con propósitos de mejora, deben ser diseñadas y validadas a partir de modelos cognitivos que permitan identificar de manera confiable los procesos de respuesta asociados a cada ítem.

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 et al., 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). A lo largo del presente trabajo, se utilizará el Ítem identificado con el ID PMA01 para ilustrar los procedimientos aplicados y los resultados obtenidos.

La genealogía curricular de los ítems, la evaluación de la congruencia y la alineación de los ítems y el modelamiento de los procesos de respuesta subyacentes, se realizó a partir de un análisis cognitivo-reticular con apoyo de un panel de expertos con experiencia en la enseñanza de las Matemáticas a nivel primaria.

De forma puntual, el panel de expertos estuvo conformado por tres psicólogos con experiencia en atención a estudiantes de primaria y secundaria, un especialista en análisis cognitivos y modelamiento de procesos cognitivos del pensamiento lógico-matemático, y un docente con amplia experiencia en la enseñanza de las Matemáticas en educación básica. La selección de los expertos se hizo de acuerdo con las recomendaciones de Rupp, Templin y Henson (2010) respecto a la elección de profesionales con un conocimiento profundo de los procesos de solución de problemas que utilizan los individuos en el dominio de interés, las diferentes vertientes para el desarrollo y aprendizaje de los componentes o atributos de la prueba y los contextos en los que los examinados adquieren y utilizan dichos atributos.

Una vez seleccionados los expertos, se les capacitó en la aplicación de métodos de análisis cognitivo-reticular, técnicas de pensamiento en voz alta concurrentes y retrospectivas, y en métodos de modelado matemático de sub-tareas de respuesta.

Para el análisis cognitivo-reticular de los ítems en estudio, se utilizaron la tabla de especificaciones de la prueba y su respectivo análisis reticular, acompañado de una revisión exhaustiva de los libros de texto diseñados para los alumnos y la guía del maestro del plan de estudios 2011[[1]](#footnote-1). La Figura 1 presenta como ejemplo el diagrama de la genealogía curricular correspondiente al ítem PMA01.

****

*Figura 1*. Diagrama de la genealogía curricular y de la alineación del ítem PMA01 de la prueba de Matemáticas de primaria del PLANEA ELCE 2015

También se aplicaron técnicas de pensamiento en voz alta concurrentes y retrospectivas con los expertos, para identificar los procesos de respuesta asociados a cada ítem y se realizaron diversas actividades para aplicar el modelado matemático de sub-tareas. Primero, mediante un proceso inductivo-deductivo los expertos identificaron y categorizaron los procesos y atributos cognitivos vinculados con cada uno de los ítems, apoyándose en los reportes verbales de su propia ejecución con los mismos, en los descriptores de los conocimientos, habilidades previas y procesos de respuesta declarados en la matriz de especificaciones de la prueba, así como en el sistema de categorías de procesos de respuesta típicos utilizados por estudiantes de educación básica en ítems de Matemáticas (Pérez-Morán, 2014). Como ejemplo, en la **Tabla 4** se pueden observar tres modelos de respuesta hipotéticos elaborados por expertos para el ítem PMA01 de PLANEA ELCE 06. Nótese que los tres modelos presentan diferentes niveles de granulación y que sólo el primero contempla la comprensión de problemas matemáticos contextualizados.

**Tabla 4**. Modelos hipotéticos del proceso de respuesta subyacentes al ítem desde la perspectiva de expertos y docentes

|  |  |
| --- | --- |
| **No.** | **Modelos hipotéticos del proceso de respuesta subyacente al ítem desde la perspectiva de especialistas y docentes** |
| 1 | a) Leer detalladamente el reactivo, b) observar la figura con sus respectivos valores, c) comprender qué es lo que se está solicitando (integrar el contexto del problema) d) determinar el tipo de cálculo a desarrollar, e) tomar los valores proporcionados en la figura para desarrollar el cálculo elegido, e) buscar entre las opciones de respuesta el valor calculado, y f) seleccionar la opción. |
| 2 | a) Leer detalladamente el ítem, b) comprender el objetivo de la tarea evaluativa, c) observar la figura d) recordar la fórmula para obtener el área de un triángulo, e) identificar las medidas que permitan calcular el área de la figura, f) determinar los valores que se utilizarán como base y altura del triángulo, g) desarrollar la fórmula para obtener el área de la figura h) obtener el resultado, y i) seleccionar la respuesta que coincida con el cálculo realizado. |
| 3 | a) Observar la imagen, b) leer la base del reactivo, c) comprender la pregunta, d) desarrollar el cálculo que determine el área de la figura, e) comparar el resultado con las opciones de respuesta, y f) seleccionar la respuesta. |

Finalmente, un equipo de expertos trabajó con los elementos de análisis del modelo para la Evaluación del Diseño Universal (EDU) propuestos por Thompson, Johnstone y Thurlow (2002), que ha demostrado ser de gran utilidad para el desarrollo de evaluaciones más accesibles para los examinados (Johnstone, 2003) y para minimizar la varianza irrelevante del constructo originada por problemas en el diseño, formato y sesgos culturales presentes en los ítems (Haladyna, Downing, y Rodríguez, 2002). Al aplicar la EDU con el fin de evaluar el diseño de los ítems, se trabajó con categorías y códigos específicos de análisis con base en los análisis propuestos por Thompson y colaboradores (2002): (a) inclusión poblacional, (b) definición precisa del constructo (c) accesibilidad e imparcialidad (d) acomodación flexible de los contenidos, (e) procedimientos e instrucciones simples, claras e intuitivas, (f) máxima legibilidad, y (g) comprensibilidad.

**Fase II**

**Piloteo y aplicación de protocolos verbales**

Durante la segunda fase, se pilotearon los ítems contenidos en la prueba y se obtuvieron reportes verbales de los procesos de respuesta empleados por un grupo de estudiantes, utilizando una vez más protocolos de pensamiento en voz alta con técnicas concurrentes y retrospectivas (Ericsson y Simon, 1984, 1993; Leighton, 2009; Leighton y Gierl, 2007b). Como análisis complementarios, 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). La implementación de estas medidas complementarias, se considera de gran ayuda para obtener información en los casos en que se presentan dificultades para evocar el reporte verbal de los procedimientos seguidos por estudiantes de corta edad, o bien, de procedimientos que suceden en cuestión de tan sólo algunos segundos y que, por tanto, no es posible su introspección (Sternberg, 1977). Todas estas técnicas complementarias, permiten una mejor verificación de la relación entre el modelo cognitivo elaborado por los expertos y los procesos cognitivos utilizados y reportados por los examinados para responder los ítems (Messick, 1989b).

Para el diseño de los protocolos verbales con estudiantes, se tomaron en cuenta los ocho pasos generales para la aplicación de las técnicas de pensamiento en voz alta recomendados por Leighton (2009) y se tomaron en cuenta las características del formato de los ítems en estudio y las características de los participantes.

Los criterios de selección para los estudiantes participantes en el piloteo y los protocolos verbales fueron elaborados a partir de las recomendaciones de Ericsson y Simon (1984, 1993), quienes proponen incorporar tanto a participantes novatos como a expertos en el dominio de interés, por lo que se tomaron como referencia variables tales como el rendimiento escolar, grado educativo y la recomendación del profesor. La selección de estudiantes expertos se realizó tomando como referencia un promedio mayor a 8.5 y el reconocimiento de sus profesores como estudiantes sobresalientes en el dominio de las Matemáticas. Los estudiantes novatos fueron seleccionados siempre y cuando tuvieran un promedio mayor a 6.0 pero menor a 8.0 y fueran referidos por sus profesores como estudiantes con bajo desempeño o no sobresalientes en el dominio de las Matemáticas. Se buscó que la muestra final quedara conformada en un 50% de estudiantes expertos y con un 50% de mujeres. En total se realizó el piloteo con 8 participantes por ítem.

Para agilizar la recolección de información y reducir en la medida de lo posible la duración total de cada piloteo, se aplicaron 25 de los 50 ítems en cada sesión, obteniendo una duración aproximada entre los 75 y 125 minutos. Durante la aplicación de los protocolos concurrentes de respuesta en voz alta se recolectaron datos del diseño de los ítems y se verificó que el proceso de respuesta utilizado por los participantes ante los ítems de la prueba estuviera representado en el modelo cognitivo elaborado por los expertos. Por su parte, los protocolos retrospectivos permitieron realizar preguntas a los participantes inmediatamente después de contestar el ítem con el fin de complementar la información obtenida en los protocolos concurrentes.

Para la captura de los datos, se montó un laboratorio cognitivo acorde a las necesidades específicas del estudio. Se utilizó el software CAMTASIA STUDIO versión 5.0.1 (TechSmith, 2008) porque permite grabar las verbalizaciones de los examinados, la imagen del ítem en la interfaz de la computadora junto con todas las acciones ocurridas en ella durante los protocolos verbales, el sendero del indicador del mouse y el tiempo de latencia de cada una de las actividades realizadas por el examinado. Además, al final de la aplicación de las técnicas de pensamiento en voz alta, se puede obtener y editar un video con todos los datos mencionados.

Una vez registrada la respuesta de los estudiantes, se capacitó a expertos en la verificación de los procesos de respuesta reportados con base en los modelos definidos previamente por el grupo de expertos conformado por docentes y especialistas en Matemáticas. Para ello, se trabajó de forma directa aplicando técnicas de análisis inductivo-deductivo. En la Figura 3 se puede observar como ejemplo, el proceso analítico realizado en el ítem PMA01 de la prueba de Matemáticas de sexto de primaria del PLANEA ELCE 2015, donde como resultado de la verificación de los modelos hipotéticos del proceso de respuesta subyacente al ítem propuesto por el panel de expertos y los reportes verbales de los estudiantes, se desarrolló el modelo cognitivo sintetizado del ítem.



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

En esta misma figura, también se pueden observar tres problemas de diseño y sesgo identificados por los expertos en el marco del EDU, correspondiendo a las categorías de: a) Problemas de inclusión poblacional, b) comprensibilidad de la base del ítem, y c) definición imprecisa del constructo, asociado este último a su vez a problemas de alineación del ítem al currículo de primaria.

**Fase III**

**Desarrollo y definición del modelo cognitivo**

Durante la tercera fase de nuestro marco metodológico, se buscó definir, estructurar y desarrollar el modelo cognitivo que diera cuenta de las habilidades, conocimientos y procesos cognitivos evaluados o requeridos por la prueba. 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).

Mediante un análisis inductivo-deductivo realizado por diadas de expertos, se sintetizaron los modelos cognitivos asociados a los diversos procesos de respuesta para desglosarlos en operaciones cognitivas concretas que permitieran un mayor nivel explicativo del comportamiento de cada ítem. 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 4).



*Figura 4*. Diagrama del modelo de un proceso de respuesta erróneo subyacente al ítem PMA01 de la prueba de Matemáticas (06) del PLANEA ELCE 2015

**Fase IV**

**Aplicación del análisis componencial**

La cuarta fase de nuestro marco metodológico consideró la evaluación de la calidad técnica de la prueba en estudio, aplicando el modelo DINA con la matriz Q diseñada por los expertos en los estudios cognitivos y evaluando la posibilidad de mejorar la matriz, explorando los resultados individuales y a lo largo de diferentes tipos de agrupaciones.

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. Se seleccionó una muestra aleatoria de 5000 estudiantes para los distintos análisis. de la calidad técnica y de validación de la prueba.

Los indicadores de la TCT fueron obtenidos con la paquetería CTT del programa libre R 2.15.1. (Ihaka, R. y Gentleman, R., 1996) 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).

El primer indicador, el índice de dificultad de cada ítem señala qué proporción de la muestra evaluada obtuvo un acierto, de acuerdo a la siguiente ecuación:

,

donde es la cantidad total de aciertos registrados para este reactivo y es la cantidad total de respuestas registradas para este reactivo (aciertos y errores).

Para obtener los *índices de discriminación* altos y bajos se utilizó la siguiente ecuación:

 ,

donde es la cantidad de aciertos registrados en el ítem por el 27% de los examinados que obtuvieron las puntuaciones más altas en el examen; es la cantidad de aciertos registrados en el mismo reactivo por el 27% de examinados que obtuvieron las puntuaciones más bajas en el examen, y es la cantidad de personas en el grupo más cuantioso ( o ).

La ecuación que se utilizó para obtener el coeficiente de correlación puntual-biserial () fue:

,

Donde  es la media de las puntuaciones totales de aquellos que respondieron correctamente el ítem,  es la media de las puntuaciones totales de aquellos que respondieron incorrectamente el ítem,  es la desviación estándar de las puntuaciones totales, es la cantidad de respuestas correctas registradas en el ítem, es la cantidad de respuestas incorrectas registradas y es igual a .

Finalmente el coeficiente de consistencia interna (**α** de Cronbach) del instrumento fue computado a partir de la siguiente ecuación:

,

donde es la cantidad de ítems de la prueba,  es la varianza de las puntuaciones de la prueba y  es la sumatoria de las varianzas de los reactivos.

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). Los indicadores de ajuste absoluto que se emplearon corresponden a la raíz de los residuos cuadráticos promedios estandarizados (SRMR) y la raíz del error cuadrático promedio de aproximación (RMSEA). Tanto el SRMR como el RMSEA deben ser inferiores a 0,05 para obtener 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 (ver **Apéndice 1**). 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.

Fase V Estimacion del Modelo DIGNA

En lo que respecta a la aplicación del modelo DINA, se utilizó la librería CDM implementada en el programa libre R 2.15.1. (Ihaka, R. y Gentleman, R., 1996). Los parámetros de adivinación y desliz del modelo DINA permiten interpretar la calidad técnica de cada ítem en función de las operaciones cognitivas de los procesos de respuesta (de la Torre, 2009).

Resultados: GDRR y JCPM

Discusión: JCPM, GDRR, RVL

Conclusiones: JCPM, GDRR, RVL

Referencias: RVL

Pérez-Morán, J. C., Vázquez-Lira, R., y Rojas, G. (Coords.) (2017). Análisis de bases de datos y de validez de pruebas y de los patrones de respuesta PLANEA. Ciudad de México: INEE.

1. Libro de texto del alumno: Desafíos matemáticos. Libro para el alumno. 6° Matemáticas. Secretaría de Educación Pública. URL: <http://libros.conaliteg.gob.mx/content/restricted/libros/carrusel.jsf?idLibro=1269>

   Libro de texto del maestro: Desafíos matemáticos. Libro para del maestro. 6° Matemáticas. Secretaría de Educación Pública. URL: <http://libros.conaliteg.gob.mx/content/restricted/libros/carrusel.jsf?idLibro=1270>

   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> [↑](#footnote-ref-1)