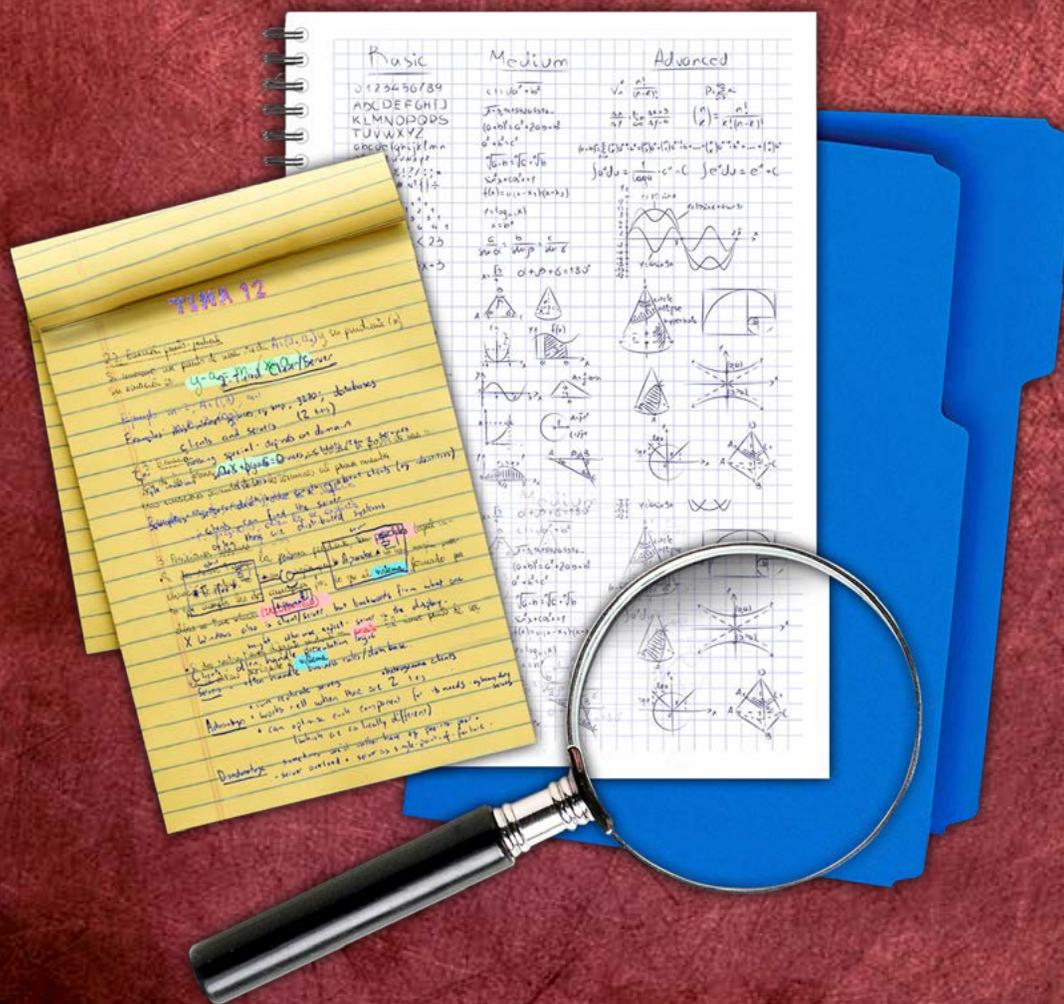


# Apuntes de investigación educativa

Alicia Alelí Chaparro Caso-López, Juan Carlos Rodríguez Macías  
y Juan Carlos Pérez Morán (COORDINADORES)



***Esta obra ha sido dictaminada por:***

Rubí Surema Peniche Cetzal, Universidad Autónoma de Aguascalientes

Mónica Fulgencio Juárez, Universidad Michoacana de San Nicolás de Hidalgo

Ana María Méndez Puga, Universidad Michoacana de San Nicolás de Hidalgo

***Apuntes de investigación educativa***

1era. edición, abril 2019

ISBN 978-607-98270-9-0

DOI 10.29410/QTP.19.05

D.R. © 2019. Qartuppi, S. de R.L. de C.V.

Villa Turca 17, Villas del Mediterráneo

Hermosillo, Son. 83220 México

<http://www.qartuppi.com>

Edición y Diseño: Qartuppi, S. de R.L. de C.V.

Diseño de portada: Stefanya Ríos Padilla



Esta obra posee una Licencia Creative Commons  
Atribución-NoComercial 4.0 Internacional.

# Índice

## PRIMERA PARTE: *EDUCACIÓN BÁSICA*

- 6** *Prólogo*  
Coral González Barbera
- 13** *Capítulo 1 - Aprendizaje de la escritura*  
Alicia Alelí Chaparro Caso-López e Irma Gloria Arregui Eaton
- 47** *Capítulo 2 - Aprendizaje de las matemáticas*  
Alicia Alelí Chaparro Caso-López y César Gómez-Monarez
- 72** *Capítulo 3 - Relevancia de la interacción en el aula desde la perspectiva psicológica*  
Amy Nayeli Vargas Ceseña, Alicia Alelí Chaparro Caso-López y Juan Carlos Rodríguez Macías
- 92** *Capítulo 4 - Efectividad de prácticas directivas en contextos escolares vulnerables*  
Genoveva Gutiérrez Ruiz, Noé Mora Osuna y Alicia Alelí Chaparro Caso-López
- 110** *Capítulo 5 - Evaluación del diseño de la prueba de matemáticas de primaria (06) del PLANEA ELCE 2015*  
Juan Carlos Pérez-Morán, Juan Carlos Rodríguez Macías, Ramsés Vázquez Lira y Alina Alejandra Ramos Vargas

## SEGUNDA PARTE: *EDUCACIÓN SUPERIOR*

- 134** *Capítulo 6 - Relación Universidad-desarrollo tecnológico en el marco de la sociedad y la economía del conocimiento*  
Rodolfo García-Galván y Juan Carlos Rodríguez Macías
- 167** *Capítulo 7 - Cumplir con los otros: una revisión documental sobre la colaboración UABC-Tercer Sector*  
Ricardo Lindquist Sánchez, Rodolfo García-Galván y Alicia Alelí Chaparro Caso-López
- 188** *Capítulo 8 - Contexto institucional de la cooperación IES-sector productivo: marco general para México y Baja California*  
Marcela Morales Páez, Rodolfo García-Galván y Juan Carlos Rodríguez Macías
- 210** *Capítulo 9 - La cooperación tecnocientífica IES-SP en Baja California desde la percepción de los profesores-investigadores*  
Marcela Morales Páez, Juan Carlos Vázquez Almaraz y Juan Carlos Rodríguez Macías
- 224** *Capítulo 10 - ¿Qué valoran los empleadores en los egresados de educación superior en Baja California?*  
Dalia Lizzett Acosta Cárdenas, Juan Carlos Rodríguez Macías, Rodolfo García-Galván y Alicia Alelí Chaparro Caso-López
- 238** *Capítulo 11 - Análisis del procedimiento para evaluar programas educativos en la educación superior de México a partir del modelo CIPP de Stufflebeam*  
Nancy Noemí Gutiérrez Anguiano, Leticia Berenice Castillo García, Alicia Alelí Chaparro Caso-López y Rodolfo García-Galván

# Prólogo

Escribir el prólogo de un libro es un verdadero honor y una enorme responsabilidad, puesto que se trata de convencer al lector de que las páginas que siguen son merecedoras de su tiempo y esfuerzo. En este caso, ambas sensaciones, honor y responsabilidad, las comparte quien suscribe, pero la tarea de convencer al lector resulta extremadamente fácil debido al interés que suscita cada capítulo en sí mismo, desde su título hasta sus referencias.

El Instituto de Investigación y Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California es, sin duda, el principal referente en materia de investigación educativa del Estado y uno de los primeros en México. En los más de 26 años desde su creación, se han realizado cientos de investigaciones y se ha formado a numerosas cohortes de estudiantes de posgrado.

Como bien es sabido, la universidad tiene varias misiones, las cuales han ido cambiando desde su creación con el devenir de los tiempos y los intereses de la sociedad. Sin embargo, es indudable que los centros universitarios actuales deben cumplir con dos objetivos principales: formar a sus estudiantes y difundir conocimientos científicos y rigurosos logrados con sus investigaciones. El IIDE por sus peculiares características, es un ejemplo de cómo integrar ambos objetivos.

*Apuntes de investigación educativa* es una prueba irrefutable de la anterior afirmación. Se trata de la compilación de estudios y revisiones documentales, diseñados y llevados a cabo por estudiantes e investigadores del IIDE, bajo el amparo de sus programas de posgrado de Maestría o Doctorado en Ciencias Educativas. Es pues, el producto final de procesos de formación e investigación desarrollados íntegramente en el instituto.

Los citados trabajos, en forma de capítulos, se agrupan bajo dos temáticas; por una parte, los cinco primeros, abordan cuestiones relacionadas con la etapa básica y obligatoria del sistema educativo. Por otra, los seis restantes, se centran en aspectos vinculados con la educación superior.

Los dos primeros capítulos, de estructura similar, se centran respectivamente en los aprendizajes más importantes de los primeros años del sistema educativo puesto que son el cimiento en el que apoyar muchos otros: la escritura y las matemáticas. Tras una extensa revisión bibliográfica, se repasan ambos procesos de aprendizaje y los problemas más frecuentes que se desprenden de los citados procesos. Cabe destacar que, tanto el capítulo 1 (Chaparro y Arregui) como el 2 (Chaparro y Gómez-Monarez), hacen hincapié en la evaluación, análisis y corrección de los principales errores cometidos. Aportan informaciones muy completas sobre ambos procesos de aprendizaje y permiten extraer conclusiones que afectan de manera directa a su enseñanza así, aunque el núcleo central es el aprendizaje, se puede afirmar que serán de interés y utilidad para todos aquellos implicados en el proceso de enseñanza-aprendizaje de la escritura o de las operaciones fundamentales de matemáticas, especialmente para docentes y padres.

Los capítulos 3 y 4, a pesar de tener estructura y objetivos diferentes presentan un denominador común: ambos profundizan en cuestiones vinculadas con las relaciones interpersonales que se establecen en los centros educativos de la etapa básica.

En el capítulo 3 (Vargas, Chaparro y Rodríguez) se presenta una revisión del estado del arte referida a la interacción en el aula entre estudiantes y estudiantes-docentes. Asimismo, se exponen los paradigmas que, desde la Psicología, han surgido para explicar las interacciones en el contexto educativo y de qué manera pueden influir en el desarrollo emocional e incluso en el desempeño académico de los estudiantes. Tras leerlo, cualquier docente tendrá herramientas para enfrentarse a su práctica diaria, siendo consciente de que numerosas investigaciones concuerdan en dos ideas: (1) que una interacción positiva y responsable con los estudiantes influye de manera decisiva la consecución de aprendizajes significativos y (2) que lograr una interrelación entre estudiantes basada en el respeto y el buen clima ayudará a tener estudiantes más motivados en su aula.

El capítulo 4 (Gutiérrez, Mora y Chaparro) se centra en una figura esencial de las escuelas de educación básica: el director o directora. Basándose en la literatura sobre el tema, se destaca la importancia que las prácticas directivas tienen en todos los contextos educativos y, especialmente en los contextos más desfavorecidos y vulnerables. La conclusión fundamental se basa en la idea de que las labores directivas sólo tendrán un efecto positivo en los estudiantes si se adaptan al contexto sociocultural y económico en el que se ubica la escuela. Este breve trabajo hará reflexionar a quienes lideran las escuelas puesto que les permitirá analizar sus tareas y funciones bajo el prisma del contexto en el que se encuentran inmersas.

Por último, en esta primera parte del libro se incluye un quinto capítulo (Pérez-Morán, Rodríguez, Vázquez y Ramos) con un enfoque totalmente diferente puesto que se presenta un interesante análisis de contenido de ítems liberados de una importante prueba que pretende medir los conocimientos matemáticos de los estudiantes mexicanos de 6º de Educación Primaria. Partiendo del llamado Modelo para la Evaluación del Diseño Universal (Thompson, Johnstone y Thurlow, 2002) y mediante una metodología basada en dos procedimientos, panel de expertos y protocolos verbales con estudiantes, presenta un pormenorizado análisis llegando a resultados enormemente interesantes sobre los ítems concretos sobre los que trabaja. Este capítulo puede ser muy clarificador y relevante para los encargados de diseñar las pruebas a gran escala que, en ocasiones, tienen implicaciones de gran impacto para estudiantes e instituciones. Además, puede ayudar a todos aquellos que diseñan pruebas para evaluar contenidos a un grupo reducido de estudiantes, ya que se aborda cuestiones que cualquier docente debería tener en cuenta al elaborar sus exámenes.

La segunda parte compila seis capítulos más, cuyo factor común es la Educación Superior. El primero de ellos, el capítulo 6 (García-Galván y Rodríguez) presenta una discusión teórica y conceptual sobre la relación de la universidad con su entorno. Los autores buscan de una manera sencilla mostrar al lector las diferentes aristas que comprenden los procesos implicados en la relación entre estos dos entes: universidad y sociedad.

El capítulo 7 (Lindquist, García-Galván y Chaparro), parte de la preocupación por conocer de qué manera la Universidad Autónoma de Baja California está colaborando con el llamado Tercer Sector. Para llegar a la principal conclusión, consistente en que hay que seguir trabajando en estrechar la colaboración puesto que aún es muy superflua, hace un interesante repaso por el origen y evolución de la universidad, deteniéndose en las funciones que se le ha otorgado a lo largo de la historia hasta concretar en la importancia de colaborar con el Tercer Sector (elemento que también se detiene a explicar). Cualquiera persona vinculada con la universidad o con organizaciones y comunidades sociales, disfrutará del enfoque y revisión de este capítulo.

De manera complementaria, el capítulo 8 (Morales, García-Galván y Rodríguez) se centra en la relación o cooperación entre las instituciones universitarias y el sector productivo o empresarial. Revisando el concepto de institución y analizando los modos existentes de colaboración con las empresas, se sintetizan las principales normativas, nacionales y estatales, que rigen actualmente este tipo de colaboración. El principal hallazgo que extraen es que, la legislación, sobre el papel, contempla elementos y fuentes de colaboración de gran relevancia para ambos estamentos, universidad y empresas, sin embargo, no todos están implementados en la práctica. Los profesionales encargados de las prácticas externas de los estudiantes, los que ocupan cargos de gestión vinculados con la relación entre estudios universitarios y sector productivo, ya sea en la universidad o en la empresa, reflexionarán con la lectura del capítulo sobre su labor y el modo de desarrollarla para favorecer esta necesaria colaboración.

En el Capítulo 9 (Morales, Vázquez y Rodríguez) se presenta un estudio empírico realizado con profesores investigadores, sobre las condiciones y procesos relacionados con la cooperación tecnocientífica (CT) entre las instituciones de educación superior y el sector productivo. Entre los hallazgos relevantes destaca que, en general, existe una predisposición positiva hacia la realización de actividades de CT. El estudio da cuenta de la necesidad de que el gobierno adopte una actitud de mayor apoyo hacia la CT, tanto a través de financiamiento directo como de la generación de políticas públicas que coadyuven al fortalecimiento de esta. Por lo tanto, el estudio presentado en este capítulo puede resultar relevante para los interesados el tema de la cooperación tecnocientífica.

En el capítulo 10 (Acosta, Rodríguez, García-Galván y Chaparro) se aborda, en esencia, la misma cuestión que el capítulo 8, pero concretando en las valoraciones de algunos empleadores del estado de Baja California. Basándose en entrevistas semiestructuradas a encargados de recursos humanos de tres importantes empresas, recogen información sobre los procesos de selección, las características que buscan de los futuros empleados y la opinión que les merece la Educación Superior en cuanto a la formación que consigue en sus egresados. Entre las interesantes conclusiones destaca la importancia que los empleadores otorgan, no tanto a los conocimientos de los solicitantes, que los presuponen, como a las actitudes que demuestran y destacan la responsabilidad, el compromiso, la iniciativa y la capacidad para resolver problemas, entre otras. Este capítulo puede resultar de especial interés para aquellos estudiantes o recién egresados de la universidad que se encuentran en el proceso de búsqueda de empleo.

Para finalizar, el capítulo 11 (Gutiérrez, Castillo, Chaparro y García-Galván) se centra en un campo esencial de la evaluación, la evaluación de programas. Tomando como referencia uno de los modelos teóricos más conocidos y valorados, el modelo CIPP de Stufflebeam, analiza los procesos que siguen los dos organismos encargados de la evaluación de los programas y planes de estudios de todas las instituciones universitarias mexicanas, el CIEES y el COPAES. El fundamental hallazgo que se desprende de la comparación y el análisis llevado a cabo es que, en esencia, ambos organismos cumplen con las etapas y fases que el modelo teórico de referencia apunta. El lector interesado en la calidad universitaria, así como en el diseño y evaluación de planes de estudios, disfrutará de una excelente integración del modelo teórico citado y de los modelos prácticos implementados en la actualidad para la evaluación de programas de Educación Superior en México.

Sin otra pretensión que potenciar la motivación hacia la lectura de este libro de investigación que compila diversos intereses vinculados con la educación básica y superior, no me queda más que la satisfacción que produce la sensación de imaginar que lo habré logrado.

Mis felicitaciones al IIDE y en especial a todos y cada uno de los autores de esta obra y a aquellos que, sin ser autores, hacen posible la edición anual de trabajos de posgrado que contribuyen a la difusión y dan visibilidad a investigaciones educativas en el ámbito nacional e internacional.

***Dra. Coral González Barbera***

Profesora de Métodos de Investigación y Diagnóstico en Educación

Facultad de Educación

Universidad Complutense de Madrid

# Capítulo 5

## Evaluación del diseño de la prueba de matemáticas de primaria (06) del PLANEA ELCE 2015

Juan Carlos Pérez-Morán, Juan Carlos Rodríguez Macías,  
Ramsés Vázquez Lira y Alina Alejandra Ramos Vargas

Encontrar formas para mejorar el diseño de las pruebas educativas de gran escala es una necesidad recurrente para los sistemas de evaluación, organizaciones y equipos encargados de su desarrollo y validación a nivel nacional e internacional (Ravela et al., 2008). En México, con los cambios recientes en la legislación federal, se puso mayor énfasis en la rendición de cuentas y la evaluación del avance del Sistema Educativo Nacional (SEN) a través de pruebas de logro educativo a gran escala. Actualmente, con la presencia del Instituto Nacional para la Evaluación de la Educación (INEE, 2018) se espera que las evaluaciones de logro educativo contempladas en el *Plan Nacional para las Evaluaciones de los Aprendizajes* (PLANEA) cuenten con una serie de características sobre la calidad técnica, atendiendo los criterios para el desarrollo de pruebas del INEE (2017) y los estándares para pruebas psicológicas y educativas de la *American Educational Research Association, American Psychological Association y National Council on Measurement in Education* (AERA, APA y NCME, 2014).

Con la implementación del PLANEA, se espera recabar información que sea útil para la mejora de los aprendizajes de todos los estudiantes de educación básica obligatoria, incluyendo a aquellos estudiantes con discapacidades y con una lengua materna diferente al castellano. Con el alto nivel de participación de los estudiantes

de todo el país en el PLANEA y de su obligatoriedad de implementación, se crea una gran responsabilidad social y política a los responsables de los Sistemas Estatales de Educación (SEE), dado que los resultados se informan públicamente para la rendición de cuentas. Con ello, hay una necesidad crítica del INEE por asegurar la calidad del diseño de las pruebas nacionales.

Actualmente, en los estándares de la AERA, APA y NCME (2014), así como investigadores y desarrolladores de pruebas reconocidos a nivel internacional (Gierl y Lai, 2018; Johnstone, 2003; Pérez-Morán, Rojas y Vázquez-Lira, 2018; You, Li, Zhang y Liu, 2018; Dolan, Hall, Banerjee, Chun y Strangman, 2005; Ketterlin-Geller, 2005; entre otros), recomiendan la aplicación del Modelo para la Evaluación del Diseño Universal (MEDU) desarrollado por Thompson, Johnstone y Thurlow (2002) para el diseño de las pruebas, garantizando que sean accesibles a cualquier persona, sin necesidad de una adaptación posterior a su desarrollo. Aunque el MEDU es un concepto que comenzó en el campo de la arquitectura, se expandió rápidamente hacia iniciativas ambientales, recreación, arte, salud y, por supuesto, educación (Mace, 1985). Por tal motivo, las pruebas diseñadas con un enfoque universal ofrecen un panorama prometedor para proporcionar condiciones de evaluación adecuadas para todos los estudiantes evaluados, brindándoles una oportunidad comparable para decantar su logro a través de instrumentos de evaluación estandarizados (Anderson, Lai, Alonzo y Tindal, 2011). Es por ello, que el MEDU presenta un replanteamiento metodológico para el análisis del sesgo en las evaluaciones educativas de gran escala, que ayuda a garantizar que la información sea de calidad en términos de justicia de las pruebas.

### **Normatividad de la evaluación en México**

El INEE, como organismo constitucional autónomo, tiene por objeto coordinar el Sistema Nacional de Evaluación Educativa (SNEE), así como evaluar la calidad, el desempeño y los resultados del SEN, en lo que se refiere a la educación básica y obligatoria, tanto pública como privada, en todas sus modalidades y servicios (Gobierno Federal, 2013). Aunado a ello, y de forma especial, el INEE tiene la responsabilidad de diseñar y realizar

mediciones y evaluaciones que correspondan a componentes, procesos o resultados del SEN respecto a los atributos de alumnos, docentes y autoridades escolares, así como de las características de las instituciones, las políticas y los programas educativos.

Como respuesta a los compromisos establecidos, en especial aquellos relacionados con la evaluación de la calidad, el desempeño y los resultados del SEN, el INEE diseñó y aplicó por primera vez en el año 2015 el Plan Nacional para las Evaluaciones de los Aprendizajes (PLANEA). Como antecedente, en el año 2013, el Instituto solicitó a un comité de expertos un estudio de validez y confiabilidad de las pruebas ENLACE (Evaluación Nacionales del Logro Académico en Centros Escolares) y las pruebas EXCALE (Exámenes para la Calidad y el Logro Educativo), administrados por la Secretaría de Educación Pública (SEP) y el INEE respectivamente (Martínez-Rizo, 2015). Entre las recomendaciones que señalaron los expertos del estudio de validez destacan cinco grandes compromisos que el INEE debía de asumir para cuidar que el nuevo plan de evaluaciones nacionales cuente con las características de calidad técnica esperados: (1) rigor metodológico del diseño, construcción y aplicación de las pruebas; (2) correspondencia de los resultados con los propósitos que se plantearon durante su diseño; (3) justicia de la prueba con la reducción del sesgo de las evaluaciones que pudiera resultar de la diversidad lingüística, de género, étnica y cultural del país; (4) control en los mecanismos de aplicación, indicando los casos en que los resultados pudieran no ser confiables; y (5) precisión y comparabilidad de las medidas, para que sus resultados pudieran ser confiables a lo largo del tiempo.

A partir de dicho estudio, el Instituto diseñó –en coordinación con la SEP– el PLANEA con tres modalidades de evaluación para conocer el aprendizaje de los estudiantes con la intención de mantener las principales fortalezas de ENLACE y EXCALE y superar sus debilidades: Evaluación del Logro referida al Sistema Educativo Nacional (ELSEN), y la Evaluación del Logro referida a los Centros Escolares (ELCE). Con dichas modalidades del PLANEA, tanto el INEE como la SEP, esperan obtener distintos niveles de desagregación de la información de los aprendizajes logrados por los estudiantes en el país, integrar las evaluaciones nacionales para un desarrollo lógico y articulado,

mejorar la coordinación entre SEP e INEE en temas de evaluación, e incorporar a las autoridades estatales y las escuelas al sistema de monitoreo de los avances del SEN (INEE, 2018). En especial, con el PLANEA ELCE se espera cumplir con el propósito de ofrecer información contextualizada para el monitoreo, planeación, programación y operación del sistema educativo y de forma específica información para la mejora de los procesos de enseñanza a nivel de los centros escolares (unidades de análisis).

Con base en la revisión de las experiencias de evaluación anteriores, particularmente ENLACE y EXCALE (Martínez-Rizo, 2015), así como de la búsqueda de diseños que cumplieran con los propósitos mencionados, el INEE definió elementos esenciales de lo que ahora conocemos como el PLANEA, dentro del cual para lo que compete al presente capítulo se resalta “[Incorporar] procedimientos para identificar el sesgo de las evaluaciones que pudiera resultar de la diversidad de género, étnica, cultural y lingüística, y para evitar comparaciones injustas” (INEE, 2015, p. 12). Esto con el fin de que las evaluaciones con las que cuenta el INEE, posean las características técnicas que le permitan tener certeza de que se está atendiendo el principio de imparcialidad y equidad. Por lo tanto, es viable y muy recomendable, adoptar un enfoque más integral para abordar estas limitaciones de las pruebas, por lo que el objetivo de aplicar los principios de diseño universal a las evaluaciones a cargo del INEE, es poder diseñar y desarrollar evaluaciones que permitan la participación de la gama más amplia de estudiantes y que den como resultado conclusiones con un grado más alto de validez sobre su desempeño.

### **Modelo para la Evaluación del Diseño Universal**

En el Modelo para la Evaluación del Diseño Universal (MEDU) se parte de un enfoque basado en los principios de accesibilidad para una amplia variedad de usuarios finales. Thompson et al. (2002), describieron los elementos que debían contener las evaluaciones diseñadas universalmente. En su informe, esta metodología se alineó a evaluaciones a gran escala, resaltando la necesidad imperante sobre atender la necesidad de que “a todos los examinados se les brinde una oportunidad equitativa y comparable para demostrar su dominio en el(los) constructo(s) de la prueba a aplicar” (p. 5).

Para ello, los Estándares para Pruebas Educativas y Psicológicas (AERA, APA y NCME, 2014) también abordan esta necesidad, el tratamiento equitativo también incluye factores tales como las condiciones de prueba apropiadas y la igualdad de oportunidades para familiarizarse con el formato de la prueba, los materiales de práctica, etc., con base en las categorías de análisis propuestos por Thompson et al. (2002): (1) inclusión poblacional, (2) definición precisa del constructo, (3) accesibilidad e imparcialidad (4) acomodación flexible de los contenidos, (5) procedimientos e instrucciones simples, claras e intuitivas, (6) máxima legibilidad, y (7) máxima comprensibilidad (p. 6).

Por ejemplo, Johnstone (2003) generó un diseño experimental en donde 231 estudiantes de bajo desempeño resolvieron dos pruebas, el orden de presentación se contrabalanceó entre los participantes. Utilizó una prueba (diseñada tradicionalmente) integrada por los reactivos liberados de una evaluación a gran escala, la cual se presentó en formato estándar, por su parte, una segunda prueba se creó utilizando los constructos de la prueba tradicional, pero incluyó características del MEDU; esta segunda prueba aseguró la correspondencia del contenido (uno a uno) entre los constructos de los reactivos de ambas pruebas. Los resultados de los 231 participantes en ambas pruebas se contrastaron y demostraron que los estudiantes obtuvieron puntuaciones significativamente más altas en la prueba diseñada bajo el MEDU. Estos hallazgos tienen implicaciones entorno al grado de validez cuando se evalúan a estudiantes con condiciones desafiantes o desprivilegiadas, en donde pese a que tuvieran condiciones que pudieran ser consideradas una barrera (p. ej. origen étnico, grado de marginación de su localidad, género, entre otras); por ello, las evaluaciones al ser generadas a través del MEDU, permiten eliminar o matizar ciertas barreras alrededor de los métodos de evaluación, extrayendo la parte sustantiva de los aprendizajes evaluados y generando una mayor inclusión social y, por ende, una disminución de la inequidad por parte de las evaluaciones, y en un país como México, donde existe tanta riqueza cultural y sociodemográfica, el poder apoyarse del MEDU para brindar un sistema educativo de calidad y equidad acerca mucho más hacia el Derecho Humano a una Educación Obligatoria de Calidad del país.

Cabe resaltar que el MEDU debe tenerse presente a lo largo de todas las fases del proceso de medición (desde el diseño y desarrollo de la prueba hasta la administración, puntuación, interpretación y uso de resultados), ya que una característica fundamental, es que las evaluaciones elaboradas bajo este esquema, se refinan continuamente, permitiendo la participación activa de una gran variedad de estudiantes lo que da como resultado inferencias más robustas sobre su desempeño, debido al alto grado de validez que guardan. Se basan en la premisa de que cada estudiante en la escuela es parte de la población a la que se va a realizar la prueba, y que los resultados de esta no deben verse influenciados por variables tales como discapacidad, género, grupo étnico o idioma. Por lo tanto, las evaluaciones diseñadas universalmente pueden reducir la necesidad de adaptaciones al eliminar las barreras de acceso asociadas con las pruebas en sí mismas. Así, el MEDU es una vía para aumentar la participación en las evaluaciones enfocadas al SNE, ya que se basan en la misma ética de equidad e inclusión que se espera para las personas con discapacidades, grupo étnico o género, debido a que el desempeño en la prueba no se debe ver afectado por dichas variables. Por otro lado, es importante recordar que el MEDU no aborda las deficiencias en la instrucción, ya que los estudiantes que no hayan tenido la oportunidad de aprender los conceptos o temas evaluados estarán en desventaja durante la prueba, sin importar cuán universal sea el diseño de la evaluación.

### **Propósito del estudio**

*“Evaluar el diseño de los ítems liberados de la prueba de Matemáticas de primaria (06) del PLANEA ELCE 2015 bajo el MEDU”.* Cabe destacar que existe una ventaja respecto a la prueba seleccionada, ya que pertenece a un campo ampliamente estudiado con métodos similares (Brown y Burton, 1978; Chen y Macdonald, 2011; Gierl et al., 2009; Ma, Çetin y Green, 2009; Pérez-Morán, 2014; Pérez-Morán et al., 2018; Pérez-Morán, Larrazolo, Backhoff y Guaner, 2015; Revuelta y Ponsoda, 1998; Romero, Ponsoda y Ximénez, 2008). Con ello, se espera recabar información que ayude a los desarrolladores de las evaluaciones a mejorar sus procesos de diseño y a los tomadores de decisiones

a interpretar de forma más justa los resultados de las evaluaciones, además de aportar a la discusión y debate en el ámbito disciplinar. Asimismo, se pretende mostrar los beneficios de la aplicación de métodos cognitivos para identificar posibles sesgos en los ítems de las pruebas mediante el análisis de los procesos de respuestas de un grupo de expertos y estudiantes aportando evidencias, validez de constructo y de equidad de la prueba (Messick, 1989; Snow y Lohman, (1989).

## **Método**

### ***Tipo de estudio***

Psicométrico y de validación, en donde se complementaron los métodos cognitivos de panel de expertos y protocolos verbales con estudiantes, así como aspectos principales del MEDU propuestos por Thompson et al. (2002). En investigaciones basadas en el MEDU se ha encontrado que los diseñadores de pruebas pueden desarrollar evaluaciones más accesibles para los examinados mediante su aplicación (Johnstone, 2003). También, se ha encontrado que con la aplicación del MEDU, el apego a estrategias de diseño eficaz y el uso de métodos cognitivos se puede minimizar la varianza irrelevante del constructo originada por problemas en el diseño, formato y sesgos culturales presentes en los ítems de las pruebas (Haladyna, Downing y Rodríguez, 2002; Ercikan et al., 2010; Smith-Castro y Molina, 2011).

En especial, los métodos del panel de expertos y los protocolos verbales con estudiantes son de gran ayuda para lograr una evaluación más granulada del diseño de los ítems y aumentar la validez de las evidencias convergentes recabadas en ambos métodos, al contrastar las hipótesis de los expertos respecto al diseño de los ítems y los reportes verbales con las evocaciones del proceso de respuesta de los estudiantes ante los ítems en análisis. Por ejemplo, contar con hipótesis de expertos referentes a los requerimientos para contestar un ítem puede ayudar a prever de explicaciones en los casos en donde, durante las técnicas de pensamiento en voz alta, se les dificulta a los estudiantes de corta edad verbalizar lo que piensan, o en casos en donde se suscitan procesos en tan sólo algunos segundos y que, por lo tanto, no es posible su

introspección (Sternberg, 1977). Asimismo, contar con información granulada de los procesos de respuesta de examinados ante los ítems de una prueba puede ayudar a verificar o corregir las hipótesis de los expertos con información muy pobre o escasa referente a los requerimientos para responder los ítems (Ericsson y Simon, 1984, 1993; Leighton, 2004, 2017; Leighton y Gierl, 2007; Messick, 1989; Snow y Lohman, 1989).

### *Participantes*

Se trabajó con dos tipos de participantes en el estudio, correspondientes a las dos técnicas cognitivas aplicadas. Para el panel de expertos se determinó trabajar con cinco especialistas con experiencia en la enseñanza de las Matemáticas: tres psicólogos con experiencia en atención a estudiantes de primaria y secundaria con problemas de aprendizaje en matemáticas, un especialista en análisis y modelamiento de procesos cognitivos en el campo del pensamiento lógico-matemático, y un docente con amplia experiencia en la enseñanza de las Matemáticas en educación básica. Para la selección de los expertos se siguieron las recomendaciones de Rupp, Templin y Henson (2010), con 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, de las diferentes vertientes para el desarrollo y aprendizaje de los componentes o atributos de la prueba y de los contextos en los que los examinados adquieren y utilizan dichos atributos.

Por su parte, para los protocolos de pensamiento en voz alta aplicando técnicas concurrentes y retrospectivas se contó con la participación de ocho estudiantes voluntarios de sexto de primaria (cuatro mujeres y cuatro hombres), tanto de rendimiento alto y bajo. En cuanto a la estimación de la cantidad de participantes requeridos para el análisis de protocolos, Nielsen (1994) menciona que puede ser variada según sea el propósito del estudio. Para el análisis de protocolos llevado a cabo en esta investigación sólo se requirió de un grupo pequeño de participantes que aportara suficiente información de sus procesos de respuesta ante los ítems estudiados con el fin de verificar si estos están representados en el modelo cognitivo elaborado previamente por los expertos.

Para el establecimiento de los criterios de selección de los participantes de los protocolos verbales se tomaron en cuenta las recomendaciones de Ericsson y Simon (1984; 1993). Dichos autores proponen incorporar al análisis del proceso de respuesta tanto a novatos como a expertos en el dominio de interés. Por lo tanto, se establecieron como criterios para la selección de los participantes las variables de rendimiento escolar, grado educativo y la recomendación del profesor. Además, se estableció que, del total de participantes, 50% fueran hombres y 50% fueran mujeres.

Como criterio de selección de estudiantes novatos y expertos se estableció un promedio de rendimiento escolar mayor a 8.5 y que, además, fueran referidos por los profesores como estudiantes sobresalientes en el dominio de las Matemáticas. Dichos estudiantes conformaron el grupo de estudiantes expertos para el piloteo. De igual forma, se estableció la selección de estudiantes de sexto grado de primaria que presentaran un promedio mayor a 6.0 pero menor a 8.0 y que fueran referidos por los profesores como estudiantes con bajo desempeño o no sobresalientes en el dominio de las Matemáticas. Dichos estudiantes conformaron el grupo de estudiantes novatos. De la misma forma, se determinó que del total de participantes, 50% fueran estudiantes expertos y 50% fueran estudiantes novatos.

### ***Instrumentos***

Los instrumentos para el análisis del diseño de los ítems mediante métodos cognitivos fueron: el sistema de categorías para la evaluación del diseño universal adaptado especialmente para los ítems liberados de la prueba de Matemáticas primaria (06) del PLANEA ELCE 2015 (ver Apéndice A) y el formato del protocolo verbal concurrente y retrospectivo elaborado por Pérez-Morán (2014). Es importante mencionar que se adaptaron y desarrollaron categorías y códigos de análisis para la evaluación del diseño de los ítems en estudio tomando en cuenta los aspectos básicos y categorías del MEDU propuestos por Thompson et al. (2002). Asimismo, es necesario resaltar que expertos en el campo de la medición (Ericsson y Simon, 1984; 1993; Leighton y Gierl, 2007; Messick, 1989; Nichols, Kobrin, Lai y Koepfler, 2016) recomiendan ampliamente el uso de métodos

cognitivos y, de forma muy particular, de paneles de expertos y reportes verbales dados sus beneficios y aportaciones al tema de validez y equidad de las pruebas.

### **Procedimiento del estudio**

El procedimiento del estudio se estructuró en tres fases. En la primera fase, se seleccionaron los métodos cognitivos, se habilitó un laboratorio cognitivo y se diseñaron los formatos de la guía para los protocolos verbales con estudiantes, así como el sistema de categorías para el análisis de los ítems tomando en cuenta los elementos del MEDU. En la segunda fase, el panel de expertos realizó el análisis cognitivo-reticular de los ítems, mapearon los modelos teóricos de respuesta y elaboraron hipótesis de posibles problemas de diseño de cada uno de los ítems de la prueba. Posteriormente, durante la misma fase, se aplicaron los protocolos verbales con técnicas de pensamiento en voz alta (concurrentes) y entrevistas cognitivas (retrospectivas) a estudiantes. Para la tercera y última fase, los expertos verificaron sus hipótesis planteadas de los procesos de respuesta de los estudiantes, así como de los posibles problemas de varianza irrelevante o algún tipo de sesgo presentes en los ítems.

### **Resultados**

En los resultados de la evaluación del diseño universal obtenidos por el panel de expertos y los protocolos verbales con estudiantes se encontró que, de los 50 ítems liberados de la prueba de Matemáticas 06 del PLANEA ELCE 2015, 46 ítems (92%) presentaron al menos un problema de diseño universal. En total, se detectaron 101 problemas, donde la moda corresponde a 2 problemas por ítem y el ítem con más problemas es el PMB14 con 6 de ellos relacionados con los elementos del MEDU de *inclusión poblacional, definición precisa del constructo y máxima comprensibilidad* (ver Apéndice A). Respecto a los resultados de la evaluación del diseño universal de los ítems en cada uno de los seis elementos del MEDU, se encontró que: 9 ítems (18%) presentaron al menos un problema relacionado con el elemento *I. Inclusión poblacional*; 34 ítems (68%) con el elemento *II. Definición precisa del constructo*; 3 ítems (6%) con el elemento *III. Accesibilidad e imparcialidad*;

en el elemento IV. *Acomodación flexible de los contenidos* no se encontraron ítems con problemas; 8 ítems (16%) presentaron problemas relacionados con el elemento V. *Procedimientos e instrucciones claras, simples e intuitivas*; y, por último, 10 ítems (20%) presentaron problemas relacionados con el elemento VI. *Máxima comprensibilidad*. Por su parte, en los resultados de la evaluación de los ítems clasificando los problemas en los rubros de: (I) *alineación curricular*, (II) *sesgo* y (III) *diseño*, se encontró que, en el primer rubro 34 ítems (68%) presentaron problemas, 12 ítems (24%) en el segundo y 31 ítems (62%) en el tercer rubro.

### ***Problemas de alineación curricular***

Del 68% de los ítems que presentaron problemas de *alineación curricular*, el ítem PMA18 y PMA01 son dos ejemplos que pueden ilustrar de forma clara el problema señalado. El ítem PMA18 fue diseñado para evaluar el contenido: *Resolver problemas aditivos con números decimales*. La incógnita es el resultado, que corresponde al bloque I del libro de texto: *Desafíos matemáticos* de sexto grado de primaria. Sin embargo, los resultados de los métodos cognitivos aplicados señalan que los modelos de enseñanza del libro de texto no coinciden con los procesos de respuesta requeridos para contestar el ítem. Tras los análisis realizados, se encontró que el ítem PMA18 se encuentra claramente relacionado con el contenido 4.4.3. *Resolución de sumas o restas con números decimales en diferentes contextos* del bloque IV del libro de texto: *Desafíos matemáticos* de cuarto grado de primaria; mismo de donde se desprende la intención didáctica: *Que los alumnos interpreten la parte decimal de un número en el contexto de la medición, para resolver problemas aditivos* (figura 1).

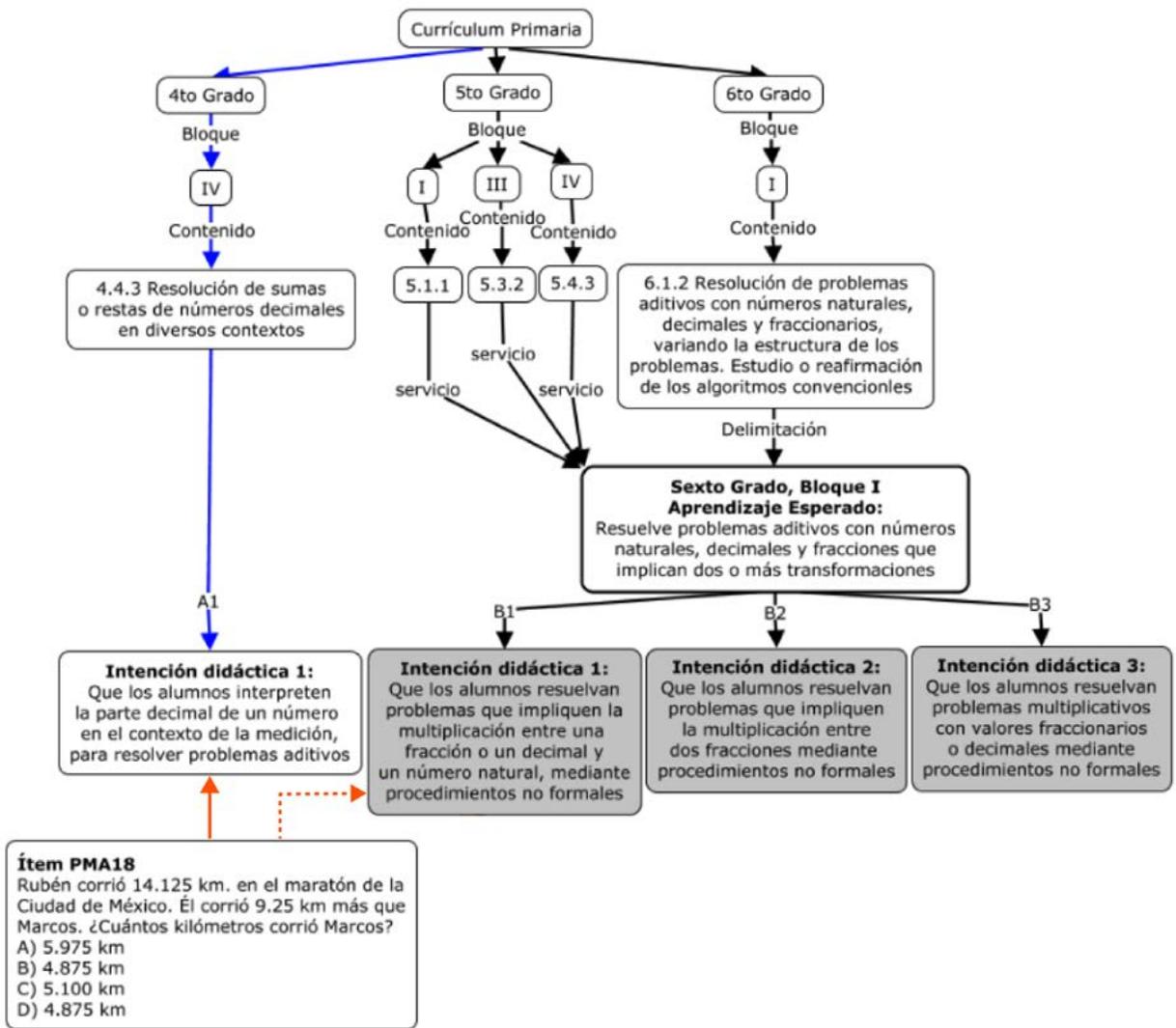


Figura 1. Diagrama de la genealogía reticular y problema de Alineación curricular del ítem PMA18.

Otro ejemplo emblemático relacionado con el problema de *alineación curricular* es el caso del ítem PMA01 (figura 2). Los expertos encontraron un problema de *sobre-simplificado* en la demanda de dicho ítem. En la especificación del ítem PMA01 se define como contenido a evaluar el *cálculo Del área de triángulos* que corresponde al *Aprendizaje esperado: Calcula el perímetro y el área de triángulos y cuadriláteros*, perteneciente al eje:

*Forma, espacio y medida* de 5° grado de primaria. Sin embargo, se puede ver, revisando los procesos de respuesta evocados por los estudiantes, que se requiere para contestar correctamente dicho ítem otras habilidades más complejas como la *comprensión de problemas matemáticos contextualizados, la deducción de los elementos del contexto requeridos para aplicar un modelo aritmético y el cálculo del área de un triángulo con rotación de la base*.

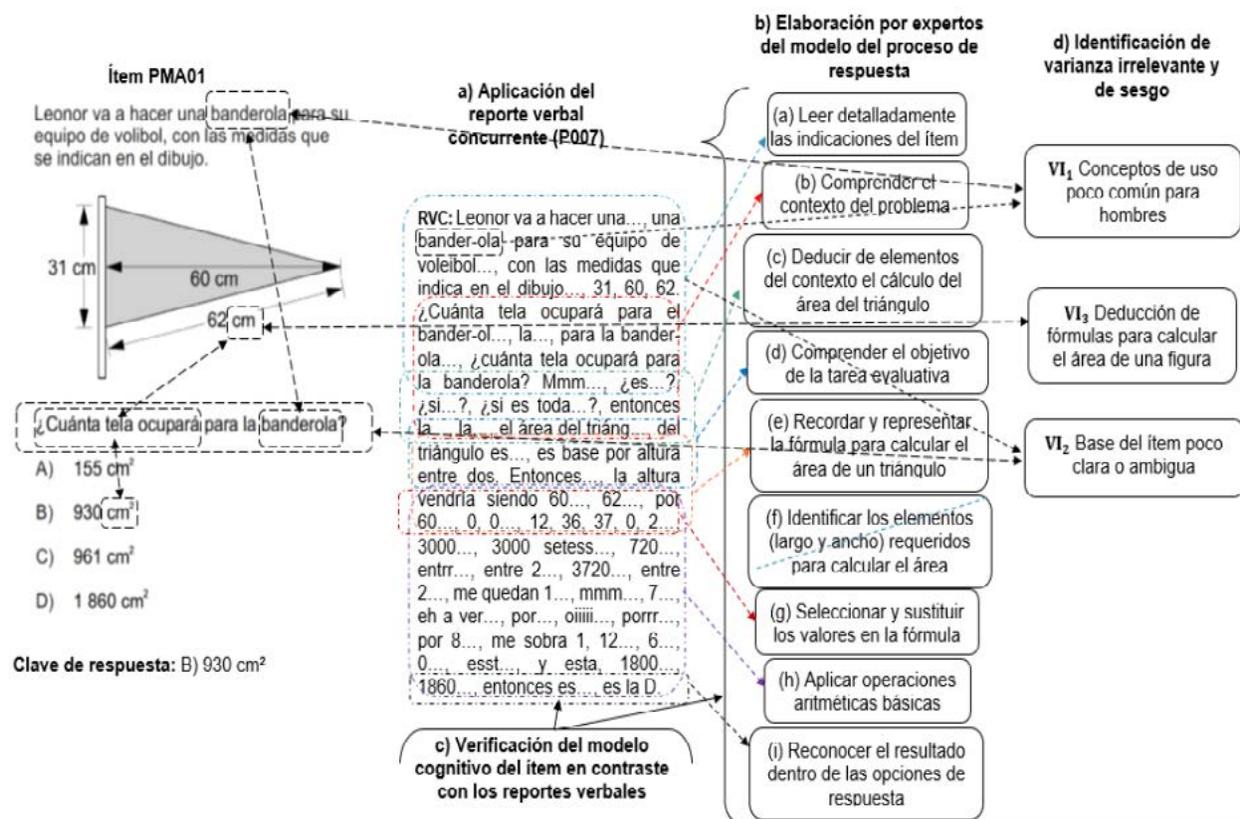


Figura 2. Diagrama del modelo de un proceso de respuesta erróneo subyacente al ítem PMA01.

También, durante el análisis de los documentos curriculares oficiales, asociados con el caso del ítem PMA01, los expertos encontraron que el estándar curricular en la *Guía del maestro* del Plan de estudio 2011 no coincidía con el *Aprendizaje esperado* establecido en la especificación (figura 3). Nótese que el estándar curricular presenta

una descripción diferente a la especificación, señalando como estándar curricular el uso de fórmulas para calcular perímetros y áreas. Sin embargo, no se establece nada referente al aprendizaje de la comprensión de problemas matemáticos contextualizados o a la deducción de elementos de un contexto particular para el cálculo del área del triángulo, ni nada relacionado con el aprendizaje del cálculo del área de un triángulo con base rotada.

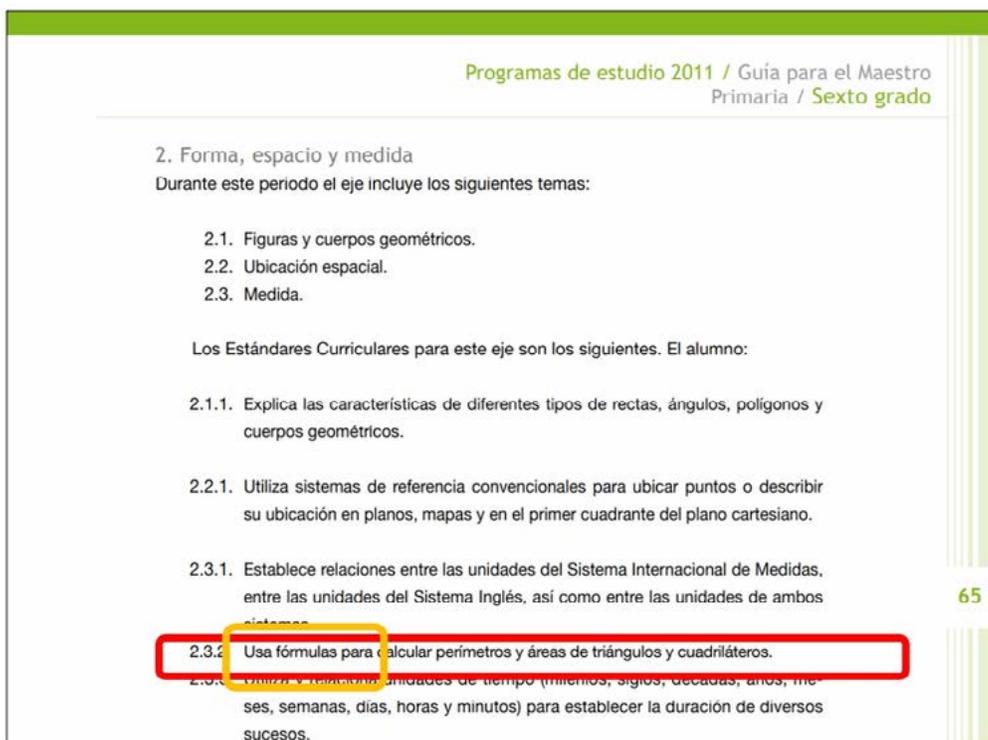
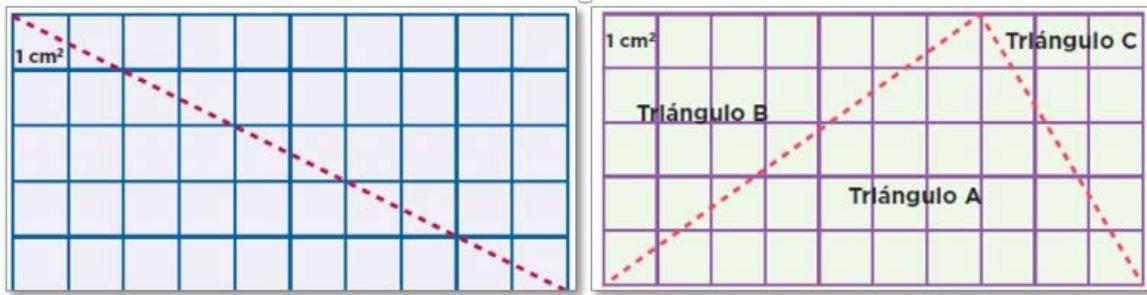
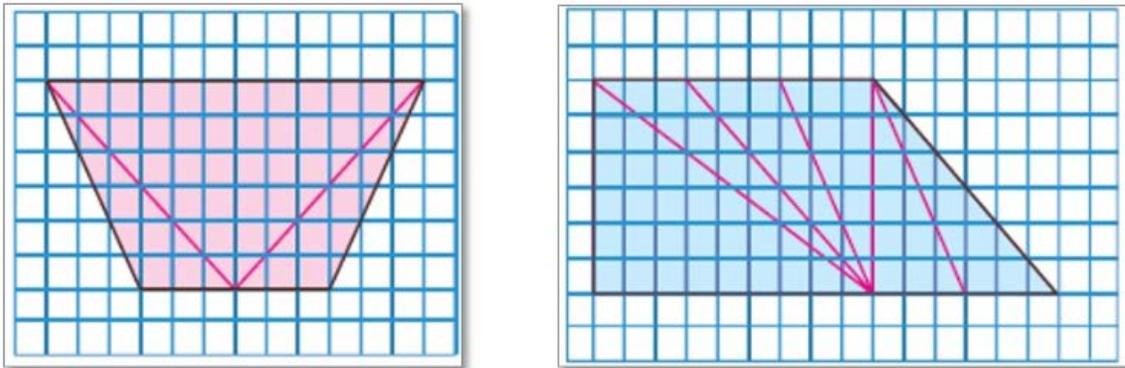


Figura 3. Estándar curricular al que se alinea la especificación del ítem PMA01 de la prueba de Matemáticas de primaria del PLANEA ELCE 2015.

De igual forma, los expertos analizaron los libros de texto del maestro como los del alumno para poder diseñar la genealogía curricular de cada uno de los ítems de la prueba en estudio. Se encontró que el ítem PMA01 se encuentra asociado especialmente con las intenciones didácticas 1 y 2 del contenido 5.3.6. del libro de texto del alumno de 5° de primaria. En la figura 4, puede observarse de forma resumida los modelos de enseñanza de las intenciones didácticas en las que se encontró mayor alineación con el ítem PMA01 y su especificación.



**Modelo del proceso de enseñanza de la intención didáctica 1:** a) En parejas leer las actividades a realizar, b) utilizar material de apoyo (pág. 209 del libro de texto), c) leer el primer problema, d) trazar una línea diagonal en el rectángulo (figura 1), e) leer la primera pregunta, f) obtener el área del rectángulo, g) responder el cuestionamiento, h) leer la segunda pregunta, i) superponer los triángulos obtenidos y responder cómo son, j) leer la tercer pregunta, k) obtener el área de cada triángulo, l) responder la pregunta, m) leer la cuarta pregunta, n) describir como se obtiene el área de un triángulo, o) leer segundo problema, p) trazar dos rectas como lo indica la figura y recortar, q) superponer los triángulos y determinar el área de cada uno, y r) responder cuál fue el área de cada triángulo.



**Modelo del proceso de enseñanza de la intención didáctica 2:** a) Leer en parejas las actividades a realizar, b) observar imagen que contienen las figuras, c) calcular el área de cada triángulo y de la figura que los contiene, d) leer la primera pregunta, e) responder cómo son la base y la altura de cada uno de los triángulos que forman el romboide, f) responder cómo son las áreas de tales triángulos, g) responder cómo son la base y la altura de cada uno de los triángulos que forman el trapecio, h) responder cómo son las áreas de tales triángulos, i) escribir conclusiones, j) formar equipos para realizar la segunda actividad, y k) calcular las áreas de cada triángulo y las áreas de las figuras completas que se muestran en las ilustraciones.

*Figura 4. Modelos de enseñanza de las Intenciones didácticas 1 y 2 del contenido 5.3.6 de 5° de primaria.*

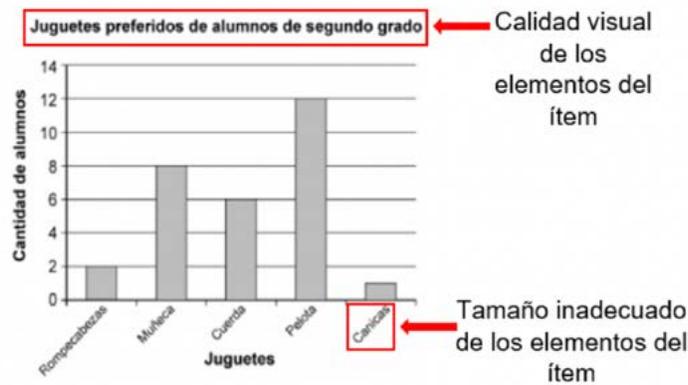
### *Problemas de sesgo*

En cuanto a los resultados de la evaluación de los ítems en lo referente a los problemas de sesgo identificados, se puede decir que, el haber complementado el análisis con el panel de expertos y protocolos verbales con estudiantes ayudó a obtener evidencias convergentes que permitieron tener mayor certeza de las hipótesis planteadas. En la figura 2, correspondiente al ítem PMA01, se observa la convergencia de los resultados de los dos métodos utilizados. En especial nótese el contraste entre las hipótesis de varianza irrelevante y sesgo con las verbalizaciones de uno de los participantes. En síntesis, fueron tres los principales problemas de diseño y sesgo que presenta el ítem PMA01: a) *problemas de inclusión poblacional*, b) *definición imprecisa del constructo* (referente a la alineación de este), y c) *problemas de comprensibilidad de la base del ítem*.

En resumen, durante el análisis cognitivo, los expertos detectaron que 12 ítems presentaron problemas de sesgo, correspondiente al 24% del total de ítems de la prueba. Los problemas detectados están relacionados con (1) *tamaño inadecuado de los elementos del ítem*, (2) *baja calidad visual de los elementos del ítem*, (3) *uso de conceptos con nivel bajo de inclusión de nivel socioeconómico* y (4) *uso de conceptos asociados con grupos, culturas o lugares particulares*. Los problemas de sesgo mencionados pueden ocasionar que un grupo minoritario o en condiciones de vulnerabilidad (por ejemplo, alumnos con debilidad visual) tengan problemas para contestar adecuadamente los ítems.

Otro buen ejemplo de los ítems de la prueba identificado con problemas de sesgo fue el PMA14 (figura 5). El problema particular de dicho ítem tiene que ver especialmente con el uso de letras pequeñas y borrosas, dificultando la tarea de contestar el ítem a las personas con debilidad visual. Durante el laboratorio cognitivo los sustentantes tuvieron que acercarse a la prueba para leer las letras de la gráfica; por unanimidad, los 8 estudiantes participantes en los protocolos verbales manifestaron que las letras eran muy pequeñas. Especialmente, perdieron tiempo descifrando las palabras e inclinaron la cabeza a la izquierda para leer la señalización “Cantidad de alumnos”.

La maestra de segundo grado preguntó a sus alumnos, cuáles eran sus juguetes favoritos y con los datos obtenidos elaboró la siguiente gráfica.



¿Cuántos alumnos tiene el grupo de segundo grado?

- A) 5      B) 12      C) 28      D) 29

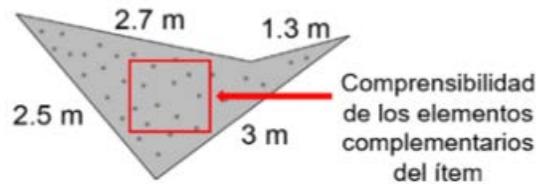
Figura 5. Problemas de sego presentes en el ítem PMA14.

### Problemas de diseño

Por su parte, 31 ítems (62) de la prueba fueron detectados por el panel de expertos con *problemas de diseño*. Los ítems que clasificados en dicho rubro presentaron problemas relacionados con (1) *falta de comprensibilidad total o parcial de las instrucciones*, (2) *base del ítem poco clara o ambigua*, (3) *uso ambiguo de términos y conceptos en la base del reactivo*, (4) *desorganización lógica de las ideas de la base del ítem* y (5) *elementos complementarios del ítem que ocasionan procesos de distracción en los sustentantes*.

En particular, un claro ejemplo del rubro relacionado con los *problemas de diseño* es el presentado por el ítem PMB08 (figura 6), ya que dicho ítem cuenta con un polígono irregular donde se incluyen elementos de sombreado y punteado que hacen énfasis en su relleno (área). Durante el laboratorio cognitivo los sustentantes manifestaron que el sombreado de la figura y los puntos (que señalaron como abolladuras, puntos, marcas) los hacían pensar que se les solicitaba obtener el área; por lo que tuvieron que regresar a las indicaciones, hasta por más de una ocasión, para asegurar la comprensión de la tarea solicitada. Algunos sustentantes invirtieron tiempo en encontrarle forma a la figura (pájaro, nave, entre otras), perdiendo tiempo valioso para la realización de la prueba.

Se requiere pintar el contorno de la siguiente figura geométrica. ¿Cuántos metros habrá que pintar?



- A) 9.5 metros
- B) 8.2 metros
- C) 5.7 metros
- D) 3.75 metros

Figura 6. Problemas de diseño presentes en el ítem PMB08.

### Conclusiones

A modo de conclusión, se puede decir que se logró con éxito la evaluación del diseño universal de los ítems de la prueba de Matemáticas de primaria (06) del PLANEA ELCE 2015, al identificar, con base en evidencias convergentes resultantes del panel de expertos y de los protocolos verbales con estudiantes, que el 98% de los ítems de la prueba presentan problemas asociados a los elementos que componen el MEDU, propuesto por Thompson et al. (2002). También, se puede decir que fue un acierto el seguir las sugerencias de especialistas en el campo de la medición (Haladyna et al., 2002; Ercikan et al., 2010; Johnstone, 2003) en cuanto a complementar la implementación del MEDU con el uso de métodos cognitivos para identificar evidencias convergentes y causales de varianza irrelevante del constructo originada por problemas de diseño, alineación al constructo y sesgos culturales de los ítems de las pruebas en análisis.

Realizar estudios que aporten evidencias de validez y equidad para mejorar de la calidad técnica de las pruebas es una actividad obligada para los desarrolladores y los tomadores de decisiones, en especial, para que las pruebas cuenten con las evidencias

teóricas y empíricas que fundamenten las interpretaciones, consecuencias y usos de resultados de forma justa y ética en los distintos contextos y procesos de aplicación (AERA, APA y NCME, 2014; Messick, 1989). En conclusión, se puede decir que se logró recabar información para que los desarrolladores de las evaluaciones nacionales mejoren sus procesos de diseño y los tomadores de decisiones del Sistema Educativo Nacional (SEN) puedan interpretar de forma más justa y correcta los resultados del PLANEA ELCE.

## Referencias

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. Washington, D. C.: American Educational Research Association.
- Anderson, D., Lai, C., Alonzo, J., & Tindal, G. (2011). Examining a Grade-Level Math CBM Designed for Persistently Low-Performing Students. *Educational Assessment, 16*(1), 15-34.
- 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.
- Dolan, R.P., Hall, T.E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment, 3*(7).
- Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice, 29*(2), 24-35.
- Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis: verbal reports as data*. Cambridge: MIT Press.
- Ericsson, K.A., & Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Gierl, M. J., & Lai, H. (2018). Using Automatic Item Generation to Create Solutions and Rationales for Computerized Formative Testing. *Applied Psychological Measurement, 42*(1), 42–57.

- 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*. New York: College Board.
- Gobierno Federal (2013). Decreto por el que se expide la Ley del Instituto Nacional para la Evaluación de la Educación. *Diario Oficial de la Federación, DOF 11-09-13*. Recuperado de [http://www.dof.gob.mx/nota\\_detalle.php?codigo=5313842&fecha=11/09/2013](http://www.dof.gob.mx/nota_detalle.php?codigo=5313842&fecha=11/09/2013)
- 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.
- Instituto Nacional para la Evaluación de la Educación (INEE). (2015). *Plan Nacional para la Evaluación de los Aprendizajes (PLANEA)*. México: INEE. Recuperado de <http://planea.sep.gob.mx/content/general/docs/2015/PlaneaDocumentoRector.pdf>
- Instituto Nacional para la Evaluación de la Educación (INEE). (2017). *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*. México: INEE.
- Instituto Nacional para la Evaluación de la Educación (INEE). (2018). *Plan Nacional para la Evaluación de los Aprendizajes (PLANEA)*. México: INEE.
- Johnstone, C. (2003). *Improving validity of large-scale tests: Universal design and student performance (Technical Report 37)*. Minneapolis: National Center on Educational Outcomes.
- Ketterlin-Geller, L.R. (2005). Knowing what all students know: Procedures for developing universally designed assessments. *Journal of Technology, Learning, and Assessment, 4*(1).
- Leighton, J.P. (2004). Avoiding Misconception, Misuse, and Missed Opportunities: The Collection of Verbal Reports in Educational Achievement Testing. *Educational Measurement: Issues and Practice, 23*(4), 6-15.
- Leighton, J.P. (2017). *Using think-aloud interviews and cognitive labs in educational research*. New York: Oxford University Press.

- Leighton, J., & Gierl, M. (2007). Verbal Reports as Data for Cognitive Diagnostic Assessment. En J. Leighton, & M. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education* (pp. 146-172). USA: Cambridge University Press.
- Mace, R. (1985). Universal Design, Barrier Free Environments for Everyone. *Designers West*, 33(1), 147-152.
- Ma, L. Çetin, E., & Green, K. (2009). *Cognitive assessment in Mathematics with the Least Squares Distance Method*. Congreso anual de la AERA 2009, San Diego.
- Martínez-Rizo, F. (Coord.). (2015). *Las pruebas ENLACE y EXCALE. Un estudio de validación*. México: INEE.
- Messick, S. (1989). Validity. En R.L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.
- Nichols, P. D., Kobrin, J.L., Lai, M., & Koepfler, J. (2016). The role of theories of learning and cognition in assessment design and development. En A.A. Rupp, & J.P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 297-327). Chichester, England: Wiley-Blackwell.
- Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies*, 41, 385-397.
- 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., Larrazolo, N., Backhoff, E., & 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.
- Pérez-Morán, J.C., Rojas, G., & Vázquez-Lira, R. (Coords.). (2018). *Análisis de validez y del diagnóstico cognitivo de las pruebas de Matemáticas de primaria (06) y secundaria (09) de PLANEA ELCE 2015. Reporte técnico*. México: INEE.
- Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer G., Martínez Rizo, F., Aylwin, M., & Wolff, L. (Eds.). (2008). *Las evaluaciones educativas que América Latina necesita*. Santiago: PREAL.

- Revuelta, J., & 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., & Ximénez, 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., Templin, J., & Henson, R. A. (2010). *Diagnostic assessment: Theory, methods, and applications*. New York: Guilford.
- Smith-Castro, V., & Molina, M. (2011). La entrevista cognitiva: Guía para su aplicación en la evaluación y mejoramiento de instrumentos de papel y lápiz. *Cuaderno metodológico*, *5*. Recuperado de [https://www.academia.edu/2175997/Entrevista\\_Cognitiva\\_2011\\_](https://www.academia.edu/2175997/Entrevista_Cognitiva_2011_)
- Snow, R., & Lohman, D. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational measurement* (pp. 263-331). New York: Macmillan.
- Sternberg, R. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Oxford: Lawrence Erlbaum.
- Thompson, S., Johnstone, C., & Thurlow, M. (2002). *Universal design applied to large scale assessments (Synthesis Report 44)*. Minneapolis, MN: National Center on Educational Outcomes.
- You, X., Li, M., Zhang, D., & Liu, H. (2018). Application of a Learning Diagnosis System in Chinese Classrooms. *Applied Psychological Measurement*, *42*(1), 89–94.

Qartuppi, S. de R.L. de C.V. está inscrita de forma definitiva en el Registro Nacional de Instituciones y Empresas Científicas y Tecnológicas (RENIECYT) con el número 1600052.

Qartuppi, S. de R.L. de C.V. es miembro activo de la Cámara Nacional de la Industria Editorial Mexicana (CANIEM) con número de registro 3751.

*Apuntes de investigación educativa*

ISBN 978-607-98270-9-0

DOI 10.29410/QTP.19.05

Esta publicación digital se terminó de producir en abril de 2019.

Su edición y diseño estuvieron a cargo de:

**Qartuppi**<sup>®</sup>

Qartuppi, S. de R.L. de C.V.

<http://www.qartuppi.com>