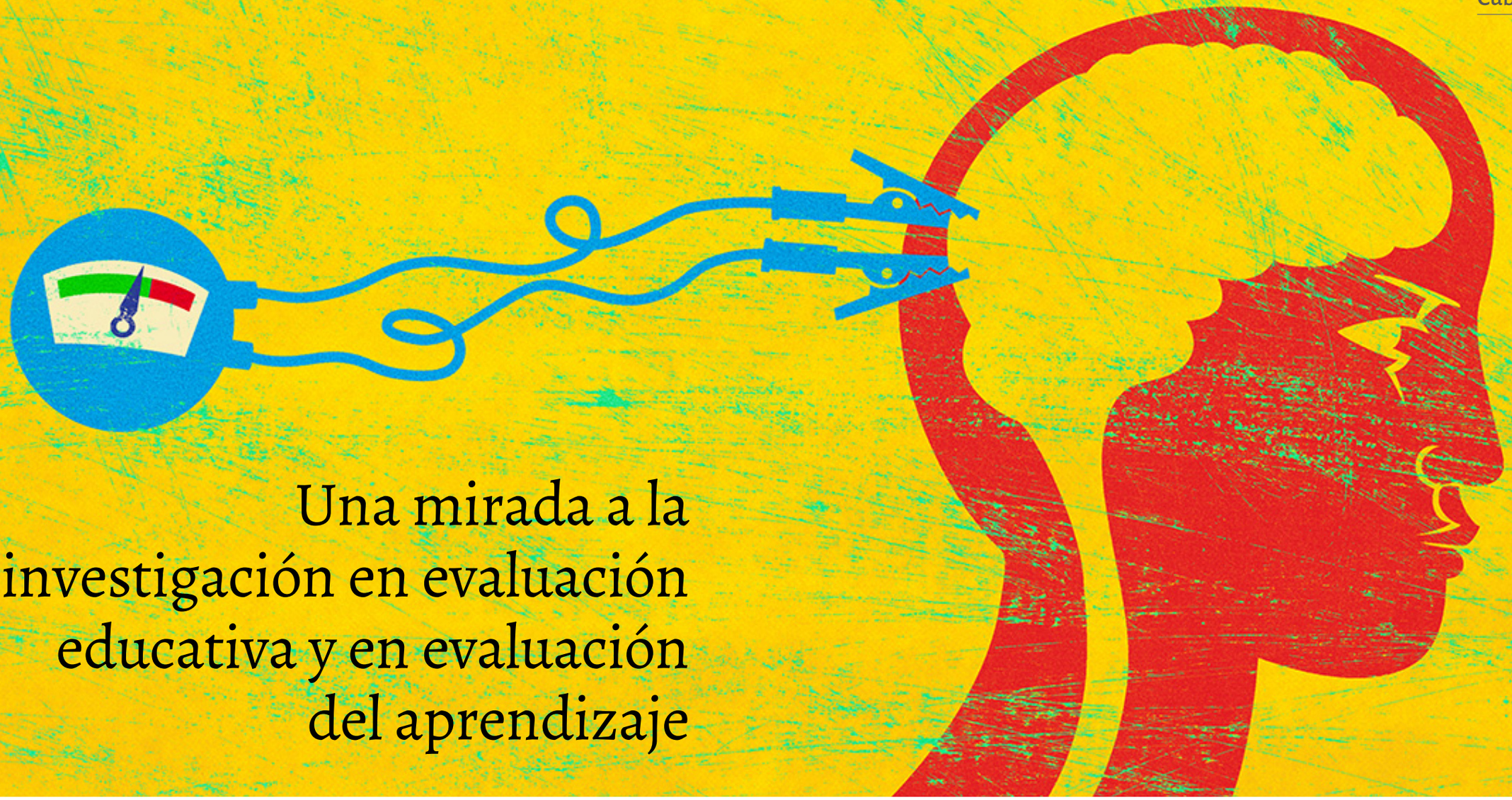


Una mirada a la investigación en evaluación educativa y en evaluación del aprendizaje

LUIS ÁNGEL CONTRERAS NIÑO

¿Por qué es importante la investigación en esta materia?

Existen muchas respuestas a esta pregunta, por ejemplo: porque buscamos procedimientos para mejorar nuestro programa educativo o nuestra práctica docente, o el aprendizaje de nuestros estudiantes. Estas y otras declaraciones implican la idea de cambio, lo que a su vez involucra costos en tiempo, dinero, esfuerzo, capacidad, voluntad, oportunidad y otros que, mientras más grandes, difíciles o importantes son, requieren de mayor y mejor evidencia para justificar el cambio.

Destaca el papel estratégico que pueden desempeñar ambos procesos como fundamento conceptual, metodológico y técnico para llevar a cabo los trabajos que realizan cotidianamente distintos actores que participan en el contexto educativo, en particular, como medio para facilitar la comunicación entre ellos, y como apoyo al trabajo docente y al aprendizaje de los estudiantes, así como a la asesoría y la actualización de profesores, la gestión escolar, la planeación educativa, la rendición de cuentas y la propia evaluación.

Toda acción educativa debe ser ejemplar, por lo que conviene que esté basada y sea orientada por las mejores prácticas derivadas de los resultados de la investigación, tanto los que se obtienen en la evaluación educativa en general como los que se logran en particular en la evaluación del aprendizaje.

Para tener una idea más amplia y precisa de lo que implican ambos ámbitos de la investigación, así como de las complejidades de las tareas que involucran, esbozaremos los principales componentes, focos evaluativos y procedimientos típicos para obtener resultados en la evaluación educativa y en la evaluación del aprendizaje.

Investigación en el ámbito de la evaluación educativa (IEE)

Existen muchos modelos que han orientado la IEE desde mediados del siglo pasado, por ejemplo, los desarrollos seminales de Tyler; Campbell y Stanley; Cronbach; Stufflebeam; Scriven; Stake; Suchman; Alkin; Guba y Lincoln; Provus; Eisner; Cook y Reichardt; y del *Joint Committee on Standards for Educational Evaluation*, entre otros citados por Stufflebeam (2001). Una síntesis de los principales elementos investigados se muestra en la figura 1.

Figura 1
Componentes que se investigan en evaluación educativa

En seguida, describiremos dichos componentes, e ilustraremos los principales focos evaluativos de interés en cada caso, y algunos procedimientos que comúnmente se utilizan para obtener resultados de la investigación evaluativa.

Los objetivos educativos: el interés se centra en indagar sobre la filosofía subyacente (concepciones de educación, del educando, de la actividad científica, etcétera), las políticas educativas (curriculares, hacia estudiantes y personal académico, líneas de investigación), el marco legal (leyes, estatutos, disposiciones administrativas), las necesidades sociales (planes de desarrollo, acceso a la cultura e información; actitudes y expectativas sociales), el avance científico y tecnológico de las disciplinas (cambios de paradigma, desarrollo de principios y métodos, integración teórica) y las necesidades de los educandos (perfiles de ingreso y egreso, ampliación e igualdad de oportunidades, mejoramiento de la calidad de vida, atención a la diversidad). Por lo general, estos asuntos se investigan mediante el análisis curricular, encuestas a actores sociales y educativos clave, análisis de contenido de documentos y entrevistas a autoridades y académicos destacados, entre otros procedimientos.

El programa educativo: incluye evaluar los componentes de las áreas de docencia (currículo; actuación, organización y formación docente; administración académica; atención a alumnos), investigación (apertura y consolidación de líneas de investigación; financiamiento y producción de la investigación), extensión y vinculación (intercambio académico, venta de bienes y servicios, eventos y publicaciones académicos) y administración (planeación, operación y evaluación del desarrollo institucional, apoyo administrativo, financiamiento). Usualmente tales componentes se investigan mediante técnicas de análisis curricular; la aplicación de entrevistas con académicos, funcionarios y estudiantes; el análisis de contenido de documentos, y la valoración de estados financieros y contables, entre otros.

La organización: se evalúan principalmente la estructura orgánica (funcionamiento organizacional, creación de estructuras, sistemas de información) y las disposiciones reglamentarias (manuales de funciones, manuales de procedimientos y relaciones, reglamentos, calendario escolar). Para obtener los resultados sobre estos componentes la investigación educativa utiliza, entre otros, entrevistas a funcionarios, docentes y alumnos, organización de paneles de expertos en administración y legislación educativas, técnica Delphi y el análisis de contenido de la documentación de referencia.

Los recursos: respecto éstos se evalúan los humanos (docentes, directivos, administradores, técnicos, personal eventual), materiales (aparatos, equipos, instalaciones, locales, materiales educativos), financieros (fuentes de sostenimiento, recursos propios, apoyos extraordinarios) y de información (libros, revistas, bases de datos, *software* educativo, acceso a Internet). Los resultados de la evaluación sobre estos elementos comúnmente se obtienen mediante inventarios, entrevistas con funcionarios y la aplicación de cuestionarios de opinión a alumnos, profesores, directivos y otros usuarios.

El contexto: se evalúan indicadores de impacto proximal (relación maestro-alumno; clima escolar; tipo, ubicación y tamaño de la escuela y de la comunidad en que se sitúa; estatus socioeconómico y cultural) e indicadores de impacto distal (adscripción a niveles y modalidades educativas, tipos de control, planes sectoriales, reformas educativas). Los resultados de la investigación del contexto educativo se obtienen empleando técnicas de triangulación, estudios etnográficos, análisis de contenido de documentos, estudios de casos, análisis de sistemas, entrevistas con informantes clave, análisis de datos secundarios y técnica Delphi.

Las relaciones: esta clase de IEE considera que los componentes no están aislados, sino que existen entre ellos múltiples vínculos

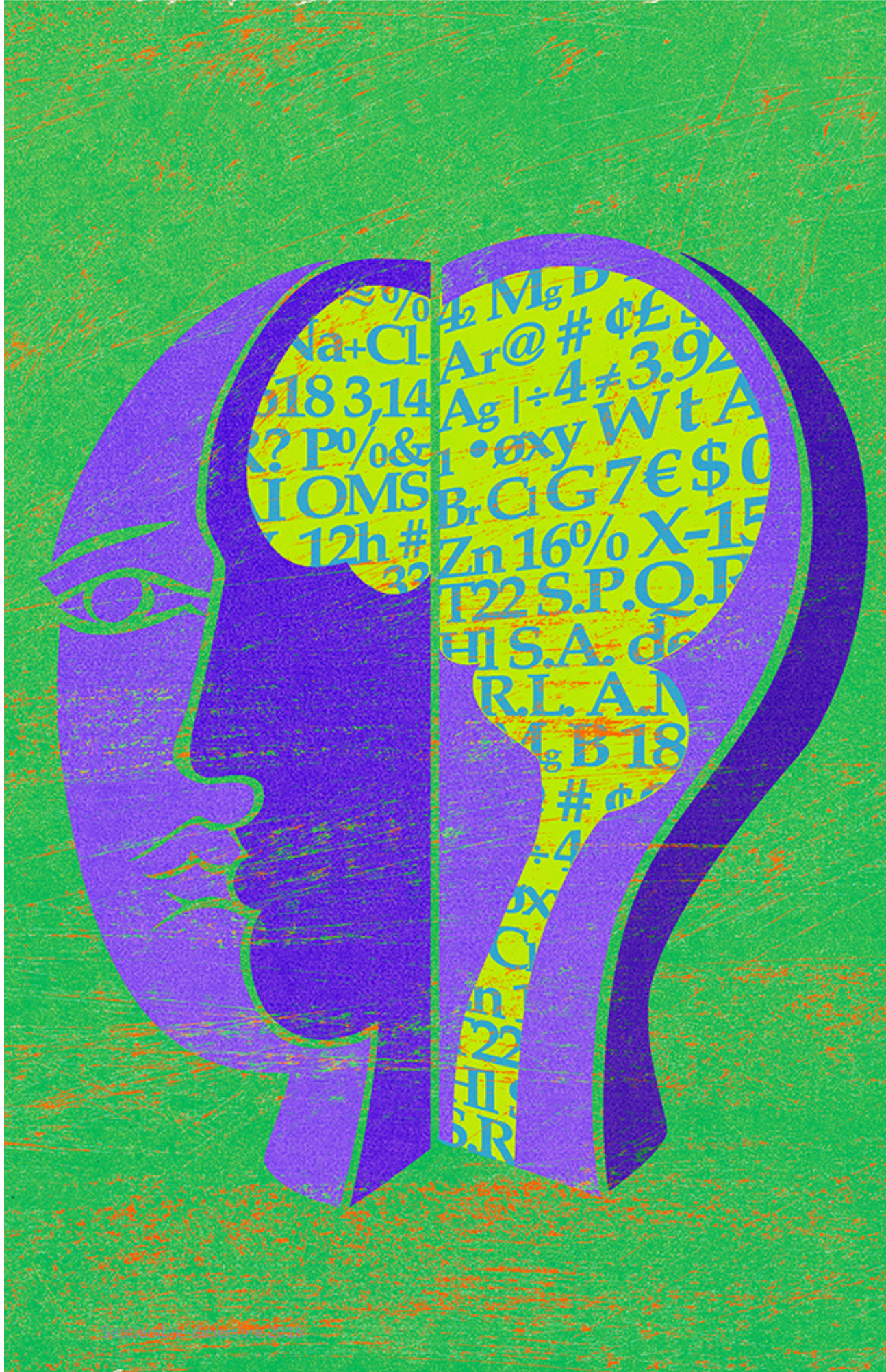
significativos que deben hacerse explícitos y valorarse. En su expresión más general, esto significa que las acciones de un programa educativo se orientan al logro de sus objetivos, y para ello requieren de una organización que las lleve a cabo, la cual a su vez demanda recursos para operar, y todo ello se realiza en un contexto que influye en el desarrollo de la agencia educativa. Los resultados de las investigaciones sobre el sistema de relaciones entre los diversos aspectos que constituyen el ámbito de la evaluación educativa usualmente se obtienen mediante estudios etnográficos, triangulación de instrumentos, análisis de sistemas, paneles de expertos, estudios de casos y análisis de contenido de documentos.

> La evaluación y sus resultados ejercen una influencia poderosa sobre procesos como la instrucción, el aprendizaje y el currículo.

Investigación en el ámbito de la evaluación del aprendizaje (IEA)

En el contexto de la IEA, ya sea la que se practica en el aula, pero sobretodo la que se aplica a gran escala, se tiene bien establecido que la evaluación y sus resultados ejercen una influencia poderosa sobre procesos como la instrucción, el aprendizaje y el currículo, lo que en el peor de los escenarios ha dado origen a metáforas con connotación negativa como que *se aprende lo que se evalúa*, o bien, que *la cola de la evaluación mueve al perro del currículum*, pero que en el mejor escenario posible implica el alineamiento virtuoso de currículo-evaluación-instrucción-aprendizaje, donde la evaluación actúa como interface que propicia y asegura la correspondencia. En consecuencia, la investigación en este ámbito se orienta tanto a buscar nuevos y mejores procedimientos para garantizar dicho alineamiento, como a aplicar los resultados de dicha indagación al desarrollo de instrumentos de evaluación que resulten válidos, o bien, a obtener evidencias de validez de los ya desarrollados —ver, por ejemplo: *Las pruebas Enlace y Excale. Un estudio de validación*.

Para los tres casos y para las modalidades de evaluación en el aula y la estandarizada, se han desarrollado estándares nacionales e internacionales que orientan la indagación y proporcionan criterios para el "desarrollo y evaluación de pruebas y prácticas de evaluación, así como guías para evaluar la validez de las interpretaciones de los puntajes en la prueba para los usos pretendidos del instrumento" (AERA, APA y NCME, 2014: 1). Las principales áreas de interés de la IEA, así como los componentes y procedimientos típicos que son empleados para



obtener evidencias de validez, confiabilidad, comparabilidad y justicia incluyen los siguientes:

- Alineación al referente de la evaluación. Aquí la IEA se centra en tres procesos:

    1) Planeación de la prueba, que incluye la determinación de lo que se va a evaluar y de los procedimientos y medios para hacerlo. En el primer caso, normalmente se trata de la adquisición de competencias o del dominio del currículo (aprendizaje de conceptos y procedimientos, desarrollo de actitudes, y otras metas especificadas en los objetivos del programa educativo). En el segundo caso, la indagación se enfoca en procedimientos que permitan identificar blancos curriculares de primer orden, determinar en qué medida los ítems representan el dominio a evaluar, y garantizar el alineamiento entre lo que se evalúa y lo que debe aprenderse. En ambos casos las evidencias se obtienen mediante el análisis de contenido de documentos; retículas, mapas conceptuales y otras estructuras que permiten representar los contenidos y sus relaciones de servicios; un documento de justificación del contenido a evaluar; la tabla de especificaciones del examen, y entrevistas con informantes clave, entre otras técnicas de análisis de información o análisis curricular.
    2) Diseño de especificaciones de ítems, que incluye evaluar las descripciones técnicas necesarias para construir las tareas evaluativas mediante las cuales se va a constatar el dominio curricular que fue determinado. La especificación de ítems detalla los atributos de los estímulos que es válido presentar en el ítem, y los atributos de las respuestas que se ofrecerán como opciones al examinado cuando lo responda. Comúnmente, incluye especificar información contextual relativa al contenido cuyo dominio se evalúa (ubicación en el programa, relaciones con contenidos previos y posteriores, importancia curricular y disciplinaria, definiciones de precisión), la estrategia para evaluar su dominio (qué y cómo evaluar, número de ítems, foco evaluativo de cada ítem), la especificación de elementos de apoyo para responder el ítem (instrucciones particulares para responder, características de la información tabular, gráfica o textual a emplear), la especificación del nivel de demanda cognitiva al cual se probará el dominio del contenido a evaluar (dimensión de proceso cognitivo y dimensión de conocimiento involucradas según alguna taxonomía), la especificación de los componentes del ítem (características de la base del ítem, de la respuesta correcta, de los distractores) y la elaboración de un ítem muestra que ilustre la correcta aplicación de los aspectos especificados, entre otros posibles componentes de la especificación. Estos tipos de evidencias se obtienen mediante el análisis de los protocolos de evaluación que utilizan paneles de expertos integrados por docentes en servicio, psicómetras, expertos disciplinarios, representantes de culturas diversas y otros actores más, quienes son capacitados para formular juicios sobre la exactitud, la completitud y la correspondencia de los elementos de la especificación de ítems elaborada, respecto a los componentes de la planeación de la prueba que les dieron origen.
    3) Elaboración y validación de ítems, las cuales comprenden la evaluación de tres componentes: el cumplimiento de los requisitos del proceso (integrar un comité de especialistas en la disciplina, docencia, teoría cognitiva, medición, lingüística y representantes de grupos con potencial afectación por sesgo y ofensa; la capacitación del comité para producir ítems efectivos; la elaboración de un manual técnico de redacción de ítems y de formatos para consignarlos), el propio diseño de las tareas evaluativas que se especificaron para integrar la prueba (elaboración del ítems de conformidad estricta con su especificación y con apego a las normas técnicas de redacción de ítems), y la validación de ítems (integrar un comité de especialistas con un perfil similar al anterior; elaborar un manual técnico para la evaluación de ítems y de protocolos de evaluación, capacitar al comité para juzgar el alineamiento de cada ítem con la correspondiente especificación que lo produjo y con los productos de la planeación de la prueba, y para detectar y corregir errores conceptuales o de redacción técnica o sesgos y ofensa o complejidad cognitiva no especificada).

- Aplicación de la prueba, la cual incluye dos procesos:

    1) Aplicación piloto de ítems y versiones de la prueba, sobre la cual se evalúan la estructuración de la prueba (ensamblaje

de ítems en versiones de la prueba, contrastación de ítems y versiones contra la tabla de especificaciones del examen para garantizar su integridad y equivalencia), definición de la muestra para la aplicación piloto (tipos de examinados y tamaño de la muestra de ensayo), aplicación piloto del instrumento (selección y capacitación de aplicadores, manual para la capacitación, protocolo de actuación *in situ*, documentación para el acceso), revisión lógica y empírica de resultados de la aplicación piloto (análisis de incidencias de la aplicación, estructuración de la base de datos con resultados de la prueba, análisis psicométricos de ítems y versiones, nueva revisión lógica y empírica de la representatividad curricular de ítems y versiones por comité de validación, estructuración final de versiones de la prueba con variables de contexto a explorar).

2) Aplicación de la prueba a la población destinataria, sobre la cual se evalúan las condiciones para la aplicación estandarizada de la prueba (edición, impresión, transporte, acopio y resguardo seguros de cuadernillos y hojas de respuestas), la definición de la población a evaluar (tipos de examinados, tamaño y distribución de la población), la aplicación estandarizada del instrumento (selección y capacitación de aplicadores, manual para la capacitación, protocolos de actuación *ex ante*, *in situ* y *post facto*, documentación para el acceso al campo y el registro de incidencias, monitoreo de la aplicación).

- Análisis de resultados y elaboración de reportes a destinatarios de la información, procesos cuya evaluación se enfoca en cinco aspectos:

  1) Preparación de los datos del examen para su análisis (lectura de hojas de respuestas; estructuración, depuración y validación de base de datos a partir de los resultados de la aplicación y con base en la estructura del examen).
  2) calificación del examen (uso de rúbricas y claves de respuestas; establecimiento de estándares de ejecución y puntos de corte para interpretar la ejecución en el examen).
  3) Análisis psicométricos de ítems y versiones de la prueba (cálculo de la dificultad, poder discriminativo, funcionamiento de distractores y funcionamiento diferencial de ítems; análisis de conducta de copia; cálculo de la confiabilidad de las versiones de examen, validez de constructo y de criterio, estimación de error de medida, etcétera).
  4) Análisis de la ejecución de los examinados en el examen (total de aciertos en la prueba, aciertos por área o dominio curricular evaluado, resultados según variables de contexto exploradas).
  5) Elaboración de reportes de resultados (determinar los usuarios del examen y la información apropiada a sus necesidades; preparar textos, tablas y figuras que sean suficientes, claros y significativos; preparar argumentos y evidencias empíricas basados en los resultados de la prueba que respalden las consecuencias y los usos previstos de los resultados así como señalar los no previstos; elaborar reporte de resultados por destinatario y hacérselo llegar; apoyar a usuarios para interpretar y utilizar correctamente los resultados del examen).

Cabe señalar que las acciones y el análisis descritos corresponden a conceptualizaciones, principios y procedimientos que provienen de la teoría clásica de los test y de la teoría de la respuesta al ítem.

> Una asignatura pendiente de la investigación sobre la evaluación del aprendizaje es desarrollar una teoría de la medición centrada en los propósitos y usos en el aula

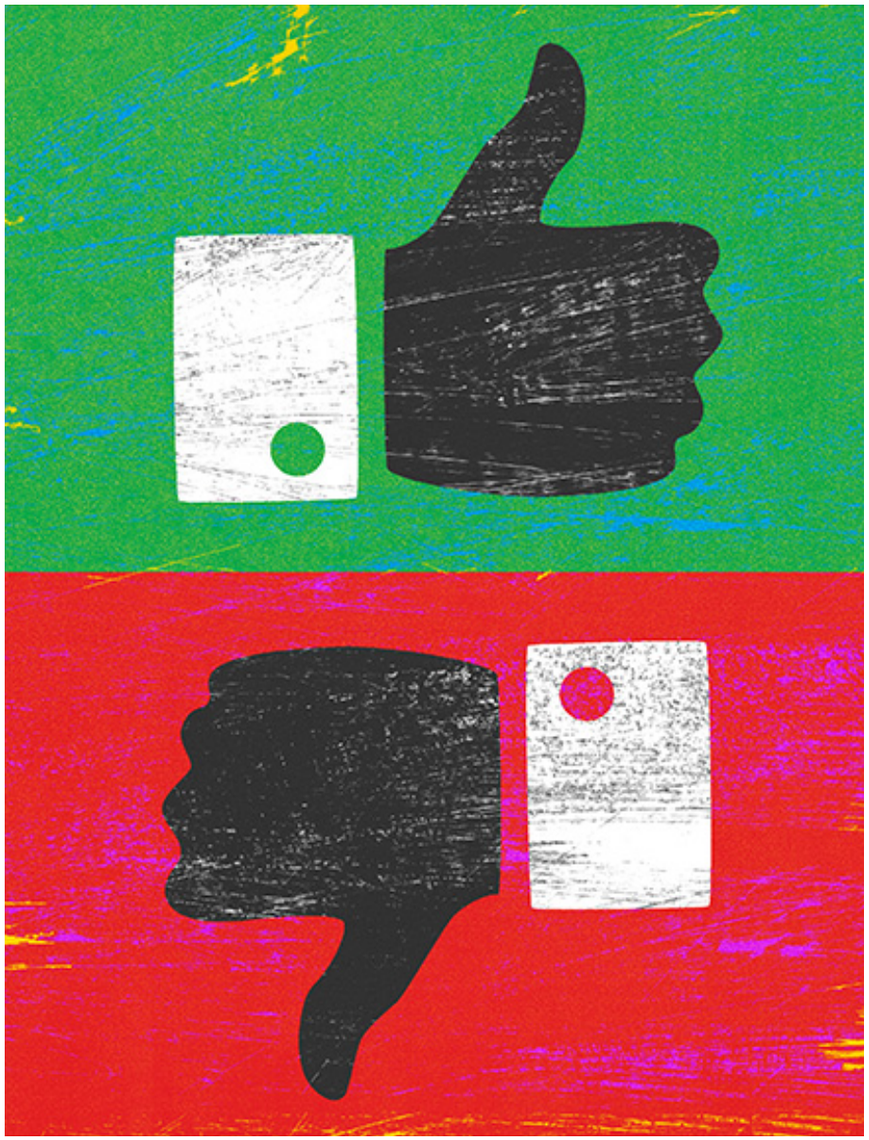
Aunque la indagación sobre las áreas de interés, los componentes y los procedimientos antes mencionados hace mucho más sentido para el caso de una prueba estandarizada que se aplica a gran escala, no obstante, también pueden ser referentes significativos para el caso de un examen departamental o una prueba de aula, ya sea que se estructuren con ítems de selección o ítems de producción. Al respecto, cabe señalar que la IEA no sólo busca descubrir nuevas formas de evaluar válidamente el aprendizaje de los estudiantes, o aplicar sus resultados para desarrollar instrumentos específicos o para evaluarlos, sino que también investiga la manera de generalizar y aplicar los resultados de la indagación en contextos auténticos para mejorar la calidad de la evaluación del aprendizaje. Es decir, en todo caso la indagación se orienta a buscar y promover la obtención de evidencias de validez de las pruebas, aquellos tipos de evidencias

cuyo estatus ha quedado bien establecido por la investigación evaluativa. De hecho, muchos de estos elementos pueden ser utilizados como guías de actuación, y ser traducidos y adaptados al contexto del aula. No obstante, es claro que una asignatura pendiente de la investigación sobre la evaluación del aprendizaje es desarrollar una teoría de la medición centrada en los propósitos y usos en el aula; una "teoría *aulamétrica*" como la denomina Susan Brookhart (2003).

La investigación en evaluación educativa y la evaluación del aprendizaje en los contextos internacional y nacional

Es necesario hacer notar que la mayor parte de la IEE y de la IEA se realiza en países desarrollados que cuentan con una gran tradición, que pueden financiarla a gran escala y que tienen gran capacidad para difundir sus resultados en libros, revistas, eventos, páginas web y otros medios. Ya se mencionó que en el caso de la IEE los principales modelos surgieron a mediados del siglo pasado, y en el caso de la IEA los primeros desarrollos relevantes se iniciaron desde de principios del siglo XX.

En México, el surgimiento de ambos tipos de investigación evaluativa es relativamente reciente. Martínez Rizo y Blanco (2012) documentan antecedentes



de dichas actividades a partir de la década de 1970. Sin embargo, no es sino hasta el inicio del presente siglo que se ha dado un vertiginoso desarrollo de organizaciones e instituciones que llevan a cabo proyectos en las áreas de IEE y IEA; entre ellas destacan las siguientes:

- En el ámbito de la Secretaría de Educación Pública, la Dirección General de Evaluación de Políticas (DGEP, antes Subdirección de Evaluación y Acreditación), cuyo foco principal ha sido la evaluación del aprendizaje a gran escala en el ámbito nacional. Esta dependencia ha desarrollado varias pruebas importantes, como la del Factor de Aprovechamiento Escolar y la de Estándares Nacionales, pero sin duda las más relevantes por su poderoso impacto fueron las del Examen Nacional del Logro Académico en Centros Escolares (ENLACE), que se aplicaron a cerca de 100 millones de niños de la educación básica nacional durante el periodo 2006-2013 (véase: http://www.enlace.sep.gob.mx/ba/aplicacion/).
- El Centro Nacional para la Evaluación de la Educación Superior (CENEVAL), asociación civil que desarrolla desde 1994 múltiples proyectos evaluativos como los Exámenes Nacionales de Ingreso (EXANI), los Generales para el Egreso de Licenciatura (EGEL), los diagnósticos de licenciatura, los de certificación de competencias laborales y los que desarrolla para la educación básica y media superior, entre otros. En la sección para académicos e investigadores de su página web (http://www.ceneval.edu.mx/) están disponibles varios estudios y publicaciones que ha realizado.
- Consejo Mexicano de Investigación Educativa (COMIE), asociación civil creada en 1993 que aglutina a cerca de 500 investigadores educativos en nuestro país, y entre ellos a quienes llevan a cabo IEE y IEA. Dicha organización tiene como objetivo promover la investigación educativa, y cuenta con cuatro fuentes de información importantes, tanto sobre los resultados de la investigación educativa como en materia específica de IEE y de IEA: La *Revista Mexicana de Investigación Educativa*, la colección Investigación Educativa en México 1992-2002, la colección general y las memorias del Congreso Nacional de Investigación Educativa que organiza bianualmente desde 2003 (Ver publicaciones).
- Instituto Nacional para la Evaluación de la Educación (INEE), organismo público autónomo creado en 2002 que tiene como propósito evaluar la calidad, el desempeño y los resultados del Sistema Educativo Nacional en la educación básica y la media superior, y que para ello evalúa los componentes, procesos y resultados del sistema, establece lineamientos sobre las funciones de evaluación a que deben sujetarse autoridades educativas federales y locales, genera y difunde información, y emite directrices para mejorar la calidad de la educación y su equidad, además de coordinar el Sistema Nacional de Evaluación Educativa. Entre los múltiples proyectos de IEA que ha desarrollado, destacan los Exámenes de la Calidad y el Logro Educativo (EXCALE), que se aplicaron durante el periodo 2005-2014 (ver su amplia e importante colección de publicaciones).

- También existen centros, departamentos y áreas que se especializan o desarrollan IEE y IEA en las universidades públicas y privadas, así como en secretarías de Educación de los gobiernos estatales. Varias de esas instituciones se han asociado para publicar los resultados de sus investigaciones, como en el caso de la *Revista de Evaluación Educativa* (*REVALUE*, disponible aquí), radicada en la Universidad Anáhuac, o bien lo han hecho por sus propios medios y con propósitos más generales, como en los casos de la *Revista Electrónica de Investigación Educativa de la Universidad Autónoma de Baja California* (*REDIE*, disponible aquí), la *Revista de Investigación Educativa de la Universidad Veracruzana* (CPU-e, disponible aquí) y la revista *Perfiles Educativos de la Universidad Nacional Autónoma de México* (disponible aquí), por citar sólo algunas.

Para que en nuestro país continúen desarrollándose a corto y mediano plazos IEE y IEA rigurosas y con alta calidad técnica, hay desafíos considerables y dos retos prioritarios: la formación de investigadores de alto nivel y alta especialización, lo que ya empezó a darse en algunas universidades, y, en relación con ello e igualmente relevante, el financiamiento de la investigación evaluativa, sobre todo para llevarla a cabo fuera del contexto institucional del reducido grupo de dependencias y universidades donde se practica. De ahí la importancia de que existan convocatorias como la del Fondo Sectorial de Investigación para la Evaluación de la Educación, lanzada recientemente por el INEE y el Consejo Nacional de Ciencia y Tecnología (CONACYT). Al respecto urge estimular en nuestro país, más de lo que se ha hecho, el interés por la investigación y el desarrollo de la cultura evaluativa, sobre todo entre los actores que tienen a su cargo la formación de los estudiantes.

Referencias bibliográficas

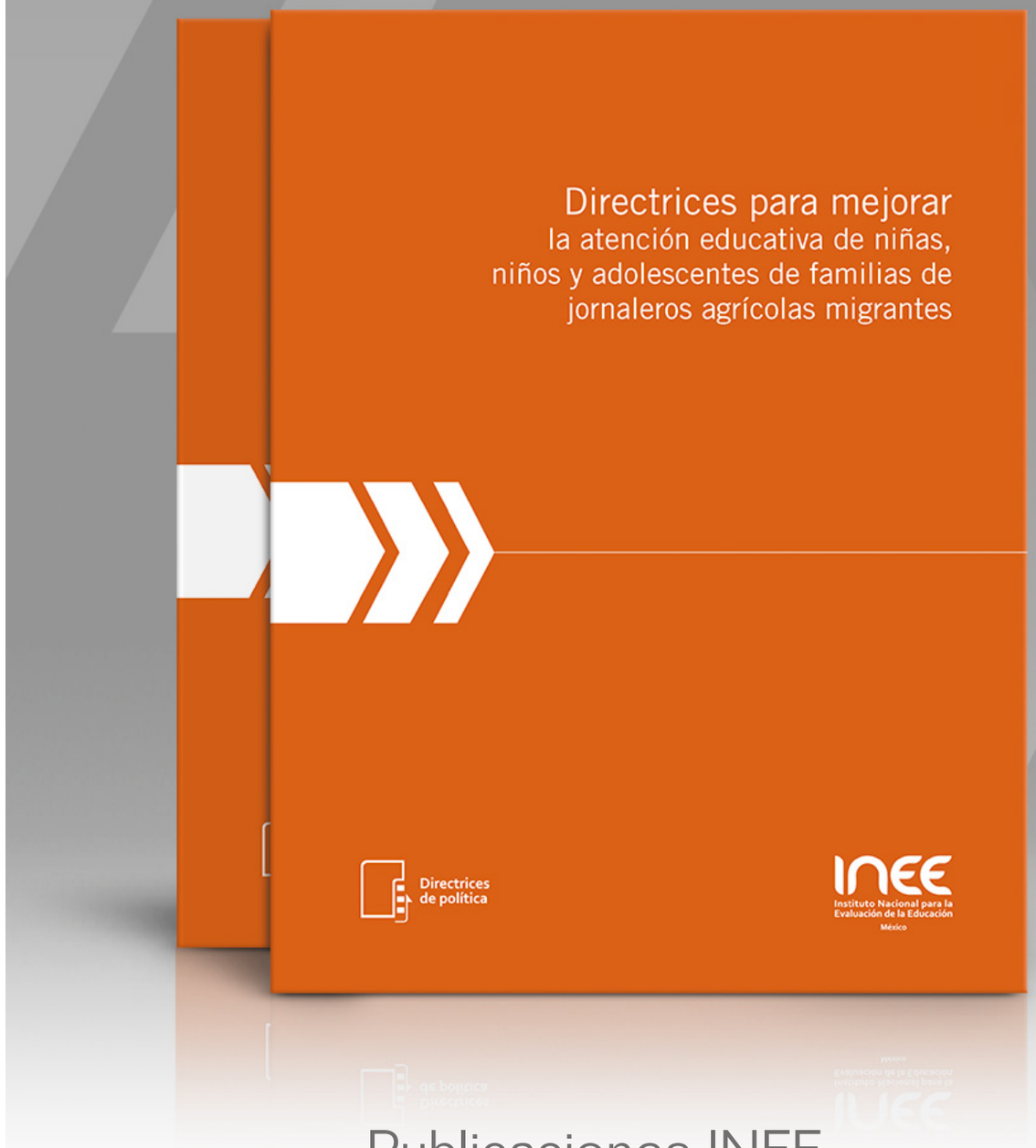
AERA, APA y NCME. American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, D. C.: American Educational Research Association.

Brookhart, S. M. (2003), Developing Measurement Theory for Classroom Assessment Purposes and Uses. *Educational Measurement: Issues and Practice*, *22*, pp. 5-12.

Martínez Rizo, F. y Blanco, E. (2012). La evaluación educativa. Experiencias, avances y desafíos. En Ordorica, M. y Prud'homme, J. F. (coords.). *Los grandes problemas de México. Edición abreviada* (pp. 145-148). México: El Colegio de México.

Stufflebeam, D. L. (2001). Evaluation Models. *New Directions for Evaluation*, *89*.

Propuestas de distintos actores clave, expertos en el tema, y representantes de organizaciones de la sociedad civil.



Publicaciones INEE