

VALOR EDUCATIVO Y FACTIBILIDAD DE LA EVALUACIÓN DE LA CALIDAD DE LA ENSEÑANZA MATEMÁTICA EN LA EDUCACIÓN PRIMARIA MEXICANA

CLAUDIA ZÚÑIGA GASPAR / JOSÉ LUIS CORTINA MORFÍN

Resumen:

Este trabajo reporta los resultados de una investigación cuyo objetivo fue explorar la utilidad y factibilidad de evaluar la enseñanza de las matemáticas en escuelas primarias mexicanas. Se evaluaron 40 sesiones de clase, de 20 diferentes aulas ubicadas en tres escuelas distintas de la Ciudad de México. El instrumento utilizado fue el de Evaluación de la Calidad de la Enseñanza (*Instructional Quality Assessment*). Los resultados muestran que este tipo de evaluaciones tienen un gran potencial para informar el diseño de acciones de mejora educativa, tanto a nivel sistémico como de un centro escolar específico. Se concluye que sí son factibles, una vez que su valor educativo compensaría los costos y limitaciones significativos, implícitos en su aplicación.

Abstract:

This study reports on the results of research aimed at exploring the usefulness and feasibility of evaluating the teaching of mathematics in Mexican elementary schools. An evaluation was made of forty class sessions in twenty classrooms in three schools in Mexico City. The instrument used was the Instructional Quality Assessment. The results show that this type of evaluations has great potential in determining the design of actions in educational improvement, at both the systemic and specific school level. The conclusion is that they are feasible, since their educational value would compensate for the significant costs and limitations that are implicit in their use.

Palabras clave: evaluación educativa, enseñanza de las matemáticas, calidad de la educación, instrumentos de evaluación, educación básica, México.

Keywords: educational evaluation, mathematics teaching, quality of education, evaluation instruments, elementary education, Mexico.

Claudia Zúñiga Gaspar: estudiante del Doctorado Interinstitucional en Educación en la Universidad Iberoamericana-Ciudad de México. Prolongación Paseo de Reforma 880, Lomas de Santa Fe, 01219, Ciudad de México, México.
CE: clauzu@msn.com

José Luis Cortina Morfín: profesor titular en la Universidad Pedagógica Nacional, unidad Ajusco, México.
CE: jcortina@upn.mx

Introducción

Está ampliamente aceptado que las evaluaciones –además de cumplir con estándares de validez técnica y confiabilidad– deben ser útiles (AERA, APA y NCME, 2014; Herman y Choi, 2012). Esto implica que sus resultados puedan ser utilizados para cuestiones como identificar el cumplimiento de estándares y objetivos, diseñar acciones para la mejora y monitorear el progreso. Darling-Hammond, Herman y Pellegrino *et al.* (2013) emplean el término *valor educativo* para denominar a esta característica de una evaluación en el campo de la educación.

La evaluación ha distinguido el caso del sistema educativo mexicano durante el presente siglo, convirtiéndose en una prioridad de la política educativa (Santiago *et al.*, 2012). Se espera que la instrumentación de evaluaciones permita la identificación de problemáticas así como el monitoreo de los esfuerzos que se realicen para resolverlas; asimismo, que las evaluaciones sirvan de referente tanto sobre el tipo de prácticas en las que se quiere se involucren docentes y directivos como acerca del perfil con el que éstos deben de cumplir para ejercer su profesión y, finalmente, que la evaluación sea un medio de rendición de cuentas, que favorezca el empoderamiento de la sociedad civil respecto de los diferentes actores del sistema educativo (INEE, 2013).

Hasta ahora, los esfuerzos se han orientado a evaluar el desempeño de los alumnos –principalmente en lengua escrita y matemáticas–, además de diferentes aspectos del perfil profesional de los profesores. Al parecer, tanto las autoridades educativas como diferentes actores de la sociedad civil le reconocen valor educativo a este tipo de evaluaciones. En otras palabras, le consideran de utilidad para impulsar el mejoramiento de la calidad de la educación que se imparte.

Algo que no ha sido priorizado en el sistema educativo mexicano es la evaluación del acto educativo mismo, la enseñanza, en el lugar y en el momento en que ocurre. Ello, probablemente debido a las complejidades técnicas y logísticas que representa evaluar la enseñanza de manera válida, confiable y útil.

En el presente artículo reportamos los resultados de un estudio que tuvo como principal objetivo indagar sobre el valor educativo de las evaluaciones de la calidad de la enseñanza matemática, en educación primaria, del sistema educativo mexicano. La finalidad fue ayudar a determinar su *factibilidad*, a través de recabar evidencia que pudiera servir para hacer un

balance entre el potencial de este tipo evaluaciones para procurar la mejora en el sistema y los costos y limitaciones que implicaría su instrumentación (Herman y Baker, 2005).

Comenzamos el artículo enmarcando nuestra investigación en los consensos internacionales respecto del aprendizaje matemático escolar; en particular, los que señalan que los alumnos deben comprender conceptos y desarrollar habilidades para resolver problemas, pensar lógicamente, comunicar ideas y argumentarlas (cf., NCTM, 1989; OCDE, 2003; SEP, 2011). Destacamos, además, el importante papel que desempeña la enseñanza en lograr estos objetivos, dado que es el aspecto del fenómeno educativo escolar con más capacidad de influir en el aprendizaje de los alumnos (Hiebert y Grouws, 2007).

Después, presentamos los resultados de una revisión de seis instrumentos para evaluar la calidad de la enseñanza matemática. De esta revisión se desprende que la aplicación a gran escala de este tipo de instrumentos puede ser relativamente costosa, además de que tiene limitaciones importantes, relativas a que la validez y confiabilidad de los resultados sean insuficientes.

En la parte final del artículo analizamos los resultados del estudio, en el cual se evaluaron 40 lecciones, impartidas en 20 aulas diferentes de tres escuelas primarias distintas. El análisis devela el gran potencial que, en términos de valor educativo, pueden tener las evaluaciones de la enseñanza matemática en el sistema mexicano de educación primaria. Con base en estos resultados, concluimos que este tipo de evaluaciones sí serían factibles, ya que su valor educativo compensaría los costos y limitaciones implicados en su aplicación. Sin embargo, destacamos tres cuestiones sobre las que sería importante saber más antes de que una evaluación de la calidad de la enseñanza de las matemáticas sea realizada a gran escala.

Aprendizaje matemático

Una de las áreas de aprendizaje más evaluadas en México ha sido las matemáticas. De acuerdo con tendencias internacionales, se les ha reconocido como un área de conocimiento esencial para la formación académica del alumnado así como para el desarrollo económico y democrático de las sociedades actuales (INEE, 2015; OCDE, 2003). Se considera que el aprendizaje matemático juega un papel muy importante en la formación de ciudadanos informados y reflexivos, que además sean consumidores inteligentes, y que cuenten con las destrezas necesarias para integrarse a las actividades productivas del mundo de hoy.

Hay que tener presente que el tipo de habilidades que los instrumentos recientes de evaluación tratan de medir no son las que típicamente se han asociado con el aprendizaje matemático en la escuela (cf., OCDE, 1999). Así, en pruebas internacionales como la del Programa Internacional para la Evaluación de los Estudiantes (PISA, por sus siglas en inglés; OCDE, 2014) y nacionales como el Plan Nacional para la Evaluación de los Aprendizajes (Planea; INEE, 2015), se le da poca importancia a evaluar el nivel de destreza con el que los alumnos ejecutan ciertos algoritmos, su capacidad de cálculo mental, y al conocimiento de definiciones y fórmulas. En cambio, se priorizan cuestiones como la capacidad de “analizar, razonar y comunicar ideas de un modo efectivo, al plantear, formular, resolver e interpretar problemas matemáticos en diferentes situaciones” (OCDE, 2003:27). En general, las pruebas estandarizadas actuales buscan ser consistentes con las tendencias internacionales hacia la educación matemática en las que se considera importante que los estudiantes comprendan los conceptos y que desarrollen habilidades que les permitan formular, representar y resolver problemas matemáticos, así como pensar lógicamente y justificar sus razonamientos (NCTM, 2014).

Tomando en cuenta estos antecedentes, los resultados de aprendizaje matemático del alumnado mexicano muestran un panorama por demás preocupante. Según lo reportado en la evaluación más reciente (INEE, 2015), el desempeño matemático de 60.5% de los alumnos de sexto grado fue insuficiente, lo que implica la existencia de deficiencias fundamentales en la formación de estos jóvenes, mismas que impactarán negativamente en sus oportunidades futuras de desarrollo académico y laboral y en su participación ciudadana. Esto con consecuencias previsibles para el desarrollo de la sociedad a la que pertenecen.

Es importante aclarar que, como ha sido ampliamente documentado, el desempeño de los estudiantes en matemáticas (al igual que en otras áreas como la lectura) está fuertemente relacionado con factores de índole económica, social y cultural, sobre los que la escuela prácticamente no tiene manera de incidir. Por ejemplo, Backhoff *et al.* (2007) calcularon que 66% de la varianza en los resultados de matemáticas de los estudiantes mexicanos, en una evaluación muestral y representativa, dependía de variables relacionadas con características socioculturales de sus madres.

Así pues, sería incorrecto atribuir directamente los pobres resultados del alumnado mexicano en matemáticas a la calidad de los servicios educativos

que ofrece el sistema educativo. Sin embargo, sí sería apropiado esperar que la mejora de esta situación se diera a través de acciones que tuvieran lugar en las aulas de las escuelas mexicanas.

Está bien documentado en la literatura que existen factores de índole netamente escolar, que tienen el potencial de influir significativamente en el aprendizaje matemático de los estudiantes. Entre éstos destaca la enseñanza que, cuando es de buena calidad, puede lograr que en el mediano plazo se reviertan las desventajas que pudiera implicar la situación económica, social y cultural de las familias de los alumnos (OCDE, 2009; Sanders y Horn, 1994).

Enseñanza de las matemáticas

Los académicos preocupados por el tema de la enseñanza matemática han considerado importante diferenciarla de otros aspectos del fenómeno educativo escolar que, si bien están relacionados con ella, de ninguna manera la agotan. Entre estos otros aspectos destacan el desempeño matemático de los alumnos, el currículo respectivo y el perfil de los docentes (cf., Hiebert y Grouws, 2007).

En relación con el desempeño matemático de los alumnos, como ya explicamos, aunque la enseñanza puede tener un papel central en él, existen múltiples factores de índole económica, social y cultural que tienen una fuerte relación con éste y que exceden el ámbito escolar. Por ello, no sería apropiado considerar que la calidad de la enseñanza que recibe un grupo de alumnos que se desempeña mejor que otro sería, necesariamente, superior (cf., Martínez Rizo, 2014).

En cuanto al currículo matemático, se reconoce que sin duda influye en las oportunidades de aprendizaje de los alumnos, simplemente porque aquello que no se enseña difícilmente se aprende. Sin embargo, el currículo no determina a la enseñanza; su instrumentación implica siempre un proceso de interpretación por parte del docente (Cobb, Zhao y Visnovska, 2008). Este proceso conlleva, entre otras cosas, decidir qué temas son más importantes que otros, cuánto tiempo se le debe dedicar a cada uno, cómo y cuándo interrogar a los alumnos y qué tipo de respuestas aceptar. En la práctica, un mismo currículo puede ser interpretado de formas diferentes y llevar a que sean muy distintas las experiencias educativas que tengan diversos grupos de estudiantes durante su aplicación.

En el caso del perfil docente, es razonable esperar que la instrumentación eficaz de ciertas formas de enseñanza requiera que los profesores

tengan una formación particular, cierto dominio del contenido, e incluso que cuenten con cierta experiencia frente a grupo. Sin embargo, no hay por qué esperar que maestros con perfiles parecidos se impliquen en formas de enseñanza similares; ni tampoco, que maestros con perfiles significativamente diferentes –en cuanto a su formación inicial o a sus años de servicio– enseñen de maneras muy disímiles. Para los especialistas en el campo, si bien el perfil incide en el tipo de enseñanza en que se implica un docente, de ninguna manera lo determina (cf., Hiebert y Grouws, 2007).

Para los propósitos de este artículo, se retoma la propuesta de Hiebert y Grouws (2007) de cómo entender la enseñanza matemática, por ser –en términos generales– consistente con los desarrollos existentes relativos a la evaluación de su calidad. Estos autores proponen entender esta enseñanza como un sistema que se constituye en las interacciones, entre los docentes y alumnos, en torno al contenido, cuyo fin es el logro de objetivos de aprendizaje matemático. Lo importante a destacar de esta propuesta es que hace que la evaluación de la calidad de la enseñanza tenga que ir más allá de evaluar las conductas del docente, para incluir lo que sucede como resultado de las interacciones entre los integrantes de un aula.

Metodología

La investigación que se reporta en el presente artículo se condujo en dos fases. La primera, de naturaleza documental, implicó identificar y examinar diversos instrumentos que, a nivel internacional, han sido desarrollados con el propósito de evaluar la calidad de la enseñanza matemática. El análisis se centró en sus aspectos generales y que serían importantes tomar en cuenta al hacer un balance de factibilidad, considerando tanto los costos que implicaría aplicarlos como las limitaciones implícitas en términos de la validez y confiabilidad de los resultados obtenidos. Además, esta fase sirvió para seleccionar el instrumento de evaluación que sería utilizado en el estudio empírico que se implementó en la segunda fase y que resultó ser el de Evaluación de la Calidad de la Enseñanza (*Instructional Quality Assessment*, IQA) (Matsumura *et al.*, 2006). Más adelante se especifican las razones que llevaron a elegirlo.

La segunda fase de la investigación consistió en la instrumentación de un estudio empírico en el que se evaluó la calidad de la enseñanza matemática en tres escuelas primarias urbanas.

Muestra

Se utilizó una muestra intencional (Cea D’Ancona, 2001) que implicó documentar la enseñanza en 20 aulas diferentes de tres escuelas primarias de la Ciudad de México. Dos eran públicas y una privada.

El criterio para seleccionar estas escuelas fue que las condiciones generales en que operaban permitieran la instrumentación de formas comunes de enseñanza matemática. Así pues, se buscaron aquellas que tuvieran un profesor por grupo y en las que ninguno estuviera a cargo de la operación administrativa de las escuelas. Además, que las aulas no tuvieran números desproporcionadamente altos de alumnos, y que no atendieran a cantidades importantes de estudiantes pertenecientes a poblaciones especiales (flotantes, o aquellas cuya primera lengua no fuera el español). Asimismo, que las escuelas contaran con la infraestructura y el mobiliario adecuados que permitieran el desarrollo favorable de las clases. Finalmente, que éstas no estuvieran involucradas en algún programa especial de enseñanza de las matemáticas. Algunas características de las primarias que conformaron la muestra se presentan en la tabla 1.

TABLA 1

Características generales de las escuelas primarias

Escuela	Turno	Total de grupos (1° a 6° grados)	Grupos por grado	Promedio de alumnos por grupo	Nivel socioeconómico de las familias de los alumnos
Pública 1	Matutino	17	3*	29.72	Medio bajo
Pública 2	Matutino	12	2	19.63	Medio bajo
Particular	Matutino	6	1	22.83	Alto

*Excepto el 6° grado que solo contaba con dos grupos.

La participación en el estudio tanto de las escuelas como de los docentes fue voluntaria. Se evaluaron dos sesiones consecutivas en cada una de las aulas. Originalmente se planeó evaluar a todas las aulas de tercero, cuarto, quinto y sexto grados, de las tres escuelas. Ello solo fue posible en dos

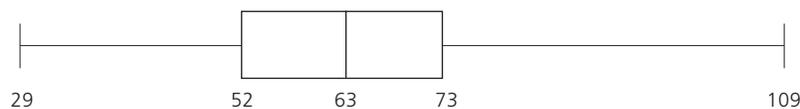
(pública 1 y particular); en la pública 2, únicamente aceptaron participar cinco de los ocho profesores cuyas aulas fueron seleccionadas. La tabla 2 presenta el número total de sesiones por grado que fueron incluidas en la muestra.

TABLA 2
Sesiones por grado que fueron videograbadas

Escuela	Tercer grado	Cuarto grado	Quinto grado	Sexto grado	Total
Pública 1	6 (3 grupos)	6 (3 grupos)	6 (3 grupos)	4 (2 grupos)	22
Pública 2	4 (2 grupos)	2 (1 grupo)	4 (2 grupos)	0 (0 grupos)	10
Particular	2 (1 grupo)	2 (1 grupo)	2 (1 grupo)	2 (1 grupo)	8
Total	12	10	12	6	40

La duración de las sesiones varió de manera significativa. La más larga fue de 109 minutos y la más corta, de 29. La mediana estuvo en 63 minutos (figura 1).

FIGURA 1
Distribución de la duración en minutos de las sesiones que formaron parte de la muestra (n=40)



En total participaron 20 profesores en el estudio (15 mujeres y 5 hombres). En promedio, tenían 20.2 años de servicio; el de menor tiempo, 6 años y el de mayor, 37 años. De los maestros, 7 estudiaron la educación normal a nivel bachillerato, de ellos, solo 3 continuaron con estudios superiores en áreas no directamente vinculadas con la educación primaria (por ejemplo, Historia); 9 eran licenciados en Educación primaria; y los 4 restantes tenían

una licenciatura en el área de Pedagogía. Finalmente, de los 20 profesores, solo dos contaban con estudios de posgrado.

Registro y codificación

Todas las clases fueron videograbadas por la primera autora de este artículo, durante los meses de marzo y abril del 2015. Dos evaluadores (los autores del presente artículo) codificaron cada una de las videograbaciones utilizando la matriz del IQA, de manera independiente. La coincidencia entre los evaluadores fue de 81%, con un coeficiente *kappa de Cohen* de .76; las discrepancias después fueron revisadas y la codificación definitiva fue consensuada.

Metodología de análisis

Dado que en el estudio se utilizó una muestra no representativa, el análisis de los datos se realizó sin la expectativa de que los resultados pudieran ser generalizables al sistema. En cambio, el objetivo principal del análisis fue reconocer el valor educativo de los datos obtenidos. Específicamente, se buscó conocer la utilidad de los datos para identificar el cumplimiento de estándares y objetivos relativos al tipo de enseñanza que se considera necesaria a nivel internacional para lograr que los alumnos desarrollen las habilidades matemáticas que pruebas de desempeño como PISA y Planea procuran evaluar, y que los propios programas de estudio oficiales de México retoman y proponen.

Dicho brevemente, se trata de un tipo de enseñanza compleja, que promueve la resolución de problemas y el razonamiento matemático, que involucra a los estudiantes en actividades que implican resolver, relacionar y argumentar, y que favorece que emerjan múltiples maneras de abordar los problemas y una variedad de estrategias de resolución. Fomenta que los alumnos analicen y argumenten diferentes formas de resolución, así como diversos aspectos de un concepto o de un procedimiento, además de que relacionen los distintos saberes que se van abordando (cf., NCTM, 2014; SEP, 2011).

También se buscó conocer la utilidad de los datos para identificar posibles diferencias en las oportunidades de aprendizaje matemático que pudieran existir entre una y otra aula, como resultado del tipo de enseñanza que se instrumentaba. Además, se buscó identificar aspectos específicos de la enseñanza que debieran ser tomados en cuenta en la elaboración de una agenda de mejora que tuviera como eje central el desarrollo profesional docente.

Tomando como base los objetivos planteados, los datos se analizaron en dos niveles. El primero implicó considerar que los resultados formaban parte de un único conjunto. Esto es, los datos se analizaron como si todas las aulas evaluadas conformaran un pequeño sistema educativo. El segundo nivel de análisis implicó analizar los resultados escuela por escuela.

Resultados de la primera fase del estudio

Como se mencionó, durante la primera fase de la investigación se revisó la literatura con el objetivo de identificar los principales instrumentos que han sido desarrollados para evaluar la calidad de la enseñanza de las matemáticas. En este análisis fue importante que en los instrumentos identificados se valorara el tipo de enseñanza matemática descrito anteriormente.

En la literatura se identificaron seis instrumentos que, en términos generales, cumplen con el perfil deseado. Éstos fueron desarrollados por diferentes equipos: cuatro por investigadores estadounidenses, uno por un grupo inglés, y otro por uno español (tabla 3). En todos ellos, las evaluaciones se realizan haciendo uso de matrices de evaluación o “rúbricas” (Martínez Rizo, 2014); esto es, tablas de doble entrada en las que se representan dos dimensiones. Un eje contiene las categorías a ser evaluadas y el otro, los niveles de desempeño en una escala ordinal.

TABLA 3

Instrumentos desarrollados para evaluar la calidad de la enseñanza matemática

	Instrumento	País
1	Instructional Quality Assessment (IQA; Evaluación de la Calidad de la Enseñanza; Matsumura <i>et al.</i> , 2006)	EUA
2	Mathematical Quality of Instruction (Calidad Matemática de la Enseñanza; Hill <i>et al.</i> , 2008)	EUA
3	Inside the Classroom (Dentro del Aula; Weiss <i>et al.</i> , 2003)	EUA
4	Instrumento de Evaluación de los Procesos Matemáticos en Educación Infantil (Alsina y Coronata, 2014)	España
5	Instrument for Ranking Classes (Instrumento para Clasificar Sesiones de Clase; Brown <i>et al.</i> , 2001)	Reino Unido
6	Scoop Notebook (Cuaderno de Recolección; Borko, Stecher y Kuffner, 2007)	EUA

Costos y limitaciones

Los instrumentos identificados fueron examinados con el propósito de reconocer en ellos aspectos que pudieran impactar tanto en los costos de aplicarlos, como en la posibilidad de que la validez y confiabilidad de los resultados obtenidos pudieran ser cuestionadas. En la revisión se reconoció que los seis instrumentos comparten una característica que hace que los costos de aplicarlos puedan ser relativamente elevados. Se trata del alto nivel de inferencia que requieren de parte de los evaluadores (Rosenshine, 1970).

Rosenshine (1970) plantea que los instrumentos que demandan altos niveles de inferencia se vuelven necesarios cuando lo que se evalúa son conductas complejas, como las que se espera que entren en juego en el tipo de enseñanza matemática que se considera necesaria en el mundo de hoy. Sin embargo, su aplicación puede ser relativamente costosa. Estos instrumentos requieren de evaluadores con perfiles especializados, quienes además deben ser cuidadosamente capacitados. Otra desventaja es que en su aplicación puede ser difícil obtener niveles satisfactorios de fiabilidad entre jueces. Esto puede llevar a que la validez y confiabilidad de los resultados sea cuestionada.

Por otra parte, se reconocieron dos características más, que cinco de los seis instrumentos comparten: requieren tener contacto directo con el fenómeno evaluado y usan como unidad de análisis la lección diaria. Ambas son relevantes en tanto que pueden incidir en los juicios de validez y confiabilidad que se hagan a los resultados obtenidos, como se explica a continuación.

Proximidad con el fenómeno

La enseñanza de las matemáticas es una actividad en la que participa una comunidad formada por un grupo de alumnos y, generalmente, un solo maestro, quienes cotidianamente conviven (Hiebert y Grouws, 2007). Colocar a un observador en el aula para evaluar directamente al fenómeno conlleva el riesgo de perturbar a la comunidad y, más importante, que la evaluación sea vista por el docente como algo invasivo y amenazante. Esto puede originar que lo que se observa no sea representativo de lo que cotidianamente acontece en el aula (Martínez Rizo, 2014) o, incluso, que los docentes busquen la manera de evitar que se realice la evaluación.

En el caso de los seis instrumentos revisados, el único que no requiere de un contacto directo con el fenómeno evaluado, ya sea a través de tener

evaluadores en las aulas, o de hacer videograbaciones de las lecciones, es el *Scoop Notebook* (Borko, Stecher y Kuffner, 2007). Este instrumento evalúa conjuntos de artefactos que formaron parte de una clase de matemáticas, a lo largo de una semana. Los artefactos son recolectados por los docentes en un cuaderno de notas e incluyen materiales generados antes, durante y después de cada clase. Evaluadores entrenados aplican una matriz de evaluación a estos cuadernos.

Lección diaria

Con excepción del *Scoop Notebook*, la segunda característica que comparten estos instrumentos de evaluación es que usan como unidad de análisis a la lección diaria. Según Hiebert y Grouws (2007), esta unidad tiene la ventaja de ser lo suficientemente grande como para que entren en juego distintos aspectos clave de la enseñanza matemática que tiene lugar en el aula, y lo suficientemente pequeña para poder ser analizada a plenitud. Sin embargo, también tiene la desventaja de que el aprendizaje matemático no ocurre de manera segmentada, lección por lección. Además, como lo señala Martínez Rizo (2014), se corre el riesgo de que la lección evaluada no sea representativa de la enseñanza que cotidianamente tiene lugar en el aula.

En síntesis, el análisis de los instrumentos de evaluación de la calidad de la enseñanza matemática mostró que su aplicación a gran escala puede ser relativamente onerosa y riesgosa. Por una parte, el nivel de inferencia que demandan hace que sea necesario contar con evaluadores bien formados y entrenados, lo que sin duda implica costos relativamente elevados. Además, tienen limitaciones importantes en términos de que, en última instancia, los resultados puedan no ser considerados lo suficientemente válidos y confiables, ya sea por la perturbación que causa la presencia de los observadores en las aulas, porque exista un bajo nivel de fiabilidad entre jueces, o porque se cuestione la representatividad de los datos obtenidos.

Selección del instrumento de evaluación

Un segundo propósito de la revisión de la literatura que se realizó fue elegir el instrumento que se utilizaría en el estudio empírico. El instrumento seleccionado fue el IQA; fue diseñado para medir la calidad de la enseñanza de las matemáticas y del inglés (como primera lengua). Cada materia se mide por separado (cf., Boston y Wolf, 2006; Matsumura *et al.*, 2006). En el caso de matemáticas, el instrumento se enfoca en la demanda cognitiva,

considerándola como un factor clave para lograr que los alumnos alcancen niveles más altos de desempeño en matemáticas (Stein y Lane, 1996).

En general, el IQA es consistente con el tipo de enseñanza matemática ya descrito; valora el uso de actividades que implican la resolución de problemas y el razonamiento matemático, así como la discusión colectiva tanto de ideas matemáticas, como de formas de resolver problemas.

La decisión de usar el IQA en el estudio empírico tomó como criterio principal de validez la aceptación que ha tenido por parte de la comunidad de investigadores. Fue el único de los seis instrumentos cuyo uso había trascendido al grupo de investigación que originalmente lo desarrolló.

El IQA ha sido utilizado no solo en estudios realizados por sus desarrolladores (Boston, 2012), se ha empleado en el proyecto denominado *Middle-school Mathematics and the Institutional Setting of Teaching* (cf., Cobb *et al.*, 2013), para evaluar la calidad de la enseñanza matemática en cuatro distritos educativos de Estados Unidos. Además, actualmente está siendo utilizado, también a gran escala, en un proyecto encabezado por investigadores de la Universidad del Centro de Florida (cf., Childs *et al.*, 2016).

El IQA fue traducido al español. Se solicitó a dos académicos, independientes del proyecto de investigación, y con amplia trayectoria en el estudio de la enseñanza matemática en México, que dieran su opinión sobre la pertinencia de aplicar el instrumento en escuelas del sistema educativo mexicano y que revisaran su traducción. Ambos consideraron que era pertinente y sugirieron algunas modificaciones menores a la versión en español.

La matriz del IQA utiliza cuatro categorías: potencial de la actividad de enseñanza que se plantea, instrumentación de la actividad de enseñanza, conversación de los estudiantes a raíz de la actividad de enseñanza, y expectativas del profesor. Cada una se evalúa con base en una escala ordinal, de 0 a 4, tomando como referente la demanda cognitiva. A continuación se presentan los resultados obtenidos de la evaluación de las 40 sesiones de clase que formaron parte de la muestra del estudio, categoría por categoría.

Resultados de la segunda fase del estudio

Potencial de la actividad de enseñanza

En el IQA una actividad de enseñanza se define como un problema, o un conjunto de ellos, que toca una idea o concepto matemático. La primera categoría del instrumento evalúa el potencial de la actividad que se le plantea al grupo.

En el nivel más alto (4) se ubican a las actividades que implican el uso de un pensamiento complejo y no algorítmico por parte de los alumnos y procuran que exploren y entiendan la naturaleza, ya sea de conceptos matemáticos, procedimientos o relaciones. Las actividades pueden requerir que los estudiantes resuelvan un problema genuino y desafiante, que comprendan por qué ciertos procedimientos o fórmulas funcionan o que hagan conjeturas y soporten sus conclusiones con evidencia matemática.

En el nivel 3 están las actividades que tienen también un alto potencial, pero que no alcanzan el 4 porque: o se procura que los estudiantes identifiquen patrones pero sin insistir en que generalicen; que utilicen múltiples estrategias, sin hacer énfasis en que las relacionen; o que formulen conjeturas, pero no que proporcionen evidencia matemática que soporte sus conclusiones.

En el nivel 2 se encuentran aquellas actividades cuyo potencial se limita a que los estudiantes utilicen un procedimiento específico. Se centran en la producción de respuestas correctas más que en la comprensión de cuestiones matemáticas.

En el nivel 1 están las actividades cuyo potencial se limita a que los estudiantes memoricen o reproduzcan información, reglas o fórmulas.

Existe además un nivel 0, aquí se ubican las actividades de enseñanza que no demandan trabajo matemático alguno.

La tabla 4 muestra los resultados en esta categoría, obtenidos al analizar las sesiones que formaron parte de la muestra.

TABLA 4

Potencial de la actividad de enseñanza

	No presente Nivel 0	Demandas de bajo nivel		Demandas de alto nivel	
		Nivel 1 (%)	Nivel 2 (%)	Nivel 3 (%)	Nivel 4 (%)
Muestra total (n=40)		3 (7.5)	19 (47.5)	11 (27.5)	7 (17.5)
Pública 1 (n=22)		2 (9.0)	13 (59.0)	5 (22.7)	2 (9.0)
Pública 2 (n=10)			4 (40.0)	2 (20.0)	4 (40.0)
Particular (n=8)		1 (12.5)	2 (25.0)	4 (50.0)	1 (12.5)

Los resultados sobre el potencial de la actividad de enseñanza muestran una imagen de gran diversidad en las sesiones observadas. Hay presencia tanto de actividades que implican altos y bajos niveles de demanda cognitiva. Vale la pena señalar que, aunque las concentraciones son distintas en cada escuela, en ninguna se nota una marcada tendencia a presentar cierto tipo de actividades.

Como ejemplo de las actividades de alta demanda cognitiva, está la que se utilizó en un aula de tercer grado. Estuvo contextualizada en una carrera atlética de corta distancia. A los niños, en parejas, se les repartieron dos tiras de papel de aproximadamente 30 cm, las cuales representaban el trayecto total de la carrera para dos competidores. Se les indicó, además, la fracción del trayecto que cada uno había recorrido después de un minuto de carrera ($3/4$ y $5/8$). Se les pidió que colaboraran entre compañeros para identificar el lugar (en cada tira) en el que creían que estaría cada corredor y que estuvieran listos para justificar su respuesta.

Un ejemplo de las actividades de baja demanda cognitiva se realizó en otra aula de tercero. Consistió en producir un apunte sobre qué son las fracciones. Se indicó a los niños que copiaran en sus cuadernos el texto escrito en el pizarrón. Después, se les repartieron tres hojas de papel a cada uno. Siguiendo instrucciones precisas, se les pidió que colorearan totalmente una de las hojas para que representara un entero. Después, que doblaran otra y colorearan una mitad para que representara un medio. La tercera hoja debía ser doblada en cuatro partes iguales y coloreada una de ellas para que representara un cuarto. Por último, se les solicitó que anexaran las tres hojas al apunte que habían copiado.

Instrumentación de la actividad de enseñanza

En esta categoría se evalúa la demanda cognitiva de la actividad de enseñanza al ser instrumentada. La escala de evaluación es similar a la de la anterior. En el nivel 4 se ubican las actividades en las que los estudiantes utilizan pensamiento complejo y no algorítmico, al explorar y comprender la naturaleza ya sea de conceptos matemáticos, procedimientos o relaciones.

En el nivel 3 se encuentran las actividades en las que los problemas, conceptos o procedimientos requieren menor grado de complejidad en el pensamiento de los alumnos que en el nivel anterior o en las que aunque se

utilizan algunas con nivel 4 de potencial, los estudiantes solo se involucran moderadamente con las demandas de alto nivel de la actividad planteada.

En el nivel 2 están las actividades que se enfocan en aplicar y ejecutar correctamente un procedimiento para obtener una respuesta correcta. En el 1 están aquellas en que los estudiantes solamente reproducen información, fórmulas o reglas. En el nivel 0 se ubican las que no involucran a los alumnos en actividad matemática alguna.

La tabla 5 muestra los resultados, en esta categoría, obtenidos al analizar las sesiones que formaron parte de la muestra.

TABLA 5
Instrumentación de la actividad de enseñanza

	No presente <i>Nivel 0</i>	Demandas de bajo nivel		Demandas de alto nivel	
		<i>Nivel 1 (%)</i>	<i>Nivel 2 (%)</i>	<i>Nivel 3 (%)</i>	<i>Nivel 4 (%)</i>
Muestra total (n=40)		4 (10.0)	22 (55.0)	14 (35.0)	
Pública 1 (n=22)		3 (13.6)	15 (68.2)	4 (18.2)	
Pública 2 (n=10)			5 (50.0)	5 (50.0)	
Particular (n=8)		1 (12.5)	2 (25.0)	5 (62.5)	

Comparando los resultados en esta categoría (ver tabla 5), con la anterior (ver tabla 4) se observa un corrimiento hacia los niveles inferiores. Se puede notar que el nivel 4 está desierto, mientras que la cantidad de sesiones ubicadas en los niveles 3, 2 y 1 se incrementa.

Estos resultados sugieren que, en general, el nivel de demanda cognitiva de las actividades de enseñanza evaluadas fue menor durante su instrumentación, que en el momento en que le fueron planteadas a los grupos.

La instrumentación de la actividad de alta demanda cognitiva que se mencionó antes sirve de ejemplo de cómo se dio esto. Poco después de que comenzó su instrumentación, el maestro sugirió a los alumnos cómo doblar las tiras para que estuvieran divididas en cuatro y ocho partes iguales. Además, aceptó como adecuadas las respuestas de los alumnos que ubicaron correctamente a los corredores, aunque no ofrecieran justificación alguna.

Conversación de los estudiantes a raíz de la actividad de enseñanza

En la tercera categoría se evalúa la demanda cognitiva presente en la conversación plenaria, en la que se revisa públicamente el trabajo realizado por los alumnos durante la instrumentación de la actividad. En el nivel 4 se ubican las sesiones en las que los estudiantes muestran o describen su trabajo escrito y argumentan coherentemente por qué la estrategia o procedimiento que utilizaron son válidas.

En el nivel 3 están las sesiones en las que, aunque los estudiantes muestran y argumentan su trabajo, las explicaciones son incompletas, poco coherentes o carecen de precisión. En el 2 se encuentran aquellas en que los alumnos muestran su trabajo, pero no argumentan por qué sus estrategias o procedimientos funcionan y por qué son apropiados para el problema. También están en este nivel las sesiones en las que se revisa solo un método de resolución.

En el nivel 1 se ubican las sesiones en las que los estudiantes responden de manera muy breve cuando se les pregunta qué hicieron o dan respuestas que no son matemáticas. Finalmente, en el 0 están las sesiones en las que no hay conversación plenaria.

TABLA 6

Conversación de los estudiantes a raíz de la actividad de enseñanza

	No presente Nivel 0 (%)	Demandas de bajo nivel		Demandas de alto nivel	
		Nivel 1 (%)	Nivel 2 (%)	Nivel 3 (%)	Nivel 4 (%)
Muestra total (n=40)	4 (10.0)	17 (42.5)	8 (20.0)	10 (25.0)	1 (2.5)
Pública 1 (n=22)	3 (13.6)	11 (50.0)	6 (27.3)	2 (9.1)	
Pública 2 (n=10)	1 (10.0)	4 (40.0)	1 (10.0)	4 (40.0)	
Particular (n=8)		2 (25.0)	1 (12.5)	4 (50.0)	1 (12.5)

Los resultados en esta categoría muestran la mayor diversidad (ver tabla 6). Se trata también de la categoría en la que las sesiones evaluadas fueron ubicadas en los niveles más bajos. En la única sesión que fue clasificada en el nivel 4, un grupo de tercero se involucró en una conversación colectiva en la que no solo hubo diálogo matemático entre la maestra y los niños,

sino entre los mismos niños. En general, la conversación versó sobre cuántas decenas hay en total en cantidades como 350 o 526. Los participantes justificaron sus respuestas y elaboraron distintas explicaciones para ayudarle a entender a dos estudiantes, quienes expresaban que no les quedaba claro el porqué de las respuestas que ofrecían sus compañeros.

En contraste, hubo cuatro sesiones en las que los alumnos nunca hablaron, lo que hace que esta sea la única categoría en la que el nivel 0 estuvo presente.

Expectativas del profesor

En esta última categoría se evalúa el nivel de demanda cognitiva en las expectativas que el maestro comunica al grupo, ya sea de manera explícita o implícita, sobre lo que deben de lograr a lo largo de la sesión. En el nivel 4 se ubican las sesiones en las que se nota que el profesor constantemente procura que los estudiantes piensen y razonen de manera compleja, que exploren y comprendan conceptos matemáticos, procedimientos, o que establezcan relaciones entre ellos. Todo esto, como resultado de trabajar en la actividad planteada.

En el nivel 3 están las sesiones en las que, aunque parece que el profesor procura lo anterior, no se logra gran complejidad, ya sea porque sus expectativas son menores al del potencial de la actividad planteada o ésta carece de la complejidad necesaria.

En el nivel 2 se colocan las sesiones en que las expectativas del profesor se centran en que los estudiantes desarrollen habilidades que, si bien son pertinentes para el aprendizaje del concepto estudiado, no implican un pensamiento complejo. En el 1 están las sesiones en las que las expectativas no se centran en un contenido matemático sustantivo, sino en cuestiones como la correcta ejecución de algoritmos o en el que el trabajo escrito cumpla con ciertas especificaciones. En el nivel 0 se encuentran las sesiones en las que las expectativas del docente no son de naturaleza matemática o están centradas en cuestiones como el esfuerzo de los alumnos en terminar la tarea asignada.

Los resultados en esta categoría (ver tabla 7) también muestran un nivel general de desempeño relativamente bajo, aunque no tanto como en la categoría anterior. El nivel 4 vuelve a quedar desierto, y el corrimiento hacia los inferiores es más acentuado que en la categoría instrumentación de la actividad de enseñanza (tabla 5). En general, se nota que la demanda

cognitiva en las expectativas que los docentes manifestaron a sus grupos fue menor al de las actividades que presentaron, durante las sesiones evaluadas.

TABLA 7
Expectativas del profesor

	No presente	Demandas de bajo nivel		Demandas de alto nivel	
	Nivel 0 (%)	Nivel 1 (%)	Nivel 2 (%)	Nivel 3 (%)	Nivel 4 (%)
Muestra total (n=40)		12 (30.0)	16 (40.0)	12 (30.0)	
Pública 1 (n=22)		9 (40.9)	10 (45.5)	3 (13.6)	
Pública 2 (n=10)		1 (10.0)	5 (50.0)	4 (40.0)	
Particular (n=8)		2 (25.0)	1 (12.5)	5 (62.5)	

Análisis: valor educativo

Incorporación de estándares

Los resultados obtenidos en nuestro estudio muestran la utilidad de instrumentos como el IQA para monitorear el cumplimiento de estándares internacionales para la enseñanza de las matemáticas. Como sistema, los resultados indican que los mayores avances se han dado en el tipo de actividades que los docentes presentan a sus grupos (ver tabla 4). En otros aspectos, el cumplimiento es menor.

En general, los datos revelan que, en este pequeño sistema, el potencial de las actividades no es aprovechado durante su instrumentación (ver tabla 5). Hay también rezagos en lograr que los docentes consistentemente procuren que sus estudiantes piensen y razonen de manera compleja, y que exploren y comprendan conceptos matemáticos (ver tabla 7).

En la orquestación de conversaciones plenarias es donde se nota el mayor rezago (ver tabla 6). Fueron relativamente pocas las aulas donde los alumnos tuvieron la oportunidad de presentar y argumentar la validez de su trabajo. En contraste, se identificaron varias sesiones en las que ellos no hicieron público su trabajo.

Respecto de las escuelas, los resultados sugieren la utilidad de este tipo de instrumentos para identificar el nivel de adopción de los estándares en

cada institución. La escuela que más los ha adoptado es la privada, donde más de la mitad de las sesiones evaluadas fueron ubicadas en los niveles 3 y 4, en todas las categorías. En contraste, en la pública 1 es donde se nota que ha habido el menor avance. Ahí, más de la mitad de las actividades fueron ubicadas en el nivel 2 o inferiores.

La escuela pública 2 es difícil de juzgar porque, como se mencionó, en ella solo se evaluaron a 5 de los 8 maestros seleccionados. Los datos muestran dos tipos de sesiones diametralmente diferentes. Aparentemente, en dos de las aulas evaluadas los estándares han sido incorporados de manera notable. En contraste, hay otras dos en las que no es así y una más que estaría intermedia.

En relación con aspectos del perfil de los docentes que pudieran estar relacionados con la incorporación de estándares internacionales, en los resultados no se notó patrón alguno. Ni los años de servicio, el tipo de formación o el nivel máximo de estudios aparecieron como un factor que ayudara a identificar a los profesores que realizaron las sesiones mejor (o peor) calificadas.

Oportunidades de aprendizaje

Los resultados también sugieren la utilidad de instrumentos como el IQA para detectar diferencias importantes en las oportunidades de aprendizaje que pueden tener los estudiantes por estar en un aula u otra. Como sistema, se nota que mientras en algunas los alumnos tienen acceso a actividades que les abren la posibilidad de explorar y entender la naturaleza de conceptos matemáticos; en otras, sus oportunidades de aprendizaje matemático se limitan a dominar la ejecución de un procedimiento. Mientras que en unas, pueden hacer público su pensamiento, de articularlo y argumentarlo; en otras, sus posibilidades de exponer su quehacer matemático son muy restringidas o de plano, inexistentes.

En relación con las escuelas, en todas se nota la existencia de diferencias importantes en las oportunidades de aprendizaje de los estudiantes, por estar en una u otra aula. Ello sugiere que –al menos en este pequeño subsistema– las oportunidades de los alumnos dependerían más del aula que de la escuela a la que asisten.

De manera global, la calidad general de la enseñanza fue superior en la escuela particular. Sin embargo, los datos muestran que, en al menos una de las aulas de esta escuela, las oportunidades de aprendizaje matemático de los alumnos pueden ser muy limitadas.

En contraparte, en la escuela pública 1 es donde los alumnos, en general, parecen contar con las menores oportunidades de aprendizaje matemático. Sin embargo, como se hace evidente en los datos, hay al menos un aula donde las posibilidades son tan buenas como en las mejores de las otras dos escuelas.

En la escuela pública 2 se da el mayor contraste en las oportunidades de aprendizaje que pueden tener los alumnos por estar en un aula u otra. Hay dos de ellas donde las oportunidades parecerían ser muy favorables, y dos en las que serían relativamente desfavorables.

Acciones para la mejora

Quizá lo más importante de los resultados de una evaluación en términos de valor educativo sea su utilidad para informar el diseño e instrumentación de una agenda de mejora. En este aspecto, los resultados obtenidos sirven no solo para dimensionar el reto que implicaría mejorar la enseñanza en las aulas evaluadas, sino también para especificar áreas en las que focalizar los esfuerzos, particularmente cuando éstos impliquen el desarrollo profesional docente.

Como sistema, los resultados muestran que es necesario ir más allá de asegurar que las actividades que se presentan en sus aulas impliquen la resolución de un problema interesante, que invite a los alumnos a reflexionar y a encontrar diferentes formas de resolverlo. Se necesita avanzar en lograr que, de manera consistente, se procure y valore el razonamiento complejo de los alumnos, a lo largo de toda la sesión.

El área que más trabajo requiere es la relativa a la comunicación abierta del quehacer matemático de los alumnos. Como los datos muestran, son relativamente pocas las aulas en las que los alumnos presentan al grupo su trabajo y argumentan la validez del mismo. En la gran mayoría, el discurso matemático lo centraliza el maestro. A los alumnos, en el mejor de los casos, se les permite comunicar qué hicieron, pero no cómo ni por qué.

Al nivel de las escuelas, los resultados muestran su utilidad en el diseño de una agenda de mejora focalizada para cada una. En los resultados se nota que es en la pública 1 donde menor avance ha habido en adoptar las recomendaciones internacionales sobre cómo enseñar matemáticas. Así, una agenda de mejora deberá procurar la adopción de estas recomendaciones por parte de la gran mayoría de los maestros en esa institución. Esto implicaría una transformación en la manera de involucrar a los estudiantes en las actividades de enseñanza, para promover en ellos el razonamiento,

la resolución de problemas, y la argumentación matemática (cf., NCTM, 1989; SEP, 2011).

Es importante señalar que los datos revelaron que en una de las aulas de la escuela pública 1 estas recomendaciones han sido adoptadas de manera notable. Haberla identificado puede ser valioso en el diseño de una agenda de mejora para esta escuela, una vez que lo que sucede en ella puede usarse como evidencia para sus maestros de que el cambio es posible y como ejemplo de las prácticas de enseñanza que se busca promover.

Los datos recabados también son útiles para proyectar una agenda de mejora para la escuela particular, donde el reto sería notablemente diferente. Como lo muestran los datos, en la mayoría de las aulas de esta escuela, las actividades que se presentan a los alumnos tienen un alto potencial, además de que se les brinda la oportunidad de mostrar y argumentar su quehacer matemático. Sin embargo, los datos también revelan que el nivel típico de demanda cognitiva es 3, en todas las categorías.

El reto general de mejorar la enseñanza aquí implicaría elevar la demanda de las actividades que se presentan a los alumnos y asegurarse de que el nivel se mantenga durante la instrumentación. También, avanzar en las formas en las que los estudiantes argumentan la validez de su quehacer matemático, de manera que sean más completas, coherentes y precisas. Además, se necesitaría que los docentes elevaran sus expectativas sobre el nivel de complejidad en el razonamiento matemático que pueden y deben alcanzar sus alumnos en una sesión de clase. Finalmente, sería necesario brindarle atención especial al aula donde se identificó que las actividades presentadas a los alumnos implicaron demandas de más bajo nivel y donde sus oportunidades de expresar su quehacer matemático fueron muy limitadas.

El caso de la escuela pública 2 se parece al de la 1 en tanto que, aparentemente, las recomendaciones internacionales de cómo enseñar matemáticas han sido incorporadas en algunas de las aulas. Sin embargo, para esbozar una agenda de mejora, sería necesario evaluar la enseñanza que tiene lugar con los tres maestros de esta escuela que declinaron participar en el estudio.

Conclusión

Como mencionamos al principio de este artículo, en el sistema educativo mexicano, los esfuerzos evaluativos se han concentrado en el desempeño de los alumnos y en el perfil de los maestros. A pesar del alto valor que tanto la autoridad educativa como otros actores del sistema parecen atribuirle a

este tipo de evaluaciones, su aportación no es del todo clara. Como lo señalan Santiago *et al.* (2012), en México se ha priorizado el uso de las evaluaciones como instrumentos de control y rendición de cuentas, en lugar de como un insumo para cuestiones más estrechamente vinculadas con el diseño, instrumentación y seguimiento de acciones de mejora.

Pese a que el desempeño del alumnado en matemáticas ha sido una de las áreas más evaluadas en los últimos tres lustros, el panorama actual no es alentador. Según los últimos resultados, la tendencia en el desempeño general del alumnado es negativa. En la aplicación que realizó el INEE de los Exámenes de la Calidad y el Logro Educativo (Excale-06) en matemáticas, a una muestra nacional alumnos de sexto grado de primaria, el porcentaje de estudiantes ubicado en el nivel I –el más bajo de cuatro niveles– creció en 5.8 puntos; pasó de ser 38.02 en 2013 a 43.82 en 2015 (cf., INEE, 2015).

La investigación aquí reportada devela el alto valor educativo que puede tener la evaluación de la enseñanza matemática para un sistema como el mexicano. Como se ha explicado, los resultados permiten ir más allá de la jerarquización de la calidad de la enseñanza que tiene lugar en las aulas. Hacen que además sea posible hacer juicios sobre el nivel de incorporación que, en aulas y escuelas, se está haciendo de estándares específicos para la enseñanza. También permite detectar, de manera general, las oportunidades de los alumnos para lograr los conocimientos y las habilidades matemáticas necesarios en el mundo de hoy (INEE, 2015; OCDE, 1999; OCDE, 2014). Pero lo más importante de este tipo de evaluaciones es la utilidad de los datos que aportan en el diseño de acciones de mejora, que pueden ser específicas para cada centro escolar.

Creemos, además, que este tipo de instrumentos puede ser útil en el sistema educativo mexicano para darle seguimiento a los esfuerzos que se realicen para mejorar la enseñanza y servir como un referente del tipo de prácticas que se espera que tengan lugar en las aulas (Boston y Wolf, 2006). Aunque hay que aclarar que nuestro estudio no tocó estos dos aspectos del valor educativo.

Con base en nuestra investigación concluimos que las evaluaciones de la calidad de la enseñanza matemática serían factibles en un sistema educativo como el mexicano, una vez que su alto valor educativo justificaría los costos junto con las limitaciones que implicarían su instrumentación. Dicho esto, es necesario señalar tres cuestiones importantes sobre las que se necesita saber más antes de aplicar un instrumento como el IQA a gran escala.

La primera de ellas tiene que ver con la fiabilidad entre jueces. Como se menciona, los evaluadores en el estudio fuimos los propios autores de este artículo. A pesar de nuestra experiencia en la observación de lecciones de matemáticas, nuestro conocimiento del IQA y formación previa en el campo, los niveles de fiabilidad alcanzados no fueron muy altos. Surge entonces la duda sobre qué implicaría, en términos de selección y formación de evaluadores, contar con un número suficiente de ellos, capaz de evaluar la enseñanza en una cantidad importante de escuelas, alcanzando niveles aceptables de fiabilidad.

La segunda cuestión se refiere a la resistencia de los docentes a que sus aulas sean evaluadas. En nuestro estudio, su participación fue voluntaria; sin embargo, hubo tres en una escuela que decidieron no intervenir, a pesar del compromiso de los investigadores de no revelar su identidad ni la de su escuela. Queda pues la duda de si una evaluación con aval institucional garantizaría la participación de los docentes o si, por el contrario, enfrentaría aún más barreras por parte de ellos.

La tercera y última cuestión se refiere al tipo de escuelas en que se realizó el estudio, las cuales eran urbanas y operaban en circunstancias institucionales relativamente favorables. Faltaría entonces saber sobre el valor educativo que tendría la evaluación de la enseñanza matemática en las muchas escuelas del sistema educativo mexicano que operan bajo condiciones institucionales poco favorables. Esto es, faltaría saber sobre el valor educativo de estas evaluaciones en las escuelas con aulas multigrado, en las que uno de los docentes es responsable de las tareas administrativas de director, en las que no dominan la lengua materna de los alumnos, y en donde los índices de ausentismo de maestros y alumnos son significativamente altos.

Referencias

- AERA, APA y NCME (2014). *The Standards for educational and psychological testing*, Washington, DC: American Educational Research Association/American Psychological Association/National Council on Measurement in Education.
- Alsina, A. y Coronata, C. (2014). "Los procesos matemáticos en las prácticas docentes: diseño, construcción y validación de un instrumento de evaluación", *Edma 0-6: Educación Matemática en la Infancia*, vol. 3, núm 2, pp. 23-36.
- Backhoff, E.; Bouzas, A.; Contreras, C.; Hernández, E. y García, M. (2007). *Factores escolares y aprendizaje en México. El caso de la educación básica*, Ciudad de México: Instituto Nacional para la Evaluación de la Educación.

- Borko, H.; Stecher, B. M. y Kuffner, K. (2007). *Using artifacts to characterize reform-oriented instruction: The scoop notebook and rating guide*, Los Angeles, CA: CRESST.
- Boston, M. (2012). "Assessing instructional quality in mathematics", *The Elementary School Journal*, vol. 113, núm. 1, pp. 76-104.
- Boston, M. y Wolf, M. K. (2006). *Assessing academic rigor in mathematics instruction: The development of the Instructional Quality Assessment toolkit*, Los Angeles, CA: CRESST.
- Brown, M.; Askew, M.; Rhodes, V.; Denvir, H.; Ranson, E. y Wiliam, D. (2001). "Magic bullets or chimeras? Searching for factors characterising effective teachers and effective teaching in numeracy", ponencia presentada en *The British Educational Research Association Annual Conference*, septiembre, Bath, University of Leeds.
- Cea D'Ancona, M. A. (2001). *Metodología cuantitativa: Estrategias y técnicas de investigación social*, Madrid: Síntesis.
- Childs, K. J.; Sutherland, M.; Sahin, N.; Gault, R. y Dixon, J. K. (2016). "Examining an instrument for assessing primary grades mathematics classrooms", ponencia presentada en *The 2016 National Council of Teachers Mathematics Research Conference*, San Francisco, CA, abril de 2016.
- Cobb, P.; Zhao, Q. y Visnovska, J. (2008). "Learning from and adapting the theory of realistic mathematics education", *Education et Didactique*, vol. 2, núm. 1, pp. 55-73.
- Cobb, P.; Jackson, K.; Smith, T.; Sorum, M. y Henrick, E. (2013). "Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale", *National Society for the Study of Education Yearbook*, vol. 112, núm. 2, pp. 320-349.
- Darling-Hammond, L.; Herman, J.; Pellegrino, J.; et al. (2013). *Criteria for High-Quality Assessment*, Stanford, CA: SCOPE.
- Herman, J. L. y Baker, E. L. (2005). "Making benchmark testing work", *Educational Leadership*, vol. 63, núm. 3, pp. 48-54.
- Herman, J. L. y Choi, K. (2012). *Validation of ELA and mathematics assessments: A general approach*, Los Angeles, CA: CRESST.
- Hiebert, J. y Grouws D. A. (2007). "The effects of classroom mathematics teaching on students' learning", en Frank K. Lester (ed.), *Second handbook of research on mathematics teaching and learning*, vol. 1, Charlotte, NC: IAP, pp. 317-404.
- Hill, H. C.; Blunk, M. L.; Charalambous, C. Y.; Lewis J. M.; Phelps G. C.; Sleep L. y Ball, D. L. (2008). "Mathematical knowledge for teaching and the mathematical quality of Instruction: An exploratory Study", *Cognition and Instruction*, vol. 26, núm. 4, pp. 430-511.
- INEE (2013). *Prácticas pedagógicas y desarrollo profesional docente en preescolar*, Ciudad de México: Instituto Nacional para la Evaluación de la Educación.
- INEE (2015). *Plan Nacional para la Evaluación de los Aprendizajes (Planea). Resultados nacionales 2015, 6º de primaria y 3º de secundaria, lenguaje y comunicación, y matemáticas*, Ciudad de México: Instituto Nacional para la Evaluación de la Educación. Disponible en: <http://www.inee.edu.mx/index.php/resultados-nacionales-2015> (consultado: 30 de noviembre de 2015).

- Martínez Rizo, F. (2014). "El estudio de las prácticas docentes", en Guadalupe Ruiz (ed.), *Acercamientos empíricos a las prácticas de evaluación en el aula en la educación básica*, Aguascalientes: Universidad Autónoma de Aguascalientes, pp. 17-71.
- Matsumura, L. C.; Slater, S. C.; Junker, B.; Peterson, M.; Boston, M.; Steele, M. y Resnick, L. B. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment*, Los Angeles, CA: CRESST.
- NCTM (1989). *Curriculum and evaluation standards for school mathematics*, Reston, Virginia: National Council of Teachers of Mathematics.
- NCTM (2014). *Principles to actions: Ensuring mathematical success for all*, Reston, Virginia: National Council of Teachers of Mathematics.
- OCDE (1999). *PISA. La medida de los conocimientos y destrezas de los alumnos: un nuevo marco para la evaluación*, Madrid: Ministerio de Educación, Cultura y Deporte.
- OCDE (2003). *Marcos teóricos de PISA 2003: La medida de los conocimientos y destrezas en matemáticas, lectura, ciencias y resolución de problemas*, Madrid: Instituto Nacional de Evaluación y Calidad del Sistema Educativo.
- OCDE (2009). *Los docentes son importantes: atraer, formar y conservar a los docentes eficientes*, París: Organización para la Cooperación y el Desarrollo Económicos.
- OCDE (2014). *Resultados de PISA 2012 en Foco: lo que los alumnos saben a los 15 años de edad y lo que pueden hacer con lo que saben* (vol. I), París: Organización para la Cooperación y el Desarrollo Económicos.
- Rosenshine, B. (1970). "Evaluation of classroom instruction", *Review of Educational Research*, vol. 40, núm. 2, pp. 279-300.
- Sanders, W. L. y Horn S. P. (1994). "The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment", *Journal of Personnel Evaluation in Education*, vol. 8, núm. 3, pp. 299-311.
- Santiago, P.; McGregor, I.; Nusche, D.; Ravela, P. y Toledo D. (2012). *Revisiones de la OCDE sobre la Evaluación en Educación. México 2012*, Ciudad de México: INEE.
- SEP (2011). *Programas de estudio 2011. Guía para el maestro. Educación básica. Primaria. Sexto grado*, Ciudad de México: Secretaría de Educación Pública.
- Stein, M. K. y Lane, S. (1996). "Instructional tasks and the development of student capacity to think and reason: an analysis of the relationship between teaching and learning in a reform mathematics project", *Educational Research and Evaluation: An International Journal on Theory and Practice*, vol. 2, núm. 1, pp. 50-80.
- Weiss, I. R.; Pasley, J. D.; Smith, P. S.; Banilower, E. R. y Heck, D. J. (2003). *Looking Inside the Classroom: A study of K-12 mathematics and science education in the United States*, Chapel Hill, NC: Horizon Research Inc.

Artículo recibido: 14 de octubre de 2016

Dictaminado: 1 de febrero de 2017

Segunda versión: 3 de marzo de 2017

Aceptado: 3 de marzo de 2017