



EL FUTURO DE LA EVALUACIÓN EDUCATIVA

Felipe Martínez Rizo

Currículo: licenciado en Ciencias Sociales. Profesor-investigador del Departamento de Educación, Centro de Ciencias Sociales y Humanidades, Universidad Autónoma de Aguascalientes. Sus líneas de investigación incluyen la educación superior y la básica, en temas de calidad, planeación, evaluación y desigualdad.

Recibido: 11 de septiembre de 2012. Aceptado para su publicación: 10 de diciembre de 2012.

Como citar este artículo: Martínez, F. (enero-junio, 2013). El futuro de la evaluación educativa. *Sinéctica*, 40. Recuperado de http://www.sinectica.iteso.mx/articulo/?id=40_el_futuro_de_la_evaluacion_educativa

Resumen

Después de precisar la noción de evaluación, y acotarla en relación con la que tiene como objeto la calidad de los sistemas de educación básica, el artículo revisa la situación que prevalece en México y América Latina en lo que se refiere a cinco tipos particulares de evaluación: pruebas de aprendizaje, indicadores estadísticos, evaluaciones de docentes, evaluaciones de escuelas y evaluaciones de instituciones de educación superior. Luego, se esbozan dos escenarios que se podrán presentar en el futuro: el que se podría esperar en el caso de que la situación y las tendencias que prevalecen en la actualidad no se modifiquen sustancialmente, y otro deseable, en el caso de que se corrijan los defectos que se identifican en la actualidad y se potencien los puntos positivos.

Palabras clave: evaluación educativa, México, tendencias.

Abstract

The paper specifies the notion of evaluation & assessment, in relation to the quality of education systems at the elementary and middle levels. The situation in Mexico and Latin America is reviewed in regard to five specific types of evaluation: learning tests, statistical indicators, evaluations of teachers, evaluations of schools and of higher education institutions. Finally two future scenarios are outlined: one which could be expected if trends currently prevailing are not substantially altered; another one if negative aspects are corrected and positive are enhanced.

Keywords: educational evaluation, assessment, Mexico, trends.

INTRODUCCIÓN

Este trabajo se plantea tres preguntas estrechamente relacionadas: ¿cuál es la situación actual de la evaluación educativa en México y, de manera más amplia, en América Latina? ¿Qué retos debe enfrentar? ¿Hacia dónde parece dirigirse y hacia dónde sería deseable que se dirigiera? Antes de tratar de responder tan complejas cuestiones, es necesario delimitar bien nuestro objeto de interés.

La evaluación se puede definir como *el juicio de valor que resulta de contrastar el resultado de la medición de una realidad empírica con un parámetro normativo previamente definido* (INEE, 2006, p. 18). Esta definición se aplica a cualquier objeto, e implica la noción de *calidad*, pues siempre se evalúa la calidad de algo.

En nuestro caso se tratará de la evaluación de la calidad educativa, pero todavía son necesarias precisiones adicionales, ya que es frecuente que la noción se identifique con la que tiene por objeto el aprendizaje que alcanzan los estudiantes, cuando en realidad se puede referir a otros *evaluandos*, como los docentes, las escuelas o el sistema educativo mismo. En este trabajo, la noción se entenderá referida a este último objeto. Se abordará, pues, la evaluación de la calidad de los sistemas educativos y, en particular, de educación básica.

Ahora bien, un sistema educativo es un conjunto de muchos elementos: alumnos, docentes y escuelas, pero también planes y programas de estudio, libros de texto y materiales didácticos, directores de plantel, supervisores, autoridades y, además, normas, reglamentos, políticas, organización, etcétera.

Es importante distinguir evaluaciones que buscan llegar a juicios sobre individuos (sean éstos alumnos, docentes, centros escolares u otros), en contraste con las que buscan alcanzar conclusiones sobre un sistema educativo en conjunto. En el caso de evaluaciones de aprendizaje, unas sirven para dar resultados de cada alumno, sea para decidir si pasa al grado siguiente o debe repetir el mismo curso; sea para tener elementos para ajustar sus esfuerzos para mejorar el aprendizaje ulterior. Otras evaluaciones, en cambio, sirven para informar sobre la situación promedio del aprendizaje de los alumnos de cierto nivel, y para sustentar decisiones de política que afecten al conjunto del sistema educativo.

En el primer caso, es necesario aplicar las pruebas a todos los alumnos (censo), mientras que en el segundo basta con hacerlo a una muestra representativa, lo que permite usar instrumentos que midan aspectos más complejos.

El nivel de aprendizaje es, sin duda, un aspecto fundamental de la calidad de un sistema educativo, pero no el único. Si una parte de los niños que deberían asistir a determinados grados está fuera de la escuela, el sistema educativo no podrá considerarse “de buena calidad”, aunque los que sí asistan tengan buenos niveles de aprendizaje. Tampoco podrá considerarse bueno un sistema educativo cuyas escuelas ofrezcan un currículo obsoleto o irrelevante, que no dé a los alumnos lo que necesitarán para una vida adulta plena; ni uno en el que algunos alumnos alcancen buenos resultados, pero muchos no, o uno que no cuente con los recursos básicos para operar adecuadamente.

Para evaluar cada una de las dimensiones anteriores de la calidad de un sistema educativo se necesitan distintos acercamientos. Las pruebas de aprendizaje sólo sirven para evaluar ese aspecto, pero para otros se deberán manejar indicadores basados en estadísticas; o bien, se requerirán estudios especiales, a veces de gran complejidad, como los que son necesarios para la evaluación de docentes.

En este artículo se revisará primero la situación actual de la evaluación educativa en México y América Latina en lo que se refiere a varias de esas modalidades (pruebas de aprendizaje, indicadores estadísticos, evaluaciones de docentes, evaluaciones de escuelas y evaluaciones de instituciones de educación superior), tratando de identificar puntos fuertes y débiles. En seguida, y en relación con los mismos puntos ya considerados, se esbozarán dos escenarios que se podrán presentar en el futuro: el que se podría esperar en el caso de que la situación y las tendencias que prevalecen en la actualidad no se modifiquen sustancialmente, y otro deseable, en el caso de que se corrijan los defectos que se identifican en la actualidad y se potencien los puntos positivos.

LAS EVALUACIONES EN MÉXICO Y AMÉRICA LATINA EN 2012

Pruebas de aprendizaje

Al inicio de los años noventa, la experiencia de México en pruebas estandarizadas se reducía a exámenes de ingreso a normal, el IDANIS (Instrumento de Diagnóstico para Aspirantes de Nuevo Ingreso a Secundaria) y el Estudio de Evaluación de Primaria que la Secretaría de Educación Pública (SEP) comenzó a aplicar en la década de 1970. En los noventa, esas actividades recibieron un fuerte impulso debido a la conjunción de varios elementos (Martínez Rizo y Santos del Real, 2009):

- La firma del Acuerdo Nacional para la Modernización de la Educación Básica (ANMEB, 1992) entre las autoridades y el Sindicato Nacional de Trabajadores de la Educación (SNTE), que incluyó el inicio del programa de estímulos de Carrera Magisterial. Para asignar esos estímulos, se decidió tomar en cuenta, entre otros aspectos, el aprendizaje alcanzado por los alumnos de los maestros que aspiraban a recibir los apoyos, por lo que fue necesario aplicar pruebas de rendimiento a números considerables de chicos. La primera aplicación fue en 1993 e involucró a más de cuatro millones de alumnos. Esas evaluaciones del llamado “Factor de Aprovechamiento Escolar” se siguieron aplicando hasta 2005, cuando el número de alumnos evaluados se acercó a ocho millones.
- En 1991, el gobierno mexicano, con apoyo del Banco Mundial, puso en marcha un programa compensatorio en beneficio de cuatro estados: el Programa para Abatir el Rezago Educativo (PARE), cuya evaluación incluyó la aplicación de pruebas de aprendizaje. Luego, continuaron otros programas compensatorios, siempre con evaluaciones que incluían aplicar pruebas de rendimiento.
- El ingreso de México a la OCDE, en 1994, incluyó el interés por participar en el proyecto PISA, de la OCDE, cuya primera aplicación fue en el año 2000. Por la misma época, comenzó la participación en otros proyectos de evaluación educativa de alcance internacional: en 1995, el Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS); en 1996, el primer estudio del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE).
- A partir de la reforma curricular de principios de la década, en 1996 comenzó un trabajo de definición de estándares curriculares, junto con el desarrollo de instrumentos de evaluación del aprendizaje alcanzado por los alumnos en relación con esos estándares. Esos instrumentos se denominaron Pruebas de Estándares Nacionales, y se aplicaron por primera vez en 1998. Hasta 2004 se aplicaron cada año a muestras nacionales de alumnos de primaria y, desde 2000, también de la secundaria básica. Las pruebas fueron acompañadas por cuestionarios sobre el contexto de la enseñanza y el hogar.
- Además, se siguió aplicando el IDANIS y se desarrollaron pruebas para premiar a alumnos destacados en el marco de las Olimpiadas del Conocimiento.

En la primera década del siglo XXI se dieron nuevos pasos con la creación, en agosto de 2002, del Instituto Nacional para la Evaluación de la Educación (INEE), como instancia dotada de autonomía técnica.

Tras constatar la existencia de serias fallas técnicas de las Pruebas de Estándares Nacionales aplicadas de 1998 a 2003, en 2004 el INEE decidió emprender el desarrollo de nuevos instrumentos que cubrieran de manera más amplia las áreas del currículo con pruebas de diseño matricial aplicadas a muestras representativas de alumnos en un ciclo de cuatro años (un año se evaluaría a los alumnos de 3° de secundaria, otro a los de 6° de primaria, uno más a los de 3° de primaria, y otro a los de 3° de preescolar).

Las primeras de estas nuevas pruebas (Exámenes de la Calidad y el Logro Educativo, EXCALE) se aplicaron en 2005. Ese mismo año, con apoyo técnico del INEE, la SEP decidió emprender el desarrollo de otras pruebas nuevas, que se denominaron Exámenes Nacionales del Logro Educativo en Centros Escolares (ENLACE) y se utilizaron por primera vez en 2006. La aplicación, sin embargo, no fue como se había planeado, ya que en lugar de usar un ciclo multianual similar al de las pruebas EXCALE, se decidió aplicarlas anualmente a los cuatro últimos grados de primaria y el último de secundaria básica. Tampoco dejaron de emplearse todas las otras pruebas que se pretendía sustituir con las nuevas (Martínez Rizo y Santos del Real, 2009).

Durante la administración federal 2007-2012, el trabajo del INEE siguió mejorando su calidad técnica con el desarrollo de pruebas de respuesta abierta para evaluar expresión escrita y niveles cognitivos complejos; pruebas de ejecución para niños de preescolar; métodos complejos de muestreo; modelos estadísticos sofisticados para el análisis de los resultados, como los modelos lineales jerárquicos y los de ecuaciones estructurales, además de los de respuesta al ítem. En el ámbito internacional, la participación en PISA fue cada vez más activa, tanto en la preparación de las pruebas como en cuanto al análisis de los resultados.

En cuanto a las pruebas ENLACE, desde los primeros resultados prevaleció la tendencia a interpretarlos simplistamente, y usarlos para hacer ordenamientos de escuelas, supuestamente según su calidad, con base sólo en los resultados de los alumnos, sin tener en cuenta las condiciones de la escuela, el número de los alumnos evaluados, entre otros aspectos. Estas tendencias se acentuaron tras las aplicaciones de 2007 y 2008, al grado que las pruebas ENLACE, con las de la OCDE, se volvieron el referente básico de la política educativa. Su propósito formativo inicial se ha desvirtuado por usos que las hicieron de gran impacto, lo que hace que surjan prácticas negativas de preparación para las pruebas y reducción del currículo en función de su contenido. Han comenzado incluso a aparecer pruebas preparatorias a nivel estatal (ENLACE intermedio) (Martínez Rizo y Santos del Real, 2009; Martínez Rizo y Blanco, 2010).

En América Latina, las pruebas en gran escala son también omnipresentes y tienen un peso cada vez mayor en las políticas. Hace tres décadas, ningún país de la región tenía pruebas censales o muestrales que dieran resultados sobre el sistema educativo en conjunto. Chile fue el primer país de la región que estableció un sistema nacional, y en la década de 1980 algunos países, como México, avanzaron tímidamente en esa dirección (Martínez Rizo, 2010a; Ferrer, 2006; UNESCO, 2008; Wolff, 2004). No obstante, fue a partir de 1990 cuando tales sistemas surgieron en la mayoría de los países de la región (como ocurrió en Europa, cfr. AEAEAC, 2009), aunque muchas veces sin continuidad, dependiendo de asesoría y recursos externos y con serias limitaciones técnicas. Un rasgo más era la escasa difusión de los resultados y el limitado uso que se hacía de ellos.

En 2010, casi todos los países contaban ya con evaluaciones muestrales en edu-

cación básica y, en diversa forma, Argentina, Brasil, Chile, Colombia, Costa Rica, la República Dominicana, Ecuador, El Salvador, Guatemala, México y Uruguay tienen o han tenido aplicaciones con carácter censal. La participación en evaluaciones internacionales aumentó también con las del LLECE, las de la IEA y las de PISA (Martínez Rizo, 2010a).

Pese a su débil tradición psicométrica, en varios países el sistema de evaluación se consolida técnica y organizacionalmente. También hay riesgos: por una parte, la tradicional falta de continuidad que suele acompañar a los cambios políticos; por otra, la aparición de efectos perversos para la calidad, que se explican por el concurso de varios factores: desconocimiento de los alcances y límites reales de las pruebas por parte de los tomadores de decisiones; presión de los medios de comunicación que privilegian interpretaciones simplistas y escandalosas; peso de sectores privados que desconocen las circunstancias de las escuelas públicas y esperan demasiado de las privadas, entre otros.

Un aspecto particular es el excesivo peso que se da a las pruebas censales que, por sus características, y contra lo que suele pensarse, dan resultados menos precisos que buenas pruebas muestrales. La razón de esta aparente paradoja es que el error no derivado del muestreo tiende a aumentar cuando lo hace el tamaño de la aplicación, que es cada vez más difícil controlar, de suerte que puede ser bastante mayor al error de muestreo, que se puede estimar con precisión.

Ejemplos de lo anterior es, en México, el peso abrumador de ENLACE frente a la poca atención que se presta a las pruebas del INEE; en Brasil, donde la PROVA Brasil (censal) desplaza al SAEB (muestral); en Perú, donde un valioso trabajo de pruebas muestrales ha dejado el lugar a pruebas censales hechas en forma apresurada; en la República Dominicana, donde desde los años 1990 se daba gran valor a los resultados de pruebas censales que tenía serias deficiencias, sólo recientemente reconocidas. En Colombia el ICFES maneja pruebas censales e insiste en la necesidad de no hacer usos de alto impacto y evitar la difusión de ordenamientos de escuelas, postura que debe enfrentar fuertes presiones en contrario (Martínez Rizo, 2010a).

La experiencia latinoamericana, al igual que la mexicana, muestra que la proliferación de pruebas, sobre todo de tipo censal y su excesivo peso en las políticas debido a su alto impacto, trae ya consecuencias negativas serias:

- Banalización del debate público sobre la calidad educativa, reducido a superficiales debates de ordenamientos discutibles.
- Mercadotecnia engañosa de las escuelas, sobre todo de sostenimiento privado, que buscan atraer alumnos basadas en esos ordenamientos.
- Empobrecimiento del currículo, por la tendencia a enseñar para las pruebas y descuidar aspectos fundamentales que no serán evaluados.
- Cansancio y desaliento en escuelas que, pese a su esfuerzo, no consiguen resultados comparables con planteles de condiciones más favorables.
- Rechazo de los alumnos ante una educación centrada en prepararlos para la prueba.
- Empobrecimiento de las políticas públicas, que tienden a buscar soluciones fáciles a problemas complejos y descuidan aspectos fundamentales, en particular la equidad (Martínez Rizo, 2010b).

Indicadores

El sistema educativo mexicano, como los de los países de América Latina y otros, producen regularmente datos estadísticos sobre aspectos básicos (como número de escuelas, alumnos y maestros) y estadísticas sociales que incluyen aspectos educativos, como analfabetismo y escolaridad. Con esos datos se construyen desde hace décadas indicadores elementales, como las tasas brutas de cobertura o los años promedio de escolaridad de la población adulta. Tanto las estadísticas como los indicadores tradicionales, sin embargo, adolecían de serias fallas técnicas, que han comenzado a corregirse en épocas muy recientes.

Al constatar las deficiencias de las estadísticas educativas tradicionales, en la década de 1980 la OCDE comenzó el desarrollo del sistema de indicadores INES (Indicators of National Education Systems), que desde 1991 produce anualmente el volumen denominado *Education at a Glance*. La Unión Europea impulsó por la misma época un sistema para los países que la integran, y en 1999 la UNESCO estableció el Instituto de Estadística. Los tres esfuerzos se conjuntan ahora para producir estadísticas de mejor calidad, con el sistema UEO (UNESCO, Unión Europea, OCDE) (Martínez Rizo, en prensa).

En México, desde 2002, el INEE comenzó a desarrollar un sistema de indicadores que inicialmente cubría sólo los niveles de educación preescolar, primaria y secundaria, y a partir de 2009 incluyó también la media superior. El sistema se distingue por su solidez conceptual y metodológica; tiene en cuenta la perspectiva del derecho a la educación y procura el desarrollo de indicadores complejos que den cuenta cada vez mejor de las diferentes dimensiones de la calidad.

En 2005, el INEE inició un trabajo con la Unidad de Planeación y Evaluación de la Política Educativa (UPEPE) de la SEP para producir en conjunto un sistema de indicadores, lo que dio lugar a la publicación de un primer anuario conjunto (INEE-SEP, 2006). A partir de 2007, ese esfuerzo no tuvo continuidad.

Posteriormente, los trabajos relativos a indicadores han incluido un esfuerzo conjunto con el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) para la medición multidimensional de la pobreza; el INEE es la instancia a cargo de la dimensión de educación.

En América Latina, además de los avances promovidos por la UNESCO a través de su Instituto de Estadística, especialmente en relación con el programa de Educación para Todos (Education for All, EFA), deben mencionarse al menos otros tres esfuerzos:

- El Proyecto Regional de Indicadores Educativos (PRIE), promovido por la Cumbre de las Américas de 1998 y coordinado por la Oficina Regional de la UNESCO para América Latina y el Caribe (OREALC), en colaboración con el Ministerio de Educación de Chile y luego con la SEP mexicana (PRIE, 2003 y 2011).
- El proyecto Sistema de Información de Tendencias Educativas en América Latina (SITEAL) de la oficina de Buenos Aires del Instituto Internacional de Planeamiento Educativo (IPE) de la UNESCO (SITEAL, s.f.).
- Y el proyecto Metas 2021 de la Organización de Estados Iberoamericanos (OEI), que incluye un sistema de indicadores para monitorear los avances de los países de la región (OEI, 2010 y 2011).

Evaluación de docentes

Con variantes, los salarios de los docentes representan con mucho el renglón más importante del gasto en todos los sistemas educativos, debido a que la educación es un *servicio intensivo en mano de obra calificada* en el que, al menos hasta ahora, el trabajo de los docentes no puede ser sustituido por el de una máquina.

En general, hay sistemas de evaluación del trabajo docente, pero suele haber gran insatisfacción al respecto, porque se basan simplemente en la escolaridad y antigüedad en el servicio de los maestros, así como en evaluaciones hechas por directores y supervisores, elementos que no parecen los mejores para distinguir los buenos docentes de los que no lo son.

Tal insatisfacción es justificada, sin duda, aunque no se debe olvidar que el uso de esos criterios fue un avance respecto a prácticas anteriores. La “escala única de salarios” (*single-salary schedule*), establecida a principios del siglo XX en Estados Unidos, buscaba erradicar la práctica de pagar al maestro según el nivel educativo en que trabajara (en primaria menos que en secundaria); según su género (a los maestros más que a las maestras); y pertenencia a minorías étnicas: a los blancos más que a otros grupos (Prostik, 1996). En la Nueva España, los requisitos para ser admitido en el gremio de los maestros eran primero demostrar “pureza de sangre” y “ser cristiano viejo”, y sólo después dominar los conocimientos básicos de lectura, escritura y aritmética (Martínez Rizo, 2011).

El interés por tener formas de estudiar y evaluar la *efectividad* docente con mayor validez y confiabilidad que la que permiten los sistemas vigentes está llevando a desarrollar sistemas que incluyen evaluaciones estandarizadas de conocimientos, pero además portafolios de evidencias de la práctica y observación directa de ésta. La creciente disponibilidad de resultados de pruebas de aprendizaje que se aplican censalmente está llevando también a utilizar los resultados de los alumnos como indicador de la competencia del docente, si bien los expertos coinciden en señalar que las técnicas disponibles distan mucho todavía de tener la precisión y confiabilidad necesarias para sustentar decisiones importantes sobre maestros individuales (Martínez Rizo, 2011).

En América Latina destaca Chile, con el Sistema de Evaluación del Desempeño Profesional Docente (SEDPD), cuyo desarrollo comenzó en 2003, con la definición de los estándares del Marco para la Buena Enseñanza. El trabajo técnico está a cargo del Centro de Medición de la Universidad Católica de Chile (MIDE-UC), con el Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (CPEIP) del Ministerio de Educación de ese país.

Se usan cuatro instrumentos: una autoevaluación; una entrevista por un evaluador par; un informe de referencia de terceros (el director y el jefe de la Unidad Técnico Profesional de la escuela); y un portafolio de desempeño pedagógico. Cada uno tiene pautas precisas aplicadas por personal calificado. El cuidadoso trabajo de diseño y validación de los instrumentos del SEDPD hace de él un caso ejemplar en la región, que se compara favorablemente con los sistemas análogos de los sistemas educativos más desarrollados (Manzi, González y Sun, 2011).

En México, a partir del ANMEB de 1992, el sistema escalafonario se modificó con el esquema de estímulos de Carrera Magisterial que, además de escolaridad, antigüedad y la evaluación del director y el colectivo docente, incluía los factores de preparación profesional y aprovechamiento escolar. En 2008, tras la firma de la Alianza por la Calidad de la Educación por la SEP y el SNTE, se anunció que el

acceso de los aspirantes a las plazas de maestro de educación básica dejaría de ser decidido discrecionalmente para dejar lugar a concursos de oposición, que comenzaron a implementarse en forma desigual según las entidades. En 2011 se dieron dos pasos más: la modificación del peso de los factores considerados en Carrera Magisterial, en los que el factor aprovechamiento escolar, ahora basado en los resultados de las pruebas ENLACE, pasó a tener el 50% del total, y el establecimiento de la llamada "Evaluación universal", que se implementó por primera vez en 2012, con muchas irregularidades.

A solicitud de la SEP, en 2010 la OCDE entregó a México un documento con recomendaciones para implementar un nuevo sistema de evaluación de maestros con rasgos similares al de Chile, pero tal recomendación no ha sido atendida. El reto de contar con un buen sistema en este campo sigue presente.

Evaluación de escuelas

De manera similar a lo que ocurre con la evaluación de docentes, la valoración de la calidad de los centros escolares no se puede basar solamente en los resultados que obtienen los estudiantes en pruebas de rendimiento porque, además de las limitaciones que tienen esos instrumentos, hay muchos factores que inciden en la calidad y, sobre todo, esa noción tiene otras dimensiones importantes, además del aprendizaje de los alumnos. Por ello, la evaluación de centros escolares implica el uso de una gama de herramientas, incluyendo portafolios de evidencias, estudios de los recursos de la escuela y de los procesos que tienen lugar en ella, como el trabajo en equipo del colectivo docente con el liderazgo del director, que producen cierto clima y la participación de los padres de familia.

Tradicionalmente, la función de los supervisores o inspectores escolares incluye de alguna manera la evaluación de las escuelas a su cargo, pero esa función se realizaba en forma poco sistemática. En las últimas décadas se han desarrollado modelos para ese propósito, unos derivados de los estudios de escuelas eficaces y otros en el marco de programas de acreditación o certificación de calidad, a veces relacionados con la familia de normas ISO.

En México, la SEP promovió trabajos de esta naturaleza utilizando modelos de España y Escocia. En el INEE se han realizado también estudios de evaluación en gran escala de recursos y procesos de las escuelas (Martínez Rizo, 2002).

EL FUTURO DE LAS EVALUACIONES

Pruebas de aprendizaje

El uso de pruebas en gran escala para evaluar el nivel de aprendizaje que logran alcanzar los alumnos de diferentes grados y niveles de la educación básica es, probablemente, un elemento que se volverá permanente, tanto en México como en toda América Latina y en el mundo.

Lleva a pensar lo anterior la creciente disponibilidad de la tecnología necesaria para realizar los complejos operativos que implican dichas pruebas, tanto en lo relativo a los conocimientos psicométricos necesarios como en cuanto a lo que se requiere para la aplicación de las pruebas, la captura de los resultados y su procesamiento y análisis. Es previsible que en un futuro no lejano será usual la aplicación en línea en lugar de mediante lápiz y papel, lo que, además de reducir de manera notable los costos, en principio permitirá manejar pruebas de mejor calidad, en particular para la evaluación de niveles cognitivos complejos.

Los riesgos de malos usos de este tipo de evaluaciones, sin embargo, están presentes y no son menores, por el desconocimiento de las cuestiones técnicas por parte de muchos tomadores de decisiones, junto a la presencia de intereses ajenos a lo que tendría que ver estrictamente con la mejora de la calidad educativa en el sentido más completo de la expresión, incluyendo las dimensiones de relevancia y equidad.

Por lo anterior, es posible considerar dos futuros posibles: uno que prolongaría los aspectos negativos actualmente presentes en el uso de pruebas estandarizadas, y otro que, además de aprovechar el potencial de la tecnología, en particular de la informática, enriquecería la visión de los alcances y límites de las evaluaciones en gran escala, viéndolas no como sustituto del maestro, sino como complemento de la tarea de unos docentes profesionales, cuyo trabajo es irremplazable tanto para promover el desarrollo cognitivo e integral de los alumnos como para evaluar de manera integral también su desempeño, en especial con un enfoque formativo.

Indicadores

En este ámbito puede mantenerse durante cierto tiempo la situación que prevalece en muchos países, en el sentido de contar con estadísticas que tienen limitaciones claras y con indicadores elementales basados en ellas. Los avances en el plano internacional, sin embargo, junto con la creciente disponibilidad de capacidades técnicas y de infraestructura, hacen previsible que, en el mediano plazo, en todos los países se producirán avances importantes, gracias a los cuales la calidad de las estadísticas mejorará sustancialmente, y será posible desarrollar indicadores de segunda generación. En México, el trabajo del INEE y el del Coneval hacen pensar que la perspectiva más optimista se puede hacer realidad en corto plazo.

Evaluación de docentes y de escuelas

La tecnología necesaria para poner en marcha sistemas avanzados de evaluación de docentes está disponible, pero implica tiempos largos y una gran capacidad organizativa, además de superar obstáculos de carácter político. Por ello, parece probable que los insatisfactorios sistemas que prevalecen en la mayor parte de los países de la región latinoamericana, incluido México, no se modificarán de manera sustancial en el corto plazo. Sería de esperar que se multipliquen experiencias positivas en sistemas educativos de dimensiones no muy grandes, como los de algunos países o regiones de países grandes. La generalización de sistemas avanzados llevará seguramente más de una década. Algo similar parece probable en lo relativo a evaluación de escuelas.

CONCLUSIÓN

En México y, al parecer, en otros países de la región, en muchos sectores sociales prevalece una gran preocupación por la calidad educativa, y sin duda hay motivos para ello, aunque también abundan lecturas de los datos disponibles que destacan de manera unilateral los puntos débiles e ignoran tanto los fuertes, que también existen, como los elementos contextuales e históricos que deben considerarse para poner en perspectiva unos y otros.

Las visiones negativas de la situación educativa suelen tener una idea también demasiado simple de la evaluación y de lo que habría que hacer para conseguir que la calidad educativa mejore. Muchos críticos no tienen conciencia de lo difícil

que es alcanzar buenos resultados con alumnos de origen social desfavorable, y es frecuente que vean con simpatía estrategias simplistas, como pensar que las fallas de la escuela pública se podrían corregir con facilidad en escuelas privadas como las que atienden a sus hijos, e ignoran que menos de diez por ciento de los mexicanos asisten a ellas. Por eso, hay opiniones de que bastará con aplicar pruebas masivamente y tomar medidas correctivas simples para que la calidad de la educación mejore de modo substancial.

En otras palabras, las estrategias simplistas de mejora parten de un supuesto falso: hacer buena educación en cualquier contexto es fácil:

Los sistemas de rendición de cuentas basados en pruebas se basan en la creencia de que la educación pública puede mejorar gracias a una estrategia sencilla: haga que todos los alumnos presenten pruebas estandarizadas de rendimiento, y asocie consecuencias fuertes a las pruebas, en la forma de premios cuando los resultados suben y sanciones cuando no ocurra así (Hamilton, Stecher y Klein, 2002).

La evaluación educativa —con pruebas estandarizadas, sistemas de indicadores u otros instrumentos— puede contribuir a que la calidad mejore, pero ello no se produce automáticamente, ya que en algunos casos la evaluación puede tener consecuencias contraproducentes. Lo indispensable para que la calidad mejore es un trabajo intenso y bien orientado por parte de maestros y escuelas, con apoyo del resto de la sociedad. Para ello, ayudan muchas cosas, incluyendo buenas evaluaciones. Lo que no hace falta es un uso de los resultados que desorienta a tomadores de decisiones y sociedad, lleva a unos a jactarse de logros dudosos y a otros, entre los que se cuentan personas serias y comprometidas, a perder incluso la poca esperanza que aún tenían (Martínez Rizo, 2012).

REFERENCIAS BIBLIOGRÁFICAS

- Agencia Ejecutiva en el Ámbito Educativo, Audiovisual y Cultural (AEAEC) (2009). *Pruebas nacionales de evaluación del alumnado en Europa: objetivos, organización y utilización de los resultados*. EURYDICE-EACEA-Ministerio de Educación.
- Ferrer, G. (2006). *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Washington: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, PREAL.
- Hamilton, L., Stecher, B. & Klein, S. (Eds.) (2002). *Making sense of test-based accountability in education*. Santa Mónica, CA: Rand Corporation.
- INEE (2006). *Plan Maestro de Desarrollo 2007-2014*. México.
- INEE-SEP (2006). *Sistema de Indicadores Educativos de los Estados Unidos Mexicanos. Conjunto básico para el ciclo 2005-2005*. México.
- Manzi, J., González, R. y Sun, Y. (2011). *La evaluación docente en Chile*. Santiago: MIDE-Universidad Católica.
- Martínez Rizo, F. (en prensa). Los indicadores y su construcción: qué esperar y qué cuidar. En Roca, E. y Kisilevsky, M. (Coords.). *Indicadores, metas y políticas educativas*. Madrid: OEI-Santillana.
- _____. (2012). Usos y abusos de la evaluación educativa. Experiencias de México. En Gavía, J. L. (Ed.). *Liber Amicorum Arturo de la Orden*. Madrid.
- _____. (2011, octubre). *La evaluación del profesorado*. Presentada en VII Jornadas de

- Cooperación Educativa Iberoamericana sobre Educación. México.
- ____ (2010a). *Assessment in the context of educational policy: Latin American Countries*. En Paterson, P., Baker, E. & McGaw, B. (Eds.). *International Encyclopedia of Education* (3rd Ed.). Oxford-New York: Elsevier.
- ____ (2010b, agosto). Usos y abusos de la evaluación. *Este País* (232), 24-27.
- ____ (2007). Marco conceptual y metodológico. En *Propuesta y experiencias para desarrollar un sistema nacional de indicadores educativos* (pp. 59-88). México: INEE.
- ____ (2002). Anexos 1 y 2. Revisión de literatura y bibliografía. En Varios. *La calidad de la educación en México: perspectivas, análisis y evaluación* (pp. 305-454). México: Miguel Ángel Porrúa-SEP.
- Martínez Rizo, F. y Blanco, E. (2010). La evaluación educativa. Experiencias, avances y desafíos. En Arnaut, A. y Giorguli, S. (Eds.). *Educación. Los grandes problemas de México*, vol. V (pp. 89-123). México: El Colegio de México.
- Martínez Rizo, F. y Santos del Real, A. (2009). Consideraciones sobre la evaluación educativa. En de Alba, A. (Ed.). *¿Qué dice la investigación educativa?* (pp. 265-304). México: COMIE.
- OEI (2010). *Metas 2021. La educación que queremos para la generación de los bicentenarios*. Madrid: Organización de Estados Iberoamericanos.
- ____ (2011). *2011. Miradas sobre la educación en Iberoamérica*. Madrid: Organización de Estados Iberoamericanos.
- PRIE (2003). *La experiencia del Proyecto Regional de Indicadores Educativos 2000-2003*. Santiago: Ministerio de Educación-OREALC.
- ____ (2011). *Panorama Educativo 2010: desafíos pendientes*. México: SEP-OREALC-OEA.
- Prostik, J. (1996). History of teacher pay and incentives reform. *Journal of School Leadership*, 6 (3), 265-289.
- SITEAL (s.f.). *Boletín de difusión*. Recuperado de http://www.siteal.iipe-oei.org/sites/default/files/gacetilla_siteal.pdf
- UNESCO (2008). *Education for All. Global Monitoring Report 2008*. París. *National learning assessments by región and country* (pp. 208-220); *Latin America and the Caribbean* (pp. 214-215).
- Wolff, L. (2004). Educational Assessments in Latin America: The State of the Art. *Applied Psychology: An International Review*, 53 (2), 192-214.