FELIPE MARTÍNEZ RIZO*

El sistema nacional de evaluación de méxico

MÉXICO

Introducción

En este trabajo se presenta primero un panorama del desarrollo de la evaluación educativa en México hasta fines del siglo XX; después, se informa sobre los avances que han tenido lugar desde el año 2001 al 2006; luego, se presentan los proyectos en curso para integrar los diversos esfuerzos de evaluación en un verdadero sistema nacional. En la conclusión, y a partir de la experiencia del país, se esbozan los rasgos de un sistema de evaluación deseable para un país como México.

Se desarrolla con mayor amplitud lo relativo a la educación básica, con menos información sobre la evaluación de la educación superior y la que en México se denomina media superior (alumnos de 15 a 17 años). Se incluyen reflexiones sobre la difusión y uso de resultados, por considerar que la evaluación no constituye un fin en sí misma, sino que cobra sentido en la medida en que contribuya a la mejora de la calidad educativa.

1. La evaluación educativa en México hasta 2000

Como ocurre en otros sistemas educativos, en México la evaluación del aprendizaje alcanzado por los alumnos ha sido una práctica a cargo de cada maestro. Cuando el sistema educativo comenzó a experimentar el vertiginoso crecimiento que la explosión demográfica trajo consigo, en la segunda mitad del siglo XX, los maestros utilizaban ampliamente pruebas de tipo objetivo que, aunque eran técnicamente rudimentarias, eran también homogéneas. Se difundió la noción elemental de evaluación referida a la norma estadística, con la expresión de *calificar por curva*. En una escala en la que la nota mínima es 5 y la máxima 10, eso hacía que en cada grupo hubiera una media cercana a 8, con menos notas de 9 y 7 y muy pocas de 6 y 10.

Precedida por trabajos realizados en instituciones de educación superior, en la década de 1970 comenzó a surgir la aplicación de pruebas estandarizadas en gran escala en educación básica. La Secretaría de Educación Pública (SEP) desarrolló evaluaciones de este tipo para el ingreso de alumnos a secundaria y luego para valorar el aprendizaje de los alumnos de primaria mediante pruebas estandarizadas, aplicadas a muestras nacionales de alumnos.

Los resultados, sin embargo, no se difundían y se usaban poco para retroalimentar la toma de decisiones. Las evaluaciones tenían también serias deficiencias técnicas que impedían asegurar la comparabilidad de versiones y la estabilidad de escalas; por ello no podían hacerse en forma consistente comparaciones de subsistemas estatales u otros y, menos aún, análisis longitudinales.

* Director General del Instituto Nacional para la Evaluación de la Educación - México

[1] Este texto se basa en documentos anteriores del autor, cuyas referencias pueden verse al final (Martínez Rizo 2003, 2005a, 2005b y 2005c), así como en un trabajo preparado por Banegas y Blanco, por parte del INEE, para la VI Reunión Ministerial del Grupo E-9 (Monterrey, México, del 13 al 15 de febrero de 2006).

A partir de 1992, y en el marco de un importante proceso de descentralización del sistema nacional de educación básica, el desarrollo de sistemas de evaluación en gran escala se fortaleció. Inicialmente se desarrollaron pruebas de aprendizaje para evaluar los programas compensatorios que comenzaron a implementarse con apoyo de organismos financieros internacionales; luego, este tipo de evaluaciones se extendió a otros ámbitos del sistema educativo. Comenzaron también a hacerse evaluaciones cualitativas, se impulsó la creación de Áreas de Evaluación en las secretarías de educación de las 32 entidades federativas del país y se constituyó un mecanismo de colaboración entre ellas.

En 1993 se estableció el programa de estímulos económicos para maestros llamado Carrera Magisterial, en cuyo marco comenzaron a desarrollarse instrumentos para evaluar tanto a los propios docentes como a sus alumnos, en números muy grandes: en la actualidad, se aplican anualmente pruebas a unos ocho millones de alumnos y alrededor de medio millón de maestros. En 1998 comenzaron a aplicarse anualmente otras pruebas (Estándares Nacionales de Lectura y Matemáticas) a muestras de alumnos de primaria y secundaria de todo el país, con propósitos de evaluación del sistema como tal.

En la segunda mitad de la última década del siglo XX comenzaron también a fortalecerse los sistemas de evaluación de algunos estados de la República y dio inicio la participación de México en proyectos internacionales de evaluación, con las pruebas del Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS) en 1995; las del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE) en 1997; y las del Programa para la Evaluación Internacional de Estudiantes (PISA) de la Organización para la Cooperación y el Desarrollo Económico (OCDE) en 2000.

En educación superior, en 1994 se creó el Centro Nacional para la Evaluación de la Educación Superior, CENEVAL, que hace pruebas de ingreso a educación media superior, licenciatura y posgrado, así como de egreso de licenciatura. En 1991 habían surgido los Comités Interinstitucionales de Evaluación de la Educación Superior (CIEES), para la evaluación de programas e instituciones; a ellos siguieron diversas agencias acreditadoras especializadas y, a fines del 2000, el Consejo para la Acreditación de la Educación Superior, COPAES.

Pese a los avances alcanzados hasta fines del siglo pasado, los resultados de las evaluaciones seguían sin difundirse y las limitaciones técnicas de instrumentos y procesos persistían.

2. De 2001 a 2006: avances de la SEP e inicio de actividades del INEE

En los primeros años del siglo XXI el desarrollo de la evaluación educativa continuó. La Dirección General de Evaluación de Políticas de la SEP continuó manejando los proyectos anteriormente mencionados y a desarrollar uno más, llamado *evaluación diagnóstica*, que busca ofrecer instrumentos de buena calidad a las escuelas y los docentes de todo el país para que se hagan aplicaciones masivas no controladas, cuyos resultados puedan contrastarse con los de muestras nacionales controladas. La SEP comenzó también a ampliar la difusión de resultados. En el marco del importante Programa de Escuelas de Calidad comenzó a desarrollarse un ambicioso proyecto de evaluación cualitativa y algunos estados fortalecieron sus propias evaluaciones y comenzaron a difundir resultados.

Por otra parte, con base en un anteproyecto elaborado desde la segunda mitad de 2000, en agosto de 2002 se creó el Instituto Nacional para la Evaluación de la Educación (INEE), con la tarea de asumir las evaluaciones cuyo propósito es valorar la calidad del sistema educativo como tal, por lo que pasaron a su ámbito de acción las anteriores pruebas de Estándares Nacionales. Al INEE corresponde también lo relativo a la participación de México en proyectos internacionales de evaluación educativa.

El INEE difundió un primer informe a fines de 2003, con resultados de las pruebas de Estándares

de lectura y matemáticas aplicadas en junio de ese año a muestras nacionales de alumnos de los grados 3^0, 5^0 y 6^0 de primaria y 1^0, 2^0 y 3^0 de secundaria (7^0, 8^0 y 9^0). Tanto el desarrollo como la aplicación de los instrumentos habían sido hechos todavía por la SEP, cubriendo en total a unos 300,000 alumnos.

Además de presentar los resultados de dichas aplicaciones, y un conjunto de indicadores del contexto y los recursos de las escuelas, el primer informe del INEE incluyó un riguroso análisis de la calidad de los instrumentos utilizados en el pasado, detectando sus debilidades, que impedían la comparación de los resultados de 2003 con los obtenidos en las aplicaciones de los años anteriores. Además de problemas con la confiabilidad de los instrumentos, el hecho de que las pruebas fueran elaboradas con un enfoque normativo y sin una clara alineación al currículo impedía hacer juicios sólidos sobre el nivel de rendimiento alcanzado por los alumnos del sistema educativo nacional.

El INEE señaló también que, para la evaluación del sistema educativo como tal, no era necesaria la aplicación de un gran número de pruebas en seis grados cada año, sino que era más adecuado un enfoque parsimonioso, con ciclos adecuados de evaluaciones en grados y áreas curriculares clave, con muestras representativas del sistema nacional y sus principales subsistemas. El Instituto señaló, además, la necesidad de racionalizar y articular el gran número de evaluaciones existente, que representan un esfuerzo muy grande y ofrecen resultados de dudoso valor.

Las legítimas demandas de diversos actores por tener información confiable sobre la educación, y el escaso conocimiento de aspectos técnicos, llevaron a una proliferación de proyectos de evaluación cuyo diseño no siempre es el apropiado para el propósito perseguido. Es frecuente que el propósito mismo no esté claro, o que se quiera usar un instrumento elaborado con cierto propósito, para otros para los que no resulta adecuado.

A partir de la experiencia de su primer año de trabajo, el INEE se propuso desarrollar, a lo largo de los tres años siguientes, varias tareas precisas:

- Desarrollar una nueva generación de pruebas para la evaluación del sistema nacional de educación básica de manera válida y confiable.
- Hacer un Plan de Desarrollo que defina las evaluaciones a realizar a mediano plazo.
- Impulsar de manera decisiva la difusión de resultados para su uso por diversos actores: autoridades, maestros y sociedad.
- Promover la integración de los diversos esfuerzos de evaluación existentes en un auténtico sistema nacional de evaluación educativa, que no constituya una nueva y pesada estructura, pero sí incluya una normatividad básica común y mecanismos eficientes de coordinación y control de calidad.

En 2004 se aplicaron, por última vez, instrumentos desarrollados con la metodología anterior, y los resultados de esa aplicación se difundieron a fines del mismo año, junto con los resultados de PISA 2003. El informe nacional de estos últimos se difundió al mismo tiempo que los resultados internacionales, el seis de diciembre de 2004.

Al mismo tiempo, se avanzó en las cuatro direcciones apuntadas:

- Se desarrollaron instrumentos nuevos para la evaluación de lectura y matemáticas de 6^0 de primaria y 3° de secundaria, con una metodología rigurosa e involucrando a decenas de maestros, especialistas en contenidos y expertos en psicometría. Estas pruebas se han llamado *Exámenes de la calidad y el logro educativo* (Excale).
- Se preparó y aprobó un Plan Maestro de Desarrollo, que precisa los proyectos de evaluación a desarrollar hasta 2016, y define de manera explícita los criterios de calidad a utilizar y los mecanismos de control que garantizarán que se respeten.

- Se sistematizaron los esfuerzos de difusión de resultados iniciados desde 2003, incluyendo presentaciones a las más altas autoridades del país y de las 32 entidades federativas, así como a la sociedad en general, en especial a través de los medios de comunicación, para cuyo personal se llevan a cabo sistemáticamente actividades de capacitación.
- Para atender la legítima demanda de la sociedad de conocer resultados escuela por escuela, lo que escapa a su ámbito de competencia, el INEE comenzó a apoyar a algunas entidades para la difusión de resultados propios llegando hasta ese nivel de desagregación.
- El trabajo para conseguir la integración del Sistema Nacional de Evaluación incluyeron la elaboración de un documento de Política Nacional de Evaluación y múltiples gestiones ante los responsables de diversos proyectos, las autoridades educativas de las entidades federativas y los legisladores.

En 2005 los esfuerzos en las cuatro líneas de trabajo continuaron:

- Se aplicaron por primera vez las pruebas Excale hechas en 2004, y se desarrollaron otras para la evaluación de lectura, matemáticas, ciencias naturales y ciencias sociales en 3^0 de primaria, a aplicarse en 2006. Se hizo además una aplicación especial de las pruebas aplicadas en el año 2000 para poder hacer una comparación de los resultados de ese año con los obtenidos con los mismos instrumentos en 2005.
- Se cumplió la primera etapa anual del Plan Maestro de Desarrollo.
- Continuó el trabajo de difusión de resultados. En diciembre se presentaron los resultados de las evaluaciones de 2005 al Presidente de la República, el Secretario de Educación y los secretarios de las entidades federativas. Prácticamente todos los diarios, estaciones de radio y de televisión importantes de la capital del país entrevistaron a los directivos del INEE.
- Quedó terminado el proyecto para integrar formalmente el Sistema Nacional de Evaluación Educativa, estando pendiente su aprobación en la forma de un acuerdo secretarial, un decreto presidencial o una ley.

3. Balance a principios de 2006 y perspectivas para el futuro

3.1. Cantidad y calidad de las evaluaciones nacionales

En términos cuantitativos, las actividades relacionadas con evaluación educativa que se han realizado en México anualmente en la última década son muy numerosas. Además de las estadísticas educativas que todo director de escuela debe llenar al inicio y fin de cada ciclo escolar, la Dirección General de Evaluación de la SEP ha elaborado y aplicado cada año diversas pruebas a poblaciones que van de varios miles a siete u ocho millones de alumnos. Las evaluaciones cualitativas del Programa de Escuelas de Calidad y otras implican muchos miles de horas hombre cada año. Sin desconocer los avances respecto a la situación prevaleciente hace 15 años, hay todavía de deficiencias técnicas, dada la ausencia de una tradición mexicana en el campo de la evaluación y de limitaciones derivadas de cuestiones burocráticas.

La creación del INEE no significó un aumento en la cantidad de las evaluaciones. Como se ha dicho, el nuevo Instituto retomó uno de los proyectos previamente existentes y redujo tanto el número de grados evaluados como la frecuencia de las evaluaciones. En cambio, el INEE emprendió un importante esfuerzo de mejora técnica, que a principios de 2006 ha llevado ya a la primera aplicación de una nueva generación de pruebas de calidad claramente mejor a las que se habían desarrollado previamente.

3.2. La participación de México en evaluaciones internacionales

La decisión tomada por México, a mediados de la década de 1990, de comenzar a participar en evaluaciones internacionales, fue un paso correcto en la dirección de fortalecer la evaluación. La ventaja más obvia es que nuestro país puede comparar su situación

con la de otros. Además, y dado que esas evaluaciones son hechas por grupos que reúnen a destacados especialistas, una participación activa puede ser un vía muy eficiente de capacitación y aprovechamiento de la experiencia acumulada sobre el tema.

México ha participado en las pruebas del TIMSS, del LLECE y PISA. La participación de México en esos ejercicios, sin embargo, ha tenido deficiencias importantes, que redujeron sus beneficios potenciales. Hasta 2003, la participación mexicana no incluyó intervenir en la planeación de las evaluaciones, como pueden hacer todos los países involucrados.

La limitada participación se manifestó también en la ausencia de análisis propios de los resultados obtenidos por los alumnos mexicanos en esos ejercicios, con lo que dejaba de obtenerse el beneficio de permitir que se sacaran de ellas conclusiones y experiencias útiles para mejorar la calidad de la educación nacional.

A partir de 2003 la situación comenzó a cambiar, al considerarse que la participación de México no debía ser de simple operador de pruebas desarrolladas en otros lugares. Además del reducido impacto en el sistema educativo de una participación limitada en evaluaciones internacionales, el desarrollo de la capacidad técnica nacional gracias a la interacción con los especialistas de otros países, tampoco se da. Tanto en lo que se refiere a la aplicación de PISA en 2006, como a la de LLECE, la participación de México es ahora muy activa.

3.3. La difusión de los resultados de las evaluaciones

Como la evaluación es un medio para la mejora de la calidad, debe reiterarse la importancia de la escasa difusión de resultados que prevaleció hasta hace poco y, en consecuencia, su casi nula utilización para la toma de decisiones desde el aula hasta el sistema nacional. Esta deficiencia explica la pobreza de la cultura de la evaluación que se observa tanto entre alumnos y maestros como entre directivos escolares, padres de familia, medios de comunicación y los diversos sectores sociales, e incluso entre tomadores de decisiones de alto nivel.

La no difusión contribuye, además, a que la calidad de las evaluaciones no mejore, ya que ni siquiera se detectan sus deficiencias. Una consecuencia más, que se deriva de las anteriores y, a su vez, las refuerza, es la ausencia de programas para formar especialistas en evaluación.

Una buena difusión no puede reducirse a poner las bases de datos y los análisis que se hagan de los resultados de las evaluaciones al alcance del público, por ejemplo por medio de una página web. Es necesario tener en cuenta las necesidades y posibilidades específicas de los diversos destinatarios y usuarios potenciales que son, al menos, las autoridades a nivel nacional y de las entidades federativas; las de nivel medio, como jefes de sector y supervisores; los directores de escuela y los maestros, así como las instituciones en que se preparan éstos; los investigadores y estudiosos; los padres de familia; y la sociedad en general, para llegar a la cual tienen un lugar especial los medios de comunicación.

Hasta finales de la década de los 90 las evaluaciones hechas en el sistema educativo mexicano se mantuvieron como material reservado a unos cuantos altos funcionarios, lo que significó también un uso muy reducido para la toma de decisiones. A partir de 1999 la situación comenzó a cambiar, lo que ha sido notorio a partir de 2003, gracias en especial al trabajo del INEE, tanto en dirección de las autoridades como de la sociedad.

En el primer sentido, destacan las presentaciones ya mencionadas de informes de los resultados de evaluaciones anuales a las autoridades educativas nacionales y de las entidades federativas. En lo relativo a la sociedad en general, sobresalen los

esfuerzos, también ya mencionados, por involucrar a los medios de comunicación, con talleres especiales para periodistas y un acercamiento sistemático con ellos, que se ha visto reflejado en una cobertura cada vez más amplia y, al parecer, también más objetiva.

La difusión más difícil es la que se refiere a los maestros, a quienes no basta dar resultados globales y análisis generales, sino que requieren materiales que les permitan utilizar los resultados de las evaluaciones para mejorar su propio trabajo docente. Los primeros esfuerzos del INEE en este sentido comienzan a difundirse solamente a fines de 2005, en la forma de una obra titulada *PISA para docentes. La evaluación como oportunidad de aprendizaje*.

Se editaron 250,000 ejemplares de esa obra, que se han hecho llegar a todas las escuelas de enseñanza media básica y superior del país, además de ponerse en la página web del INEE. El libro presenta las unidades de reactivos liberadas de PISA 2000 y 2003, con resultados nacionales e internacionales que permiten al docente contrastar los resultados de sus propios alumnos con los de grupos de interés; se incluyen además ejemplos de análisis de reactivos desde una perspectiva pedagógica, para ilustrar las implicaciones que deberían tener los resultados de PISA sobre el trabajo docente. El INEE pretende continuar trabajando en esa dirección, difundiendo anualmente materiales para docentes derivados de las pruebas Excale.

3.4. El análisis de los resultados y su utilización para la toma de decisiones

Hasta 2003, además de las fallas técnicas y la falta de difusión ya mencionadas, la información derivada de los trabajos de evaluación (pruebas, series de indicadores, estudios cualitativos de escuelas, proyectos internacionales) no era procesada e integrada para formar un todo coherente, que fuera base sólida para tomar decisiones. Tanto a nivel nacional como, todavía más, en el ámbito de las entidades federativas, había muy pocos análisis de resultados de estadísticas y evaluaciones enfocados expresamente a la formulación o reformulación de políticas.

Para subsanar esta deficiencia, y además de elaborar los informes anuales a los que se ha hecho mención, el INEE ha promovido, desde 2004, la elaboración de informes similares en cada una de las entidades federativas, aprovechando no sólo los resultados de las evaluaciones del propio Instituto, sino también los de otros proyectos de evaluación de la SEP, sus indicadores y estadísticas educativas, y los datos del contexto socioeconómico que pueden encontrarse en las bases de datos y las publicaciones del Instituto Nacional de Estadística, Geografía e Informática (INEGI) y otras fuentes. Para impulsar la preparación de tales informes se organizaron talleres para el personal de las entidades y se ha conseguido que en 30 de las 32 existentes se elaboren.

En 2005 comenzaron a verse casos de decisiones de política educativa a nivel nacional basadas en los resultados de las evaluaciones del INEE. El primer ejemplo fue un programa especial de apoyo a las escuelas telesecundarias, lanzado a principios de 2005, con una partida especial de 500 millones de pesos.

Al recibir los resultados de las evaluaciones de este último año, en diciembre de 2005, el Presidente de la República anunció un nuevo programa de apoyo, esta vez para el conjunto de las secundarias, con recursos especiales por seis mil millones de pesos.

Se anunció también que se emprendería una revisión en profundidad del modelo de cursos comunitarios, modalidad de servicios educativos que mostró resultados insatisfactorios en la evaluación comparativa 2000-2005 hecha por el INEE, y que atiende al sector de la población que tiene condiciones de vida más desfavorables, ya que se trata de quienes viven en el gran número de localidades de menos de 250 habitantes.

3.5. Balance y prospectiva

Una apreciación global de la situación de la evaluación educativa en México hasta principios

del siglo XXI, distingue dos grandes etapas, a las que puede añadirse una más, que no es otra que la que se considera deseable para el futuro.

En la primera etapa, que comprende desde la década de 1970 hasta fines del siglo XX o principios del XXI, el INEE sostiene que, *con cierta crudeza*, puede afirmarse que *en México hay muchas actividades relacionadas con la evaluación educativa pero no hay una verdadera evaluación*, en el sentido de un conjunto de juicios de valor objetivos y mesurados sobre el estado del sistema educativo, que integre los datos derivados de mediciones y estadísticas, los contraste con parámetros pertinentes, explore explicaciones y utilice ampliamente los resultados para sustentar decisiones de mejora en todos los niveles del sistema.

En la segunda etapa, que comprende el primer lustro del siglo XXI, la situación ha comenzado a cambiar, tanto en lo relativo a los aspectos técnicos (calidad de los instrumentos y los análisis de resultados), como en cuanto a la difusión y uso de éstos y también a la capacitación del personal dedicado a la evaluación y a la articulación de los diversos esfuerzos evaluativos en un verdadero sistema. La tercera etapa se concibe como una en que todos los aspectos se habrán desarrollado cualitativamente en la forma deseable. La situación se resume en este cuadro:

Tres etapas en la evaluación nacional

ETAPAS	Calidad de los instrumentos	Calidad de los análisis	Difusión y uso de los resultados	Capacitación del personal	Articulación del sistema
ETAPA I (Hasta 2000)	Noción de calidad restringida. Poca atención a factores contexto. Escalas simples. Resultados no comparables en el tiempo.	Inexistentes o escasos. Cuando los hay, solamente descriptivos, con técnicas primarias.	Difusión casi nula. No se utilizan los resultados para diseñar políticas o mejorar prácticas pedagógicas.	Baja competencia del personal de evaluación, tanto a nivel federal y con mayor razón en las entidades.	Evaluaciones no articuladas. No hay un sistema de indicadores a nivel nacional. Poca colaboración entre actores.
ETAPA II (2000-2005)	Noción de calidad multi-dimensional, trasciende pruebas. Planeación sistemática. Atención al contexto. Escalas complejas.	Descriptivos y asociativos. Se identifican factores asociados.	Difusión amplia. Interpretación superficial. Uso ocasional para decisiones. Uso limitado a nivel de escuelas.	Profesionalización en progreso en el nivel federal. Acciones de capacitación para las entidades.	Colaboración frecuente entre SEP-INEE-AEE. Sistematización incipiente de evaluaciones y de sistema nacional de indicadores.
ETAPA III (futuro deseable)	Además de II: Marcos teóricos sólidos-explícitos. Resultados que se pueden comparar en el tiempo.	Complejos-varios. Adecuados al objeto (multinivel, longitudinal). A cargo de los organismos de evaluación y de investigadores	Difusión amplia y adecuada a las necesidades de todos los actores. Uso en todos los niveles para mejorar la calidad de la educación.	Profesionalización alta en el nivel federal y estatal. Capacitación de directores y maestros para uso de evaluaciones y autoevaluaciones.	Sistema integrado y eficiente. Sistema nacional de indicadores complejo y dinámico. Colaboración permanente. independientes.

Por la importancia de la difusión y uso de los resultados, en un cuadro más se presenta con mayor detalle la situación prevaleciente en las dos etapas transcurridas, hasta fines del siglo XX y en el primer lustro del XXI, así como en la tercera etapa, que se visualiza como deseable. En el cuadro se distinguen los distintos públicos destinatarios de la difusión, que son a la vez actores para usos específicos de los resultados.

Tres etapas en la difusión y el uso de las evaluaciones en México

ETAPAS	Autoridades Educativas nacionales y estatales	Supervisores, Directores y Maestros	Familias	Sociedad en general
ETAPA I (Hasta 2000)	Difusión muy limitada. No se usan resultados para tomar decisiones.	No se difunden resultados.	No se difunden resultados.	No se difunden resultados.
ETAPA II (2000-2005)	Difusión amplia. Resultados informan decisiones de política en algunos casos.	Difusión limitada. Evaluación no se utiliza sistemáticamente para mejorar las prácticas pedagógicas.	Difusión limitada. Primeras iniciativas a nivel de las entidades.	Difusión amplia. Resultados, en ocasiones, se interpretan de forma incorrecta o sesgada.
ETAPA III (futuro deseable)	Difusión amplia. Resultados son un insumo fundamental para formular políticas y tomar decisiones.	Difusión amplia. Evaluación retroalimenta de manera permanente las prácticas pedagógicas.	Difusión amplia. Los resultados permiten rendir cuentas a las familias y participar en los procesos escolares.	Difusión amplia. Resultados generan un debate público informado y responsable.

Conclusión: hacia un sistema deseable de evaluación educativa

El amplio consenso que rodea la idea de que la evaluación es importante, no debe hacer perder de vista que no cualquier evaluación es positiva para la educación. Mal concebida o realizada, la evaluación puede ser irrelevante, en el mejor de los casos, o incluso destructiva. Los sistemas educativos no necesitan evaluaciones de cualquier tipo, aunque sean destructivas, sino unas adecuadas, que contribuyan efectivamente a que la calidad educativa mejore.

La preocupación por la calidad de la educación es justificada. En los países en vías de desarrollo los resultados educativos tienen niveles inferiores a los de países más avanzados, lo cual es explicable, pero no puede dejar indiferente, ya que un mejor nivel educativo es condición necesaria, aunque no suficiente, de mejores niveles de desarrollo económico y social.

Algunas personas preocupadas por la educación tienen una visión simplista tanto de las causas de la situación como de la forma de valorarla y corregirla. Los diagnósticos simplistas no toman en cuenta los factores de las escuelas y su contexto que influyen en la calidad; los medios simplistas de valorar ésta se reducen a la aplicación de pruebas de rendimiento, sin distinguir su enfoque y nivel técnico; y las estrategias para remediar la situación se reducen a proponer que se asignen recursos a las escuelas en función de los resultados de sus alumnos en las pruebas.

En realidad, los aspectos de la calidad educativa y los factores que la explican son complejos, por lo que una valoración adecuada de ellos no

puede alcanzarse con un solo acercamiento. Las estrategias de mejora deben ser complejas también, y los resultados sólo podrán observarse en el mediano y largo plazo, como fruto de esfuerzos serios y sostenidos.

Como consecuencia de comprensiones superficiales, en muchos países se observa una proliferación de pruebas con las que se pretende valorar la calidad de las escuelas, aunque los instrumentos no hayan sido diseñados de manera apropiada para tal propósito y aunque, en muchos casos, adolezcan de deficiencias técnicas. Aparecen, entonces, dos tipos de reacciones: los maestros se ven impulsados a orientar su trabajo pedagógico en función de los contenidos de las pruebas, ya que serán el medio privilegiado para valorar su trabajo; por otra parte, entre los mismos maestros y entre los estudiosos de la educación, es frecuente un rechazo absoluto a las pruebas, por considerarlas intrínsecamente inadecuadas para valorar correctamente la calidad.

La tesis central de esta conclusión es que la evaluación que necesitan los sistemas educativos no puede reducirse a pruebas, aunque debe incluirlas, a condición de que sean de buena calidad y se usen de manera parsimoniosa, pero que debe haber indicadores de otras dimensiones de la calidad, que utilicen datos numéricos sobre aspectos más fácilmente cuantificables, como la cobertura, junto con acercamientos cualitativos, apropiados a la naturaleza de los sujetos y de los procesos pedagógicos. Así, y sólo así, la evaluación será una herramienta que contribuya efectivamente al mejoramiento de las escuelas, y que sea apreciada como tal por los maestros, los estudiosos y, en general, todas las personas preocupadas por la educación.

La calidad educativa

Para que pueda haber buena evaluación de la calidad educativa, la primera condición es que se parta de un concepto adecuado de esta compleja noción. Un concepto pobre de calidad puede traer consigo consecuencias graves, al poner el énfasis en unos aspectos en detrimento de otros.

Hay, por ejemplo, quienes oponen *cobertura educativa*, como algo meramente *cuantitativo*, a *calidad educativa*, refiriendo este concepto al aprendizaje de los alumnos. Se dice: en el pasado ha prevalecido el interés por la cantidad (cobertura); ahora debe preocuparnos la calidad. En realidad un sistema educativo no puede considerarse de buena calidad si atiende sólo a una fracción de la demanda, aunque consiga que sus alumnos tengan alto nivel de aprendizaje. En el mundo actual, al menos en los niveles obligatorios, un sistema educativo sólo podrá considerarse bueno si consigue altos niveles de aprendizaje de todos los alumnos de cada cohorte o, al menos, de una muy alta proporción de ellos.

Aplicada a un sistema educativo, la noción de calidad implica, al menos, los aspectos de pertinencia y relevancia, eficacia interna y externa, impacto, eficiencia y equidad. Concebida en esta forma amplia la noción, se sostiene que un sistema educativo es de buena calidad si:

- Establece un currículo adecuado a las circunstancias de la vida de los alumnos (**pertinencia**) y a las necesidades de la sociedad (**relevancia**).
- Logra que la más alta proporción de destinatarios acceda a la escuela, permanezca en ella hasta el final del trayecto y egrese alcanzando los objetivos de aprendizaje establecidos en los tiempos previstos para ello. (**Eficacia interna**).
- Consigue que los aprendizajes sean asimilados en forma duradera y se traduzcan en una vida adulta más plena, no sólo en lo económico, sino también en lo cívico y lo cultural, con lo que la educación será fructífera para sociedad e individuos. (**Eficacia externa e impacto**).

- Cuenta con recursos humanos y materiales suficientes, y los usa de la mejor manera posible, evitando derroches y gastos innecesarios. (**Eficiencia**).
- Tiene en cuenta la desigual situación de alumnos y familias, las comunidades y las escuelas, y ofrece apoyos especiales a quienes lo requieren, para que los objetivos educativos sean alcanzados por el mayor número posible de estudiantes. (**Equidad**).

Evaluación y medición

Para tener buenas evaluaciones no basta partir de un concepto rico de calidad educativa. Es necesario tener también un concepto adecuado de evaluación, lo que no siempre se da, ya que suele haber concepciones limitadas al respecto. Si se aplican pruebas para medir el aprendizaje, por ejemplo, suele pensarse que se ha *evaluado*, cuando en realidad sólo se ha *medido*. *Evaluar* exige algo más: comparar el resultado de la medición con un punto de referencia que establezca lo que se debería saber, para llegar a un juicio sobre lo adecuado o inadecuado del aprendizaje.

El concepto de evaluación que se propone es más amplio. Establece que *la evaluación es el juicio de valor que resulta de contrastar el resultado de la medición de una realidad empírica con un parámetro normativo previamente definido*.

Una buena evaluación no debe caracterizarse solamente por cualidades técnicas como confiabilidad y validez. Estas dos características son indispensables en toda buena medición, y como evaluar implica medir, la evaluación también las requiere. Pero como evaluar va más allá de medir, una buena evaluación implica también otras características:

- *Carácter comprensivo de la conceptualización que la sustente*, que deberá atender todas las dimensiones de la calidad educativa, y no sólo algunas, como el nivel de aprendizaje. Es por ello que un sistema de evaluación no puede limitarse a pruebas de rendimiento.
- *Alto nivel técnico de las mediciones en que se base*, para garantizar su confiabilidad y validez, mediante diseños sólidos; el uso de enfoques diversos y complementarios en modelos e instrumentos; la selección cuidadosa de muestras representativas; rigurosos procesos de recolección de datos; y análisis cuidadosos de los datos obtenidos.
- *Pertinencia de los referentes que se definan como parámetros* para contrastar con ellos los resultados de la medición, de manera que las comparaciones tengan sentido. Los referentes se definen normativamente, no se derivan de los datos empíricos. Pueden ser óptimos: ideales con que se compara una situación; promedios de los individuos que se evalúa; y mínimos, con los menores valores aceptables. Cada uno arroja cierta luz sobre lo evaluado y ninguno es suficiente. Convendrá usar los tres tipos de parámetro, para una mejor apreciación.
- *Mesura de los juicios de valor* derivados de contrastar mediciones y parámetros, que evitarán excesos triunfalistas o derrotistas y tendrán siempre en cuenta el valor de equidad, considerando el contexto de los alumnos y las escuelas.
- *Amplitud, oportunidad y transparencia de la difusión de resultados*, que deberá llegar a los sectores involucrados en versiones adecuadas a las características de cada uno.

La evaluación educativa necesaria: un decálogo de características

Para que la evaluación sirva realmente para mejorar la calidad deberá tener los rasgos definidos en el punto anterior: *atender todas las dimensiones de la calidad educativa; ser técnicamente sólida con acercamientos variados y complementarios; usar referentes pertinentes; culminar en juicios mesurados y contextualizados; y difundirse amplia, oportuna y transparentemente*. De manera más precisa, se postula que la evaluación educativa que necesitamos deberá ser:

I. *Constructiva*, porque:

- No deberá tener otro propósito que la mejora de la calidad, en función de la cual deberá diseñarse y realizarse.

- Cada evaluación tendrá un propósito específico -como certificar las competencias adquiridas en un programa, la selección de alumnos para otro, o la asignación de estímulos- pero todos deberán concebirse en función de la mejora de la calidad.
- La evaluación del rendimiento de los alumnos, en especial, deberá verse como parte del proceso docente, como una oportunidad de aprendizaje en sí misma, y no como un elemento extraño y amenazante.
- En educación básica, las evaluaciones de rendimiento tendrán, fundamentalmente, propósitos formativos, buscando que ofrezcan retroalimentación sobre el avance de los alumnos para la aplicación oportuna de medidas remediales; el uso de la repetición de grado como estrategia correctiva será excepcional.

II. *Sistémica*, ya que:

- Incluirá evaluaciones de los alumnos, pero también de docentes y directores, así como de planteles, programas y políticas educativas.
- Utilizará pruebas de rendimiento, pero también estadísticas e indicadores de todas las dimensiones de la calidad.
- Será realizada por diversas instancias, articuladas en un conjunto que forme un verdadero sistema de evaluación.

III. *Confiable* por su calidad técnica, ya que:

- Se basará en diseños adecuados al propósito que se persiga.
- Empleará instrumentos y procedimientos debidamente validados, congruentes también con su propósito, sea que se trate de pruebas estandarizadas o de otros instrumentos, como cuestionarios, indicadores, observaciones en aula u otros.
- Cuando sea el caso, asegurará la representatividad de las muestras empleadas.
- Analizará cuidadosamente los resultados, empleando metodologías apropiadas según la naturaleza de los datos disponibles.

IV. *Válida*, en el sentido más amplio del término:

- Se basará en un concepto rico de calidad, que incluya resultados de aprendizaje, pero también otras dimensiones de calidad, como la relevancia de los currículos; la cobertura del sistema y las tasas de terminación; el impacto de la educación en la vida adulta; la disponibilidad de recursos suficientes y su utilización eficiente; la existencia de procesos de gestión ordenados tanto en el sistema en conjunto como en sus demás niveles, hasta la escuela singular y el aula.
- En cuanto a los resultados de aprendizaje, no se limitará a los que pueden medir las pruebas estandarizadas más simples, entendiendo que el currículo no debe supeditarse a la evaluación y estimulando a los maestros a atender los aspectos difíciles de evaluar en forma estandarizada, como las habilidades cognitivas superiores y las dimensiones actitudinales y valorales.
- Será conciente de sus límites y actuará de manera congruente con ello. No se usarán evaluaciones diseñadas con un propósito, para otro diferente para el que no sean apropiadas. Si se usan resultados de pruebas de rendimiento de alumnos como elemento para la evaluación de escuelas, por ejemplo, las pruebas empleadas deberán ser sensibles a la instrucción.
- Contrastará los resultados de las mediciones con parámetros pertinentes. En lo relativo a niveles de rendimiento se basará en estándares curriculares claros. Podrá usar comparaciones con países avanzados, perspectiva que se complementa con otras, como las comparaciones con países similares.
- Empleará el pasado como referente mínimo y el futuro como óptimo, valorando si se alcanzan o no las metas fijadas en cierto horizonte temporal.
- Será prudente para tomar decisiones con base en la evaluación, evitando que las

medidas de alto impacto se tomen sin sustento adecuado, que nunca consistirá sólo en resultados de evaluación en gran escala, sino que deberá complementarse con evaluaciones micro, a cargo de instancias cercanas a la situación.

V. *Validada externamente* también, de manera que:

- La evaluación interna se complementará con controles externos, sin subordinar los propósitos educativos al interés de la instancia externa.
- La evaluación que hace cada maestro de sus alumnos se complementará y validará con la del equipo docente y el director, así como con las evaluaciones en gran escala a cargo de instancias externas.
- La evaluación de cada plantel a cargo de su director, se enriquecerá con la de supervisores y autoridades locales, y en su caso con la de instancias externas.
- Las evaluaciones que hagan las entidades subnacionales podrán compararse y validarse gracias a las evaluaciones nacionales.
- Las evaluaciones nacionales se enriquecerán con la participación en proyectos internacionales, en forma activa y en función de las necesidades de cada país.

VI. *Útil*, porque será aprovechada por cada actor para sustentar decisiones apropiadas:

- Por las autoridades, para la definición de políticas.
- Por los maestros, para retroalimentar sus prácticas docentes.
- Por los alumnos mismos, para orientar mejor su propio trabajo, cuando permite resultados a nivel individual.
- Por los padres de familia y la sociedad, para sustentar sus exigencias de rendición de cuentas y orientar sus esfuerzos de apoyo.
- En todos los casos, porque aportará elementos para comprender mejor las causas de las situaciones detectadas.

VII. *Accesible* para diversos públicos interesados en la educación, porque:

- Sus resultados se difundirán de manera amplia, oportuna e inteligible.
- La difusión se concebirá como parte integral de la tarea de evaluar, con enfoques apropiados para que sea pertinente para cada público.
- A las autoridades se darán resultados globales sobre aspectos satisfactorios o deficientes del sistema educativo, con elementos que ayuden a interpretarlos incluyendo, en la medida en que lo permita la información, explicaciones de las situaciones detectadas.
- A la sociedad en general se darán también resultados e interpretaciones generales, claras pero no simplistas.
- A maestros, directores y supervisores, además de resultados generales, deberá darse información más detallada sobre aquellos aspectos del sistema educativo que dependen de su trabajo, para retroalimentar las prácticas docentes y las de gestión en el nivel escolar, que es donde, en última instancia, se da o no se da la calidad.
- A los investigadores y estudiosos de la educación deberá ofrecerse acceso a las bases de datos y a toda la información metodológica y técnica, que les permita replicar análisis previos y hacer otros, así como emprender estudios que aporten elementos adicionales tanto descriptivos como explicativos.
- Será valiosa para construir consensos entre actores con puntos de vista no coincidentes, que podrán acercarse más fácilmente si hay información creíble para todos.

VIII. *Parsimoniosa*, porque se verá a sí misma como medio y no como fin, por lo cual:

- Será respetuosa del trabajo de todos los actores, usándose tanto cuanto contribuya a la mejora de la calidad.
- La evaluación de aprendizajes, en particular, no se verá como sustituto de la labor del

maestro sino como complemento de ella, entendiendo que evaluaciones macro y micro deben apoyarse mutuamente.

- Las evaluaciones externas tratarán de interferir lo menos posible en el trabajo cotidiano de maestros y escuelas.

IX. *Eficiente*, porque:

- Buscará optimizar los recursos de que se pueda disponer.
- Las evaluaciones en gran escala generalmente se basarán en muestras de escuelas y alumnos, y se aplicarán en ciclos, para cubrir paulatinamente áreas curriculares y grados clave. La decisión de llevar a cabo evaluaciones censales se tomará sopesando cuidadosamente costos y beneficios.
- Concebirá su propio desarrollo como tarea de largo plazo, por lo que deberá comenzar con elementos como indicadores básicos y aplicaciones muestrales de pruebas en grados y áreas curriculares clave.

X. *Legítima*, finalmente, porque habrá ganado credibilidad y respeto gracias a:

- La calidad técnica de su trabajo.
- La independencia y seriedad de las instancias que la lleven a cabo.
- La utilidad de sus resultados para los diferentes actores, incluyendo a las autoridades educativas, pero también a los maestros y la sociedad en general.

Consideraciones finales

Para tener un sistema educativo de buena calidad es requisito necesario, aunque no suficiente, tener sistemas de evaluación muy buenos también. Una evaluación deficiente no es útil para que la calidad mejore: lleva fácilmente a juicios excesivos, triunfalistas o derrotistas: ni unos ni otros son base sólida para definir estrategias de mejora. El triunfalismo lleva a la continuidad de lo existente, cuyas fallas no se detectan. Y una idea derrotista de que no hay nada positivo en el pasado no es buen antecedente para avanzar.

El único punto de partida correcto para iniciar un proceso de mejora es el que reconoce con objetividad logros y limitaciones, teniendo en cuenta que *un sistema educativo de calidad es, finalmente, el que siempre mejora respecto a sí mismo, sin idealizar el pasado y con metas ambiciosas pero realistas para el futuro*.

En efecto: una vez conocidos los resultados de las evaluaciones, las autoridades y los maestros deberán sacar las conclusiones pertinentes para mantener y mejorar los puntos fuertes y corregir los deficientes. Conviene precisar que las decisiones no corresponden a los evaluadores, sino a la autoridad, desde los niveles de decisión más altos hasta el director de escuela y el maestro. La sociedad, por su parte, tendrá la última palabra en cualquier sistema democrático, y por ello la evaluación es herramienta indispensable para el ejercicio del derecho a la rendición de cuentas.

Por otra parte, hay que señalar un riesgo que trae consigo efectos contraproducentes: el de tomar decisiones que ofrezcan resultados espectaculares, que pretenden corregir en poco tiempo deficiencias ancestrales.

Establecer metas inalcanzables condena de antemano a la frustración y se basa en un supuesto falso: que hacer buena enseñanza con cualquier grupo de alumnos y en cualquier contexto fácil. Si así fuera, el que los resultados sean bajos sólo podría ser el resultado de una gran negligencia y mejorar debería ser sencillo. Si se entiende, al contrario, que los bajos resultados educativos son el producto de un complejo entramado de factores, se entenderá también que la mejora requiere de tiempos largos y estrategias complejas.

Aún si se diseñan estrategias integrales de mejora, que tomen en cuenta los diversos factores e involucren a los actores relevantes,

el tiempo necesario para que los esfuerzos den frutos es largo. Si esto no se tiene presente es fácil cometer errores graves de voluntarismo, pensando que cualquier meta puede alcanzarse en poco tiempo, por ambiciosa que sea, si se hace el esfuerzo necesario.

Señalar la complejidad y lentitud de los procesos de mejora educativa no implica una postura fatalista. Al fijar metas para la educación hay que apuntar alto, pero teniendo claro que mejorar en profundidad la educación sólo podrá conseguirse mediante estrategias integrales que tengan metas ambiciosas, sin olvidar que los caminos, por largos que sean, se recorren paso a paso. No hay que olvidar que, en educación, una década es corto plazo.

Por último, deberá evitarse un riesgo que acecha a los sistemas de evaluación contemporáneos: limitarlos a la aplicación masiva de pruebas de rendimiento de dudosa calidad. Esta manera de ver las cosas sobreestima las posibilidades de la evaluación en gran escala y pierde de vista que el maestro es y será la pieza clave en la búsqueda de la mejora educativa.

Un buen sistema de evaluación reforzará la profesionalización del maestro y el papel de los padres de familia de varias maneras: aportando materiales que apoyen el uso inteligente de resultados; complementando las evaluaciones externas con otras que utilicen instrumentos similares y sean aplicados cada escuela; y capacitando a las autoridades, los directores, maestros y padres para que aprovechen los resultados de las evaluaciones para retroalimentar su trabajo y conozcan el nivel de aprendizaje que cada alumno ha alcanzado, para mejorarlo.

Referencias bibliográficas

- BANEGAS GONZÁLEZ, ISRAEL, y E. BLANCO BOSCO (2006). *Políticas y sistemas de evaluación educativa en México. Avances, logros y desafíos*. México. Instituto Nacional para la Evaluación de la Educación.
- MARTÍNEZ RIZO, FELIPE (2005a). "La evaluación educativa en la actualidad. Una perspectiva latinoamericana". Ponencia magistral presentada en las *Jornadas de Cooperación Educativa con Iberoamérica*. Cartagena, Colombia, noviembre 21.
- MARTÍNEZ RIZO, FELIPE (2005b). "La evaluación educativa: ¿oportunidad o peligro?" Presentación en el *VIII Congreso Nacional de Investigación Educativa*. Hermosillo, México, octubre 31.
- MARTÍNEZ RIZO, FELIPE (2005c). Sobre la difusión de resultados por escuela. *Cuadernos de Investigación*, N° 17. México, Instituto Nacional para la Evaluación de la Educación.
- MARTÍNEZ RIZO, FELIPE et al. (2003). *La calidad de la educación básica en México*, 2003. México. Instituto Nacional para la Evaluación de la Educación.