

LA EVALUACIÓN EDUCATIVA EN MÉXICO: EXPERIENCIAS, AVANCES Y DESAFÍOS

Felipe Martínez Rizo

Universidad Autónoma de Aguascalientes

Emilio Blanco

El Colegio de México

INTRODUCCIÓN

La evaluación educativa en México constituye, en la actualidad, una institución relativamente consolidada. Más de tres décadas han transcurrido ya desde las primeras iniciativas, muy incipientes, que buscaban conocer de manera sistemática procesos y resultados del sistema educativo. En la última década se han producido avances muy importantes, los cuales han modificado nuestra imagen del sistema educativo y sobre cómo debe evaluarse.

Hoy, si bien con diferencias en cuanto a su concepción, la mayor parte de los actores educativos coincide en la importancia de realizar evaluaciones periódicas de diferentes aspectos del sistema educativo. Resulta difícil imaginar la posibilidad de reflexionar o intervenir sobre la educación nacional sin tener en cuenta esta herramienta. En un período relativamente breve nos hemos habituado a la existencia de múltiples evaluaciones, cuyos resultados son públicamente difundidos y concitan, aunque brevemente, la atención del conjunto de la sociedad. La mayor parte de estas evaluaciones se realiza de acuerdo a estándares conceptuales y metodológicos aceptados por la mayoría de los actores involucrados. Sin embargo, el camino de la evaluación aún enfrenta desafíos. En el caso de México, por una parte, se asiste a una expansión acelerada de distintas iniciativas de evaluación que podría superar la capacidad del sistema educativo para asimilar y aprovechar sus resultados. Del otro lado, este proceso aún está incompleto. El vínculo entre datos y decisiones en los distintos niveles del sistema es aún débil, y las evaluaciones no impactan de manera decisiva en las políticas y programas. La distribución de atribuciones entre los organismos de evaluación no está claramente definida, y éstos aún no cuentan con la autonomía suficiente para garantizar su independencia conceptual y técnica. En muchas entidades del país no se han desarrollado aún las capacidades para producir e interpretar

información de manera rigurosa. Muchos actores en la base del sistema, principalmente maestros y directores, mantienen una fuerte resistencia a las evaluaciones estandarizadas.

Todo esto muestra la importancia de recapitular el proceso de institucionalización de la evaluación en México, a fin de comprender su situación actual, sus perspectivas y retos. En la primera sección de este capítulo se ofrecen algunas definiciones básicas sobre lo que se considera una evaluación educativa de calidad, y se detallan las condiciones sociales e institucionales necesarias para que ésta cumpla sus objetivos. También se enumeran algunos riesgos asociados a los usos inadecuados de la evaluación. En la segunda sección se realiza un recorrido por la historia de la evaluación en México entre 1970 y 2002, período en el que se implementaron iniciativas con diverso grado de éxito. La tercera sección se enfoca en el período que va desde 2002 hasta nuestros días, durante el cual se delinearon las características actuales del sistema de evaluación educativa. Se pone especial énfasis en los antecedentes y las condiciones en que se creó el *Instituto Nacional para la Evaluación de la Educación* (INEE), así como en su estructura, sus principales iniciativas y sus logros. También se destacan las evaluaciones realizadas por la SEP, especialmente la prueba ENLACE. Finalmente, en la última sección se realiza un balance de los logros obtenidos hasta el momento, así como de los desafíos que enfrenta la evaluación educativa en México. No parece superfluo advertir que este capítulo se refiere únicamente a la evaluación del sistema educativo nacional en lo relativo a la educación de tipo básico, lo que implica que no se tratará lo relativo a la educación media superior ni la superior, que ha tenido un desarrollo diferente y, de hecho, anterior al de la evaluación de la educación básica.

LA EVALUACIÓN EDUCATIVA: DEFINICIONES, CONDICIONES Y RIESGOS

ALGUNAS DEFINICIONES EN TORNO A LA EVALUACIÓN EDUCATIVA

La evaluación educativa abarca un amplio conjunto de prácticas e implica un gran arsenal conceptual y metodológico. No obstante, toda evaluación, para ser considerada de calidad, debe contar con un conjunto de características básicas.

Rigor conceptual y metodológico

La evaluación debe ser rigurosa en lo conceptual y lo metodológico, lo que supone la participación de expertos provenientes de múltiples disciplinas. Pero el rigor científico de una

evaluación no es un producto exclusivo de la capacidad de los equipos encargados de realizarla: también son necesarias condiciones institucionales que garanticen la continuidad de los esfuerzos y la autonomía de los organismos evaluadores. Son imprescindibles, entonces, un marco normativo adecuado, así como el autocontrol por parte de las autoridades educativas (permanentemente tentadas hacia la colonización política de los espacios técnicos, al menos en el caso mexicano). Asimismo, es importante que el proceso de diseño, implementación y análisis de las evaluaciones esté abierto al escrutinio público y a la revisión por parte de especialistas no directamente vinculados a estos organismos.

Definición de referentes

Toda evaluación se basa en algún tipo de referente. Es imposible evaluar sin disponer de un criterio sobre qué resultado es deseable o indeseable, suficiente o insuficiente, excelente o mediocre. Anclada, en última instancia, en referentes socialmente construidos, valorativos e ideológicos, la evaluación es un campo de permanente debate y negociación. La “objetividad” de las evaluaciones es, en realidad, una objetividad con referencia a ciertas normas. La calidad educativa, en particular, es un concepto susceptible de numerosas definiciones, en ocasiones contrapuestas (Edwards 1991; UNESCO 1995). Esto supone un elemento de tensión, en tanto no existe una definición unánime sobre los objetivos deseables del sistema educativo; cualquier juicio sobre la calidad educativa puede ser cuestionado, sin importar el rigor conceptual o metodológico de las evaluaciones. Mantener esto en mente es un seguro contra discusiones estériles acerca de si las pruebas reflejan o no “la calidad” del sistema, como si existiese independientemente de los instrumentos utilizados para observarla.

Amplitud de objetivos

Por lo general, la evaluación educativa se piensa principalmente como evaluación de aprendizajes. Esta dimensión, si bien importante, no es la única. Los aprendizajes son sólo una parte de los resultados del sistema educativo. La educación no constituye solamente un mecanismo de producción de conocimientos o habilidades para el mercado de trabajo, sino que también es una agencia fundamental de socialización, esto es, de formación en valores y actitudes compatibles con un modelo de sociedad.

También debe tenerse en cuenta que una evaluación integral del sistema educativo supone atender también a los insumos, los actores, las condiciones, y especialmente, a los procesos educativos. Es necesario, por lo tanto, elaborar indicadores válidos y confiables de aspectos como los recursos disponibles en las escuelas; las capacidades, representaciones y prácticas pedagógicas de los maestros; las condiciones de vida de los alumnos; la gestión y el clima escolar. Sólo atendiendo a estos aspectos es posible pasar de la descripción a la explicación de los fenómenos, ampliando así las posibilidades de intervención. Otros aspectos importantes a evaluar son los programas de estudio de los alumnos, la formación docente, o el funcionamiento del sistema. Asimismo, es fundamental una evaluación sistemática de las políticas y programas educativos, en particular de su impacto.

Adecuación de la difusión y el uso de la información

La difusión de resultados es otra dimensión fundamental de una buena evaluación; de hecho, podría afirmarse que constituye el sentido último de todo el sistema. Una difusión amplia, adecuada a las exigencias y necesidades de los diferentes actores educativos, es indispensable para la formulación de políticas educativas eficaces, la rendición de cuentas, y la toma de decisiones .

Garantizar el acceso a un gran volumen de información no es, entonces, suficiente. Deben ofrecerse formatos y herramientas adecuadas para su interpretación, así como es necesario evitar usos inapropiados. Esto implica considerar la diversidad de atribuciones, necesidades y capacidades de los potenciales usuarios de la información.

Las autoridades educativas requieren evaluaciones diagnósticas generales sobre el sistema educativo en su conjunto, o sobre sectores del mismo, de manera que esto les permita elaborar políticas de largo y mediano plazo, y rendir cuentas a la sociedad. Las necesidades del sector intermedio del sistema son más específicas, acotadas a la detección de situaciones problemáticas que requieren intervención en determinadas regiones o escuelas, o a la formulación de estrategias de apoyo técnico-pedagógico a las escuelas de una zona. En la base del sistema, a su vez, directores y maestros requieren de información particular sobre sus centros para diseñar estrategias y proyectos orientados a mejorar los resultados educativos; también debe utilizarse esta información para rendir cuentas a las familias de los alumnos. Éstas, por su parte, requieren que los resultados sean reportados con la máxima claridad posible, a fin de incrementar sus posibilidades de participación en los procesos educativos.

Muy importante es también la distinción que se puede hacer entre evaluaciones de consecuencias fuertes o alto impacto, y las de bajo impacto o consecuencias débiles. Ejemplo de las primeras son las evaluaciones con base en las que se decide si un alumno es promovido o no al curso siguiente, o es aceptado en una institución, o las que llevan a contratar a una persona como maestro o a modificar su salario. Las pruebas que pretenden sólo informar sobre la situación del sistema educativo para orientar el diseño de políticas, por su parte, son típicamente ejemplos de evaluaciones de consecuencias débiles, aunque puede ocurrir que la opinión pública las haga adquirir una importancia que las vuelve de alto impacto.

CONDICIONES DE LA EVALUACIÓN EDUCATIVA

Existen numerosos argumentos a favor del establecimiento de sistemas de evaluación educativa sólidos y confiables, los cuales son de uso común, por lo que no es necesario incluir aquí más que un breve resumen.

- Toma de decisiones: la evaluación es indispensable para el diseño de políticas y programas educativos eficaces, así como para la innovación a nivel escolar.
- Rendición de cuentas: la evaluación es fundamental para el control social e institucional de las decisiones gubernamentales sobre la educación.
- Distribución de incentivos: la evaluación puede utilizarse para establecer un sistema de incentivos que premie las prácticas eficaces y fomente su difusión en el sistema.
- Participación social y elección: la evaluación es útil para establecer mecanismos de control social directo sobre la educación, sea a través de la participación social o de la elección de escuelas en función de sus resultados.

Estos argumentos no carecen de atractivo. Sin embargo, al contrastarlos con la realidad mexicana, resulta evidente que la evaluación no logra, consolidar estos procesos por sí misma. Actualmente, la evaluación no se utiliza de manera sistemática en el diseño de políticas o la implementación de iniciativas de mejora escolar. En lo referente a la rendición de cuentas, la mera disponibilidad de información no ha sido suficiente para que la sociedad formule y canalice efectivamente sus demandas. La complejidad de los problemas y el hecho de que los resultados de las reformas sólo se ven en el mediano y largo plazo hacen particularmente difícil este proceso.

Por otra parte, una relación positiva entre evaluación y mecanismos de participación social y elección requiere de condiciones sociales específicas, relativamente débiles aún en

México. Estas condiciones son: a) conciencia de los problemas educativos y capacidad de demanda de una educación de calidad; b) disponibilidad a participar de manera organizada para resolver problemas educativos a nivel local y societal; c) capacidad de interpretar la información proveniente de las evaluaciones y de tomar decisiones con base en ellas.

La probabilidad de encontrar estas condiciones en una sociedad fuertemente estratificada como la mexicana, además, es menor en los grupos más empobrecidos. Las capacidades de decodificación de la información no se distribuyen de manera equitativa. En algunos casos, como en los que se refieren a la elección de una escuela por parte de los padres, con base en los resultados de las evaluaciones, se pierde de vista que para ese propósito se necesitarían evaluaciones mucho más precisas y complejas que las existentes, y que para miles de alumnos en México sólo existe una opción real de educación: la única escuela situada en las cercanías de la comunidad.

RIESGOS DE LA EVALUACIÓN EDUCATIVA

A lo anterior debe agregarse que, bajo ciertas condiciones, la evaluación incluso entraña ciertos riesgos para la calidad y la equidad educativas. Por “riesgos” nos referimos a consecuencias o usos no deseados de la evaluación. No se trata de riesgos inherentes a esta actividad, sino de riesgos derivados de lógicas institucionales modificables, pero que pueden tener efectos realmente contraproducentes.

Inadecuación en la difusión y el uso de la información

Lograr un balance adecuado entre simplicidad y precisión de la información es tal vez el desafío más grande en cuanto a la difusión. Simplicidad, a fin de que sean correctamente interpretadas por el mayor número posible de personas. Precisión, para evitar la difusión de una imagen sesgada de los fenómenos educativos.

Si las autoridades malinterpretan el tipo de conclusiones que es válido extraer a partir de determinado tipo de información, o cómo deben ser comunicados los resultados a fin de evitar errores de interpretación, pueden obtener y comunicar al resto de la sociedad conclusiones espurias. En el caso de las pruebas censales, por ejemplo, los resultados suelen ser difíciles de comprender, no sólo para los padres, sino también para muchos maestros, lo cual genera cuestionamientos hacia su validez.

Los medios de comunicación también tienen una cuota de responsabilidad en este punto. Uno de los ejemplos más claros de difusión inadecuada es la publicación de listas de escuelas ordenadas en función de los resultados promedio obtenidos en las pruebas censales. El supuesto detrás de estas listas es que representan fielmente las diferencias en la “calidad” de las escuelas. Por lo general, las conclusiones suelen ser idénticas cada año: la mayoría de las escuelas están “reprobadas”; los niveles de aprendizaje “no mejoran”; las escuelas privadas son las que tienen mejores resultados.

En líneas generales, algunas de estas conclusiones pueden ser correctas. Sin embargo, en su prisa, no consideran que de todas las dimensiones posibles de calidad educativa sólo se están considerando unos pocos indicadores; que una parte muy importante de los resultados no depende de lo que la escuela hace, sino de su entorno sociocultural (Fernández 2007; Blanco 2008); que, finalmente, las variaciones interanuales en los resultados deben ser interpretadas con mucha cautela, pues pueden obedecer a numerosos factores no relacionados con la calidad de la enseñanza.

La “información” así transmitida por los medios, y que por lo general tiende a dejar mal paradas a las escuelas públicas, se alimenta de extendida percepción relativa a la ineficacia de la educación en este sector. Con ello, lejos de producir una reacción positiva para la mejora de las deficiencias reales, se ponen en riesgo los aspectos positivos de la educación pública y se contribuye al deterioro de la relación entre escuelas y familias.

Finalmente, existe el problema del uso inadecuado de los resultados, en particular para la toma de decisiones que afectan directamente a los actores educativos. Una vez que los datos están disponibles, es difícil evitar la tentación de utilizarlos para fines diferentes a los originales. No es inusual que, en México, algunas autoridades educativas quieran “premiar” o “castigar” a escuelas o maestros, sin fundamento sólido, y existen casos de supervisores o directores que utilizan los resultados individuales para seleccionar a los “mejores” alumnos e impedir la entrada quienes no han obtenido buenos resultados en las pruebas censales. Controlar la tentación punitiva de autoridades que desconfían de sus maestros, así como controlar las tendencias a la simulación de “buenos” resultados por exclusión de “malos” alumnos, es un imperativo en términos de equidad que implica reducir la ansiedad y el fetichismo depositado en las evaluaciones.

Saturación de información

Si bien es cierto que evaluar es bueno, evaluar más no necesariamente es mejor. Una tentación frecuente es pretender evaluar todas las dimensiones imaginables de la educación, de manera muy frecuente. La superabundancia de información así generada impide un análisis reflexivo de los resultados; el exceso de información conspira contra el conocimiento. Si a esto se agrega el hecho de que cada instancia gubernamental pretende realizar su propia evaluación, muchas veces sin objetivos claros, se comprende por qué las evaluaciones se incrementan a una tasa mayor que la de la capacidad de explotar la información.

Resistencia de autoridades y magisterio

Las actitudes de algunos actores educativos hacia la evaluación en México ha sido, y continúa siendo en ocasiones, contradictoria. La evaluación excesiva coexiste con la resistencia a las evaluaciones. En los sistemas educativos caracterizados por elevados niveles de politización, burocratización y desconfianza la resistencia a la evaluación es una característica frecuente, en todos los niveles. Este es uno de los factores que impide, por ejemplo, que los resultados entregados a las escuelas sirvan como disparadores de procesos de mejora escolar. Si bien es cierto que existen intereses políticos y corporativos espurios detrás de esta actitud (temor a la exposición pública y a los ataques por parte de adversarios políticos), también es verdad que el manejo autoritario y arbitrario del poder característicos del sistema educativo mexicano genera elevados niveles de recelo, aún entre quienes cumplen responsablemente su tarea.

A nivel de los maestros, los argumentos más frecuentes para justificar este recelo son: a) el desajuste entre lo evaluado y los contenidos de los programas; b) la desatención de las evaluaciones a los “procesos” de enseñanza a favor de los “resultados”; c) la inadecuación de las pruebas a la diversidad de alumnos y situaciones de enseñanza; d) versiones sobre falsificación de resultados por parte de colegas; e) la dificultad para tomar decisiones a nivel escolar a partir de los resultados.

Confusión entre indicadores y fenómenos

Un último riesgo atañe fundamentalmente al ámbito de autoridades y expertos en materia educativa. La evaluación puede convertirse en un fetiche, un sustituto de las realidades que se pretende conocer. Reducir la calidad educativa de una escuela a sus resultados en una o dos asignaturas, por importantes que sean, es a todas luces desacertado.

La obsesión por las pruebas puede hacer perder de vista que la educación produce algo más que lo que se mide a través de ellas. Se corre el riesgo de confundir lo fácilmente medible con lo relevante.

Es cierto que toda evaluación supone una selección, intencionada, de aquellos aspectos que se consideran más relevantes, en tanto nunca es posible observar directamente “la realidad como un todo”. El peligro es que se tome la parte por el todo; que los indicadores se confundan con la calidad educativa, y que se pretenda mejorar la calidad únicamente a través de la mejora de dichos indicadores. Algunas escuelas, al recibir resultados insatisfactorios, reaccionan “entrenando” regularmente a sus alumnos para responder instrumentos de múltiple opción. En algunos casos, semanas antes de la aplicación de alguna de estas pruebas, las actividades de clase se enfocan exclusivamente en este sentido. Este tipo de preparación ya se ha institucionalizado incluso a nivel de algunas entidades en México. La enseñanza se reduce al absurdo del entrenamiento para responder ítems. Si, además, estas evaluaciones se utilizan para tomar decisiones de alto impacto sobre el futuro de las escuelas o los alumnos (por ejemplo, a través de “premios” económicos o de barreras al acceso, los errores adquieren dimensiones de cataclismo. Estos riesgos obligan a monitorear críticamente el proceso de institucionalización de los sistemas nacionales de evaluación. Una vez que estas prácticas y representaciones se institucionalizan, se torna difícil revertirlas y pueden socavar la legitimidad de toda iniciativa.

Como puede verse, la evaluación educativa enfrenta múltiples desafíos. En México, la consolidación de evaluaciones de calidad, integrales, y que articulen adecuadamente los esfuerzos de todas las instancias involucradas, es un proceso lento y aún inacabado. En las páginas siguientes describiremos las líneas generales de dicho proceso.

LA EVALUACIÓN EDUCATIVA EN MÉXICO: 1970 – 2010

¿Cuál ha sido la evolución de la evaluación educativa en México, en lo relativo a la educación básica, y cuáles han sido las decisiones e iniciativas que han marcado esta historia? En esta segunda sección desarrollaremos las principales iniciativas que han contribuido a moldear el actual sistema de evaluación en México.

Creemos conveniente definir tres períodos:

1. El primero comprende las décadas de 1970 y 1980. A partir de 1970 la Secretaría de Educación Pública (SEP) mejoró y sistematizó significativamente la recolección de información censal para la construcción de estadísticas referidas a la educación nacional. Es también en este período cuando, en el ámbito de la SEP, se realizaron las primeras evaluaciones de aprendizajes en educación primaria.
2. El segundo periodo está comprendido entre los años 1990 - 2002, durante el cual se desarrolló un amplio conjunto de instrumentos de evaluación de aprendizajes, con distintos propósitos, particularmente en educación básica. Además de estas iniciativas, en esta década se implementaron las primeras participaciones de México en pruebas internacionales de aprendizaje, las cuales se convirtieron en instancias clave para el desarrollo de un sistema nacional de evaluación.
3. El inicio del tercer periodo, que llega hasta nuestros días, puede fijarse en el año 2002. Esta etapa se caracteriza por dos avances de importancia en relación a los periodos anteriores. En el terreno político, se confiere un papel estratégico a la evaluación de la educación, como elemento imprescindible para la planeación y la rendición de cuentas por parte de la autoridad educativa. En el ámbito institucional, se da un paso fundamental hacia la reestructuración de las instancias encargadas de la evaluación, a través de la creación del Instituto Nacional para la Evaluación de la Educación (INEE). También destaca la iniciativa de la Secretaría de Educación Pública para realizar una prueba de aprendizajes de tipo censal en educación básica (ENLACE).

PRIMERAS EXPERIENCIAS DE EVALUACIÓN: LAS DÉCADAS DE 1970 Y 1980

En esta etapa los avances principales no se produjeron tanto en el terreno de la evaluación como en el de la medición. Se mejoró de manera considerable la información censal disponible sobre el sistema educativo, con el objeto de disponer de estadísticas confiables y comparables a través del tiempo.

Las evaluaciones de aprendizajes de esta época, propiamente dichas, tuvieron un carácter incipiente y técnicamente fueron considerablemente débiles; estuvieron más orientadas a la selección a niveles superiores que a un diagnóstico de la calidad del sistema. En 1972 se aplican las primeras pruebas en gran escala, como insumo de decisión sobre el ingreso de alumnos a la educación secundaria. En el año 1974 se crea la Subdirección de Evaluación y Acreditación (hoy Dirección General de Evaluación de Políticas, DGEP), dentro

de la Secretaría de Educación Pública (SEP), y se continúan aplicando pruebas de selección. La primera evaluación de aprendizajes, en sentido estricto, se realiza durante el sexenio 1976-1982 en muestras de alumnos con representatividad nacional.

Debido en parte al estado del conocimiento sobre evaluación educativa en dicho momento, las pruebas de aprendizaje de este período aún se ubicaban lejos de los estándares actuales de validez y confiabilidad. Por otra parte, la institucionalidad y el clima político del período no favorecían la difusión de resultados, y dificultaban su incorporación al diseño de políticas. Al igual que en otros ámbitos de la gestión gubernamental, la información se manejaba discrecionalmente, a partir de consideraciones políticas (*politics*) antes que de política pública (*policy*), lejos del escrutinio de una opinión pública aún en formación.

LA EVALUACIÓN ENTRE 1990-2002: UN LARGO PROCESO DE APERTURA

La década de 1990 representó un avance significativo respecto del período anterior. La federalización del sistema educativo coincidió con un aumento en la preocupación de las autoridades educativas por desarrollar instrumentos de evaluación confiables, que pudieran ser utilizados para diseñar políticas educativas eficaces (LGE, Arts. 29 a 31). En este contexto, la DGEP se encargó del diseño y la coordinación de múltiples evaluaciones.

A nivel nacional, las iniciativas más destacables han sido:

- El factor de Aprovechamiento Escolar en el marco del Programa de Carrera Magisterial (1994).
- Las pruebas denominadas “Estándares Nacionales”, a partir de 1998.
- Las evaluaciones internacionales del *Trends in Mathematics and Sciences Study* (TIMSS) de 1995, el *Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación* (LLECE) de 1997, y la ronda 2000 del *Programme for International Student Assessment* (PISA)¹.

¹ Las pruebas PISA, como se verá, se aplican a estudiantes de 15 años de edad, por lo que, estrictamente, salen parcialmente del ámbito de la educación básica. Lo mismo ocurre con los exámenes del Centro Nacional de Evaluación para la Educación Superior (CENEVAL), creado en 1994. El mayor número de quienes sustentan los exámenes de este organismo está formado por aspirantes a entrar a una institución de educación superior, en tanto que otros hacen pruebas de salida del nivel de licenciatura. Hay también aspirantes a ingresar a bachillerato, que están terminando la secundaria, por lo que marginalmente también el CENEVAL realiza algunas evaluaciones que tienen que ver con la educación básica.

Factor de Aprovechamiento Escolar de Carrera Magisterial

El Programa Carrera Magisterial inició en 1992 con el objetivo de establecer una estructura de incentivos para maestros y directores, de acuerdo a su desempeño y al de sus alumnos. Para asignar estos incentivos se consideran, entre otras cosas, la “preparación profesional” del docente, y el “aprovechamiento” (aprendizaje) de los alumnos. En 2009 las pruebas con las que se evaluaba este último factor se aplicaban a cerca de 390,000 maestros de Educación Básica.

La DGEP tiene la responsabilidad principal en la evaluación. Su amplitud ha requerido del establecimiento de estructuras orgánicas donde se involucra tanto al personal centralizado como a las autoridades educativas estatales (Velázquez 2000: 663). Debe destacarse también que el sindicato magisterial tiene un peso importante a lo largo de todo el proceso de evaluación, lo que resta independencia y neutralidad a esta prueba.

Las pruebas utilizadas son del tipo “basadas en normas”², lo que ocasiona que no pueda conocerse en sentido estricto lo que los alumnos han aprendido con respecto a los Programas oficiales. A esto debe agregarse que no se recoge información sobre el contexto del alumno, lo que no permite valorar la posible influencia de factores escolares y extraescolares en los resultados. Estas pruebas, por tanto, no pueden ser utilizadas con fines de diagnóstico.

Estándares Nacionales: 1998-2003

Estas pruebas representaron una innovación importante dado que se orientaron a conocer el grado en que los alumnos de primaria y secundaria alcanzaban los niveles mínimos de aprendizaje propuestos en los programas oficiales. Constituyen, en buena parte, el germen de la evaluación de aprendizajes tal como se la conoce hoy en México. Entre los años 1998 y 2003, la DGEP tuvo a cargo la elaboración de las pruebas y de los mecanismos de evaluación. A partir de 2003, las pruebas se transfirieron íntegramente al Instituto Nacional para la Evaluación de la Educación (INEE).

En el diseño de esta prueba se buscó elevar la validez y confiabilidad de los instrumentos, para lo cual la SEP convocó a especialistas en diversos ámbitos. La mejora en la conceptualización y observación de los aprendizajes se reflejó, entre otras cosas, en la adopción de un modelo de evaluación basado en estándares, así como en la estimación de los

² El diseño de estas pruebas supone que tanto los reactivos más sencillos como los más difíciles sean eliminados. Asimismo, los resultados individuales se reportan por referencia a la media poblacional.

resultados mediante modelos estadísticos, en particular el que estima la probabilidad de que un alumno responda correctamente una pregunta considerando solamente la dificultad de la pregunta misma (modelos Rasch o de un parámetro), además de la capacidad del sustentante. También se aplicaron cuestionarios de contexto para los alumnos, docentes, directores.

No obstante, las pruebas también presentaron fallas. Hasta el año 2003 no se disponía de un marco conceptual definido para el diseño de los cuestionarios de contexto. En cuanto a la aplicación de las pruebas, no se realizaban pilotajes, y los levantamientos de datos no se realizaron en los mismos momentos del período escolar en todos los años, ni se cuidó que la dificultad de las pruebas aplicadas en años sucesivos fuera la misma, por lo que los resultados no podían compararse a lo largo del tiempo. Por otra parte, la información no fue aprovechada ni para fines académicos ni para la toma de decisiones (INEE 2003: 46).

Evaluaciones internacionales de aprendizajes

Las evaluaciones realizadas en el marco de los programas TIMMS, LLECE, y PISA, durante el período considerado aquí tuvieron un potencial importante para la mejora de la evaluación educativa. En primer lugar, porque podían aportar información de calidad para ubicar la calidad de la educación mexicana en el contexto internacional, en un momento histórico en el cual México apenas estaba construyendo la capacidad de generar tales datos. En segundo lugar, porque su desarrollo, aplicación y análisis eran una excelente oportunidad para desarrollar tales capacidades a nivel nacional. En tercer lugar porque podrían fortalecer actitudes favorables hacia las evaluaciones educativas y la difusión de resultados. En realidad este potencial positivo se concretó parcialmente; cada una de las participaciones aportó lecciones diferentes.

Estudio de Tendencias en Matemáticas y Ciencia (TIMSS)

TIMSS es un proyecto patrocinado por la *International Association for the Evaluation of Educational Achievement* (IEA), con el objetivo de evaluar conocimientos en Matemática y Ciencias. En el año 1995 México participó en la primera aplicación del TIMSS; fueron evaluadas muestras aleatorias representativas de alumnos de 9 y 13 años de edad de escuelas tanto públicas como privadas de todo el país (más de 20,000 y 24,000, respectivamente), quedando la aplicación a cargo de la SEP.

La participación de México en TIMMS finalizó de manera abrupta, debido a la decisión del gobierno mexicano de no dar a conocer los resultados. En el año 2000, la SEP, a través de la Dirección General de Evaluación (DGEP), decidió replicar las pruebas. No obstante, los resultados de este relevamiento tampoco fueron publicados³.

Este tipo de decisiones ejemplifican claramente la ambigüedad, cercana en el tiempo, de las autoridades políticas hacia las evaluaciones. Se era consciente de la necesidad de contar con información de calidad para la toma de decisiones, pero al mismo tiempo no se estaba preparado para aceptar las consecuencias políticas y sociales de la difusión pública de los resultados.

Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación (LLECE)

Esta instancia, que agrupa a los responsables de evaluación de los ministerios de educación de los países de América Latina, fue establecida por la Oficina Regional de Educación para América Latina y el Caribe (OREALC), perteneciente a la UNESCO. En 1995 decidió poner en marcha un Primer Estudio Regional sobre el aprendizaje en primaria, que se llevó a cabo en 1997, con la participación de 13 países de la región. Posteriormente, en el año 2006, esta iniciativa fue continuada a través del *Segundo Estudio Regional Comparativo y Explicativo* (SERCE). El Primer Estudio evaluó las competencias curriculares de muestras de alumnos de 3º y 4º grado de primaria, en Matemática y Lengua. La OREALC publicó varios informes de resultados en los años siguientes que permitieron contar con un panorama general sobre la calidad de la educación en la región.

El carácter novedoso en lo regional del LLECE derivó, lógicamente, en que la experiencia de 1997 no fuera enteramente satisfactoria. En el caso de México, la aplicación presentó fallas importantes (relacionadas principalmente con la aplicación de los cuestionarios) que limitaron las posibilidades de análisis. A pesar de ello, el LLECE representó una experiencia valiosa para México, ya que permitió mejorar las competencias nacionales en el campo de la evaluación, y arrojó datos de interés respecto de la educación

³ Hasta hoy, el único análisis de estos resultados fue realizado por el INEE en el año 2003. Sin embargo, las bases de datos presentan carencias de información que limitan su aprovechamiento. Los análisis realizados por el INEE (Backhoff y Solano 2003) son de tipo comparativo, aunque también se analizan correlaciones entre los indicadores socioculturales de las entidades y los niveles de logro alcanzados en cada una.

mexicana en el contexto latinoamericano. La difusión del informe internacional, sin embargo, tuvo escaso impacto, y el país no llevó a cabo un análisis propios de los resultados.

Programa de Evaluación Internacional de Estudiantes (PISA)

PISA es un estudio desarrollado por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) en el que han participado hasta ahora más de 60 países. Su objetivo básico es evaluar las habilidades de los estudiantes de 15 años inscritos en cualquier grado escolar, a partir del primero de secundaria, si bien la mayor parte se encuentran ya en primero de media superior a esa edad. Dado el interés del Programa por apoyar las políticas educativas de los países participantes, se busca una amplia participación de expertos nacionales en el diseño y adaptación de las pruebas. También se implementan actividades de formación de los equipos nacionales encargados de la aplicación y codificación de los resultados. Todo esto representa una contribución importante a la conformación de equipos técnicos nacionales con alto nivel de capacitación, en particular en los países subdesarrollados.

Desde el ingreso de México en la OCDE, el país ha participado en cuatro rondas de PISA (desde 2000 hasta 2009, en forma trianual). Los instrumentos aplicados tienen como objetivo “evaluar en qué medida los estudiantes de 15 años han adquirido los conocimientos y habilidades esenciales para participar plenamente en la sociedad, y hasta qué punto son capaces de extrapolar lo aprendido para aplicarlo a situaciones novedosas, tanto en el ámbito escolar como extraescolar” (INEE 2007: 15). Por este motivo, también se evalúan habilidades generales, tales como la familiaridad con las nuevas tecnologías, las estrategias de manejo de los propios aprendizajes, y la resolución de problemas.

Probablemente por falta de capacidad técnica, sin embargo, la participación de México en la primera ronda de PISA, en 2000, fue limitada: no se intervino en el desarrollo de las pruebas, la aplicación se limitó al mínimo, y no se hizo análisis alguno de los resultados, que sólo se conocieron por el informe internacional que, a diferencia de lo ocurrido años antes con los resultados del estudio TIMSS, sí incluyó los datos de México. El momento en que se dio a conocer dicho informe (diciembre de 2001) se situaba ya en otro contexto.

LA DIFUSIÓN DE LOS RESULTADOS EN EL PERÍODO 1990-2002

Durante esta década, la difusión de resultados continuó siendo un punto débil de la evaluación nacional. El caso más notorio, ya mencionado, fue la decisión de no publicar los resultados de la prueba TIMSS en 1995. Esta situación se explica, fundamentalmente, por la falta de autonomía de los organismos evaluadores, así como por condiciones político-institucionales más amplias, propias de un régimen no democrático.

Al final del período, existieron, no obstante, algunas iniciativas importantes de difusión, fundamentalmente a través de reportes basados en pruebas de aprendizaje. Un reporte de la SEP utilizó los resultados de las pruebas de Aprovechamiento entre 1997 y 1999 para ubicar a todas las escuelas participantes en cinco niveles de acierto (SEP 2000a). La DGEPI comenzó a difundir los resultados de las pruebas de Carrera Magisterial, intentando distribuirlos en forma desagregada para cada docente. Sin embargo, tal como lo mostraron Fernández y Midaglia (2003), no era habitual que estos resultados llegaran a las escuelas, debido a la reticencia o indiferencia de los actores involucrados. Además, esta Dirección, junto con las Áreas Estatales de Evaluación de numerosas entidades elaboraron informes sobre el funcionamiento de los programas compensatorios y las características de escuelas incrementales y decrementales (SEP 2000b, 2001).

Si bien debe reconocerse el mérito de estos esfuerzos, también es cierto que fueron escasos y que su repercusión en el sistema educativo fue relativamente baja. La información generada en la década no se incorporó al proceso de toma de decisiones o diseño de políticas educativas. Los reportes públicos fueron escasos y su difusión limitada. Aún no existían canales institucionalizados de difusión de la información hacia los niveles más bajos del sistema: los supervisores, directores y maestros recibían información escasa y asistemática. Era infrecuente la difusión de resultados en la prensa, e inexistente hacia las familias. El acceso público a la información en bruto, por su parte, estaba fuertemente restringido, lo que dificultaba realizar investigaciones independientes a nivel académico.

A partir del año 2002, y de manera progresiva, muchos de los aspectos anteriores experimentarían fuertes transformaciones. Se abría así una nueva etapa para la evaluación nacional, con logros encomiables, pero también nuevos desafíos.

LA EVALUACIÓN EDUCATIVA EN MÉXICO DE 2002 A 2006: EL INEE

SITUACIÓN AL COMENZAR EL SEXENIO DE VICENTE FOX

Los cambios ocurridos en el campo de la evaluación educativa en la primera década del siglo XXI deben analizarse en el contexto de las transformaciones más amplias que tuvieron lugar en el país, en especial las de naturaleza política.

El triunfo de Vicente Fox, candidato del Partido Acción Nacional, en las elecciones del 6 de julio del año 2000, representó un paso importante en el proceso de democratización de la vida política mexicana, al darse por primera vez el cambio del partido en el poder por vía electoral, la *alternancia* en el gobierno. Este fenómeno fue acompañado por una creciente demanda de participación en la vida pública por parte de diversos grupos de ciudadanos. En contraste con la ausencia de información sobre los asuntos públicos, que caracterizaba las épocas más autoritarias, una de las expresiones de la nueva dinámica política fue la demanda de *transparencia*, en cuanto al acceso a la información pública.

La transformación previa de la instancia a cargo de los procesos electorales, que dejó de ser un organismo controlado por el ejecutivo en turno para volverse uno *ciudadanizado* y autónomo (Instituto Federal Electoral, IFE) fue elemento importante en la transición democrática. Con la toma de posesión de Vicente Fox como Presidente de la República se dieron otros pasos en el mismo sentido, como la aprobación de la llamada Ley de Transparencia, y la creación del Instituto Federal de Acceso a la Información (IFAI).

En el ámbito educativo, las tendencias autoritarias del sistema político mexicano habían traído consigo, entre otras cosas, que el desarrollo de mecanismos de evaluación en gran escala del rendimiento escolar, a los que se han referido los puntos anteriores de este capítulo, no hubiera sido acompañado, como se esperaba, por la difusión de los resultados y su utilización tanto para sustentar el diseño de políticas educativas, como para apoyar la rendición de cuentas.

Sólo en octubre de 2001, y por información extraoficial difundida por uno de los diarios más importantes del país, la opinión pública se enteró de que en 1996 la Secretaría de Educación Pública había impedido que se incluyeran los resultados de México en el informe internacional del TIMSS. De hecho muchas personas creyeron erróneamente que esa información se refería a una decisión reciente de las autoridades educativas nacionales,

respecto a los resultados de las pruebas PISA de la OCDE aplicadas en el año 2000, que debían difundirse y se difundieron en diciembre de 2001.

Por otra parte, entre agosto y noviembre de 2000, especialistas de diversos campos se dedicaron a preparar elementos para el programa de gobierno del nuevo presidente, formando lo que se llamó el *equipo de transición*. El grupo encargado del área de evaluación de ese *equipo* decidió incluir, en el documento que preparó, la propuesta de crear un organismo técnico, con un grado considerable de autonomía, que se encargara de realizar evaluaciones en gran escala del sistema educativo, hasta entonces a cargo exclusivamente de la DGEP. En la propuesta se daba ya a esa instancia el nombre de *Instituto Nacional para la Evaluación de la Educación, INEE*.

La propuesta del grupo de educación del equipo de transición, formado por personas conocedoras de la problemática educativa, se debía a una preocupación ampliamente compartida por ellos respecto a la calidad de la educación, así como a la conciencia de que una buena evaluación de dicha calidad es una tarea de alta complejidad técnica, que sólo se puede llevar a cabo de manera adecuada en una instancia que, además de contar con los recursos necesarios, no esté sujeta a las presiones de las instancias interesadas, entre las que destacan las mismas autoridades educativas, y el Sindicato Nacional de Trabajadores de la Educación.

En este sentido, indirectamente, la propuesta de crear el INEE se relacionaba también con el proceso de democratización de la vida pública nacional, no sólo en su dimensión estrictamente política, sino también en otros ámbitos necesitados también de transparencia y democratización, como el educativo. Que Vicente Fox captaba esa relación se hace evidente en la expresión que comenzó a utilizar desde los primeros días de su gestión, en diciembre de 2000, cuando comenzó a referirse al futuro organismo de evaluación como *el IFE educativo*.

NACIMIENTO DEL INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN

Desde diciembre de 2000, el nuevo titular de la Secretaría de Educación Pública encomendó el desarrollo de un proyecto completo para la creación del INEE, a partir del anteproyecto del equipo de transición. El nacimiento del INEE debió esperar más de un año por las dudas que había en cuanto a la mejor alternativa legal para ello. Las opciones eran dos: la preferida por miembros de la Comisión de Educación del Senado, suponía la aprobación de una ley al respecto; la otra consistía en la expedición de un decreto presidencial. Esta

última no permitiría dar plena autonomía el nuevo organismo, pero tenía la ventaja de no implicar negociaciones para alcanzar la mayoría necesaria, lo que sí requería la primera opción, ya que ninguno de los grupos parlamentarios tenía mayoría .

Finalmente el Presidente se inclinó por la opción que dependía de su propia decisión, y expidió el decreto de creación del INEE el 8 de agosto de 2002. El hecho provocó de inmediato diversas reacciones; si bien todas manifestaban, con diversos matices, que un organismo de ese tipo era necesario, en algunos casos prevalecía la insatisfacción porque el nuevo Instituto no tuviera una autonomía mayor que la que podía darle el decreto presidencial del 8 de agosto. Las posturas críticas provenían de dos sectores distintos: algunos legisladores veían la creación del INEE por decreto presidencial como una invasión de un campo del poder legislativo; algunas personas del sector privado, por su parte, temían que la no completa autonomía del INEE respecto de la SEP se tradujera en una subordinación clara, que llevara a una situación similar a la anterior en cuanto a la falta de transparencia en lo relativo a la difusión de resultados.

Sin embargo, y a pesar de que su creación por decreto presidencial no dotaba al nuevo organismo de plena autonomía jurídica, la manera como se diseñó su estructura de gobierno, y la manera concreta como se puso en marcha, hizo que el margen de autonomía real de que gozó el INEE desde sus inicios fuera muy importante, similar al de organismos dotados de autonomía constitucional.

La independencia y el carácter plural de la Junta Directiva

La Junta Directiva, máximo órgano de gobierno del INEE es presidida por el titular de la SEP, y la integran 14 personas más; siete son funcionarios públicos, pero sólo dos tienen relación de subordinación respecto al Secretario de Educación. Los otros siete integrantes representan a sectores no gubernamentales interesados en la educación. Esta composición plural hace que las decisiones de la Junta impliquen un amplio grado de consenso por parte de sus integrantes, lo que la hace realmente independiente, ya que ni el titular de la SEP, ni el SNTE o cualquier otra instancia tiene el control de las decisiones.

La naturaleza académica y el Consejo Técnico

El Consejo Técnico está formado por 16 especialistas, mexicanos o no, reconocidos por su trabajo en evaluación o investigación educativa, que la Junta Directiva del Instituto

designa con base en sus méritos personales. Los consejeros nombrados por la Junta al nacer el Instituto, al igual que los que después se han incorporado a ese colegiado, han sido estudiosos mexicanos y extranjeros muy destacados, cuyo perfil académico asegura que el trabajo se apege a los estándares de calidad más exigentes, además de ser una garantía adicional de independencia.

La cercanía con el sistema educativo y el Consejo Consultivo

El Plan Maestro de Desarrollo del Instituto señala que su naturaleza *conlleva dos principios definitorios de su quehacer que, lejos de oponerse, se complementan: autonomía y cercanía respecto al sistema educativo*. La Junta Directiva y, a su manera, el Consejo Técnico, aseguran la autonomía del Instituto, *respecto de las autoridades educativas y cualquier otra instancia, en todo lo que se refiera a los aspectos técnicos de su trabajo y a la difusión de resultados del mismo*, de manera que el INEE sea reconocido *como autoridad técnica en el desarrollo del proceso evaluativo, lo que implica garantizar el apego a normas y prácticas reconocidas internacionalmente en el diseño y aplicación de instrumentos y metodologías de medición y evaluación, para mantener la confianza de la sociedad en la evaluación*.

El Plan Maestro introduce el principio de *cercanía*, señalando que *la autonomía no deberá traducirse en distanciamiento. Como instancia que tendrá sentido en la medida en que contribuya a la mejora de la calidad educativa, el INEE deberá tener una relación permanente y estrecha con las autoridades educativas federales y estatales y, en especial, con las áreas de evaluación correspondientes, para asegurar que los resultados de las evaluaciones se utilicen para sustentar decisiones de mejora*. El Plan precisa que *el Consejo Consultivo es la instancia apropiada para asegurar esa relación, y que la cercanía implica el compromiso del INEE de dar a conocer los resultados de su trabajo de manera transparente, oportuna y comprensible, para evitar las repercusiones negativas que puede traer consigo una interpretación inadecuada de las evaluaciones*.

La forma de nombrar al primer director general contribuyó a fortalecer la autonomía del INEE: el Presidente de la República aceptó no hacer personalmente la designación, sino encomendar la decisión a la Junta Directiva, que decidió hacer la designación con base en un concurso de méritos a partir de una convocatoria pública. Sin intervención del Secretario de Educación la Junta Directiva designó al primer titular del nuevo organismo, e hizo también el

nombramiento de los 16 primeros consejeros técnicos mediante un procedimiento abierto similar, lo que contribuyó a dar legitimidad a los designados.

PRIMEROS PASOS DEL INEE

El Estatuto Orgánico del Instituto establece cuatro áreas para atender otros tantos aspectos de las evaluaciones: construcción de indicadores, elaboración de pruebas, evaluación de recursos y procesos de las escuelas y proyectos internacionales. Se crearon también áreas para encargarse de la aplicación de instrumentos y la relación con las entidades federativas, así como para la difusión de resultados, además de las áreas de apoyo informático, de asuntos jurídicos y de administración.

Un paso indispensable, que llevó varios meses completar, fue la delimitación de las funciones precisas de evaluación que corresponderían en el futuro al nuevo Instituto y a la DGEP, que no desapareció. El resultado del proceso fue definir la misión del INEE, como la de evaluar el conjunto del sistema educativo como tal, y no sus componentes individuales, alumnos, maestros o escuelas; las evaluaciones de personas e instituciones consideradas individualmente seguirían correspondiendo a la Secretaría.

Esta delimitación, que no siempre se entiende bien, es congruente con una visión de la evaluación que la define ante todo como una herramienta que puede ser valiosa para la mejora educativa, pero que también puede ser contraproducente y traer consecuencias negativas si se utiliza mal, y que considera que, dada la complejidad del sistema educativo, una buena evaluación deberá consistir en un verdadero sistema, un conjunto con varias herramientas de enfoques y características distintas y complementarias.

A fines de 2003, el INEE publicó los resultados de sus primeros trabajos: en primer lugar, el anuario de indicadores del sistema educativo denominado *Panorama Educativo de México*. El interés del Instituto por que los productos de su primer año de trabajo incluyeran éste se entiende como un esfuerzo por contrarrestar la tendencia a reducir toda la evaluación a la que se basa en la medición del rendimiento escolar, mediante la aplicación de pruebas a los alumnos de ciertos grados, en ciertas áreas curriculares. Con sus anuarios de indicadores el INEE ha querido destacar siempre que la calidad educativa incluye otras dimensiones, como la cobertura o la eficiencia terminal, que un sistema de evaluación no puede ignorar, y que un sistema de indicadores integra.

A fines de 2003 el INEE difundió también un informe con resultados de pruebas de rendimiento, las Pruebas de Estándares Nacionales de lectura y matemáticas aplicadas por la SEP en junio de ese mismo año a muestras de alumnos de primaria y secundaria todo el país. El INEE hizo también un 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 y no permitían llegar a conclusiones sólidas sobre el nivel general del aprendizaje en el 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 representaban un esfuerzo muy grande y ofrecían resultados de dudoso valor.

Las legítimas demandas por tener información confiable sobre la educación, y el escaso conocimiento de aspectos técnicos, habían llevado a una proliferación de proyectos de evaluación cuyo diseño no siempre era el apropiado para el propósito perseguido. Era frecuente que el propósito mismo no estuviera claro, o que se usara un instrumento elaborado con cierto propósito para otros para los que no resultaba adecuado.

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

- Desarrollar una nueva generación de pruebas, de calidad superior a las anteriores, gracias a las cuales sería posible contar con información confiable y comparable en el tiempo sobre los niveles de aprendizaje alcanzados por los alumnos.
- Impulsar la difusión de resultados para su uso por autoridades, maestros y sociedad, en contraste con el ocultamiento previo de la información.
- Promover la integración de los diversos esfuerzos de evaluación existentes en un auténtico sistema nacional de evaluación educativa, reduciendo el número de pruebas y evitando duplicidades innecesarias, no como una nueva estructura, sino como una red de instancias con una normatividad básica común y mecanismos eficientes de coordinación y control de calidad.

A lo largo de los años siguientes se avanzó en las direcciones apuntadas:

- Cada año se publicó un volumen de indicadores, *Panorama Educativo de México*, con información que antes era difícil encontrar, y análisis novedosos y consistentes que permiten dimensionar los principales problemas del sistema educativo.
- Se desarrollaron nuevos instrumentos para evaluar las principales áreas del currículo de secundaria y primaria, y luego también de educación preescolar, con una metodología rigurosa e involucrando a maestros, especialistas en contenidos y en psicometría. Estas pruebas se denominaron *Exámenes de la calidad y el logro educativo* (Excale); desde 2005 se aplican, en un ciclo de cuatro años, a muestras nacionales y estatales de 3° de secundaria, 6° y 3° de primaria y 3° de preescolar. Los resultados han permitido a autoridades, maestros y público en general, tener por primera vez datos confiables sobre los niveles de aprendizaje alcanzados por los alumnos de primaria y secundaria, en relación precisa con lo que establecen al respecto los programas de estudio, así como saber con bases sólidas si esos niveles avanzan o no a lo largo del tiempo.
- Se llevaron a cabo operativos anuales para la evaluación de ciertos recursos y procesos que tienen lugar en las escuelas, cubriendo también los tres niveles de la educación básica, en un ciclo similar al de las pruebas Excale.
- Se asumió la participación de México en las evaluaciones internacionales de la OCDE (Programa para la Evaluación Internacional de Estudiantes, PISA) y la UNESCO (Laboratorio Latinoamericano de Evaluación de la Calidad Educativa), consiguiendo alcanzar un nivel de participación similar al de los países más avanzados lo que, además de mejorar la imagen del país, permitió aprovechar las mejores experiencias internacionales en evaluación, y sacar provecho de la rica información que arrojan los resultados de esas evaluaciones.
- Se continuaron y ampliaron 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 organizaron periódicamente actividades de capacitación.
- Para promover la integración del Sistema Nacional de Evaluación, se elaboró un documento de *Política Nacional de Evaluación*, que precisa el enfoque que habrá

de darse a la evaluación, y los usos que se deberán de hacer de sus resultados, para que su impacto sea positivo para la mejora de la calidad educativa. (--)

LA EVALUACIÓN EDUCATIVA EN MÉXICO A PARTIR DE 2006

En 2006, último año de la administración federal encabezada por Vicente Fox, se dieron pasos importantes en el campo de la evaluación, en seguimiento de lo hecho en años anteriores.

LAS EVALUACIONES DEL LA SEP

El paso más importante por sus consecuencias futuras, positivas o no, fue sin duda la puesta en marcha de un nuevo proyecto de evaluación, los *Exámenes Nacionales del Logro Académico en Centros Escolares*, ENLACE.

Estas nuevas pruebas, que se aplicaron por primera vez en 2006 y luego cada año, se distinguen por ser de tipo censal, o sea que se aplican a todos los alumnos de los grados que comprenden, que inicialmente fueron del 3° al 6° de primaria y el 3° de secundaria, y luego se ha planteado que se extiendan a todos los grados de la primaria y la secundaria, así como a la enseñanza media superior.

Inicialmente, la idea de desarrollar un sistema de pruebas censales como ENLACE se debió en parte a un esfuerzo por atender la demanda de algunos sectores, especialmente de los medios de comunicación y la empresa privada, de contar con resultados de cada una de las escuelas del sistema educativo, con base en los cuales se pensaba que habría bases suficientes para tomar decisiones fuertes que llevarían a la mejora de la calidad, en la forma de estímulos a las escuelas de mejor desempeño, y correctivos a las que mostraran resultados bajos.

La reflexión que hicieron las autoridades de la SEP, con participación del INEE, llevó a la conclusión de que el alcance esperable de unas pruebas censales no era tan grande como pensaban quienes pedían que se desarrollaran. El conocimiento al respecto hace ver que esas pruebas, por bien hechas que estén, tienen inevitablemente límites en lo que se refiere a la gama de contenidos y niveles cognitivos que pueden evaluar, en cuanto a la confiabilidad de los resultados que se obtienen en aplicaciones hechas en una escala masiva y, sobre todo, en lo que tienen que ver con la no consideración de muchos factores del entorno de cada escuela y del interior de cada una de ellas, para interpretar correctamente los resultados y no atribuirlos sin más a méritos o deficiencias de los docentes y las mismas escuelas.

Los resultados obtenidos por los alumnos en pruebas en gran escala pueden ser un indicador valioso entre otros, pero de ninguna manera el único para evaluar la calidad de una escuela o un docente. Para esto sería necesario, entre otras cosas, usar diseños longitudinales que permitan estimar de manera confiable el *valor agregado* por escuelas y maestros al aprendizaje de los alumnos. La experiencia de otros países muestra, además, que el hecho de asociar consecuencias fuertes a las evaluaciones en gran escala del rendimiento escolar tiende pronto a distorsionar gravemente su uso, dando lugar a prácticas que empobrecen la enseñanza y llevan acciones francamente fraudulentas.

Por todo lo anterior, la decisión tomada a principios de 2006 para poner en marcha el desarrollo de ENLACE concebía las nuevas pruebas como una herramienta de enfoque formativo, que debería servir para retroalimentar a maestros, alumnos y padres de familia sobre el avance de cada chico, pero sin asociarse directamente con medidas correctivas o supuestamente de estímulo. En particular se proponía desligar las nuevas pruebas del programa de Carrera Magisterial, y que las pruebas desarrolladas antes desaparecieran, para evitar la sobrecarga de trabajo para las escuelas.

Además de lo anterior, en 2006 se dieron avances significativos en lo relativo a la integración de los esfuerzos de evaluación existentes en un verdadero sistema nacional, con la planeación del Sistema Nacional de Indicadores Educativos, hecha de manera conjunta por el INEE y la SEP, que incluyó la difusión de un primer anuario de indicadores publicado también conjuntamente por ambas instancias. Se preparó también el marco regulatorio del Sistema Nacional de Planeación y Evaluación Educativas, y en particular de los subsistemas de indicadores y de evaluación. Estos esfuerzos, sin embargo, no culminaron en la expedición de acuerdos u otros instrumentos normativos que les dieran fuerza legal y aseguraran su continuidad, y de hecho hasta el momento en que se escribe este artículo (fines de 2009) no ha habido avances al respecto.

Por lo que se refiere a las pruebas ENLACE, desde que se difundieron los resultados de la primera aplicación, en septiembre de 2006, y pese a las intenciones en contrario, prevaleció la tendencia a interpretar simplistamente sus resultados, utilizándolos para preparar ordenamientos de escuelas, supuestamente según su calidad, con base sólo en los resultados obtenidos por los alumnos, sin matiz alguno según las condiciones de la escuela, su tamaño, el número de los alumnos evaluados, etc. Estas tendencias se acentuaron tras las aplicaciones de 2007 y 2008, al grado que las pruebas ENLACE parecen haberse vuelto el elemento decisivo

para formular cualquier juicio sobre la calidad educativa. El propósito formativo previsto inicialmente se ha visto desvirtuado por usos que han vuelto de gran impacto las pruebas ENLACE, lo que ha hecho ya 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*) con propósitos similares.

Por otra parte, la SEP continuó manejando otros instrumentos de evaluación que se habían comenzado a aplicar previamente. Aunque la idea original al planear las pruebas ENLACE había sido que éstas sustituyeran a las anteriores, varias se han seguido aplicando, aunque en menor escala, como fue el caso de las pruebas de ingreso a secundaria y las de las Olimpiadas del Conocimiento, por lo que se refiere a alumnos de educación básica.

En 2007 la Subsecretaría de Educación Media Superior (SEMS) de la SEP, en el marco de un proyecto más amplio de reforma de ese tipo educativo, planteó también, junto con la Unidad de Planeación y Evaluación de Políticas Educativas (UPEPE), el inicio de un nuevo proyecto de evaluaciones censales, esta vez para alumnos de educación media superior, proyecto que se designó informalmente con la expresión *ENLACE de bachillerato*. El desarrollo de estas pruebas se encomendó al Centro Nacional de Evaluación para la Educación Superior (Ceneval), que en 2008 aplicó por primera vez esas pruebas, que se planea aplicar anualmente.

Por otra parte, se siguieron aplicando también las evaluaciones de maestros, en concreto las pruebas para valorar el Factor Preparación Profesional del programa de estímulos de Carrera Magisterial. 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 por las autoridades educativas o el sindicato, para basarse en evaluaciones objetivas que, por única vez, en 2008 serían manejadas por la SEP, pero a partir de 2009 lo serían por un organismo independiente. En febrero de 2009 se ha anunciado la creación de ese cuerpo, pero todavía no hay información precisa sobre su composición y atribuciones.

LAS EVALUACIONES DEL INEE

En 2007 y 2008 el trabajo del INEE siguió adelante, plasmándose en un número creciente de publicaciones, de indicadores educativos, de resultados de las pruebas Excale y las internacionales, y de proyectos de evaluación de recursos y procesos de las escuelas⁴.

Continuaron también los esfuerzos de mejora técnica; los informes publicados evidencian que se han alcanzado niveles de calidad acordes con los estándares internacionales al respecto. Puede mencionarse el desarrollo de pruebas de respuesta abierta para evaluar expresión escrita y niveles cognitivos complejos, así como de pruebas de ejecución para alumnos de preescolar; el desarrollo de indicadores complejos; la utilización de métodos complejos de muestreo; la evaluación en gran escala de recursos y procesos de las escuelas; el uso de 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 lo relativo a la intervención en la preparación de las pruebas como en cuanto al análisis de los resultados. Se elaboraron y difundieron informes nacionales tanto de PISA 2006 como del Segundo Estudio Regional del Laboratorio Latinoamericano.

El trabajo de difusión también avanzó, con el creciente número de publicaciones y de visitas a la página web, pero el trabajo del Instituto perdió visibilidad ante el creciente impacto de las pruebas ENLACE.

En 2007, y además de lo relativo a las pruebas ENLACE para ese tipo educativo, la Subsecretaría de Educación Medias Superior emprendió un proyecto más ambicioso para desarrollar evaluaciones que cubran de manera completa las diversas dimensiones de la calidad de la educación media superior, y encomendó a diferentes instancias el desarrollo de cada componente. El INEE quedó a cargo de la evaluación del conjunto del sistema nacional de educación media superior, para lo que comenzó en 2008 a desarrollar tres elementos, similares a los previamente trabajados para educación básica: un sistema de indicadores, uno de pruebas muestrales de rendimiento, y uno de recursos y procesos de las escuelas.

Es importante añadir que, a diferencia de la situación prevaleciente hasta 2006, cuando la relación entre el INEE y la SEP en lo relativo a evaluación era muy estrecha, a partir de 2007 la relación entre ambas instituciones ha sido escasa y no exenta de tensiones, que reflejan las diferencias de perspectivas que cada parte tiene sobre la evaluación.

⁴ En la página web www.inee.edu.mx pueden verse todas las publicaciones del Instituto, que en 2008 llegaron a más de 200.

Como se ha dicho ya, tras el cambio del gobierno federal no tuvieron continuidad los trabajos previos para la integración del sistema nacional de evaluación, que no se ha constituido, y no continuó la colaboración entre el INEE y la SEP para el desarrollo del sistema de indicadores, ni en lo relativo a las pruebas ENLACE.

La evaluación en las entidades federativas

En paralelo al trabajo de evaluación de la SEP y el INEE, en las entidades federativas ha habido también actividades diversas. En algunos estados hay Institutos de Evaluación, con características diferentes, a veces en lugar de las Áreas Estatales de Evaluación que se crearon desde la década de 1990 (Chiapas), a veces además de ellas (Sonora, Nuevo León, Estado de México). Otras entidades tienen planes con diverso grado de avance para crear institutos similares (Sinaloa, Tlaxcala, Veracruz). Todas las entidades participan en las evaluaciones nacionales, tanto de la SEP como del INEE, así como en las internacionales. Nuevo León participó como un país más en el Segundo Estudio Regional del Laboratorio Latinoamericano de la UNESCO. Algunas entidades, además (como Aguascalientes, Baja California, por medio de un convenio con la Universidad Autónoma del Estado, Coahuila, Colima y Sonora) manejan sus propias evaluaciones, también con diversos enfoques y calidad. En este sentido, puede afirmarse que a nivel estatal el proceso de institucionalización de los sistemas de evaluación es, en la mayoría de los casos, incipiente. Se observan, además, disparidades importantes en el grado de consolidación técnica e institucional de estos sistemas, las cuales son un reflejo de las disparidades regionales de México.

CONCLUSIÓN. BALANCE Y PERSPECTIVAS

SITUACIÓN ACTUAL

Un análisis realizado en el INEE sostenía que, *con cierta crudeza*, era posible afirmar que, desde la década de 1970 hasta fines del siglo XX, en México había *muchas actividades relacionadas con la evaluación educativa pero no una verdadera evaluación*. Había, en efecto, muchos datos, algunos derivados de la aplicación de pruebas, pero no se llegaba a formular juicios de valor sobre el estado del sistema educativo, integrando los datos derivados de mediciones y estadísticas y contrastándolos con parámetros pertinentes. Mucho menos se llegaba a explorar explicaciones y utilizar los resultados para sustentar decisiones de mejora.

En la primera década del siglo XXI la situación ha cambiado mucho. El balance que parece posible a fines de 2009 sobre la situación actual de la evaluación educativa en México tiene, sin embargo, aspectos favorables y desfavorables.

En el lado positivo se deben mencionar, al menos tres puntos:

- Los avances técnicos conseguidos, y la formación de un número pequeño pero significativo de técnicos y especialistas de muy buen nivel, en contraste con la casi total ausencia de personas preparadas que prevalecía hace pocos años.
- La creciente conciencia ciudadana del derecho a conocer los resultados de las evaluaciones, que contrasta drásticamente con el hermetismo anterior.
- El que algunas autoridades educativas, tanto en el nivel federal como en algunos estados, comiencen a hacer uso de los resultados de las evaluaciones (tanto de resultados de pruebas, como de otros indicadores) para toma de decisiones

También pueden mencionarse al menos tres aspectos desfavorables:

- El número ya excesivo y todavía creciente de evaluaciones que se desarrollan y aplican en las escuelas, y su desarticulación, que pesa cada vez más sobre alumnos, maestros y escuelas.
- El predominio de la evaluación en gran escala sobre la que deben hacer los maestros en el aula, viendo a la primera como sustituto de la segunda, y no como un complemento a la tarea insustituible de los maestros.
- El uso inapropiado cada vez más frecuente de los resultados, y su excesivo peso en el diseño de las políticas públicas.

Estos tres aspectos negativos están estrechamente relacionados, y tienen como fuente común el desconocimiento de los alcances y limitaciones reales de cada enfoque de la evaluación, lo que lleva a expectativas desmesuradas y a usos carentes de fundamento.

Se ha mencionado ya el cambio de enfoque de las pruebas ENLACE, de bajo a alto impacto, y su creciente peso como referente privilegiado para las decisiones de política educativa. Algo similar, aunque con matices diferentes, ha ocurrido con las pruebas PISA de la OCDE. A diferencia de lo que pasó en 2001 y 2004, cuando se difundieron los resultados de las dos primeras aplicaciones de esas pruebas, los que de la aplicación de 2006, que se dieron a conocer en diciembre de 2007, atrajeron poderosamente la atención, al grado de que

PISA se ha vuelto el referente para la primera meta del Programa Sectorial de Educación del sexenio del Presidente Calderón.

PERSPECTIVAS Y RETOS

Desde 2008 han aparecido nuevos elementos en el panorama educativo del país, en especial en lo que tiene que ver con la evaluación.

Sobresale la firma, por parte de las autoridades de la SEP y la dirigencia del SNTE, de un convenio denominado *Alianza por la Calidad de la Educación* que, entre otras cosas, contempla el establecimiento de un sistema nacional de evaluación. El documento alude también a una propuesta que el SNTE ha hecho en varias ocasiones, para que se establezca una *Norma Oficial Mexicana* que cubriría todos los aspectos de la educación, indicando los requisitos que deberían satisfacer para considerarse adecuados. Esos aspectos incluyen no sólo el aprendizaje de los alumnos, sino también contenidos curriculares, características que deberían reunir los edificios escolares así como su mobiliario y equipamiento, perfil de los docentes y directores de escuela, entre otros puntos.

El documento que desarrolla la propuesta de *Norma Oficial* del SNTE, además de deficiencias técnicas importantes, implica que todas las decisiones de carácter normativo sobre el sistema educativo, hasta ahora reservadas constitucionalmente a las autoridades educativas federales y, en algunos aspectos, estatales, sería competencia de instancias paritarias de la Secretaría y el Sindicato.

Hasta fines de 2009, sin embargo, no se ha integrado el Sistema Nacional de Evaluación Educativa, ni se ha discutido abiertamente la propuesta de Norma Oficial. Los desacuerdos entre las autoridades y la dirigencia sindical, al parecer, no han permitido que se avance en ninguno de esos terrenos. Añade confusión a la situación la aparente alianza política entre la presidencia de la República y la fuerza política del magisterio.

En la perspectiva del futuro, la evaluación en el sistema educativo nacional tiene retos considerables por enfrentar, para que los aspectos positivos de su desarrollo reciente puedan continuar, y se puedan corregir las deficiencias actuales.

El reto fundamental es el que se puede denominar como la reconceptualización de las evaluaciones para que, con base en una reflexión muy seria sobre los alcances y límites de cada enfoque, se diseñe un conjunto balanceado, en el que diferentes acercamientos aporten elementos específicos, complementándose unos a otros, y con una clara y adecuada

valoración de la evaluación en aula a cargo de los maestros, como elemento clave del conjunto, al que los elementos de gran escala apoyan, pero no pueden sustituir.

Un segundo desafío, estrechamente relacionado con el anterior, es el que consiste en el desarrollo de lo que se puede llamar la cultura de la evaluación entre todos los actores que tienen que ver con el sistema educativo incluyendo, desde luego, a los maestros y a las autoridades educativas, pero también a los padres de familia, los medios de comunicación y la sociedad en general, para que sus legítimas exigencias de mejor calidad y transparencia no se orienten por rumbos inadecuados⁵.

Un tercer reto se refiere al uso de los resultados por parte de todas las instancias: por las autoridades para sustentar mejor el diseño y la implementación de políticas, pero también por los maestros, para retroalimentar su práctica docente, y por alumnos y padres de familia, para que sus acciones apoyen el trabajo de las escuelas.

Por último, sin que ello signifique desconocer su importancia crucial, el permanente reto de mejorar las evaluaciones, para que sus resultados sean cada vez más válidos y confiables y, por lo tanto, más útiles para sustentar buenas decisiones.

En la tarea de mejorar las evaluaciones habrá que prestar especial atención a la dimensión fundamental de la validez, que es más difícil de cuidar que la confiabilidad. Una medición válida de la calidad educativa supone, como punto de partida, ponerse de acuerdo en qué es una buena escuela, que implica mucho más que conseguir que los alumnos obtengan altas puntuaciones en pruebas de lectura y matemáticas.

Sin desconocer la importancia de lo anterior, una buena escuela tiene que hacer mucho más: formar a los ciudadanos pensantes y comprometidos con el futuro que el país necesita. Y valorar la medida en que un propósito tan importante y complejo se cumple es un reto mayúsculo, que los organismos de evaluación no deberán eludir.

⁵ Una cultura de la evaluación desarrollada debería comprender al menos: a) capacidad de interpretación adecuada y uso de los resultados; b) compromiso con la transparencia, tanto en el diseño y aplicación como en la difusión de los resultados; c) expectativas positivas respecto del papel de la evaluación.

REFERENCIAS

- Backhoff, Eduardo y Guillermo Solano-Flores, (2003), *Tercer Estudio Internacional de Matemáticas y Ciencias Naturales (TIMSS): Resultados de México en 1995 y 2000*, INEE, México.
- Blanco, Emilio (2008), “Factores escolares asociados a los aprendizajes en la educación primaria mexicana: un análisis multinivel” *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en la Educación*. Vol. 6, N°1.
- Edwards, Verónica, (1991), *El concepto de calidad de la educación*, UNESCO / OREALC, Santiago de Chile.
- Fernández, Tabaré (2007), *Distribución del conocimiento escolar: clases sociales, escuelas y sistemas educativos en América Latina*. El Colegio de México.
- Fernández, Tabaré y Carmen Midaglia, (2003) *Quiénes y cómo usan los informes generados por los sistemas de evaluación de aprendizajes en la educación primaria. Los casos de México y Uruguay*, Programa de Promoción de la Reforma Educativa en América Latina (PREAL).
- INEE (INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN), (2003), *La Calidad de la Educación Básica en México. Primer Informe Anual, 2003*, INEE, SEP, México.
- INEE (INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN), (2007), *PISA 2006 en México*, INEE, SEP, México.
- LGE (Ley General de Educación), (1993), Decretada por el Congreso de los Estado Unidos Mexicanos el 13 de Julio de 1993.
- SECRETARÍA DE EDUCACIÓN PÚBLICA (SEP), (2000a), *Distribución de los planteles públicos de educación primaria y secundaria, según el nivel de aciertos de sus alumnos en los exámenes de Carrera Magisterial*, SEP, México.
- SECRETARÍA DE EDUCACIÓN PÚBLICA (SEP), (2000b), *Las escuelas primarias rurales y los apoyos de los programas compensatorios*, Reporte final del primer Estudio/Diplomado sobre bases metodológicas de investigación cualitativa. SEP, México.
- SECRETARÍA DE EDUCACIÓN PÚBLICA (SEP), (2001), *¿Cómo transformar las escuelas? Lecciones desde la gestión escolar y la práctica pedagógica*, Segundo estudio/diplomado. Reporte final. SEP, México.
- UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO), (2005), *Global Monitoring Report. Education For All – The Quality Imperative*.
- Velázquez, Víctor, (2000), “Hacia una cultura de la evaluación. 1994-2000”, En: *Memoria del Quehacer Educativo. 1995-2000* (2), SEP, México.
- Vidal, Rafael, María A. Díaz y Javier Loyola, (2003), *El Proyecto PISA: su Aplicación en México*. Cuaderno N° 9, Colección Cuadernos de Investigación, INEE, México.