

Evaluación

# Avances y desafíos en la evaluación educativa

**Elena Martín**  
**Felipe Martínez Rizo**  
Coordinadores

**Metas  
Educativas**  
2021

La educación que queremos  
para la generación de los  
Bicentenarios

Organização  
dos Estados  
Ibero-americanos



Organización  
de Estados  
Iberoamericanos

OEI

Para la Educación,  
la Ciencia  
y la Cultura

Fundación **Santillana**

Evaluación

# Avances y desafíos en la evaluación educativa

Elena Martín  
Felipe Martínez Rizo  
Coordinadores

**Metas  
Educativas**  
2021

Organização  
dos Estados  
Ibero-americanos



Para a Educação,  
a Ciência  
e a Cultura

**OEI**

Organización  
de Estados  
Iberoamericanos

Para la Educación,  
la Ciencia  
y la Cultura

Fundación **Santillana**

© Del texto: Organización de Estados Iberoamericanos  
para la Educación, la Ciencia y la Cultura (OEI)  
C/ Bravo Murillo, 38  
28015 Madrid, España  
[www.oei.es](http://www.oei.es)

Las opiniones de los autores expresadas en este libro no representan necesariamente los puntos de vista de la OEI.

La colección METAS EDUCATIVAS 2021 es una iniciativa de la OEI en colaboración con la Fundación Santillana.

Impreso en España por Intigraf, S. L.

ISBN: 978-84-7666-194-9  
Depósito legal: M-14081-2009

# Índice

Preámbulo, <i>Álvaro Marchesi</i> .....	7
Introducción, <i>Elena Martín y Felipe Martínez Rizo</i> .....	11
La evaluación educativa hoy .....	15
Evaluación y cambio educativo: los debates actuales sobre las ventajas y los riesgos de la evaluación, <i>Alejandro Tiana</i> .....	17
La evaluación de la calidad de los sistemas educativos: propuesta de un modelo, <i>Felipe Martínez Rizo</i> .....	27
Las evaluaciones internacionales, <i>Enrique Roca</i> .....	41
Las evaluaciones regionales y nacionales en América Latina, <i>Guillermo Ferrer</i> .....	53
Las múltiples caras de la evaluación .....	65
Indicadores y desafíos de los sistemas educativos en América Latina, <i>Margarita Poggi</i> .....	67
La evaluación de políticas y programas: desafíos para la gestión en educación, <i>Marta Kisilevsky</i> .....	81
Currículo y evaluación estandarizada: colaboración o tensión, <i>Elena Martín</i> .....	89
La evaluación de centros: riesgos y posibilidades, <i>Leda Badilla</i> .....	99
La evaluación del desempeño docente para el desarrollo de las competencias profesionales, <i>Pedro Ravela</i> .....	113
Evaluación de los aprendizajes en el aula: una conceptualización renovada, <i>Juan Manuel Esquivel</i> .....	127
El sentido de la evaluación .....	145
Difusión y uso de resultados de evaluaciones educativas a gran escala en América Latina, <i>Patricia McLauchlan de Arregui</i> .....	147
Los resultados de las evaluaciones y su papel en las políticas educativas, <i>Álvaro Marchesi</i> .....	161
Bibliografía .....	169
Los autores .....	179

# Difusión y uso de resultados de evaluaciones educativas a gran escala en América Latina

*Patricia McLauchlan de Arregui*

## INTRODUCCIÓN

Así como durante la etapa de expansión del acceso a la educación en Latinoamérica se dio prioridad a la recogida y utilización de indicadores de cobertura y niveles de escolaridad, cuando lograr la efectividad de las escuelas y el aprendizaje real se convirtió en el principal desafío a finales de los ochenta, y se tuvo que diseñar y justificar nuevas estrategias para la asignación de recursos, la medición de la calidad de la educación y los aprendizajes se hizo impostergable.

Así, a mediados y finales de los años noventa la mayoría de países de América Latina empezó a dar sus primeros pasos hacia el desarrollo de capacidades técnicas básicas para diseñar e implementar evaluaciones a gran escala y para emprender análisis estadísticos de los datos recogidos. Una visión general del desarrollo de los sistemas de evaluación de los logros de aprendizaje puede verse en Ferrer (2006).

Hoy, tras varios años de experiencias, en buen número de países se están dando discusiones y acciones orientadas al rediseño de los sistemas de evaluación. Una parte del debate se refiere a necesarias mejoras técnicas en el desarrollo de instrumentos y procedimientos para la recolección y el análisis de datos, así como al alineamiento de los mismos con los objetivos curriculares vigentes. Otra parte tiene que ver con la constatación de que los esfuerzos para comunicar y difundir los resultados de las pruebas no han alcanzado a muchos agentes clave, incluyendo profesores y padres (Ravela, 2002). Tanto o mayor desasosiego genera el hecho de que, aparentemente, la abundante información que se está produciendo está siendo usada de modo insuficiente para sustentar y apoyar procesos de mejora, tanto en las políticas como en la enseñanza, o como instrumento de responsabilización o *accountability*. Las evaluaciones parecerían, en opinión de algunos, no estar teniendo mucho impacto *más allá de informar, sensibilizar y crear conciencia* (Moncada *et al.*, 2005). Por supuesto, no faltan quienes están más preocupados aún por los usos no válidos o hasta perversos de las evaluaciones para el logro de objetivos para los cuales no fueron diseñadas.

La primera sección de este capítulo revisa de manera resumida la experiencia emergente de los países de América Latina en la difusión de resultados de evaluaciones, mientras que la segunda pasa revista al uso de los resultados de evaluaciones nacionales, sub-nacionales, regionales o internacionales. Por último, se presentan algunas generalizaciones y advertencias sobre los desafíos que supone una utilización más productiva de los resultados.

## LA DIFUSIÓN DE RESULTADOS<sup>1</sup>

La mayoría de los sistemas nacionales de evaluación desarrollados en América Latina durante los noventa enfatizaron sus finalidades de carácter formativo y sin consecuencias directas sobre las escuelas. Sus roles explícitos fueron generalmente constituirse en proveedores de información para la toma de decisiones por parte de las autoridades y actores del sector educativo, así como para el público sobre el estado de la educación nacional. Su cumplimiento requería, por lo tanto, *difundir* y *utilizar* la información recogida para rediseñar o ajustar políticas, planes, programas y prácticas pedagógicas y de gestión escolar. La concreción de esas intenciones, sin embargo, no ha estado exenta de una serie de limitaciones.

En primer lugar, falta hasta ahora en muchos Ministerios una *visión estratégica sobre el rol de la evaluación*. No se tiene clara la manera en que se espera que esta impacte sobre el sistema educativo. En casos extremos, parecería creerse que el mero hecho de evaluar producirá mejoras o que todos los actores comprenderán y emplearán los resultados que se divulguen. Hay unidades y actores de medición de logros educativos que consideran que su misión consiste únicamente en producir datos e información cuyo análisis no les corresponde y debe ser realizado por otras agencias o actores, pasando por alto que en realidad aún no existen muchas audiencias en condiciones de comprender y utilizar la información, por lo cual es necesario invertir tiempo en formarlas y en "modelar" análisis sólidos y válidos de los datos.

Falta aún en algunos países una definición clara de cuáles son las audiencias a las que se debe y quiere llegar, cuáles son los contenidos y mensajes que se quiere hacer llegar a cada una y cuáles son los vehículos o tipos de productos de difusión apropiados en cada caso. Incluso hoy día no es del todo extraño que se produzca un único informe de carácter predominantemente técnico, pensado desde la perspectiva de los especialistas de la unidad de evaluación y no de las demandas y preocupaciones de las audiencias, ni siquiera de las audiencias internas del sector educativo, obviando también a las familias, la opinión pública, el sistema político, el mundo académico, las organizaciones sociales y hasta a la prensa. En los casos en que sí se procura llegar con informes a estos otros actores, generalmente no se ha evaluado el grado de comprensión y uso que ellos hacen de la información recibida.

Otro rubro de problemas que ha afectado a la difusión de los resultados ha resultado de inconsistencias técnicas y desarticulaciones internas que caracterizaron las actividades iniciales de las agencias evaluadoras.

Un ejemplo de ello es la confusión generada por algunos Ministerios en torno a pruebas que son de tipo descriptivo y diagnóstico, pero que son interpretadas por la prensa y el público en términos de aprobación y reprobación de alumnos. La demanda pública parece ir en un sentido —y quizá deba ser educada para que asuma otra orientación—, pero las pruebas fueron diseñadas desde otra perspectiva, lo cual no permite satisfacer esa demanda.

Otro problema crucial que ha dificultado la difusión y aprovechamiento es la falta de articulación y comunicación entre las unidades de evaluación y de currículo al interior de los Ministerios. Suele ocurrir que las pruebas no estén adecuadamente alineadas con los objetivos establecidos

---

<sup>1</sup> Esta sección está basada fundamentalmente en el documento *¿Cómo aparecen los resultados de las evaluaciones educativas en la prensa?* del Grupo de Trabajo sobre Estándares y Evaluación del PREAL.



en el currículo o que estos últimos estén formulados en términos tan vagos y generales que es imposible realmente evaluar su logro. En estos casos, las evaluaciones no pueden responder a la pregunta principal de si los alumnos están aprendiendo lo que se espera que aprendan y qué tan bien o mal están al respecto. En su capítulo, Elena Martín analiza con más detalle las relaciones entre currículo y evaluación de rendimiento.

No menos grave ha sido el uso político de los resultados por los propios Ministerios. En parte debido a una suerte de visión ingenua que han tenido algunas altas direcciones de esas instituciones respecto a que los resultados pueden mejorar rápidamente, ellas no han estado muchas veces preparadas para manejar con solvencia los malos resultados ante la prensa. Al respecto se han observado diversas reacciones: a) las autoridades simplemente deciden no publicar los resultados o, lo que es más común en años recientes, los publican casi "subrepticamente", sin una estrategia integral de difusión; b) al no dominar el tema, transmiten mensajes que incluyen errores de interpretación importantes; c) utilizan los malos resultados para desprestigiar a la administración anterior (normalmente al inicio de un mandato) o para presionar sobre el cuerpo docente, aunque luego se esperará que los mismos mejoren rápidamente antes del término de su gestión. Como dice Pedro Ravela, "normalmente los funcionarios políticos esperan buenos resultados que les sirvan como propaganda o aval a su gestión, y los malos resultados, que suelen ser los más comunes, se tornan una 'papa caliente' que no saben cómo manejar" (2003).

Un tema vinculado a los anteriores es el dilema de la independencia política de las unidades de evaluación. Estas producen y divulgan la información sobre logros educativos, pero en muchos casos son a la vez parte interesada, lo cual coloca inevitablemente a la información bajo sospecha para los periodistas y otros actores. Esta falta de confianza se da incluso cuando la unidad de evaluación está fuera de la estructura ministerial, ya que igualmente dependen de los recursos del Estado para realizar las evaluaciones y de decisiones políticas para difundirlos.

El problema no radica tanto en el marco institucional elegido como en la cultura política y en la cultura de la evaluación. Es necesario –y en muchos países se ha logrado avances en ese sentido– que los Ministerios de Educación conciban la labor de evaluación como una de información al público y otorguen a las unidades de evaluación amplia independencia en la difusión de los resultados. Igualmente se requiere que la sociedad se acostumbre a que existe una unidad de evaluación que periódicamente informa sobre el estado del sistema educativo y de los aprendizajes de manera clara, técnicamente apropiada y sin interferencias de tipo político. Ello no impide, por supuesto, que los Ministerios deban tener y ofrecer su propia visión e interpretación de los resultados y debatan con otros sectores de opinión al respecto.

Un último grupo de problemas tiene que ver con las débiles relaciones con los medios de comunicación que mantienen algunas unidades evaluadoras. Estas parecen a veces desconocer la lógica y el *modus operandi* de los medios y no haberse trazado estrategias ni dedicar recursos para vincularse de mejor manera con ellos. Por su parte, el personal de prensa de los Ministerios está generalmente dedicado principal o casi exclusivamente a organizar y difundir las apariciones públicas de las autoridades y a cuidar su imagen, antes que a facilitar la comunicación con los medios sobre los temas educativos.

Frente a esta situación, el Grupo de Trabajo sobre Estándares y Evaluación del PREAL ha recomendado que las agencias de evaluación propongan, y los Ministerios definan, *qué preguntas pretenden responder las evaluaciones*, qué tipo de información se va a producir, quiénes y cómo se

espera que utilicen la información. Luego, que *definan una estrategia de difusión de resultados*, identificando diferentes audiencias y sus respectivas preguntas, así como el tipo de productos de comunicación apropiados para cada una. Se recomienda informar siempre de modo transparente, sin interferencias políticas, y entregar la información a actores cualificados con diferentes posturas, a los cuales pueda recurrir la prensa para obtener interpretaciones y visiones diferentes a partir de los resultados, evitando la tentación de usar los resultados con fines propagandísticos o de ocultarlos cuando no son satisfactorios.

Es recomendable que la misma unidad de evaluación explique los resultados, incentivando a las autoridades del sector a que reflexionen e informen sobre las implicancias que tienen para las políticas educativas. Ello requeriría que se preparen mensajes claros, balanceando buenas y malas noticias, evitando tanto el catastrofismo como el triunfalismo, presentando los resultados como desafíos al sistema, *como un diagnóstico y no como autopsia* (Ravela, 2003), y enfatizando las líneas de acción con que se plantea enfrentar los problemas que se ponen de manifiesto.

Para fomentar la apropiación de los resultados de la evaluación, es necesario que actores de otras oficinas del Ministerio se sientan también propietarios del proceso, participando en el diseño y validación de las pruebas, así como en la interpretación de resultados. Contribuiría también a la aceptación transmitir que el sistema de medición seguramente tiene debilidades, pero que está abierto a mejoras, para lo cual se quiere recoger la opinión cualificada de actores diversos, incluyendo a la prensa, a la cual se quiere convertir en aliado crítico.

Es conveniente que los mensajes sean pocos, claros, reiterativos y basados en la evidencia resultante de las pruebas y del análisis de los factores vinculados. Debe ejemplificarse el tipo de conclusiones que no pueden ser derivadas de los datos y las principales limitaciones de la evaluación, a la vez que se preparan explicaciones pertinentes y al alcance de diferentes audiencias acerca de aspectos técnicos centrales, como la equivalencia (o no) de las pruebas con mediciones anteriores; el significado de las métricas empleadas; la posibilidad (o no) de interpretarlas en términos de *aprobación*; el modo en que se han establecido *líneas de corte* (si se ha hecho), o el significado de los *niveles* (si se han establecido).

El tenor de los mensajes debería ser que la educación es una política de Estado y un compromiso de toda la sociedad, evitando tanto la búsqueda de culpables o responsables únicos (los docentes, las escuelas, los supervisores) como el uso de las acciones educativas con fines de propaganda política, informando con modestia y cautela sobre los avances y los desafíos pendientes.

En términos de *las relaciones con la prensa*, es necesario asegurar la disponibilidad de recursos humanos idóneos y suficientes para la labor de comunicación con ella. Deberán organizarse actividades de capacitación para periodistas con relación a los temas clave de las evaluaciones, idealmente en momentos separados de la presentación de resultados. Muchas de las explicaciones y conceptos técnicos, así como de las advertencias sobre lo que no puede concluirse válidamente de los datos o el tipo de interpretaciones que no son apropiadas, pueden resultar "sospechosos" para el periodista cuando se hacen en el marco de la difusión de resultados. Convendría a las unidades de evaluación cultivar vínculos con periodistas reconocidos o especialmente interesados en temas educativos, y mantenerlos permanentemente informados, no con el objetivo de contar con "incondicionales acrílicos", sino más bien con "incondicionales acerca de la importancia de la difusión de resultados y del debate público sobre la educación", capacitándolos en modos de hacer "noticia educativa" de manera constructiva.



## LOS USOS DE LOS RESULTADOS

Los resultados de las evaluaciones se han utilizado para varios propósitos, como sigue:

### **Elevar la conciencia ciudadana sobre los retos de la educación**

Si bien las evaluaciones del rendimiento se iniciaron generalmente a partir de la necesidad de informar a los tomadores de decisiones, el contexto de resurgimiento democrático de la época y la idea de *rendición de cuentas* que empezaba a instalarse en el discurso político también hacía necesario satisfacer los derechos de los ciudadanos y de algunas audiencias clave de conocer la situación y las tendencias de los resultados educativos. Los resultados han provisto de pruebas duras sobre los logros insuficientes y desiguales, han estimulado una preocupación pública sobre el problema y reorientado el foco de la demanda social hacia una distribución más equitativa de oportunidades de aprendizaje.

Aun así, en la mayoría de los casos la prensa recogió de los reportes oficiales solo los resultados más llamativos sobre el bajo nivel de logro general, y en el caso de las pruebas internacionales, sobre los bajos puntajes obtenidos por los participantes latinoamericanos. A menudo se puso más énfasis en las grandes diferencias entre las escuelas privadas y las públicas, y mucho menos en aquellas entre las poblaciones urbana y rural o entre niñas y niños. Rara vez se mencionaron las diferentes características socioeconómicas y culturales de las poblaciones atendidas por escuelas privadas y públicas. Por otro lado, al menos al inicio, los funcionarios públicos rara vez estaban disponibles para analizar los resultados y promover un diálogo capaz de orientar a la opinión pública a partir de las evaluaciones.

Entre las audiencias interesadas son quizá los padres de familia el grupo menos atendido, en lo que respecta a proveerles con información útil para monitorear el desempeño de las escuelas a las que asisten sus hijos o del sistema educacional en su conjunto. Los reportes producidos están más allá de las posibilidades de comprensión de la mayoría de los padres. El manejo de la información por la prensa no ha ayudado a que la información se interprete de forma precisa, y las capacidades de traducción de los sofisticados procedimientos técnicos de evaluación y análisis estadístico son aún limitadas.

### **Mejorar el diseño de las políticas educativas y la asignación de recursos**

Gracias a las evaluaciones nacionales e internacionales, se cuenta con datos más sólidos sobre el bajo rendimiento de los estudiantes, así como sobre las grandes diferencias en los resultados de varios grupos poblacionales y tipos de escuelas, lo que ha hecho posible el establecimiento y la generación de consensos sobre nuevas prioridades generales de política.

La necesidad de que los Ministerios de Educación, los administradores y las escuelas se centren en el aprendizaje de los estudiantes es ahora un mandato público generalizado en la región. Y, si bien la focalización y la discriminación positiva o acción afirmativa están hasta ahora lejos de ser el eje principal de las políticas públicas, existe una creciente conciencia de la necesidad de distribuir más equitativamente las oportunidades de aprendizaje.

Muchas políticas inspiradas por los resultados de las evaluaciones no han logrado trascender el papel donde han sido formuladas; en cambio sí se ha pasado a la acción con reformas curriculares y desarrollo de estándares, producción y distribución masiva de libros de texto y cuadernos de

trabajo, el fomento de la inclusión de estudiantes pobres en escuelas particulares subvencionadas, programas de capacitación o desarrollo profesional docente focalizados en el fortalecimiento de capacidades de comprensión lectora y de razonamiento matemático, o subvenciones diferenciadas que favorezcan a las escuelas y estudiantes más pobres.

#### **Diseñar programas de apoyo y seleccionar escuelas o estudiantes para participar**

Las evaluaciones han permitido identificar áreas más específicas donde hay problemas, analizar sus posibles causas y diseñar programas para superarlos, así como identificar y seleccionar poblaciones objetivo o instituciones educativas que puedan ser beneficiarias.

Un ejemplo de estos programas se dio en Chile, cuando los resultados de las pruebas SIMCE mostraron grandes diferencias de rendimiento entre escuelas privadas, subvencionadas y municipales y se dio inicio como una estrategia de emergencia al P-900, el Programa para 900 escuelas *Ayudando a los pobres*. El 10% de escuelas con el rendimiento más bajo en cada una de las regiones del país, con grandes proporciones de estudiantes pobres urbanos y rurales de nivel primario, fue invitado a participar en el programa, inicialmente concentrado en remodelar aulas y abastecer de equipo básico, libros de texto y bibliotecas a las escuelas, pero luego incluyendo talleres para profesores que buscaban promover la autonomía profesional y la responsabilidad, y espacios de tutoría en grupos pequeños para estudiantes con bajo rendimiento. El P-900 se convirtió más adelante en un programa permanente, centrado en proveer servicios de capacitación de profesores *in situ* y desarrollo profesional o capacitación en servicio de directores a lo largo de tres años. Empezó a incluir también escuelas de nivel secundario y abrió un subprograma especial para escuelas rurales.

El programa se fue modificando conforme la investigación y la práctica iban mostrando más elementos que necesitaban ser atendidos para promover cambios, manteniéndose la prioridad del desarrollo de capacidades fundamentales en lectura, escritura y matemáticas. Las escuelas P-900 –que llegaron a sobrepasar las 1.500– fueron el foco de programas específicos de mejora y eran supervisadas entre 8 y 16 veces al año por funcionarios provinciales y centrales.

Similares en enfoque y objetivos han sido la iniciativa de *Nivelación para la Excelencia* de Bogotá, las escuelas de jornada completa de Uruguay establecidas preferentemente en áreas urbanas empobrecidas y surgidas de la constatación que ellas habían obtenido mejores logros en la primera evaluación nacional, y la jornada completa introducida también en Chile.

Las políticas de focalización se han convertido en una característica frecuente en muchos países de América Latina. Esto requiere indicadores técnicos sólidos y socialmente aceptables con los cuales identificar las poblaciones objetivo de programas especiales de apoyo, y las evaluaciones se han constituido en un insumo importante en este sentido.

#### **Realizar investigación y evaluar el impacto de programas**

Si consideramos que el aprendizaje es la misión central de la educación, es posible afirmar que la variable dependiente más importante para la investigación educativa orientada a la formulación de políticas no estaba disponible en gran parte de Latinoamérica antes de que los sistemas de evaluación comenzaran a generar resultados. Y dado que la investigación debería ser un insumo importante para la toma de decisiones, la región mostraba un serio retraso.

Hasta 1999 en toda América Latina no se podía encontrar más de una veintena de estudios empíricos sobre posibles asociaciones entre las diferencias en rendimiento, diferencias en las características intraescolares o de aula y diferencias en factores externos tales como las características educativas y la condición socioeconómica de las familias. Apenas seis años más tarde una consulta bibliográfica rápida reveló la existencia de 150 estudios en 16 países, que exploran los efectos específicos o conjuntos sobre el rendimiento de factores tales como: a) la condición socioeconómica de los estudiantes, el trabajo durante el año escolar, la raza y la condición nutricional; b) el salario, calificaciones y prácticas de enseñanza en aula de los docentes; c) características de la escuela tales como la cultura organizacional, la autonomía, la composición socioeconómica de los distintos grupos, las prácticas de evaluación en el aula, la cobertura del plan de estudios, las políticas sobre la realización de tareas en casa, el liderazgo de los directores, la infraestructura, la disponibilidad de libros de textos, y d) el financiamiento o administración pública o privada de la escuela, la presencia de programas de bonos escolares (*vouchers*), de programas de alimentación escolar, las prácticas de retención/repetición, el número de alumnos por aula y la duración de la jornada escolar.

Los resultados de las pruebas se están también utilizando cada vez más para estimar el impacto de políticas y programas educativos y sociales. Los Ministerios incluyen muestras mayores de algunas poblaciones escolares, de manera de poder hacer algunas generalizaciones válidas sobre el impacto de intervenciones especiales, tales como escuelas bilingües indígenas (en Perú y Guatemala), programas de expansión escolar basados en la comunidad (*Nueva Escuela Unitaria* de Guatemala, escuelas PRONADE), perspectivas nuevas o experimentales para planes de estudios (*Escuelas de Excelencia* y el SIMAC en Guatemala) y mecanismos de gestión y financiamiento (escuelas privadas subvencionadas en Chile y Nicaragua). Con frecuencia también se ha permitido e incluso se ha estimulado a los investigadores a usar los resultados de las pruebas nacionales o re-aplicaciones o adaptaciones de los instrumentos de las pruebas para evaluar el impacto de programas específicos.

### Revisar expectativas curriculares o establecer estándares de aprendizaje

Hoy se entiende que elevar y aclarar las expectativas de aprendizaje constituye un elemento crucial para su mejora. Si ellas no están bien definidas, no pueden ser utilizadas para alinear de forma efectiva los libros de texto, los planes de estudio y la capacitación docente.

Cuando los sistemas de evaluación en América Latina empezaron a desarrollar pruebas referidas a criterios, experimentaron grandes dificultades para definir los tipos de tareas que se debería esperar que los estudiantes puedan realizar, si han logrado dominar un área particular de los contenidos. También confrontaron –y muchos siguen confrontando– grandes desafíos para definir los niveles de desempeño en cada criterio seleccionado. Allí donde se ató alguna consecuencia a los resultados de esas evaluaciones, o donde se han realizado esfuerzos importantes por devolver resultados a las escuelas, los contenidos de las pruebas han empezado a cumplir la función de *aclaradoras* del currículo, si bien no todos los curriculistas aprobarían el que ello sea así. La amplitud de las consultas a docentes, expertos y productores de textos que pueda realizar la agencia evaluadora también ha contribuido en algunos casos a generar un proceso de redefinición de metas curriculares nacionales.

Mayor ha sido la influencia de las pruebas internacionales. Los resultados de los países que han participado en TIMSS, PISA y otras evaluaciones internacionales no siempre fueron congruentes

con la distribución de estudiantes entre los distintos niveles de rendimiento fijados en sus propias pruebas nacionales o con los índices de aprobación<sup>2</sup>. Esto los impulsó a que dieran una nueva mirada a sus currículos buscando posibles explicaciones y, en algunos casos, a profundizar sus esfuerzos de aclarar metas de aprendizaje estableciendo estándares.

Los primeros esfuerzos fueron desplegados por los países centroamericanos en 1999, bajo la conducción de la Secretaría General del organismo de Coordinación Educativa y Cultural de la región. Desde entonces, Honduras y Guatemala han avanzado sendas propuestas de estándares para todos los grados de la educación básica, para lo cual han tomado en consideración los resultados logrados en evaluaciones nacionales anteriores, y con los cuales empiezan ya a alinear un conjunto de instrumentos de evaluación de aprendizajes. En Nicaragua, el conjunto de estándares de la SGCECC fue utilizado para revisar el currículo y sirvió de insumo para la construcción de su primera evaluación nacional, aplicada en el 2003.

Colombia desarrolló y estableció estándares en el 2003, tomando como uno de sus referentes la información disponible sobre las capacidades actuales de los estudiantes según lo evaluado por el Instituto Colombiano de Fomento de la Educación Superior (ICFES). Los estándares se centran en competencias para resolver problemas de tipo real en lenguaje, matemáticas, ciencias naturales y sociales, y más recientemente en ciudadanía. Se espera que sirvan como guías para el desarrollo de programas de mejora y que se conviertan en el conjunto básico de expectativas de aprendizaje que el ICFES utilizará para las evaluaciones en corto plazo.

En Chile, el análisis del SIMCE y de los resultados de las pruebas internacionales ha mostrado poco incremento significativo en los puntajes promedio, a pesar de los grandes aumentos en el gasto público en educación, de las reformas curriculares y de la inversión en varios proyectos a gran escala. Se introdujeron sucesivamente varios nuevos instrumentos y aclaraciones curriculares, antes de optar por la construcción de estándares de contenido y desempeño, conocidos ya como Mapas de Progreso y Niveles de Logro, respectivamente. Los Mapas y Niveles se nutrieron, entre otros elementos, de los resultados obtenidos por los alumnos chilenos en sus pruebas nacionales e internacionales. Se convertirán gradualmente en referentes básicos para la elaboración de los instrumentos de evaluación y se usan ya en el reporte de resultados del SIMCE. Se pretende que sirvan también para guiar la reflexión, la planificación y la práctica pedagógica de los docentes, en subáreas específicas de matemáticas, lenguaje, ciencias sociales, ciencias naturales e inglés como segunda lengua y, más adelante, en educación física, educación preescolar, educación técnica y arte.

#### **Promover la responsabilización, rendición de cuentas y libre elección de escuelas**

En general, las pruebas estudiantiles con altas consecuencias (aprobación, graduación de alumnos, ascenso o salida de docentes, cambio de autoridades o cierre de escuelas) no son comunes en América Latina, a excepción de pruebas de tránsito a niveles superiores del sistema educativo como en Costa Rica, República Dominicana y, más recientemente, en El Salvador. Sin embargo, también tienen consecuencias las pruebas nacionales cuyos resultados sirven para seleccionar a escuelas para recibir premios o sanciones, ya sean simbólicos o efectivos, tales como la publi-

---

<sup>2</sup> Enrique Roca, en su capítulo en este libro, revisa con detalle estas evaluaciones.

cación de *rankings*, premios a la excelencia de instituciones, incentivos para equipos docentes o ascensos individuales en la carrera magisterial.

La práctica más simple y extendida a este respecto es *la identificación de las escuelas más destacadas en rankings nacionales* que se hacen públicos ya sea intencionalmente o por iniciativa de los medios, como en Chile, El Salvador o Guatemala. En Colombia se otorga el Galardón a la Excelencia a las escuelas cuyos resultados en las pruebas presentan mayores mejoras, mientras que las escuelas públicas más destacadas son invitadas a participar de una Red de Acción para la Excelencia a través de la cual comparten información sobre sus prácticas exitosas con otras escuelas.

Se argumenta muchas veces que esta práctica suele convertirse en un incentivo a los *ganadores* para lograr progresos adicionales antes que en un estímulo a los que se encuentran en los niveles más bajos de logro. Los educadores en América Latina, como los de muchos otros países del mundo, suelen estar fundamentalmente opuestos a la publicación de *rankings* y consideran que los resultados de las pruebas solo deberían compartirse al interior de la comunidad escolar. Muchos expertos en evaluación lamentan además la manera poco refinada en que se construyen la mayoría de esos *rankings*, que proyectan como diferencias reales en calidad diferencias entre puntajes que pueden atribuirse a errores de medición.

Otro tipo de uso que se da a los resultados de las evaluaciones son *los incentivos al desempeño docente y escolar*. Dadas las dificultades para identificar exactamente qué insumos educativos "garantizan" mejores resultados de rendimiento en distintas condiciones, los diseñadores de políticas analizan varias combinaciones de incentivos y mecanismos de responsabilización con los cuales movilizar actores e instituciones clave en su propia búsqueda de las "mejores" direcciones a seguir. La investigación sobre el tema, si bien muestra que las compensaciones económicas no son el más eficaz incentivo al desempeño docente, sugiere también que estas pueden servir para tal propósito cuando se las incluye dentro de una estrategia mayor para asegurar que los profesores concentren sus esfuerzos en el rendimiento estudiantil y, particularmente, cuando requieren de un esfuerzo colectivo para lograr los premios (Odden y Kelly, 1997).

Aun así, evaluaciones de programas tales como el Sistema Nacional de Evaluación del Desempeño de Docentes (SNED) en escuelas con financiamiento público de Chile, o Carrera Magisterial en México, que han utilizado con diferente peso pero de manera directa los resultados de las evaluaciones estudiantiles como criterio de adjudicación de un incentivo, no parecen haber tenido los efectos buscados (Mizala y Romaguera, 2004, y Santibáñez *et al.*, 2007) y sí, en cambio, algunas consecuencias indeseables aunque no intencionales. Las limitaciones de estos y otros esquemas de incentivos similares basados en resultados (como indicadores de desempeño) se discuten en el capítulo de Pedro Ravela en este volumen.

También se han usado los resultados de las evaluaciones para facilitar que *los padres puedan elegir libremente a qué escuela enviar a sus hijos*, lo que de forma indirecta se supone que sirve para presionar a las escuelas a mejorar la calidad de sus servicios y sus resultados. Esto es así particularmente en aquellos lugares donde el financiamiento público se asigna en función del número de niños atendidos, como en el caso de los sistemas de subvenciones escolares en Chile. Los directores de cada escuela reciben un informe de su desempeño en cada una de las áreas curriculares que han sido evaluadas, comparándolo con el de escuelas similares y promedios nacionales, así como información sobre las tendencias observadas en el tiempo a este respecto. Además se pone a disposición del público general los resultados obtenidos en las pruebas por cada una de

las escuelas de cada región del país, junto con los promedios para cada área, información sobre los cambios observados desde las rondas anteriores y comparaciones con escuelas con características socioeconómicas similares. Se elaboran informes para la prensa y para los padres, buscando revertir la tendencia de la prensa a construir clasificaciones simplistas sobre cuáles son las *mejores o peores* escuelas o a publicar titulares escandalosos sobre la ausencia de mejoras en el rendimiento.

Las investigaciones acerca de si el acceso a la información sobre el rendimiento sirve realmente para ejercer una presión efectiva sobre las escuelas para mejorar su oferta no son concluyentes (Elacqua y Fábrega, 2004). No se sabe realmente si el esquema competitivo ha servido o no para estimular mejoras en el sistema ni en las escuelas municipales subvencionadas que atienden a los pobres, mientras que parecería que las escuelas de administración privada han tendido a responder seleccionando estudiantes con mayores probabilidades de éxito antes que con mejoras en la calidad. También se achaca a la libre elección de escuela el incremento en la brecha entre los pobres y los más ricos, la desmoralización de los “perdedores”, la “enseñanza para las pruebas” y la reducción del currículo, así como la generación de algunas prácticas corruptas.

**Mejorar la gestión pedagógica e institucional en las escuelas o áreas escolares**

#### **Uso de los resultados para diseñar planes de mejoramiento escolar**

Es parte de un discurso común asumir que las evaluaciones a gran escala tendrán un impacto en los logros de aprendizaje solamente si sirven para generar cambios positivos en la escuela y el aula, actuando como insumo para el diseño, implementación y evaluación de estrategias y planes de mejoramiento escolar. La descentralización educativa o la autonomía escolar que se viene impulsando en varios países de la región han traído consigo una serie de normas e incentivos para el diseño de planes de mejoramiento educativo a nivel subnacional, local e institucional. Estos proyectos de desarrollo institucional, planes de mejora, proyectos de innovación, etc., deben basarse, al menos parcialmente, en una lectura de los resultados de las evaluaciones.

Donde las pruebas son censales y los resultados son devueltos a cada escuela –con guías o capacitación sobre cómo interpretarlos, además de con asistencia técnica sistemática en el proceso–, la posibilidad de diseñar planes consistentes, centrados en mejoras pedagógicas y en el mejoramiento de los logros se incrementa considerablemente. También ocurre esto cuando se asocia un incentivo económico a la elaboración de buenos proyectos, lo que permitiría su implementación. Está pendiente aún un análisis sistemático de la consistencia de diagnósticos y planes de acción y de la viabilidad de estos planes, tales como se vienen elaborando en Colombia, Paraguay, Perú, Chile o Uruguay, entre otros. Está también pendiente una evaluación de su impacto sobre el mejoramiento de los aprendizajes.

#### **Uso de los resultados para proveer oportunidades de desarrollo profesional para profesores y supervisores**

Según los especialistas educativos, el uso de los resultados de las pruebas para la formación y capacitación docente es tan importante como su uso para el planeamiento escolar.

En ese sentido, una de las primeras y más sencillas formas de utilización de los resultados de las pruebas en América Latina ha sido la publicación y distribución más o menos extensa de folletos de “análisis de ítems”. La publicación de los resultados generales de las pruebas ha estado acompañada por la publicación de ítems seleccionados de las pruebas para cada una de las



áreas curriculares, usualmente reflejando distintos niveles de dificultad. Se suele presentar el ítem completo con las múltiples opciones de respuesta, seguidas normalmente por una descripción de las habilidades curriculares específicas que estaban siendo evaluadas con esa pregunta y una explicación de por qué se les consideraba lo suficientemente importantes como para ser evaluadas en la prueba. Las publicaciones suelen analizar también las razones probables por las que los estudiantes podrían elegir la respuesta correcta o alguna de las respuestas incorrectas y, a veces, ofrecen algunas sugerencias –tanto conceptuales como didácticas– sobre cómo los profesores podrían ayudar a los estudiantes a desarrollar la habilidad prevista.

En países donde las pruebas solo se aplicaron a muestras de estudiantes de algunos grados, estas publicaciones con frecuencia sirvieron como una primera oportunidad para que los profesores y supervisores se informaran sobre los propósitos y métodos de las evaluaciones nacionales. Por esto tales informes “pedagógicos” generalmente incluyen algunas páginas introduciendo a los educadores al significado y uso potencial de las evaluaciones para su propio desarrollo profesional y al análisis de su posible desempeño en este tipo de pruebas.

Asimismo, y dado que los informes generales rara vez logran llegar hasta las escuelas, las publicaciones incluyen también un resumen de los resultados nacionales o subnacionales. En la mayoría de países se esperaba que no solo todas las escuelas recibieran copias de estos materiales, sino que estas y los distritos escolares tomarían la iniciativa para analizarlos. Sin embargo, en muchos casos la distribución del material estuvo seriamente limitada por falta de presupuesto y, en la mayoría de países, no se ha hecho ningún esfuerzo para supervisar de qué modo se está utilizando estas publicaciones en la práctica, quiénes las están usando, ni mucho menos con qué resultados.

En América Latina, una vez que las evaluaciones comenzaron a estar –parcial o totalmente– guiadas por marcos referidos a criterios, los informes dirigidos a los docentes se han centrado en descripciones de los distintos niveles de desempeño en cada una de las áreas o criterios de aprendizaje, cada una con ejemplos de uno o varios ítems o tareas que los estudiantes que alcanzan ese nivel de desempeño podrían responder o realizar correctamente.

Esto ha ocurrido en Argentina, Brasil, Perú y Chile, entre otros. Sin embargo, seguimientos realizados sobre el uso de estos instrumentos sugieren que solo las mejores escuelas reportan un mayor uso, mientras que aquellas que más mejoras necesitan no están en capacidad o no están dispuestas a utilizar los resultados de las pruebas para este fin. Actualmente se está impulsando a las escuelas a organizar eventos especiales con el personal con el objetivo específico de discutir los informes y existe una convicción de que para ello es indispensable proveer asistencia técnica a las escuelas.

A pesar de haber realizado consultas con varias fuentes, no se ha encontrado ninguna información sistemática sobre maneras específicas en las cuales los resultados de las pruebas se estén utilizando para diseñar capacitaciones o programas de desarrollo profesional. De hecho es muy posible que esté habiendo mucha actividad en este sentido, dado que los funcionarios ministeriales suelen contar la capacitación de profesores entre los usos dados a los resultados de las pruebas.

Sin embargo, hay razones para creer que el contenido de muchas de tales actividades bien puede consistir solo en un análisis superficial de los resultados de las pruebas nacionales y en una introducción general para los profesores sobre el sistema de evaluación. La identificación y descripción de formas más promisorias en que los resultados de las pruebas puedan ser utilizados como

material para la capacitación docente es aún una tarea pendiente. En su capítulo en este libro, Pedro Ravela analiza en detalle la evaluación y el desarrollo profesional de los docentes.

## CONCLUSIONES

Una primera conclusión de esta rápida revisión es que los resultados de las pruebas parecen estar siendo utilizados más de lo que usualmente se sospecha. Dada la existencia relativamente corta de los sistemas de evaluación de la región y la inestabilidad del compromiso político con su desarrollo, no sorprende que distintos países utilicen los resultados en menor medida de lo que se esperaría. Sin embargo, si se toma en cuenta la grave escasez de recursos para el mejoramiento educativo y la existencia de muchos programas valiosos que valdría la pena financiar, parecería necesario redoblar los esfuerzos para dar un uso adecuado a los resultados de las evaluaciones.

Aunque la mayoría de países latinoamericanos no iniciaron el desarrollo de sus sistemas de evaluación con ideas claras sobre los usos que se podría dar a los resultados, una vez que empezaron a recoger información fueron encontrando diversas maneras de usarla. Así, por ejemplo, la información ha sido utilizada para: el diseño e implementación de políticas; la asignación de recursos; estimular cambios pedagógicos y promover la responsabilización entre varios actores e instituciones educativas. Ha habido además una considerable evolución en la forma de presentar los resultados a diversas audiencias, y cada vez se está dando mayor atención a la necesidad de incluir información sobre factores de contexto que deben ser tomados en cuenta en el momento de interpretar los resultados.

Dado que en muchos lugares la prensa se ha convertido en el principal comunicador de los resultados a profesores, padres y tomadores de decisiones, es claro que no se han realizado esfuerzos suficientes para educar a los medios de comunicación de modo que estos se conviertan en aliados de los esfuerzos para movilizar al público en apoyo de la educación, en vez de actuar, conscientemente o no, como enemigos del sistema público. Esto requerirá que la mayoría de agencias de evaluación asigne una cantidad considerablemente mayor de recursos que los que se han estado invirtiendo en esfuerzos de difusión (tiempo del personal para definir claramente los principales mensajes que deben ser transmitidos al público en general a través de la prensa; tiempo para responder a interpretaciones incorrectas o no deseables de los resultados de las pruebas; así como más dinero).

Aunque ahora está empezando, el efecto de los resultados de las pruebas en la renovación curricular y el establecimiento de estándares también parece prometedor –cabe notar que muchos países pusieron en marcha sus sistemas de evaluación al mismo tiempo que daban inicio a reformas curriculares cuya llegada a las aulas aún tenía que ser asegurada–. Sin embargo, el diseño de las pruebas aún necesita mejorarse para permitir investigaciones adicionales de las implicancias pedagógicas de sus resultados. La posibilidad de utilizar los resultados para mejorar la planificación en el ámbito de las escuelas también está emergiendo como una línea de acción interesante y requerirá de aplicaciones de tipo censal o de la provisión de instrumentos para la autoevaluación y el análisis colectivo de los resultados.

Otra manera de que las evaluaciones pueda contribuir a mejorar el desempeño docente y los logros de aprendizaje es la producción de materiales para el desarrollo profesional que estén basados en análisis detallados de los resultados de las pruebas y la generación de espacios institucionales y eventos de capacitación en los cuales estos puedan ser apropiados y usados a cabalidad.

El análisis de ítems seleccionados de las pruebas que es difundido por las agencias de evaluación ha facilitado con frecuencia entre los profesores la reflexión sobre sus propias expectativas para el aprendizaje de sus estudiantes, así como la reconsideración por parte de los especialistas de cuán adecuados son los objetivos de aprendizaje que han definido para varios grados. El acceso directo a los instrumentos de evaluación también parece una poderosa manera de aliviar los temores y prejuicios existentes sobre la gama supuestamente limitada de habilidades que las pruebas pueden evaluar. Hoy día, incluso los ítems de opción múltiple pueden explorar procesos de razonamiento complejos y pueden servir para identificar fuentes de error ligadas a las prácticas de aula, si bien no debe aspirarse a sustituir el rol de las evaluaciones formativas en aula, otra área crítica en la gestión del aprendizaje.

Sin embargo, si se quiere asegurar el éxito de las estrategias de difusión y uso de los resultados, se tendrá que prestar mayor atención a las culturas organizacionales, tanto a nivel de las escuelas como en los niveles intermedios de la administración educativa. Las directrices tradicionales de tipo vertical (*top-down*) con frecuencia han fallado en tener los resultados anticipados cuando los intereses burocráticos y profesionales aparecen amenazados por la nueva información disponible, o cuando simplemente no existen las capacidades para ser movilizadas.