

La evaluación que no evalúa

Rafael E. de Hoyos

El mandato del sistema educativo mexicano, como el de cualquier otro país, es garantizar el cumplimiento del derecho que cada niño y joven en edad escolar posee: tener acceso a una educación de calidad. La evidencia demuestra que el estricto cumplimiento de este mandato sería suficiente para alcanzar tasas de crecimiento económico más elevadas, reducir la desigualdad del ingreso, contribuir al abatimiento de la pobreza y promover la movilidad social.¹ Por otro lado, México invierte 5.3% de su Producto Interno Bruto (PIB) o más de 20% del presupuesto total del gobierno federal en la provisión de educación pública. Ésta no es una inversión menor y el contribuyente tiene derecho a saber si el sistema educativo está cumpliendo con su parte del contrato social

proveyendo una educación de calidad a todos los niños y jóvenes en México. Dado que el acceso a una educación de calidad es un derecho social, aunado a la importancia de la educación para el desarrollo social y económico del país y a la magnitud de la inversión pública en este rubro, resulta indispensable que la sociedad cuente con herramientas confiables para medir y monitorear el cumplimiento del mandato del sistema educativo.



Pero, ¿cómo medimos el cumplimiento del mandato del sistema educativo? La provisión de una educación de calidad a todos los niños y jóvenes puede ser monitoreado por tres indicadores: (1) la cobertura, es decir, la proporción de niños y jóvenes en edad escolar que están dentro del sistema

educativo; (2) la calidad media del sistema; (3) la distribución o disparidad en la calidad de los servicios recibidos entre estudiantes. Medir el acceso es relativamente fácil, se está o no se está dentro del sistema. Desde 1990 las estadísticas oficiales de la Secretaría de Educación Pública reportan la cobertura bruta del sistema y otros indicadores relacionados con la eficiencia del sistema.² Sin embargo, la definición y, por lo tanto, la medición de la calidad de los servicios educativos ha sido tema de largos debates. Una forma —simplista si usted quiere— de aproximarnos a la medición de la calidad de los servicios educativos es mediante una prueba que mida el nivel de logro en determinadas áreas del conocimiento. Esto no es nada nuevo, los docentes de todo el mundo han aplicado exámenes a sus alumnos desde hace muchas décadas. Pero así como un docente dentro de un salón de clases debe aplicar criterios de evaluación homogéneos a todos sus alumnos, lo mismo sucede cuando se trata de evaluar estudiantes en salones de clases, escuelas, municipios y estados distintos. En este caso, la dificultad, contenido, proceso de aplicación y criterios de calificación aplicados a todos los alumnos deben ser equiparados. Las pruebas estandarizadas homologan estos criterios para medir a todos

los alumnos “con la misma vara”. Cuando están bien diseñadas y cuentan con representatividad estadística o bien son censales, las pruebas estandarizadas proveen información confiable sobre los niveles de logro o conocimientos en determinadas áreas y son, por lo tanto, una buena aproximación de la calidad media del sistema y su distribución.

Ya que el punto de partida es el derecho a recibir una educación de calidad, contar con una herramienta confiable y transparente para medir la calidad es una condición necesaria para que el Estado haga valer este derecho. La medición del cumplimiento del mandato del sistema educativo es un argumento suficiente para justificar la necesidad de contar con una prueba estandarizada. Pero este argumento no es el único. Hay evidencia que demuestra que las pruebas estandarizadas *pueden* ser herramientas poderosas para elevar la calidad del sistema.

Hay, por lo menos, dos vínculos entre las pruebas estandarizadas y la calidad del sistema educativo. La mediación de las pruebas estandarizadas entre la obligación del Estado a proveer una educación de calidad y el derecho de los padres a demandarla, hace que su implementación y difusión sea suficiente para generar una dinámica positiva entre la rendición de cuentas, la participación ciudadana y los resultados. Padres más informados sobre la calidad de los servicios educativos participan activamente en el proceso de enseñanza de sus hijos y exigen mejores resultados de la autoridad educativa.³ Las pruebas estandarizadas también sirven para alinear los incentivos de todos los actores involucrados en el proceso de enseñanza, como la autoridad federal, la estatal, los supervisores, directores, docentes, estudiantes y padres de familia en torno a lo que tiene importancia, es decir, que los niños y jóvenes aprendan. Esta alineación de incentivos posibilita a los actores involucrados en el proceso de aprendizaje a tener un punto de partida común, diseñar estrategias de mejora para abordar los retos identificados y fijarse metas.⁴ Sin embargo, la mera aplicación de una prueba estandarizada no garantiza que se capitalicen todos los beneficios asociados a ella.

Durante los últimos años México ha dado pasos importantes en la aplicación de pruebas estandarizadas para medir la calidad de los servicios educativos. Gracias a pruebas internacionales como el Programa Internacional de Evaluación de Estudiantes (PISA, por sus siglas en inglés) diseñada por la OCDE o las pruebas PERSE y SERCE de la UNESCO tenemos una radiografía de la calidad de los servicios ofrecidos por el sistema educativo mexicano y su comparación con otros países del mundo.

Quizá la experiencia más relevante en la aplicación de pruebas estandarizadas en nuestro país se dio a partir de 2006 y con mayor relevancia desde 2007 con la aplicación de la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE). A partir de ese año todos los niños cursando desde el tercero hasta el sexto de primaria y de primero a tercero de secundaria participan en la prueba ENLACE, la cual mide los niveles de logro en matemáticas, español y una materia rotativa; a partir de 2008 la prueba ENLACE también evalúa a todos los jóvenes del último año de educación media superior del país, nivel que

recientemente adquirió obligatoriedad. ENLACE, como las pruebas internacionales, provee una radiografía de la calidad de los servicios educativos proveídos por el sistema, pero por su carácter censal ENLACE es capaz de generar un diagnóstico sobre los niveles de logro en cada estado, municipio, localidad, escuela, grado y estudiante en el sistema. Con la información producida por ENLACE por primera vez en nuestro país todos los agentes del sistema educativo (funcionarios del gobierno federal y estatal, supervisores, directores, docentes, estudiantes y padres de familia) partían de un diagnóstico común basado en una herramienta que mide lo que es importante, es decir, lo que aprenden nuestros jóvenes en las aulas (ver figura 1).

FIGURA 1

Ejemplo de cartel informativo con los resultados ENLACE para una escuela



Fuente: Imagen tomada de la página de internet de la prueba: <http://www.enlace.sep.gob.mx/>

El carácter censal de la prueba ENLACE abrió la posibilidad de fortalecer —en todas las escuelas del país— la dinámica entre rendición de cuentas, participación ciudadana y resultados. Cada escuela recibía un cartelón con los resultados en un formato similar al que se presenta en la figura 1. El director, docentes, padres y estudiantes de todas las escuelas en México podían ver cómo ha evolucionado su nivel de logro promedio en los últimos años en las áreas del conocimiento evaluadas. También era posible comparar el nivel de logro obtenido por la escuela con el promedio estatal y nacional para escuelas del mismo sostenimiento (público o privado) y grado de marginación. ENLACE permitía que los resultados a nivel de la escuela fueran desagregados por área del conocimiento, grado y grupo, reactivo por reactivo, para identificar los contenidos en donde los estudiantes presentaban mayores retos. La figura 2 presenta un ejemplo del diagnóstico detallado al que tenía acceso cada director y docente en México, permitiéndoles focalizar sus esfuerzos en donde se presentan las mayores falencias. Esta información también podía ser utilizada por el director y el cuerpo

docente para diseñar estrategias de mejora que involucre la capacitación de docentes en áreas del conocimiento deficitarias (ver figura 2).

FIGURA 2

Ejemplo de resultados a nivel de la escuela desagregado por grado y reactivo

Tema												
Análisis y representación de datos	080	106	115									
Figura y cuerpos	021	022	046	047	048	075	101	110				
Medidad	014	015	025	026	050	051	077	078	103	104	111	112
Números y sistemas de numeración	009	010	018	019	043	044	072	073	098	099	107	
Problemas aditivos	111	020	045	108								
Problemas multiplicativos	012	074	100	109								
Proporcionalidad y funciones	016	017	079	105	113	114						
Ubicación especial	013	024	049	076	102							

- Preguntas que contestaron incorrectamente menos del 40% de los alumnos de este grupo.
- Preguntas que contestaron incorrectamente entre el 40% y el 60% de los alumnos de este grupo.
- Preguntas que contestaron incorrectamente más del 60% de los alumnos de este grupo.

Fuente: Imagen tomada de la página de internet de la prueba: <http://www.enlace.sep.gob.mx/>

Como cualquier otra herramienta de medición, las pruebas estandarizadas tienen limitaciones. Los críticos de las pruebas estandarizadas, en particular los de la prueba ENLACE en México, aluden a los siguientes argumentos:

- a. Se crean incentivos para que los actores escolares focalicen la atención sólo en los contenidos incluidos en la prueba, descuidando otros contenidos igual o más importantes.
- b. Las pruebas estandarizadas capturan sólo parcialmente los procesos de generación de conocimientos, los cuales son muy complejos.
- c. Las comparaciones de resultados de logro académico entre estados, escuelas, comunidades, alumnos, etcétera, son injustas ya que no toman en cuenta las diferencias culturales, regionales, de contexto social, económico, entre otros.

Todos estos son argumentos válidos, pero ninguno de ellos invalida la necesidad de contar con una prueba estandarizada, más bien sirven para pensar en el correcto diseño de la prueba, una que sea capaz de producir resultados confiables. En relación a los incisos (a) y (b) arriba expuestos, nadie pudiera argumentar que la prueba estandarizada deba y pueda medir todas las áreas del conocimiento y mucho menos todos los procesos por los cuales se genera conocimiento. Las pruebas miden los niveles de logro en determinadas áreas del conocimiento, la preocupación de los

incentivos a enseñar sólo para salir bien en la prueba y la parcialidad de conocimientos medidos es algo que se puede abordar con el diseño de la prueba. Si sabemos que el sistema se focalizará en los contenidos de la prueba y éstos son, por definición, limitados, entonces la prueba tendría que medir lo que los niños y jóvenes deben saber a través de una muestra representativa de todos los contenidos relevantes por área del conocimiento y grado. Los reactivos de la prueba deben cambiar año con año pero conservando su representatividad estadística del universo de conocimientos. Si el diseño de la prueba tiene esas características, entonces el que los agentes del sistema se focalicen en salir bien en la prueba no debería ser motivo de preocupación. Es decir, muchos de los argumentos en contra de las pruebas estandarizadas son de corte metodológico o estadístico.

La comparación de resultados entre escuelas o alumnos de distintos contextos (inciso (c) arriba) es algo que se puede abordar a través de la contextualización de los resultados. En México, como se presenta en la figura 1, los comparativos de resultados de la prueba ENLACE se contextualizaban por tipo de escuela (pública o privada) y por nivel de marginación — definido por el “Consejo Nacional de Población” —.⁵ Pero la

comparación más relevante no es entre distintas escuelas, sino de una escuela contra sí misma, es decir, la evolución de los niveles de logro en una misma escuela. Por definición, esta comparación elimina todos aquellos elementos que son fijos en el tiempo (o que cambian muy lentamente) como el contexto social y económico, el nivel cultural, entre otros. Sin embargo, contextualizar los resultados es muy distinto a contextualizar los contenidos de la prueba. La exigencia por una educación de calidad para todos, y la medición de su observancia, debe ser homogénea a través de toda la diversidad cultural, social y económica del país.

Quizá la crítica más fuerte a la prueba estandarizada en México (ENLACE) se dio en torno a la falta de controles durante su implementación. En su origen ENLACE fue diseñada con pocos controles para garantizar resultados confiables ante la posible copia. El argumento central era que, al no tener repercusiones para ninguno de los actores educativos, ENLACE no generaba incentivos a la copia. Esto fue válido durante los primeros años de la implementación de la prueba cuando la sociedad no conocía ENLACE y, por lo tanto, no exigía mejores resultados a la autoridad educativa (director y cuerpo docente). A medida que la sociedad fue

familiarizándose con la prueba y la dinámica entre la rendición de cuentas, la participación ciudadana y los resultados se fue fortaleciendo, se crearon incentivos para la copia. Dado que la prueba no se adaptó a la creación de incentivos para la copia (producto de una dinámica positiva), en cierta medida el éxito de ENLACE fue un precursor de su fracaso. El otro elemento importante que generó incentivos para la copia y que acabó en una serie de anomalías relacionadas a la implementación de la prueba, fue la inclusión de sus resultados en el algoritmo utilizado para premiar salarialmente a docentes dentro del programa Carrera Magisterial. Esto fue un uso para lo cual claramente ENLACE no estaba diseñada, generando grandes incentivos a la copia, poniendo así en duda su confiabilidad.

A pesar de todos los beneficios asociados con las pruebas estandarizadas y de la necesidad de contar con éstas para medir el cumplimiento del mandato del sistema educativo, las pruebas son sólo una herramienta que *puede* ser utilizada para mejorar la calidad de los servicios. Aun la mejor herramienta no garantiza su buen uso. De poco sirve tener

una prueba que mide el nivel de logro de todos los alumnos todos los años si los actores relevantes (autoridad educativa, supervisores, directores y docentes) no utilizan estos resultados para diseñar políticas que aborden los problemas identificados y mejoren la calidad de los servicios. Entre 2007 y 2013 la Secretaría de Educación Pública produjo más resultados de los que el sistema podía digerir. Nuestros supervisores, directores y docentes no estaban capacitados, no tenían las habilidades, para sacar todo el provecho de la información que generó ENLACE año tras año.

La evidencia internacional y nacional demuestra que pautas de mejora sencillas basadas en un diagnóstico específico para cada escuela son suficientes para mejorar los niveles de logro, aun en escuelas con los peores resultados iniciales.⁶ Los efectos positivos asociados al uso correcto de los resultados de las pruebas estandarizadas, mejorando los niveles de logro representan un fuerte argumento para abogar por una prueba censal. Pero los resultados en cada escuela deben ir acompañados de las capacidades de gestión de los directores y supervisores para explotarlos para el diseño de estrategias de mejora. La otra gran asignatura pendiente relacionada con la falta de uso de las pruebas estandarizadas en México es su

análisis por parte de la comunidad académica para estudios e investigaciones que sirvan de insumos para el diseño de políticas educativas basadas en evidencia empírica. Desde el punto de vista de la investigación empírica, la base de datos generada por ENLACE es una mina de oro, permitiendo el seguimiento de las trayectorias escolares de todos los niños y jóvenes en México. Esta información es invaluable para identificar el papel que juegan los distintos insumos escolares, el contexto socioeconómico, el esfuerzo personal, entre otros, en la evolución de los niveles de logro.

México ha dado pasos importantes en el uso de pruebas estandarizadas como mecanismo para monitorear el cumplimiento del mandato del sistema educativo. El país también ha avanzado en el uso de pruebas para satisfacer la obligatoriedad del Estado a rendir cuentas a la sociedad. Estos avances me parece que hacen que hoy el debate no gire en torno a la existencia o no de una prueba estandarizada, sino más bien en las características que debe tener dicha prueba, los protocolos de implementación y, sobre todo, los usos que

le queremos dar a los resultados para mejorar la calidad de los servicios. Hay varias lecciones claras de que nos deja la experiencia de ENLACE:

1. Uno de los grandes beneficios en términos de rendición de cuentas y uso de resultados para la mejora están asociados con el carácter censal de la prueba.

2. Si la prueba estandarizada tiene éxito en generar una dinámica positiva entre la rendición de cuentas y la participación ciudadana, entonces generará incentivos a la copia y la prueba tiene que incorporar controles para prever esta posibilidad.

3. Por los incentivos que genera para la copia y otros argumentos asociados a la medición y atribución de causa y efecto, no es una buena idea utilizar los resultados de una prueba estandarizada aplicada a los alumnos para evaluar docentes.

4. Sin importar la calidad de la herramienta de medición (prueba), su uso correcto para la mejora de los servicios está sujeto al nivel de habilidades de los usuarios. Si se quiere

explotar los beneficios asociados al buen uso de la prueba es necesario acompañar su implementación con una estrategia de capacitación a directores y supervisores.

5. Es importante que la autoridad competente haga pública toda la información relacionada con las pruebas estandarizadas (respetando, claramente, la confidencialidad de la información) para incentivar su uso en estudios e investigaciones que tanta falta le hacen al país.

La reforma constitucional en materia educativa recientemente aprobada le transfiere al Instituto Nacional de Evaluación de la Educación (INEE) el diseño e implementación del sistema de evaluación del país, incluyendo, entre otras cosas, la medición de logro de los alumnos. La tarea del INEE no es sencilla, afortunadamente el instituto no parte de cero ya que cuenta con un cúmulo de experiencias relevantes que dejó la aplicación de la prueba ENLACE; aún más relevante, el instituto cuenta con capital humano del más alto nivel, capaz de diseñar una prueba que mida, informe y detone el potencial del sistema educativo.

Rafael E. de Hoyos

Economista Sénior. Trabaja en el Departamento de Educación para América Latina del Banco Mundial.

Las opiniones expresadas en este artículo son de exclusiva responsabilidad del autor y no necesariamente representan las del Banco Mundial.

¹ Ferreira *et al.* (2013): *La movilidad económica y el crecimiento de la clase media en América Latina*, Banco Mundial, Washington DC.

² Ver: <http://planeacion.sep.gob.mx>

³ Gertler, Patrinos y Rubio-Codina (2012): “Empowering parents to improve education: Evidence from rural Mexico”, *Journal of Development Economics*, Elsevier.

⁴ Boudett, City y Murnane (eds.) (2005): *Data wise: a step by step guide to using assessment results to improve teaching and learning*, Harvard University Press, Cambridge, MA.

⁵ <http://www.conapo.gob.mx>

6 De Hoyos, García y Patrinos (2013): *Increasing learning outcomes in low performing schools in Mexico*, The World Bank, Washington DC.

2014 Octubre, Agenda.

Te recomendamos leer:

Así se rompió la izquierda

www.nexos.com.mx

Powered by [AddThis](#)

