



ASEGURAMIENTO DE LA CALIDAD DE LAS PRUEBAS DE ESTÁNDARES NACIONALES PARA LA EDUCACIÓN PRIMARIA, DE LA SECRETARÍA DE EDUCACIÓN PÚBLICA

Ponencia presentada en el V
CONGRESO DE LA
INVESTIGACIÓN
EDUCATIVA, MÉXICO

Francisco Martínez Rodríguez

Sylvia Schmelks

¿Cómo fue el proceso de evaluación de la calidad en México?

¿Qué aspectos se evaluaron?

¿Qué consecuencias originó la medición?

El objetivo de la evaluación con base en estándares nacionales es medir el logro académico de los estudiantes en función de algunos aspectos que el currículo actual enfatiza: el desarrollo de habilidades fundamentales en lectura y matemáticas, para todos los grados de primaria. No se trata de pruebas de conocimiento, si bien algunos se incluyen cuando son necesarios para desarrollar alguna habilidad. Se centran en habilidades básicas de lectura y matemáticas. Pretenden medir, por un lado, cuántos alumnos alcanzan los estándares de habilidades identificados en el currículo. Por otro intentan decir algo acerca de qué habilidades poseen los alumnos y qué les falta desarrollar para alcanzar dichos estándares.

Las pruebas se aplican con expresa finalidad de que sus resultados retroalimenten al sistema educativo nacional y a los sistemas estatales. Específicamente, se considera que su aplicación proporciona información de gran riqueza para retroalimentar las funciones normativa y compensatoria. Pero además puesto que permiten que permiten decir algo acerca de la habilidades que aún no logran los alumnos, también ofrecen información valiosa a los docentes. Esta prueba se visualiza como el inicio de un ejercicio recurrente, que se aplica cada año y que proporcione información longitudinal y comparativa. También se espera que la prueba se constituya como un ejercicio acumulativo, que permita ir ampliando la información sobre habilidades básicas en secundaria, y sobre otras habilidades no complementadas en esta primera ocasión en primaria.

Puesto que su objetivo es retroalimentar al sistema educativo, no se deben tomar decisiones sobre los individuos sobre la base de los resultados de esta prueba. Si pueden tomar decisiones que afecten a las escuelas, pero siempre que sean para apoyarlas o ayudarlas, nunca para dañarlas o castigarlas.

Se trata de pruebas con referencias a criterio. Ni se construyeron ni se analizaron con referencia a una norma. En este sentido, las pruebas miden la habilidad real del alumno, y no de distancia que cada uno de ellos guarda respecto de la media poblacional. La prueba de comprensión lectora parte del supuesto de que los niños pueden llegar a leer con distintos grados de éxito. Esto dependerá del del reconocimiento que hagan las características de los textos que circulan socialmente, del grado de desarrollo que hayan logrado en las habilidades que se requieren para leer y de la complejidad del texto al que

se enfrente. De acuerdo con estas variantes, se calificó el desempeño de los alumnos en la resolución de tareas de la lectura en tres niveles acumulativos.

En el caso de la prueba de habilidades matemáticas, se aplicó una prueba diferente para cada grado escolar, con excepción de primero, segundo y tercero de secundaria, grados en que la prueba aplicada fue la misma, correspondientes a los estándares de sexto grado de primaria. No obstante, puesto que se contaba con los datos de las prueba de lectura, las pruebas de cuarto y quinto grado fueron equiparadas, de la misma manera que fueron las del sexto grado y las aplicadas en secundaria.

La prueba mide seis de las siete¹ habilidades matemáticas comprendidas en el Plan y Programas de Educación y Primaria.

Estas habilidades son las siguientes:

- a) Comunicar e interpretar ideas, cantidades o números, mediante símbolos, signos, diagramas, tablas o gráficas.
- b) Imaginar: construir ideas evocando imágenes de objetos, sucesos, relaciones, atributos y procesos.
- c) Generalizar: analizar regularidades para encontrar patrones.
- d) Medir: establecer relaciones entre magnitudes para encontrar una medida.
- e) Operar: establecer relaciones entre números o expresiones para producir un resultado.
- f) Inferir: establecer relaciones causa-efecto, o entre los datos explícitos o implícitos que aparecen en un texto, ilustración, tabla, gráfica o diagrama para encontrar un resultado.

De la misma manera que en el caso de lectura, para cada una de estas habilidades se establecieron tres niveles acumulativos de logro. La prueba se aplicó por primera vez, en octubre de 1998, a una muestra de 377,652 alumnos de 4.047 primarias y 1.079 secundarias. ² La muestra representativa en el nivel nacional de escuelas primarias privadas, públicas urbanas, públicas rurales, indígenas y cursos comunitarios, y de escuelas secundarias generales, técnicas y telesecundarias. En el ámbito estatal, las pruebas son comprobables para el caso de las escuelas primarias públicas rurales e indígenas en 20 entidades federativas, y para el caso de las secundarias generales y técnicas.

Mecanismos de seguridad de la calidad

Se ha tenido cuidado en diseñar procedimientos rigurosos para cada una de las fases de elaboración de esta prueba. Aunque se conoce que aún hay aspectos a mejorar, como resultado de esta serie de procedimientos contamos ahora con un instrumento de medición de logro de habilidades fundamentales en lectura y matemáticas que es estable (no varía de acuerdo con la población que se aplica) y que nos permite medir el nivel de logro en estas habilidades, hasta que cambie su definición. Lo que se ha realizado hasta la fecha para asegurar la calidad del instrumento es el siguiente:

1. Las pruebas fueron elaboradas por equipos mixtos: en cada equipo, responsable de las pruebas de cada ciclo, participaron especialistas en la materia, elaboradores del currículo de educación básica, y especialistas en evaluación.

2. El proceso fue coordinado y supervisado por un grupo constituido por personas de diferentes instancias de la Secretaría de Educación Pública, interesadas por la evaluación y sobre todo por sus resultados. Así participaron personas del área de la Secretaría de Educación, de la Subsecretaría de Planeación y Coordinación Educativas, de la Subsecretaría de Educación Básica, y de la dirección general de Evaluación.

3. Para elaborar las pruebas se partió de un cuidadoso análisis del currículo de español y de matemáticas para la educación primaria. Dicho análisis permitió clasificar las habilidades requeridas para el desarrollo de la comprensión lectora y para la resolución de problemas matemáticos.

4. Las primeras versiones de las pruebas elaboradas mediante este procedimiento se pilotearon con diferentes poblaciones (indígenas, población rural dispersa, rural, urbana), en más de una ocasión (dos para el caso de matemáticas, tres para el caso de lectura). Los resultados de los piloteos fueron sometidos a dos tipos de análisis: el clásico (grado de dificultad, índice de discriminación), y el análisis Rash, basado en la teoría de respuestas al reactivo (modelo de un parámetro – dificultad). Este último genera dos tipos de valores: la media de la población. Su análisis permite independizar la prueba de la población a la que se aplica (un procedimiento más potente aun que la clásica estandarizado). Con este procedimiento se obtuvieron los valores propios de la dificultad de los reactivos, así como los valores propios de la habilidad de los estudiantes. Además con estos resultados se sentaron las bases para posibles equiparaciones que habrían de darse.

5. Además de los procedimientos anteriores, la prueba fue sometida a su lectura y análisis por parte de dos expertos en evaluación: uno de los estados Unidos y otro de España.

6. Después de la aplicación definitiva de las pruebas, éstas fueron nuevamente sometidas al análisis clásico y nuevamente depuradas. Estos procedimientos se realizaron con muestras de 20.000 estudiantes, lo que representan de un 40 a un 67% de la población sustentante según la prueba en cuestión. Los resultados se expresan en medidas Rash (logito). Con los instrumentos depurados, se volvió a realizar el procesamiento para llegar a los resultados definitivos.

7. Establecidas las medidas propias de dificultad y habilidades, se definieron los niveles de desempeño considerando la medida de dificultad más – menos una desviación estándar para el nivel dos. El nivel uno quedó con los valores inferiores a este intervalo, mientras que el nivel tres abarcó los valores superiores.

8. El procedimiento de lectura de los datos fue válido mediante procedimientos sistemáticos de verificación de la lectura. Cabe señalar sin embargo, que los actuales lectores de información de la Dirección General de evaluación son altamente confiables y, al registrar cualquier marca en la zona de respuesta, permiten que se pierda menos información. Los datos fueron sometidos a procesos rigurosos de limpieza.

No hemos hablado aquí del procedimiento muestral. Sin embargo, cabe decir que la muestra fue seleccionada por un especialista (Richard Wolfe) y validada por un despacho privado también especializado. Esta muestra está diseñada para tener un nivel de precisión de 95. No admite sustituciones, por lo cual está obredimensionada sobre todo en los tipos de escuelas en los que se presentan mayores pérdidas (Cursos Comunitarios). Tampoco hemos detallado los procedimientos para asegurar rigor en la aplicación de las pruebas. Basta decir que se siguieron los procedimientos ya probados que se han implantado por cada una de las áreas de evaluación de los estados en otros proyectos de evaluación realizados.

Puesto que su objetivo es retroalimentar al sistema educativo, no se deben tomar decisiones sobre los individuos sobre la base de los resultados.

Conclusión

Con toda seguridad, el procedimiento seguido para la construcción de estas pruebas es perfectible. Aunque hemos procurado desarrollar mecanismos de aseguramiento de la calidad para cada uno de los pasos involucrados en el desarrollo de este proyecto, seguramente éstos pueden mejorarse y afirmarse. No obstante la existencia de un grupo coordinador y supervisor como el que describimos en el punto 2, arriba nos asegura que tenemos entre nuestras manos tanto un buen instrumento como un buen proceso de aplicación y análisis de él. La experiencia en este sentido ha sido tan rica por la crítica continua que hemos obtenido del proceso, que nos atrevemos a sugerir considerar la instauración de grupos similares en procesos innovadores y/o ,masivos de evaluación de la calidad de la educación, como el que aquí hemos descrito.

Notas

1. No se midió la habilidad de estimar, por la imposibilidad de saber si el niño efectivamente estimaba las mediciones a media.

2. Puesto que la aplicación fue en octubre, se aplico la prueba correspondiente al grado inmediato anterior. Así, la prueba de sexto grado se aplicó al primer grado de secundaría. se aprovechó la visita de las secundarias para aplicar la misma prueba también en segundo y tercer grado.

F. Martínez Rodríguez pertenece a la Dirección General de Evaluación, Secretaría de Educación Pública. Sylvia Schmelkes pertenece al departamento de Investigaciones Educativas-CINVESTAV-1PN.

La experiencia ha sido tan rica que nos asegura que tenemos en nuestras manos un buen instrumento

ANEXO

La evaluación de la carrera magisterial

La evaluación de la carrera magisterial es un instrumento que define el puntaje que adquieren los docentes mexicanos y, además, entre otras cuestiones, su salario. Se realiza anualmente, al fin del curso escolar, y en el puntaje total se tienen en cuenta varios tópicos. Entre ellos el plan de la clase, la forma de seguimiento que el maestro hace de cada alumno, los estudios de posgrado (maestrías y doctorados), los cursos de actualización que tienen valores de carrera los resultados de las pruebas respondidas por los alumnos y los y los conocimientos pedagógicos del maestro. Estos dos puntos son los que adquieren más importancia en el puntaje total. Y un último tópico, tal vez el más conflictivo de todos, la evaluación que hacen los compañeros respecto a cada docente.

Por ejemplo, en la evaluación de los conocimientos pedagógicos del maestro, se le pregunta no solamente sobre pedagogía sino también sobre política educativa, modificaciones en la política educativa, consideraciones curriculares, materiales educativos a los que tiene acceso.

Es interesante destacar que, respecto de las definiciones salariales, la diferencia entre el salario base y el primer nivel de carrera es de 37% . Si un docente llega hasta el final de la carrera magisterial (aunque todavía nadie llegó porque en la instrumentación de este proyecto la carrera es relativamente nueva), el nivel E puede llegar a representar un 3000% más del salario base.

Actualmente alrededor del 70% de los maestros están incorporados a la carrera.

De acuerdo con las opciones de Sylvia Schmelkes, esta metodología de estímulo tiene aspectos sumamente valiosos, entre ellos, la posibilidad de que el maestro mejore sus salarios sin necesidad de dejar la docencia, "porque—según comentó—antes era absolutamente necesario dejar la docencia para poder mejorar su salario, porque la única posibilidad de ascenso era convertirse en director o supervisor". Pero también ha tenido una serie de retos contraproducentes, uno de los más importantes sería la ruptura del sentido de equipo en las escuelas, debido al punto que tiene relación con la evaluación de los compañeros y la corrección que el director hace de esta evaluación.

"Tengo una alumna que está haciendo su tesis sobre este tema, su objetivo de estudio fue el órgano de evaluación escolar y se encontró con que el órgano no funcionaba. Es decir, detecto los mecanismos de resistencia que tienen los maestros para evitar que se rompa la posibilidad de trabajar en equipo, la posibilidad de colaboración"

¿Qué otras evidencias encontró?

– Los maestros se resisten a evaluarse a sí mismos y a tener jerarquías en el interior de la escuela, a guardarle rencor al que lo evaluó mal y al que no lo dejó entrar en la carrera. La investigadora encontró que todos los maestros se evaluaban igual. El máximo puntaje era 10, entonces todos se ponían 10 y la carrera magisterial tenía que aplicar un factor de corrección que anulaba esa evaluación. Finalmente fueron tantos los anulados, que ya no servía.

¿Se han basado en otras instituciones para instrumentar este modelo?

– Es un modelo tomado de las universidades, en donde un grupo académico le da legitimidad a otra. Pero allí los pares se definen como pares en abstracto, no como tu compañero de trabajo con el que tienes que hacer las cosas. Allí se evalúa a otro académico que tal vez no está en la misma institución, en cambio aquí se aplicó a la escuela. A lo mejor, si se aplica como mecanismo de evaluación por pares, pero a partir de documentos y no respecto del grupo específico con el que se trabaja, funcione mejor.

¿Cuál es tu posición al respecto?

– En los inicios del proyecto de carrera magisterial yo opinaba que se debían premiar escuela, no maestros; entonces, si la escuela era buena y había un maestro que no funcionaba, igual iba a resultar un buen estímulo y la presión sobre él iba a ser más fuerte. Así se aseguraba que el trabajo en equipo funcionara y, por otro lado, se fortalecía el liderazgo del director, porque tendía la oportunidad de generar que todos los maestros y él mismo fueran estimulados. Eso no pasó.

(Extraído de "Entrevista a Sylvia Schmelkes (DIE/CINVESTAV), diversos métodos para el seguimiento de los aprendizajes", Revista Novedades Educativas N° 102.)