

Tendencias negativas que se observan en educación a partir de usos inapropiados de las evaluaciones a gran escala

Negative trends seen in education from inappropriate uses of large-scale assessments

Miranda Esquer, José Francisco

Universidad del Valle de México, Campus: Hermosillo

jfrancisco_mes@yahoo.com.mx

Recibido: 03 de abril de 2012

Aceptado: 17 de mayo de 2012

Por la errónea utilización de pruebas de rendimiento estandarizadas tradicionales para evaluar la calidad de las escuelas hay cosas realmente terribles que están ocurriendo en las escuelas de nuestros niños en estos días. Una es que aspectos importantes del currículo se están haciendo a un lado, porque no son medidos por las pruebas. Otra es que los niños están siendo entrenados sin descanso para que dominen el contenido de esas pruebas de alto impacto y, en consecuencia, están comenzando a odiar la escuela. Y una más es que, en muchos casos, los maestros se dedican a preparar a sus alumnos para las pruebas, lo que se parece mucho a hacer trampa, porque están inflando las puntuaciones de los alumnos sin elevar su competencia en los aspectos que se supone miden las pruebas

James Popham

67

Existe la creencia de que la educación y el desarrollo de un país se corresponden fuertemente, es decir, que entre mejor educación se genere, mayores niveles de desarrollo y mejores niveles de ingresos serán constituidos; aunque claro, esta correspondencia no es perfecta ni automática, sí se ha encontrado que se manifiesta en un grado significativo, tanto así que la mayoría de los países están interesados por conocer el estado que guardan sus sistemas educativos para poder arribar a estándares de calidad educativa que permitan ubicarse cada vez en mejores niveles de competitividad, el principal medio para conseguir tal propósito son las evaluaciones en gran escala, que se difundieron por todo el mundo a finales de los años noventa.

En nuestro país ha sido un suceso la aplicación de estos exámenes, debido a que las primeras experiencias que se tuvieron con el examen del Programa Internacional para la Evaluación de Estudiantes (PISA, por sus siglas en inglés) en 1997 resultó desalentadora al situarse México en el penúltimo lugar, sólo arriba de Brasil, lo mismo ocurrió durante la

siguiente aplicación, tres años después, durante el año 2000; en las posteriores se han obtenido resultados similares. De esta manera el sistema educativo mexicano aparece situado en los últimos lugares, dentro de los países que evalúa este programa patrocinado por la Organización para la Cooperación y el Desarrollo Económico (OCDE).

Estas evidencias han generado la necesidad de establecer un control sobre los resultados educativos, con el propósito de asegurar una mejor calidad en la educación que se presta en nuestro país. De esta manera son ampliamente conocidas dos evaluaciones a gran escala en educación básica, que son la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) y los Exámenes de la Calidad y el Logro Académico (EXCALE); el primero es un examen censal, es decir, se aplica a toda la población estudiantil desde el tercer grado de educación primaria, los tres grados de educación secundaria y en educación media superior el último grado; mientras que el examen EXCALE son instrumentos muestrales, es decir, se aplica a una muestra del total de los estudiantes y de grados específicos: tercero de preescolar, tercero de primaria, sexto de primaria, tercero de secundaria y tercero de bachillerato; es decir, se aplica a intervalos de cada tres grados.

68

Estos exámenes son elaborados a partir de un marco teórico en donde puede verse la investigación y el conocimiento que se ha generado alrededor de estas pruebas, de tal manera que han transformado el formato de opción múltiple hasta convertirlo en reactivos capaces de evaluar contenidos de un nivel cognoscitivo considerable, algunas de las características más sobresalientes pueden verse en la siguiente cita:

Por distintas razones, especialmente por su versatilidad y características psicométricas, los reactivos más comunes son los de opción múltiple. Pueden ser utilizados en distintos contextos y con diferentes propósitos; sirven para evaluar conocimiento factual (puramente memorístico), habilidades intelectuales de alto orden, o disposiciones actitudinales y valorativas. Con ese tipo de preguntas, siempre que sean bien utilizadas, se puede medir una gran cantidad de atributos sofisticados de los estudiantes (Backhoff, Peón y Sánchez, 2005 p. 13).

Sin embargo, a pesar de los avances para elaborar cada vez mejores instrumentos y de los esfuerzos que se despliegan para no sólo diseñar exámenes, sino para aplicarlos, calificarlos, obtener hallazgos estadísticos y traducirlos a lenguaje pedagógico, a fin de que sean empleados por los profesores y directivos; también son, por otra parte, ampliamente conocidas las limitaciones que aún permanecen en este tipo de instrumentos (Shepard, L. 2006; Martínez, F. 2009), por lo que los resultados que se obtienen resultan ser un referente

útil para algunas tareas educativas, como son: evaluar el sistema educativo, proporcionar indicios para la mejora de docentes y centros escolares, entre otras; pero no resulta indicado para muchas otras tareas, de ahí la importancia del señalamiento de los autores de la cita anterior, cuando afirman: “siempre que sean bien utilizadas”, este asunto del empleo correcto de los exámenes parece estar desplazando la discusión y la polémica respecto a la validez y confiabilidad por el de la aplicación y el uso, debido a las implicaciones que se pueden derivar de usos inadecuados de estas evaluaciones; ya que tal como lo advierte Lorrie Shepard (2006), estos exámenes pueden convertirse en un arma de doble filo:

“Es preciso, sin embargo, advertir sobre un riesgo que no se puede ignorar: el peligro de que las pruebas en gran escala produzcan consecuencias negativas para la calidad educativa, si se les comprende y utiliza mal” (p. 6).

Por tal, las decisiones que se toman en relación a los usos que se darán a las evaluaciones de gran escala, implica una responsabilidad mayúscula, ya que como lo menciona la misma autora:

69

Aspectos esenciales que no evalúan las pruebas en gran escala, como la expresión escrita y oral, la formación de actitudes y valores, la educación artística, e incluso los niveles cognitivos más complejos de las áreas tradicionalmente cubiertas de Lectura, Matemáticas y Ciencias. (Shepard, p. 6)

Desde esta perspectiva se justifica el esfuerzo de las evaluaciones en gran escala, en el sentido de que permiten evaluar a grandes poblaciones, pero señala una de las limitaciones más visibles de estos exámenes, el hecho de que dejan fuera de evaluación a grandes zonas del currículo, que no son sensibles al formato de opción múltiple; aún así se puede considerar que este tipo de instrumentos es un buen coadyuvante, que junto a las evaluaciones que realizan los docentes contribuye a la mejora educativa:

Las pruebas de gran escala pueden ser una herramienta valiosa para apoyar los esfuerzos de mejora de la calidad educativa, si se las ve como un complemento de las evaluaciones a cargo de los maestros, insustituibles para evaluar de manera detallada todos los aspectos del currículo y para hacerlo de manera que puedan ofrecer a cada alumno la retroalimentación precisa sobre sus puntos fuertes y débiles, esencial para mejorar el aprendizaje (2006, p. 6).

La autora pone énfasis en el uso que se haga de las evaluaciones, que pueden convertirse en herramientas muy útiles, con la capacidad de arrojar luz sobre zonas oscuras

que frecuentemente quedan fuera de la reflexión para los profesores y para las escuelas; sobre todo si se trabajan como complementos de las evaluaciones formativas (Shepard, 2006; Martínez y Santos, 2009). Sin embargo, por otra parte los exámenes de gran escala pueden llegar a dañar el proceso educativo en un aula o escuela, tal como se señala en la cita anterior, los exámenes de gran escala no deben ser utilizados para evaluar el proceso enseñanza aprendizaje al interior de un grupo escolar, sino que se constituyen en un referente para docentes, alumnos, padres de familia, directivos y colectivo escolar para tomar nota de las áreas de oportunidad que sean señaladas a partir de estos instrumentos de evaluación; es decir, no es correcto ni ético dar un uso a los exámenes que no corresponde con el objetivo para el que fueron creados, ni con los alcances que este formato de examen presenta, el doctor Martínez Rizo aborda este punto, cuando formula una crítica al empleo de los resultados obtenidos en la prueba ENLACE :

Pese a las consideraciones anteriores, el desconocimiento de los alcances y límites de las pruebas ENLACE ha hecho que sus resultados se utilicen para propósitos para los que no son adecuados, en particular para valorar la calidad de una escuela solamente con base en el puntaje promedio obtenido por sus alumnos y, lo que es aún más inadecuado, para valorar el desempeño de los maestros.

Este señalamiento es grave, ya que en lugar de promover la mejora de los profesores y de los centros educativos, se genera una obsesión por los resultados.

A esto agrégueseles las anomalías en que se incurren durante la aplicación de los exámenes en donde los medios de comunicación reportan evidencias de copiado que resulta bastante significativo y que lógicamente afecta o golpea podría decirse, de manera contundente los resultados que se obtienen, se habla de un número de copiado de alrededor de doscientos mil alumnos, durante la aplicación de los instrumentos de evaluación (Arreguín, 2008, abril, 26), aunque es muy probable que la cifra sea significativamente más alta, dado que es difícil establecer cuáles de los exámenes son copiados o son dictados parcialmente por los docentes aplicadores y cuáles no, los casos que se mencionan son los que han sido detectados, seguramente una gran cantidad de casos de copiado pasan inadvertidos, al respecto Martínez y Santos (2009) señalan lo siguiente:

[Se] "...hace inevitable el riesgo de que algunos alumnos copien las respuestas de otros, lo que se facilita por las dimensiones del sistema educativo y el carácter masivo de ENLACE,

que dificultan el control de su aplicación, por lo que se decidió utilizar procedimientos computarizados para la detección de posibles casos de copia” (p. 270).

Evidentemente esta es un área que requiere de mayor investigación, aunque lógicamente resulta difícil aproximarse al objeto de estudio desde algunas perspectivas de investigación, si podría ser un terreno propicio para metodologías como las representaciones sociales, imaginario social, interaccionismo simbólico, etnografía o etnometodología respecto al copiado en los exámenes a gran escala. A pesar de ser este un asunto muy evidente, cualquier profesor de educación básica sabe que esto ocurre de ordinario, sin embargo, es un tema que ha quedado en un segundo plano, siendo que esta es una fuente de error muy sensible en este tipo de exámenes; y esto ocurre lógicamente debido a que los resultados de los exámenes afectan de manera positiva o negativa a los docentes y al plantel educativo en el que laboran, como consecuencia de que se hacen del conocimiento de toda la estructura educativa, desde jefes de sector, supervisores, directores de escuela y docentes; por lo que la posición obtenida en las evaluaciones es socializada en las zonas escolares, haciéndose la escuela de una especie de fama, dependiendo del lugar en que haya sido ubicado, de acuerdo a los resultados de los exámenes; es decir, se presta a que se creen sanciones sociales entre colegas y entre directivos hacia los docentes. De esta manera los profesores y los directivos desean que su grupo, escuela y zona escolar aparezcan con una buena calificación. Shepard advierte del tipo de escenarios que se generan a partir de estimular o castigar por medio de las evaluaciones en gran escala:

Otros ejemplos de manejo inapropiado de los resultados de las evaluaciones en gran escala son la asignación de estímulos a los docentes o la elaboración de ordenamientos simples de escuelas supuestamente en función de la calidad de unos y otras, sin tener en cuenta los numerosos factores que inciden en los resultados de los alumnos en las pruebas ni tener en cuenta las limitaciones de éstas. En vez de ello, para facilitar el aprendizaje, es igualmente importante que la retroalimentación esté vinculada explícitamente a criterios claros de desempeño y que se proporcione a los estudiantes estrategias de mejoramiento (Shepard, 2006, p. 19).

Es decir, cuando se socializa el lugar que obtiene la escuela y esto provoca la comparación entre docentes se crean estímulos y castigos sociales entre compañeros, apareciendo para algunos alumnos, docentes y directivos, como una experiencia desalentadora, desviándose del propósito al que toda evaluación educativa se encuentra

orientada, que es el de mejorar las prácticas a partir de la retroalimentación que proporciona, lo expuesto anteriormente es explicado por Shepard (2006) de la siguiente manera:

“Tal vez las consecuencias negativas más serias de las prácticas tradicionales de calificación provienen del uso de comparaciones normativas” (p. 36).

Al parecer, este escenario se recrudecerá, debido a que en el ciclo escolar 2009-2010, que recién ha concluido, se otorgaron estímulos a docentes y directivos por buenos desempeños en la prueba ENLACE, en el marco del Programa de Estímulos a la Calidad Docente, contraviniendo las recomendaciones de los expertos en evaluación, en el sentido de que no debe estimularse a los profesores a partir de los resultados de estos exámenes. El escenario que se genera a partir de estos usos de las evaluaciones en gran escala en nuestro país lo describe Rosa Aurora Padilla:

La enseñanza y el aprendizaje se han convertido hoy en una especie de adiestramiento para pasar la prueba. Más que preocuparse por el aprendizaje de sus alumnos, los maestros procuran entrenarlos para aprobar exámenes como ENLACE. Para que la evaluación cumpla con su función formativa de mejorar la enseñanza y el aprendizaje, es necesario pasar de la evaluación como objetivo, o más aún como indicador, a la evaluación vinculada al proceso de aprendizaje, al contenido de evaluación y a las formas de enseñanza. (2009, p. 1-2).

Esto implica desviarse de los objetivos del currículo para apropiarse de objetivos personales relacionados a la obtención de un estímulo económico, para lo cual se tendría que trabajar en función del examen:

El uso de calificaciones como recompensa contribuye a lo que Lave y Wenger (1991) llamaron la “comercialización del aprendizaje” (p. 112). Cuando no se da un valor cultural al incremento de la habilidad y la participación de alguien en un esfuerzo, la única razón para participar es obtener un conocimiento superficial que puede exhibirse para que lo evalúen. En revisiones de estudios experimentales, los investigadores descubrieron que el uso de recompensas externas puede minar realmente el interés intrínseco de los estudiantes en una tarea (Shepard, 2006, p. 36).

Parecen claras las recomendaciones encontradas en la literatura sobre el tema, en el sentido de la peligrosidad de un uso inadecuado de los resultados generados en las evaluaciones en gran escala, como son estimular a los maestros o alumnos con mejores logros, produciéndose una paradoja en donde los profesores y directivos en un afán de generar expectación y deseo por alcanzar mejores niveles de logro, toman un atajo, dejando de lado el currículo ocupándose directamente de adiestrar a los alumnos para contestar

exámenes de opción múltiple; lógicamente esta práctica no generará mejores niveles de calidad educativa, quizá la única ganancia que se obtenga de esto es que los alumnos obtendrán mejores lugares en los exámenes PISA ¿Será esto una coincidencia? O en realidad hay una intencionalidad en los altos mandos de la educación en nuestro país.

Referencias

Arreguín, A. (2008, Abril, 26). *Detectan Copiado en Pruebas de ENLACE y PISA en Querétaro*. En: Diario rotativo. Consultado el 1 de agosto de 2010. En: <http://rotativo.com.mx/queretaro/detectan-copiado-en-prueba-enlace-y-pisa-en-queretaro/6153/html>

Backhoff, E. Peón, M. y Sánchez, A. (2005). *Manual Técnico. Construcción de Reactivos*. México: INEE.

Martínez, F. (2009) Evaluación Formativa en Aula y Evaluación a Gran Escala: Hacia un Sistema más Equilibrado. En: Revista Electrónica de Investigación Educativa. Vol. 11. No. 2. P.p. 1-18. Recuperado el 1 de agosto de 2012 de: <http://redie.uabc.mx/contenido/vol11no2/contenido-mtzrizo2.pdf>

Martínez, F. y Santos, A. (2009). *Consideraciones sobre la evaluación educativa*. En De Alba A. *¿Qué dice la investigación educativa?* México. COMIE, pp. 265-304.

Padilla, R. A. (2009). *La Prueba ENLACE desde un Análisis Didáctico. Más Allá de una Política de Calidad para la Educación Básica*. En: Memorias del X Congreso Nacional de Investigación Educativa del Consejo Mexicano de Investigación Educativa. Veracruz: COMIE. Recuperado el 1 de agosto de 2012. En: http://www.comie.org.mx/congreso/memoria/v10/pdf/area_tematica_02/ponencias/1235-F.pdf

Shepard, L. (2006). *La Evaluación en el Aula*. México: INEE.

*La RED-IES es una revista electrónica semestral
publicada por la Red de Investigación Educativa en Sonora A. C. (REDIES).
Navojoa, Sonora, México. Junio, 2012.*

© REDIES, A. C.
Red de Investigación Educativa

