

# Las pruebas estandarizadas en la escuela contemporánea, ¿llave o cerrojo para la mejora de la educación?

Tiburcio Moreno Olivos\*

Recibido: 10/04/2016

Aceptado: 02/06/2016



## Resumen

Este artículo tiene como objetivo analizar, desde una perspectiva crítica, cómo operan las pruebas estandarizadas en el marco de la escuela contemporánea, y el debate sobre su contribución al mejoramiento de la educación. Considerando el poder que la evaluación ha adquirido en las últimas tres décadas en todo el mundo —el que se manifiesta en su capacidad para influir tanto en el diseño del currículum como en los resultados de aprendizajes esperados— resulta importante reflexionar acerca del funcionamiento de las pruebas estandarizadas, dado que éstas constituyen uno de los instrumentos *par excellence* que los sistemas educativos nacionales emplean para valorar el rendimiento escolar. El artículo también contrasta la situación de los países desarrollados que están viviendo una era de postestandarización, mientras que en los países en desarrollo las pruebas estandarizadas han incrementado su hegemonía e influencia. Se abordan los alcances y limitaciones de este tipo de instrumentos para el perfeccionamiento de la educación, deteniéndonos, particularmente, en el caso de México.

## Palabras clave

Pruebas estandarizadas, evaluación del aprendizaje, metodología de evaluación.

## Standardized testing in contemporary schools: key or padlock for the improvement of education?

## Abstract

The purpose of this article is to analyze, from a critical perspective, how standardized tests operate in contemporary schools and to discuss their contribution to the improvement of education. Considering the power that evaluation has acquired over the past three decades around the world —as manifested in their ability to influence both curriculum design and expected outcomes of learning—, it is important to reflect on how standardized testing works, since these type of tests are the instruments *par excellence* employed by most national educational systems from developing countries to assess school performance. The article also contrast the situation of how evaluation is conducted in developed countries with that of developing ones, where the first ones are living in an era of post-standardization while in the second, standardized tests have increased their hegemony and influence. The paper closes with an analysis of the scope and limitations of standardized testing for the improvement of education, stopping, particularly in the case of Mexico.

## Key words

Standardized testing, learning and students' assessment, evaluation methodology.

\* Universidad Autónoma Metropolitana–Cuajimalpa, México. E-mail: tmoreno@correo.cua.uam.mx

*“Llena tu escuela con profesores que sólo piensan en mejorar los resultados de los exámenes y acabarás sin tener una organización de enseñanza y aprendizaje, tendrás, más bien, una arraigada distracción de las tareas educativas más importantes en un escenario diverso.” (Hargreaves & Shirley, 2012, p. 56).*

## 1. Introducción

Desde la segunda mitad del siglo pasado —y en lo que va del presente— hemos sido testigos de cómo el poder de la evaluación se ha ido acrecentando, a tal grado, que en la actualidad parece irrefrenable su capacidad para intervenir en el funcionamiento y operación de los sistemas educativos nacionales en todo el mundo. Este impulso se explica, en parte, por el marcado interés de ciertos organismos internacionales (OCDE, UNESCO, FMI, BID, entre otros) en conseguir una mayor calidad de la educación, mediante la eficiencia y la rendición de cuentas públicas de los sistemas educativos.

Aunque el empleo de las pruebas estandarizadas es de vieja data, y pese a las innumerables críticas que han recibido a lo largo de la historia de la educación, estos artefactos, lejos de perder fuerza, continúan siendo un instrumento privilegiado para medir el rendimiento de los alumnos, los docentes y los centros educativos. Su dominio es tal, que en diversos ambientes el vocablo *pruebas* o *exámenes* es usado, con demasiada frecuencia, como sinónimo de evaluación. Hoy sabemos que «evaluación» es un término cuyo significado tiene un alcance mayor, y que las pruebas o exámenes son sólo un instrumento de evaluación entre los varios que existen. El asunto de los instrumentos de valoración remite a la dimensión metodológica, es decir, al *cómo* de la evaluación, lo cual constituye sólo un elemento de cualquier modelo evaluativo.

El uso de las pruebas estandarizadas en los sistemas educativos contemporáneos, mucho más abiertos y democráticos que antes, sólo se puede comprender si se amplía el foco y se sitúa a dichas pruebas en el paradigma de evaluación en el que se inscriben, lo que permite analizar el significado y las finalidades que cumplen en la sociedad actual.

Por el momento histórico y el contexto (inicios del siglo pasado) en que surgieron las pruebas estandarizadas, en el seno de una sociedad industrial occidental, éstas se corresponden con lo que posteriormente se denominó como paradigma tecnológico–racional de la educación. La visión de la evaluación educativa que imperaba en aquella época también se sitúa en las coordenadas de este paradigma. Se trata de una perspectiva de la evaluación como sinónimo de medición; cuando los autores de aquella época hacían referencia a la evaluación, en realidad estaban hablando de medición. No olvidemos que toda evaluación entraña una medición, pero no toda medición implica una evaluación. Como la evaluación surgió, inextricablemente, unida a la medición en el ámbito de la psicología, esta influencia ha sido tan notoria y duradera que, incluso hoy en día, en algunos ambientes escolares se pueden hallar ciertas reminiscencias de este legado.

Aquellos años también se caracterizaron por una excesiva confianza en erigir a la evaluación como una ciencia, lo que demandaba el empleo de instrumentos de valoración presumiblemente *científicos*, que fueran objetivos, válidos y confiables. Las pruebas estandarizadas representaban muy bien ese prototipo de instrumentos que se consideraban asépticos, neutrales y libres de valores, exentos de cualquier sesgo cultural, razón por la

cual se pensaba que se podían aplicar universalmente obteniéndose los mismos resultados.

Desde luego, con el paso del tiempo, esta confianza fue mermando, al reunirse cada vez más evidencias que demostraron que las pruebas estandarizadas no eran ni tan «incoloras» ni tan inocuas como se pensaba. Por ejemplo, se puso de manifiesto que evaluar a los alumnos en un idioma que no comprenden producirá puntajes bajos (e inexactos) en las pruebas.

Incluso, recientemente, algunos de los impulsores de las pruebas estandarizadas en los Estados Unidos de Norteamérica (país considerado por excelencia la cuna de la evaluación), en una especie de *mea culpa*, han confesado el error que cometieron durante tanto tiempo al promover su empleo, toda vez que las pruebas estandarizadas más que mejorar la educación representan un pesado lastre que ha producido funestas consecuencias en el aprendizaje de muchas generaciones de alumnos.

Esta apretada síntesis que hemos expuesto, asumiendo el riesgo de simplificar un proceso histórico sumamente complejo y controversial, nos permite adentrarnos en el tema que nos ocupa en este documento: las pruebas estandarizadas en la escuela contemporánea.

## **2. La supremacía de las pruebas estandarizadas**

La historia de la educación está plagada de ejemplos de prácticas escolares uniformes que cierran la puerta a una diversidad (siempre al borde y marginada) que, pese a todo, lucha por sobrevivir y ocupar su lugar en la escuela. La evaluación representa un claro ejemplo de una política que homogeniza realidades tan disímbolas. Las pruebas estandarizadas —por su propia naturaleza y diseño— no pueden reconocer la diversidad de las poblaciones escolares a las que se les suministran, aunque la información que brindan y las decisiones que se toman a partir de estos datos, sí que afectan de manera significativa la actuación de los centros escolares, así como la vida presente y futura de los sujetos evaluados.

Hemos llegado a un punto en el que resulta evidente el predominio que tiene la evaluación, y particularmente las pruebas estandarizadas, para modelar el currículum y condicionar, en gran medida, los resultados de aprendizaje. En algunos ambientes se ha denunciado cómo las pruebas estandarizadas distorsionan el proceso de enseñanza—aprendizaje, causando un severo daño a la capacidad de los alumnos para continuar aprendiendo a lo largo de la vida (*lifelong learning*). Esta competencia para aprender de forma constante es muy apreciada en la sociedad del conocimiento. Una vez que se acepta que el conocimiento evoluciona y cambia de forma tan vertiginosa que pronto se vuelve obsoleto, la necesidad de que los individuos desarrollen su capacidad para continuar aprendiendo de forma permanente se vuelve indispensable.

Pero las pruebas estandarizadas no sólo lesionan el aprendizaje, al convertir a los alumnos en sujetos pasivos, incapaces de comprometerse con él, y en sujetos ocupados con la memorización antes que con la comprensión. Estos instrumentos también afectan a la enseñanza y, a menudo, constriñen el currículum escolar, junto con sus correspondientes evaluaciones de aula, toda vez que los profesores limitan su instrucción a las exigencias de las pruebas, las cuales generalmente están centradas en conocimientos que sólo requieren el recuerdo o la evocación de ciertos hechos o datos aislados. Existe una constante presión sobre el profesorado para que oriente su esfuerzo en preparar a los alumnos para las pruebas de rendición de cuentas. En efecto:

Hoy en día, hay demasiados profesores obligados a concentrarse en exámenes de lengua y matemáticas en detrimento de otras áreas del currículum como las ciencias sociales, el medio ambiente o las artes. Pedir que las escuelas basen su labor en torno a los datos lleva a muchas de ellas a centrar su atención en los exámenes, y a culturas de ansiedad por conseguir resultados a corto plazo. En las comunidades más pobres, esto puede llevar a un estrepitoso fracaso e intervenciones no deseadas. (Hargreaves & Shirley, 2012, p. 73)

Herman y Baker (2009) resumieron los efectos que numerosos estudios han documentado, en relación con el poder de las pruebas externas de alto riesgo, que incluyen: señalar las prioridades del currículum, modelar al profesor en función de aquello que se evalúa, alinear el currículum con lo que se examina y disminuir el énfasis sobre lo que no se examina. Irónicamente, un currículum estrecho tiende a ser más riguroso con los alumnos que se encuentran en riesgo de fracaso escolar.

En este contexto vemos como la iniciativa, la creatividad, la libertad de pensamiento, la intuición, y la autonomía... otrora cualidades tan apreciadas de una buena enseñanza, ahora parecen no tener cabida, dado que los requerimientos y apremios son otros. Así, los educandos aprenden para la evaluación que, en este caso, se restringe a las pruebas. Además, se envía un mensaje distorsionado a los alumnos: sólo vale la pena aprender lo que viene en las pruebas, lo que no se incluye en ellas no merece ningún esfuerzo (Monereo, 2009). Y si admitimos la pobreza conceptual de muchas de estas pruebas, el daño causado es evidente. Esto no significa que los profesores no deban preparar a sus alumnos para salir airoso de estos exámenes, lo que se advierte es que tenemos un serio problema cuando la meta de aprendizaje consiste, fundamentalmente, en superar estas pruebas, el horizonte se limita y se pierde el sentido trascendente del proceso educativo.

Un cambio significativo, durante los pasados cincuenta años, ha sido la creciente demanda de pruebas por parte de los responsables de formular las políticas educativas, quienes se encuentran insatisfechos con la calidad de la educación y esperan pruebas que sirvan como instrumentos para el cambio de la enseñanza y el aprendizaje (Linn, 1989). En consecuencia, se especula que el surgimiento de nuevas pruebas, basadas en renovadas perspectivas constructivistas sobre la enseñanza y el aprendizaje (como se refleja en las reformas curriculares más recientes) cambiarán las prácticas de aula. Se cree que estas pruebas novedosas también podrían emplearse para facilitar el aprendizaje, brindar retroalimentación a los alumnos y a los profesores, y para mejorar la educación en general.

No obstante, las prácticas de evaluación continúan siendo normativas, en su mayor parte, con un énfasis en la selección y, aunque ha habido intentos por desarrollar nuevos procedimientos de evaluación, éstos no han sido radicalmente modificados para informar y apoyar el cambio en la enseñanza y el aprendizaje (Moreno, 2014). Estos propósitos de evaluación, junto con los procedimientos de las pruebas usadas para estos fines, parecen sugerir que la evaluación es vista como una tecnología, desarrollada por técnicos o expertos en medición, y usada por otros para tomar decisiones específicas o introducir cambios. Sin embargo, la tecnología ha sido refinada para ser más eficiente, estandarizada y codificada, y la evaluación, generalmente ha sido vista como externa y separada de la enseñanza y el aprendizaje.

### 3. Las pruebas estandarizadas en la era de la postestandarización

Aunque se reconoce el notable papel que la evaluación educativa ha adquirido en las últimas décadas, en todo el mundo, es justo precisar que no se trata de cualquier tipo de evaluación, sino de una perspectiva particular que nos retrotrae al pasado: la evaluación como sinónimo de medición, centrada en productos o resultados, homogénea y descontextualizada.

Esta declaración no debe interpretarse como un rechazo a ultranza del empleo de las pruebas estandarizadas, pues —como veremos más adelante— ellas cumplen una función que puede ser importante para valorar el funcionamiento general de los sistemas educativos. No obstante, consideramos que antes de centrar la atención en el tipo de instrumentos de valoración, deberíamos tener claro desde el inicio cuál es el sentido y la finalidad del sistema de evaluación empleado, qué se hará con los resultados obtenidos, qué decisiones se adoptarán, a quiénes afectarán esas decisiones y de qué forma lo harán. La evaluación no es un acto neutral, tiene efectos y consecuencias duraderas en la vida de las personas y de las instituciones evaluadas. Esto es algo que no puede ni debe ignorarse.

En México, la perspectiva de la evaluación como tecnología, lejos de perder fuerza, parece surgir con un énfasis renovado; esto se puede constatar por el peso que han adquirido las pruebas estandarizadas y sus puntuaciones, en las políticas educativas nacionales; contrariamente a lo que ocurre en los países desarrollados donde las evaluaciones estandarizadas parecen haber entrado en una fase de declive. Se trata de una era de postestandarización de la educación que responde a una:

(...) estandarización educativa [que] ha embrutecido nuestro currículum y ha lastrado nuestras escuelas con una burocracia aún más despótica y arbitraria cuya inflexibilidad nos está minando la adaptación al futuro. Estas viejas ideas del cambio educativo surgidas en el siglo XX están totalmente desfasadas y no sirven para este rápido, flexible y vulnerable nuevo mundo del siglo XXI. (Hargreaves & Shirley, 2012, p. 13)

Este demérito de la estandarización también lo encontramos en una férrea crítica a las ideas en boga para la reforma de las escuelas en Estados Unidos, entre las que se encuentran las pruebas estandarizadas y la rendición de cuentas punitiva, remedios que van acompañados de excesivas prescripciones y requerimientos burocráticos (Ravitch, 2011). A partir de estas ideas, se advierte de las secuelas negativas de la estandarización, la cual ha mermado la autonomía y la capacidad profesional del profesorado: *“Los niños no aprenden a pensar por sí mismos si se espera que los maestros sólo hagan lo que se les dice”* (MacKinnon, 2010). En este sentido, recuperamos las valiosas aportaciones de Hargreaves y Shirley (2012), quienes plantean la necesidad de atacar ese exceso de estandarización evaluadora que ignora la diversidad y destruye la creatividad. La estandarización —señalan— ha minado nuestra capacidad de entender y tratar con la diversidad. Y ponen como ejemplo lo que ocurre en Singapur, una nación con un alto rendimiento escolar que: *“Enseña menos, aprende más”*, y ordena un 10% de *espacio en blanco* en el horario para que los profesores aporten iniciativa individual y creatividad a sus horas de enseñanza.

En la sociedad de la información, resulta irrefutable la importancia que tienen los datos. En educación, ofrecen información para la toma de decisiones, nos permiten hacer un seguimiento del progreso y nos obligan a asumir nuestras responsabilidades y los resultados no favorables. Pero hay un problema, y es que las actuales reformas educativas han hecho

promesas exageradas sobre el alcance de los datos, lo cual ha generado confusión. Como sostienen los autores que hemos venido referenciando:

Los datos estadísticos no son siempre evidentes por sí mismos: se necesitan juicio profesional y experiencia para interpretarlos y para añadir información a los mismos. El problema del sendero de la tecnocracia no es la gran cantidad de datos, sino cómo han pasado a ser definidos y delimitados los datos y las pruebas, cómo se interpretan y se usan los datos, cómo depender demasiado de los datos distorsiona el sistema y lleva a la ignorancia y la marginalización de la importancia del criterio moral y de la responsabilidad profesional. (Hargreaves y Shirley, 2012, p. 62)

En educación es interesante prestar atención a la procedencia de los datos, que, como ya se ha señalado, se convierten en un artilugio poderosísimo para fundamentar la defensa de una evaluación como sinónimo de medición. Tales datos estadísticos proceden de los exámenes o pruebas estandarizadas que son externos a la escuela y que pueden ser nacionales e internacionales. De este modo:

En las escuelas y sistemas escolares, casi todos los datos de la nueva ortodoxia vienen de exámenes regulados en lengua y matemáticas, y a veces en ciencias. Los tecnócratas valoran lo que miden en lugar de medir lo que valoran. Reducen el currículum, priorizan las bases que examinan y les da igual que la enseñanza se centre en un «entrenamiento» para pasar los exámenes. Estos exámenes varían entre distritos, estados y naciones, y por eso las escuelas se pueden encontrar con que producen muy buenos resultados en unos exámenes y muy malos en otros. (Hargreaves y Shirley, 2012, p. 63)

#### **4. Las pruebas estandarizadas y su contribución al aprendizaje**

La investigación, en este ámbito, ha revelado la escasa influencia que tienen las pruebas estandarizadas en la mejora del aprendizaje de los alumnos, toda vez que sus resultados son de poca utilidad para los profesores y las escuelas. Esto sucede así porque los informes suelen redactarse en un lenguaje técnico que sólo resulta comprensible para los expertos en medición, y otras veces, porque la información que generan las pruebas ofrece una visión de conjunto del funcionamiento del sistema educativo, pero los resultados no guardan una relación directa con la enseñanza y el aprendizaje. En este sentido, cabe recuperar la advertencia al respecto que se nos hace desde los Estados Unidos de Norteamérica:

Los resultados de este tipo de evaluación pueden servir para informar políticas y para tomar decisiones sobre programas educativos importantes. Pero tales pruebas por sí mismas no pueden producir la deseada mejora de la escuela porque los *test* no tratan directamente con asuntos de la efectividad del profesor o de la motivación del estudiante. (Stiggins, 1999, p. 191)

En México, los resultados de las pruebas nacionales estandarizadas, parecen tener poca incidencia en la mejora de la educación básica; así lo ha expresado el Instituto Nacional para la Evaluación de la Educación (INEE), en su Informe Anual 2007 titulado *La educación para poblaciones en contextos vulnerables*. Basándose en información derivada de los cuestionarios de contexto de las pruebas EXCALE,<sup>1</sup> reconoce que pese a la dificultad para

1 La sigla se refiere a los Exámenes de la Calidad y el Logro Educativos, aplicados en México, por El Instituto Nacional para la Evaluación de la Educación, organismo público autónomo, creado en 2008 por el presidente de la época Vicente Fox Quesada. Para más información puede consultarse el siguiente Links: <http://www.inee.edu.mx/index.php/acerca-del-inee>. [Nota del editor].

distinguir con claridad el efecto de los factores del entorno y la escuela:

La gran desigualdad de los resultados escolares se explica, sin duda, por la conjunción de ambos. Los alumnos que viven en circunstancias más desfavorables en el hogar son atendidos en escuelas de mayores carencias. En vez de que la escuela compense, en alguna medida, las desventajas de los alumnos de medios desfavorables, contribuye más bien a agudizarlas. (Instituto Nacional para la Evaluación de la Educación [INEE], 2007, citado en Moreno, 2011, p. 118)

Esto ocurre porque las pruebas estandarizadas no guardan una conexión directa con la mejora del proceso enseñanza–aprendizaje, su función es otra. Se presenta un desfase cuando se le atribuyen a este tipo de instrumentos funciones de evaluación que no le competen, tales como las formativas y de motivación para el aprendizaje. Las funciones de estas pruebas son, sobre todo, sumativas y de rendición de cuentas, de ahí su escaso valor para retroalimentar el proceso formativo. Muchas veces se piensa que se puede motivar a los alumnos para que acrecienten su esfuerzo y consigan mejores resultados:

(...) estableciendo estándares académicos más altos, subiendo el listón e implementando más pruebas de alto impacto (*high-stakes tests*). Éste es el fundamento de la creencia en el poder de las pruebas estandarizadas orientadas hacia la rendición de cuentas para la mejora de la escuela. (Moreno, 2010, p. 89)

Lo que sabemos es que cuando ciertos alumnos encaran un reto más arduo mediante pruebas de alto impacto, su respuesta suele ser un incremento en su esfuerzo y así consiguen aprender más de lo que lo hubiesen hecho sin ese incentivo adicional. Pero no hay que perder de vista que esto ocurre sólo con *algunos* alumnos. La mayoría de los educandos cuando son confrontados con un desafío aún más difícil que el primero en el que ya ha tenido una experiencia de fracaso, no aumentará su esfuerzo, algo que generalmente se olvida. Estos alumnos percibirán los altos estándares y las demandas por mayores puntuaciones en las pruebas como algo inalcanzable, cayendo en la desesperanza. Lo anterior significa que el presupuesto de que las pruebas estandarizadas tendrán el mismo efecto en todos los alumnos es erróneo. Esto no sucederá así en absoluto. Algunos alumnos afrontan las pruebas con un historial personal académico fuerte y una expectativa de éxito, mientras que otros traen detrás una estela de fracaso escolar lacerante e incluso crónico: *“Algunos alumnos vienen a dar muerte al dragón, mientras que otros esperan ser devorados por él”* (Stiggins, 2002, p. 761). Consecuentemente, las pruebas estandarizadas impulsarán el aprendizaje de algunos alumnos, mientras que desmotivará a otros y causará en ellos desaliento.

Pero, ¿cuál es el papel asignado a los profesores en este tipo de evaluaciones? Desde el enfoque de una evaluación estandarizada, los docentes se convierten en *medidores* de las conductas observables de los alumnos, mientras que su trabajo se rutiniza, degrada y devalúa. La evaluación, al ser externa e impuesta por otros agentes, deja de ser un servicio de auto–información para conducir el proceso de enseñanza–aprendizaje y reorientar al sistema, transformándose en una preocupación casi exclusiva por el control de la situación y por el mantenimiento de la disciplina en el aula. En este escenario no resulta difícil advertir cuáles son las funciones preponderantes de la evaluación: selección, represión, fiscalización y potenciación de la reproducción social dominante.

## 5. Alcances y límites de las pruebas estandarizadas

La preeminencia que la evaluación ha adquirido en las últimas décadas, debido —entre otras cosas— a la estrecha vinculación que tiene con los conceptos de *calidad* de la educación y *accountability* o rendición de cuentas, ha generado cambios relevantes en los centros escolares. No obstante las críticas expresadas con anterioridad, las pruebas estandarizadas también aportan algunos beneficios significativos a la educación. Entre las transformaciones positivas podemos mencionar los importantes avances en la capacidad de la evaluación para crear nuevas pruebas estandarizadas que brindan información precisa acerca del logro del alumno, lo que permite hacer inferencias válidas sobre los conocimientos y las habilidades que posee un alumno en un área particular de conocimiento, de modo que estos logros puedan ser comparados con una muestra nacional de alumnos de la misma edad o grado escolar (Popham, 1999). Sin embargo, aunque los programas internacionales de evaluación reúnen aceptables estándares de calidad técnica, el costo de haber logrado esa alta calidad de la evaluación a gran escala ha sido enorme, no sólo en dinero invertido, sino también en oportunidades perdidas. En este sentido, Shepard (2000), por ejemplo, reconoce los efectos perversos de los *test* de rendición de cuentas y el alcance que tienen los exámenes externamente impuestos, que impiden y proscriben valiosas prácticas de evaluación en el aula.

Nosotros no negamos el valor de las pruebas estandarizadas, puesto que de alguna manera, fuerzan a los alumnos, profesores y administradores para conseguir altos estándares académicos, y en este sentido, ciertamente, pueden hacer que los esfuerzos sean más productivos para muchos. Tales pruebas proporcionan a los educadores oportunidades para reflexionar sobre qué se logró y qué no se pudo lograr, para lo cual será necesario que los docentes tengan acceso a los resultados y que éstos se les presenten en un lenguaje comprensible, considerando que ellos no están familiarizados con la jerga técnica de los evaluadores y expertos en medición.

Otra dificultad principal que se ha identificado es que la estandarización tiene la consecuencia negativa de convertir a los alumnos en agentes pasivos. En este sentido, se afirma que:

El problema fundamental del enfoque predominante de la evaluación es que construye alumnos como sujetos pasivos. Es decir, los alumnos no tienen ningún otro papel que no sea someterse a los actos de evaluación de los demás, a ser medidos y clasificados. Cumplen con las normas y procedimientos para satisfacer las necesidades de una burocracia evaluadora. (Boud, 2007, p. 17)

Tenemos entonces que las pruebas estandarizadas fomentan la pasividad de los alumnos, toda vez que ellos no se sienten implicados en el proceso de aprendizaje; más bien, son forzados a comprometerse con la memorización de los contenidos que se les enseñan. El aula se convierte en una tierra infructuosa, al tiempo que se ignora el tema de la responsabilidad de los alumnos en su propio aprendizaje. En definitiva, enseñar a los alumnos para responder las pruebas estandarizadas genera problemas en el aula y en el ambiente de aprendizaje. En este escenario:

(...) enseñar para las pruebas estandarizadas es una consecuencia inevitable de la presión administrativa por la medición de los resultados. Consecuentemente, los estudiantes son orientados para dar una respuesta «correcta» antes que para pensar por sí



mismos y convertirse en aprendices creativos. (Buckman, 2007, p. 33)

Más recientemente, ha surgido un nuevo pensamiento sobre la *confiabilidad* y la *justicia*, que han sido conceptualizadas principalmente desde los requerimientos técnicos de la evaluación a gran escala. Así tenemos que Parkes (2013) propone una reconceptualización de la *confiabilidad* que combina la medición y la enseñanza, que considera la toma de decisiones del profesor y un proceso inferencial subjetivo de los datos obtenidos, e incluye las dinámicas sociales en el aula. Este autor propone un nuevo marco conceptual para el desarrollo de la confiabilidad de la evaluación en el aula. En tanto que Tierney (2013) argumenta que la *justicia* para la evaluación en el aula, mientras sea inconsistentemente definida en la literatura, es mejor conceptualizarla como una serie de prácticas, que incluye la transparencia de las expectativas de la evaluación del aprendizaje y de los criterios para evaluar el trabajo, oportunidad para aprender y demostrar lo aprendido, coherencia y equidad para todos los alumnos, reflexión crítica, y la creación y mantenimiento de un ambiente de aprendizaje constructivo en el que la evaluación es empleada para mejorar el aprendizaje.

En el actual clima de rendición de cuentas de la escuela y la toma de decisiones en el manejo de los datos, se espera que los profesores utilicen buenas prácticas de evaluación como un medio para mejorar los resultados de aprendizaje de los alumnos. Los hallazgos de investigación más recientes revelan que los profesores necesitan integrar los enfoques de las pruebas a gran escala, con los tipos de ítems con los que evalúan a sus alumnos, y con un plan para la enseñanza futura (McMillan, 2013).

Trabajos como los de Schneider, Egan y Julian (2013) describen la naturaleza de las pruebas de alto riesgo, y explican cómo las evaluaciones de aula se pueden alinear con estas pruebas para conseguir un balance, triangulando la comprensión de lo que los alumnos saben y pueden hacer. Ellos señalan que la comprensión profunda y las habilidades del pensamiento de orden superior requieren de pruebas externas que es difícil que los profesores puedan cubrir con sus evaluaciones de aula. El desajuste resultante es problemático. Con una mayor coherencia entre todos los principales tipos de evaluación (de fin de curso, interna, de aula) los profesores tendrían una mejor comprensión de las fortalezas y debilidades de los alumnos, así como una mejor visión de las mejoras de enseñanza que se necesitan.

Por otro lado, un reconocido experto internacional en evaluación como Linn (1998), también ofrece una serie de razones del porqué la medición es un mecanismo ampliamente usado por las reformas educativas, pues la historia ha mostrado que:

1. *Las pruebas son relativamente económicas.* En comparación con cambios que implican un aumento del tiempo de enseñanza, reducción del tamaño de las clases, formación y atracción de mejores profesores, la evaluación tiene costos bajos.
2. *Los cambios en la medición pueden implementarse relativamente rápido.* Otras reformas escolares pueden tomar muchos años para implementarse, y puede tomar más tiempo aun el saber si han mejorado la escuela.
3. *Los resultados de las pruebas son visibles y llaman la atención de los medios.* Resultados pobres en el primer año de un nuevo programa de medición suelen verse seguidos por el incremento de los puntajes en los años siguientes, dando la apariencia de que las escuelas están mejorando.
4. *La medición puede generar otros cambios que serían difíciles de legislar.* La investigación ha demostrado que los requerimientos de medición y evaluación, a nivel estatal o distrital,

han motivado cambios en el currículo y en la enseñanza en los niveles de la escuela y el aula. Es mucho más difícil legislar directamente cambios en el aula.

Por supuesto, Linn (1998) también nos advierte que cuando las pruebas se usan para tomar decisiones importantes sobre las escuelas y los alumnos, estos rasgos atractivos dan lugar a problemas inesperados. Los resultados de las pruebas pueden ser incompletos o engañosos, conduciendo a malas decisiones de política. Sin embargo, el citado autor afirma que un alto interés permanente en la medición educacional está asegurado por la necesidad que tienen quienes formulan las políticas de contar con información rápida, tanto sobre el progreso de los alumnos como de la calidad de las escuelas.

## 6. Reflexiones finales

Durante la mayor parte del siglo XX, la investigación sobre la evaluación en educación estuvo centrada en el papel de las pruebas estandarizadas (Shepard, 2006; Stiggins & Conklin, 1992). La comunidad profesional de medición de la educación estaba preocupada por el papel de las pruebas estandarizadas, tanto desde una perspectiva de la evaluación a gran escala como con respecto a cómo los profesores usaban los datos de las pruebas para la enseñanza en sus propias aulas. El siglo pasado estableció lo que podría describirse como un paradigma de *medición/psicométrico*, un contexto poderoso en el que la atención sobre la evaluación se centró en los principios psicométricos de las pruebas a gran escala y los correspondientes asuntos formales, técnicos y estadísticos para todos los educadores (McMillan, 2013).

En la segunda mitad del siglo pasado las pruebas estandarizadas recibieron un cúmulo de críticas: a) miden sólo conocimiento declarativo y no procedimental, b) se centran en el resultado y no en el proceso, c) no cubren adecuadamente el dominio evaluado, d) existen diferentes habilidades e incluso inteligencias que no son evaluadas por este tipo de *tests*, e) son medidas relativas, f) la estandarización supone una muestra homogénea, g) el formato de elección múltiple limita las evaluaciones y supone otras habilidades distintas a las evaluadas y, h) se alejan de las verdaderas demandas contextuales.

Por todo lo anterior, se plantea que este tipo de evaluación no ofrece a los alumnos la oportunidad de mostrar sus verdaderas competencias. Los objetivos son homogeneizados y no representan adecuadamente los intereses de cada estudiante en su currículum; la puntuación también se centraliza y se basa en criterios establecidos por personas que no conocen a los alumnos, sus metas o sus oportunidades de aprendizaje. Como consecuencia, llegó un punto en que la palabra *tests* se había convertido en tabú (Moreno, 2012).

Aunque el contexto de investigación de la evaluación en el aula ha cambiado, alejándose de los principios específicos de las pruebas a gran escala, las ideas fundamentales de validez, confiabilidad y justicia siguen siendo una parte muy importante de la comprensión de lo que ocurre en el aula y en la investigación sobre la evaluación en el aula. Desde la perspectiva de los investigadores, las tradicionales ideas de validez, confiabilidad y justicia son empleadas para asegurar la instrumentación adecuada y la credibilidad de los datos obtenidos (McMillan, 2013).

Por otro lado, en México, desde la última década del siglo pasado, se está haciendo una fuerte inversión de recursos para establecer y perfeccionar un sistema nacional de evaluación basado en pruebas estandarizadas, con la finalidad principal de hacer comparaciones

entre escuelas y entre alumnos. Dichas comparaciones, generalmente, resultan desafortunadas, sobre todo en un país como México, con marcados y profundos contrastes sociales, económicos y culturales. Tales analogías pueden tener efectos nocivos en las escuelas que obtienen los niveles más bajos en la clasificación, toda vez que minan su confianza y motivación.

Las pruebas estandarizadas tiene un enfoque cerrado, pero también existen prácticas de evaluación con un enfoque abierto que permiten la valoración de los individuos, sólo que éstas no pueden proporcionar el tipo de comparaciones globales que un público ávido por la *productividad* educativa de sus escuelas parece querer. En estas coordenadas, Eisner (1999) menciona que una opción consiste en emplear dos tipos de evaluación diferentes. Un tipo sería el uso de pruebas estandarizadas a gran escala que buscan ofrecer datos comparativos sobre el desempeño de escuelas o zonas escolares. Un segundo tipo de evaluación sería creado para conocer los distintos talentos de los alumnos —individualmente— y los efectos de las prácticas escolares en su propio desarrollo. Se trata de evaluaciones de aula, diseñadas por el profesor y con un propósito formativo.

El sistema educativo mexicano no sólo no ha logrado superar el vetusto problema de la evaluación del alumnado, sino que todo parece indicar que lejos de avanzar hacia la senda de una *verdadera evaluación*, es decir, una evaluación auténtica, más cualitativa, centrada en los procesos de aprendizaje y en las personas como aprendices; es decir, en vez de orientarse hacia una evaluación formativa, integral y humana, lo que se observa —a juzgar por la importancia revelada en la cantidad de recursos invertidos que actualmente están teniendo las pruebas nacionales estandarizadas para evaluar tanto a los alumnos como a las escuelas y sus profesores—, es un preocupante retorno al pasado en el que la evaluación fue concebida como sinónimo de medición, lo cual redujo su alcance y coartó las posibilidades de transformación de la educación.

Contrariamente a lo que ocurre en países que tienen un sistema educativo exitoso, como es el caso de Corea del Sur, donde desde hace varias décadas las pruebas para promover a los alumnos de primaria a secundaria y de secundaria a bachillerato fueron eliminadas por considerarse enfermedades sociales, en México este tipo de pruebas están a la alza, cada vez se emplean con más frecuencia, abarcan más áreas de conocimiento e involucran a un mayor número de alumnos y escuelas.

Cuando se aborda la temática de evaluación en educación, un ejemplo recurrente suele ser el caso de Finlandia, que destaca por sus magníficos resultados de forma constante en la prueba PISA. Lo que quizá las personas no familiarizadas con este tópico ignoran, es que:

Curiosamente, Finlandia no tiene un sistema de exámenes nacionales excepto algunas muestras confidenciales con motivos de seguimiento. No hay una palabra en finlandés para *rendición de cuentas*. En su lugar, la educación pública se ve como una *responsabilidad* colectiva, social y profesional. (Hargreaves & Shirley, 2012, p. 91)

Volviendo a la situación del caso mexicano —que es justo decirlo, no es privativa de este país— en realidad estamos presenciando una idea firmemente arraigada de antaño: para que la evaluación goce de autoridad y legitimidad ha de adoptar la forma de medida. La consecuencia de esta forma de pensar es que sólo si pudiésemos perfeccionar las medidas y, quizá, sólo si los alumnos actuaran de forma más parecida a objetos inanimados a los

que pudiéramos aplicar nuestros instrumentos de medida, podríamos operar la educación como una ciencia empírico-analítica (Grundy, 1991). Esta visión preconiza un paradigma positivista de la evaluación, el cual no sólo nunca nos ha abandonado, sino que además parece cobrar un nuevo impulso en varios países latinoamericanos con las recientes reformas educativas que se fundamentan en políticas de corte neoliberal.

En esta tesitura, podríamos hacer referencia a los denominados sistemas de vigilancia, los cuales en vez de informarse por la evidencia en culturas de relaciones de confianza que profundizan y desarrollan el aprendizaje y el rendimiento; las comunidades docentes se convierten en equipos improvisados, insertos en la cultura de los resultados inmediatos. Las estrategias que suelen emplearse en estos sistemas de vigilancia consisten en más preparación para los exámenes nacionales, más clases de refuerzo, más atención a los niños que están justo por debajo de la nota de corte de los exámenes, y otros fenómenos parecidos; que apuntan a las presiones de un sistema que cree que los datos pueden hacerlo capaz de saber cómo y cuándo intervenir con cualquier alumno, escuela o aula, en cualquier momento. Con estas acciones es posible que el rendimiento en los exámenes se haya elevado pero, ¿realmente es la mejor enseñanza?, ¿qué culturas estamos creando en nuestras escuelas con este proceso?

Resulta claro que detrás de este tipo de sistemas de vigilancia, obcecado por los datos estadísticos procedentes de las pruebas estandarizadas externamente impuestas, subyace una desconfianza absoluta en la capacidad y el compromiso de los docentes y las escuelas para cumplir con su misión, por lo que necesitan ser constantemente fiscalizados para asegurarse que efectúen las tareas para las que fueron contratados. Ante un panorama controlador como el descrito antes, ni las pruebas estandarizadas ni ningún otro artefacto pedagógico puede contribuir a la mejora de la educación, sino todo lo contrario. En definitiva, en un sistema cerrado donde se inspecciona y controla a los docentes y a las escuelas, las pruebas estandarizadas terminan convirtiéndose en un cerrojo, antes que en la llave que abra la puerta a la mejora de la educación.

### Referencias bibliográficas

- Boud, D. (2007). Reframing assessment as if learning were important. En: Boud, D. & Falchikov, N. (eds.). *Rethinking Assessment in Higher Education. Learning for the longer term*. (pp. 14–25). Nueva York: Routledge.
- Buckman, K. (2007). What Counts as Assessment in the 21st Century? *Thought & Action*, 23, 29–37.
- Eisner, E. (1999). The Uses and Limits of Performance Assessment. *Phi Delta Kappa*, 80(9)658–670.
- Grundy, S. (1991). *Productos o praxis del currículum*. Madrid: Morata.
- Hargreaves, A. & Shirley, D. (2012). *La cuarta vía. El prometedor futuro del cambio educativo*. Barcelona: Octaedro.

- Herman, J. L. & Baker, E. L. (2009). Assessment policy: making sense of the babel. En: Sykes, G., Schneider, B. & Plank, D. N. (eds.). *Handbook of Education Policy Research*. (pp.176–190). Nueva York: Routledge.
- Linn, R. L. (1989). Current perspectives and future directions. En Linn, R. L. (ed.), *Educational Measurement*. (pp. 1–10). New York: Macmillan.
- Linn, R. L. (1998). *Responsabilización basada en estándares. Diez sugerencias*. Santiago de Chile: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL).
- MacKinnon, N. (2010). Comment *The Times Educational Supplement Scotland*. Recuperado de: <http://www.tes.co.uk/article.aspx?storycode=6048653>.
- McMillan, J. H. (2013). Why We Need Research on Classroom Assessment. En: McMillan, J. H. (ed.). *Research on Classroom Assessment*. (pp. 3–16). Thousand Oaks, CA: Sage Publications.
- Monereo, C. (coord.) (2009). *PISA como excusa. Repensar la evaluación para cambiar la enseñanza*. Barcelona: Graó.
- Moreno, T. (2010). Lo bueno, lo malo y lo feo: las muchas caras de la evaluación. *Revista Iberoamericana de Educación Superior*, 1(2), 88–101.
- Moreno, T. (2011). La cultura de la evaluación y la mejora de la escuela. *Perfiles Educativos*, 33(131), 116–130.
- Moreno, T. (2012). Evaluación para el aprendizaje. Perspectivas internacionales. *Revista de Evaluación Educativa*, 1(1). Recuperado de: <http://revalue.mx/revista/index.php/revalue/issue/current>.
- Moreno, T. (2014). Posturas epistemológicas frente a la evaluación y sus implicaciones en el currículum. *Revista Perspectiva Educativa*, 53(1), 3–18.
- Parkes, J. (2013). Reliability in Classroom Assessment. En: McMillan, J. H. (ed.). *Research on Classroom Assessment*. (pp. 107–123). Thousand Oaks, CA: Sage Publications.
- Popham, W. J. (1999). *¿Por qué las pruebas estandarizadas no miden la calidad educativa?* Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL). Recuperado de: [http://www.oei.es/evaluacioneducativa/pruebas\\_estandarizadas\\_no\\_miden\\_calidad\\_educativa\\_popham.pdf](http://www.oei.es/evaluacioneducativa/pruebas_estandarizadas_no_miden_calidad_educativa_popham.pdf).
- Ravitch, D. (2011). *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. Nueva York: Basic Books.
- Schneider, M. C., Egan, K. L. & Julian, M. W. (2013). Classroom Assessment in the Context of High–Stakes Testing. En: McMillan, J. H. (ed.). *Research on Classroom Assessment*. (pp. 55–70). Thousand Oaks, CA: Sage Publications.

- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, 29(7), 2–14.
- Shepard, L. A. (2006). Classroom assessment. En: Brennan, R. L. (ed.). *Educational measurement*. (pp. 623–646). Westport, CT: Praeger.
- Stiggins, R. (1999). Assessment, Student Confidence, and School Success. *Phi Delta Kappa*, 81(3), 191–198.
- Stiggins, R. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappa*, 83(10), 758–765.
- Stiggins, R. & Conklin, N. F. (1992). *In teachers' hands: Investigation the practices of classroom assessment*. Albany: State University of New York Press.
- Tierney, R. D. (2013). Fairness in Classroom Assessment. En: McMillan, J. H. (ed.). *Research on Classroom Assessment*. (pp. 125–144). Thousand Oaks, CA: Sage Publications.