

“CORRUPCIÓN DE LA MEDIDA” E INFLACIÓN DE LOS RESULTADOS DE ENLACE

EDUARDO BACKHOFF / SOFÍA CONTRERAS ROLDÁN

Resumen:

Los propósitos de este trabajo son dos: *a)* explicar cómo se corrompen las mediciones educativas y aparece el fenómeno de inflación de resultados en las evaluaciones estandarizadas y *b)* mostrar evidencias de la inflación de resultados de la prueba de la Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) (básica y media superior). Se compararon los resultados de ENLACE —en Lengua (comprensión lectora) y Matemáticas, de tercero, sexto, noveno y doceavo grados— con los de Excale y PISA que se utilizaron como pruebas control. Se concluye que la prueba ENLACE/educación básica se corrompió y sufrió una inflación en resultados debido a tres factores: la publicación de rankings de escuelas, la asociación de estímulos económicos a los docentes y la falta de control externo de su aplicación. La inflación de ENLACE/del nivel medio superior se pudo deber al uso de reactivos de menor dificultad en sus últimas versiones.

Abstract:

This article has two purposes: *a)* to explain the corruption of educational measurements and the phenomenon of inflated results in standardized evaluations; and *b)* to show evidence of the inflation of results from the National Evaluation of Academic Achievement at Schools (ENLACE) (elementary and high school). A comparison was made of the ENLACE results—in language (reading comprehension) and mathematics for the third, sixth, ninth, and twelfth grades—and the results of Excale and PISA, which were used as control tests. The conclusion is that the ENLACE/elementary education test was corrupted and suffered from the inflation of results due to three factors: the publication of schools' rankings, the association of results and economic stimuli for teachers, and the lack of external control for administering the test. Inflated results of the ENLACE/high school test may have been due to the use of less difficult questions in the test's recent versions.

Palabras clave: corrupción de la medida, inflación de resultados, evaluaciones estandarizadas, prueba ENLACE, PISA, Excale, México.

Keywords: Corruption of measure, score inflation, standardized testing, enlace test, PISA, Excale, Mexico.

Eduardo Backhoff es miembro de la Junta de Gobierno del Instituto Nacional para la Evaluación de la Educación, México. José María Velasco núm.101, colonia San José Insurgentes, 03900, México, DF. CE: ebackhoff@gmail.com
Sofía Contreras Roldán es jefa de proyecto en la Junta de Gobierno del Instituto Nacional para la Evaluación de la Educación, México. CE: scontrerasr@inee.edu.mx

Algunos estudios internacionales relacionados con la evaluación del aprendizaje de alumnos muestran que con el paso del tiempo los resultados de las pruebas estandarizadas sufren un proceso de inflación, entendido como un crecimiento desmedido de las puntuaciones que arroja una prueba (Koretz, 2005). En los sistemas de evaluación educativa se retoma uno de los aforismos más importantes de las ciencias sociales que es la ley de Campbell, el cual afirma que: “Entre más se utilice un indicador para tomar decisiones sociales, éste estará más sujeto a las presiones de corrupción y será más propenso a distorsionar y corromper los procesos sociales que pretende monitorear” (Campbell, 1975:35).

En México, las condiciones de la aplicación (sin control) de la prueba ENLACE (Evaluación Nacional de Logro Académico en Centros Escolares de la educación básica y las consecuencias asociadas a los resultados de los alumnos (rankings de escuelas y estímulos salariales a los docentes) hacen suponer que estos resultados sufrieron este fenómeno, sin embargo no existen estudios que evidencien lo anterior. Por ello, el propósito de este trabajo es doble: *a)* explicar la forma en que se corrompen las mediciones educativas y *b)* mostrar evidencias de la inflación de los resultados de ENLACE, tanto de educación básica como de media superior.

Para lograr estos dos objetivos se abordan los siguientes apartados. Primero, se hace un breve recuento de los antecedentes de las evaluaciones educativas de gran escala: desde el primer estudio de la Asociación Internacional de Logro Educativo (IEA, por sus siglas en inglés), hasta la implementación de la prueba ENLACE en la educación obligatoria. Segundo, se describe el fenómeno de la corrupción de la medida y la inflación de resultados, así como los mecanismos por los que se produce en el sector educativo. Tercero, se describen las características de la prueba ENLACE así como las condiciones de su aplicación y uso. Cuarto, se comparan los resultados de la prueba ENLACE con los de PISA (Programa Internacional para la Evaluación de Estudiantes, PISA, por sus siglas en inglés) y de los Excale (Exámenes de la Calidad y el Logro Educativos). Finalmente, se arriba a conclusiones sobre la inflación de los resultados de ENLACE y su posible explicación.

Antecedentes de las evaluaciones nacionales de aprendizaje

Han pasado cinco décadas desde que la IEA realizó el Primer Estudio Internacional de Matemáticas, en el que participaron estudiantes de 13 y

de 17-18 años de edad, provenientes de doce países (Husen, 1967; Wolf, 2004). Desde ese entonces, se han realizado diversos estudios de gran escala para conocer el nivel de aprendizaje que logran los alumnos de distintos grados y edades, así como en diversos dominios escolares (particularmente, comprensión lectora y matemáticas). En cuanto a programas nacionales, hay que destacar el de Progreso Educativo (NAEP, por sus siglas en inglés) de Estados Unidos, por ser el primero y el que más ha impactado a otros de su misma naturaleza.

Aunque estos estudios de logro educativo son financiados por los gobiernos, hasta hace muy poco tiempo los tomadores de decisiones habían mostrado poco interés en sus resultados. Sin embargo, esta actitud cambió claramente con la introducción de PISA a cargo de la Organización para la Cooperación y Desarrollo Económicos (OCDE). Desde ese entonces, los tomadores de decisiones han mostrado mayor interés en los resultados de las evaluaciones de aprendizaje y ahora creen firmemente en la importancia que tienen sus resultados para el diseño de reformas educativas (Ben-Simon y Cohen, 2004).

Con base en la exitosa experiencia de los estudios internacionales y de algunos sistemas nacionales de evaluación de logro, ciertos países han desarrollado pruebas estandarizadas propias, dando prioridad a uno de los dos grandes propósitos que tienen las evaluaciones de aprendizaje de gran escala:

- 1) informar a la sociedad y a las autoridades educativas sobre el estado que guarda la educación en el país (función de rendición de cuentas) y
- 2) brindar información oportuna y relevante a los docentes que les ayude a mejorar tanto su enseñanza como el aprendizaje de sus estudiantes (función pedagógica).

En Norteamérica y en Latinoamérica se ha dado mayor énfasis a la función informativa y de rendición de cuentas, mientras que en algunos países de Europa (por ejemplo, Holanda y Dinamarca) la evaluación estandarizada tiene una función eminentemente pedagógica.

Por su parte, la primera experiencia de México en evaluaciones de aprendizaje de gran escala se puede ubicar en 1995, con su participación con la IEA en el Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS); proyecto del que nuestro país se retiró, debido a sus malos resultados. Asi-

mismo, a finales de los años noventa, la Secretaría de Educación Pública (SEP) incursionó por primera ocasión en el desarrollo e implementación de pruebas estandarizadas para evaluar los aprendizajes de los estudiantes de educación básica, a las que denominó Estándares Nacionales.

Sin embargo, se puede decir que el gran cambio de actitud de México a favor de las evaluaciones de logro comenzó con la entrada del nuevo milenio. Son tres los proyectos que, en mayor o menor medida, han impactado la opinión de los tomadores de decisiones, de los medios de comunicación y de la sociedad en general. El primero fue la participación de México en el estudio PISA, en 2000; el segundo, la creación del Instituto Nacional para la Evaluación de la Educación (INEE), en 2002, y con él el desarrollo de las pruebas Excale, cuyos resultados se publicaron por primera ocasión en 2005; y, el tercero, de gran impacto social, el proyecto ENLACE, de la Dirección General de Evaluación (de la SEP), cuyos primeros resultados se dieron a conocer en 2006 para educación básica (ENLACE/básica) y en 2008 para el nivel medio superior (ENLACE/MS).

El uso de la prueba ENLACE/básica fue suspendida en 2013 por el Secretario de Educación Pública, Emilio Chuayffet; decisión que fue avalada por el INEE, por considerar que se contaban con algunas evidencias del robo, distribución y venta de los exámenes, así como de mecanismos que habían corrompido la prueba e invalidado parcialmente los resultados en algunas entidades federativas (INEE, 2013).

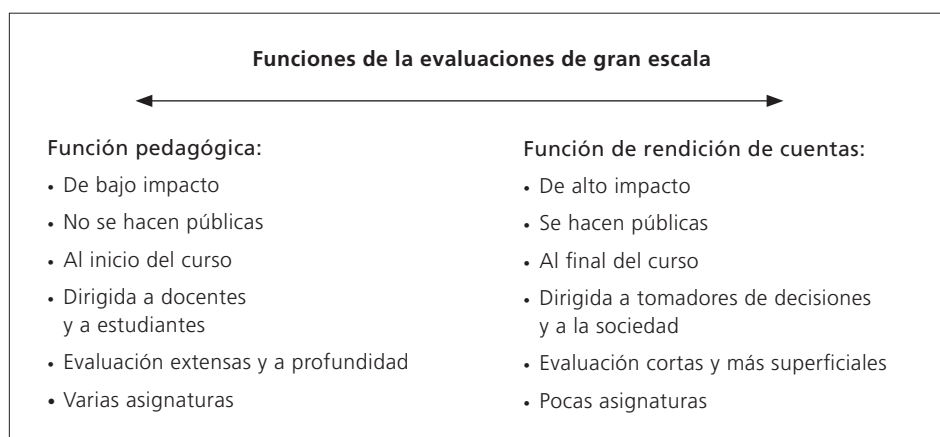
Corrupción de la medida e inflación de resultados

Los grandes propósitos que tienen las evaluaciones de aprendizaje se pueden ubicar en dos extremos, hasta cierto punto incompatibles, de acuerdo con las funciones que priorizan: la pedagógica o la de rendición de cuentas. La figura 1 muestra el continuo teórico de estas funciones y las características que distinguen a las evaluaciones respectivas. Del lado izquierdo se ubican aquéllas cuya función central es la pedagógica, que son de bajo impacto y cuyos destinatarios principales son los docentes y los estudiantes. En el extremo derecho se encuentran las evaluaciones cuya función principal es la rendición de cuentas, que son de alto impacto, y sus destinatarios fundamentalmente son los tomadores de decisiones y la sociedad en general. En medio del continuo se pueden ubicar evaluaciones intermedias que tratan de abarcar más de una función; sin embargo, siempre se deberá tener en mente que una prueba de aprendizaje se diseña con un propósito

y que no podrá atender funciones distintas sin perder eficacia en una de las dos primordiales.

FIGURA 1

Funciones y características de las evaluaciones estandarizadas de aprendizaje



Fuente: Elaboración propia.

Es importante señalar que la función de rendición de cuentas de la evaluación de logro ha adquirido gran preponderancia en los últimos años debido a la exigencia creciente de la sociedad por conocer los resultados de los programas de gobierno, así como a la creencia de algunos sectores de la sociedad de que la exposición pública de tal información sirve como "motor" para el mejoramiento de la calidad de la educación.

La exigencia de la rendición de cuentas, a través de mecanismos de evaluación de estudiantes, ha llegado en los últimos años a responsabilizar a la escuela y al docente. En consecuencia, han habido iniciativas de política educativa para vincular la ejecución de los alumnos en las pruebas estandarizadas con estímulos económicos a los maestros (por ejemplo, el programa de Carrera Magisterial), así como el reconocimiento a estudiantes, el acceso a fuentes de financiamiento de las escuelas o, inclusive, el cierre de aquellos planteles que no muestren progreso en años consecutivos (por ejemplo, la ley estadounidense No Child Left Behind).

Entre mayores sean las consecuencias que se le asocien a los resultados de las evaluaciones de logro, mayor será el impacto para escuelas, docentes y

estudiantes. El impacto más común resulta cuando sólo se publican rankings escolares, poniendo en entredicho el prestigio social de las instituciones y de los “responsables” de los bajos resultados educativos (Ravela, 2006). Esta situación genera incentivos perversos y pone en marcha diversos mecanismos que provocan corrupción de las medidas e “inflación” de los resultados de las evaluaciones.

Diversos autores han estudiado este fenómeno y señalan que la presión por obtener altas puntuaciones (o sufrir las consecuencias implícitas y explícitas) induce, por ejemplo, a que los docentes reduzcan la enseñanza del currículo a los contenidos que evalúan las pruebas, dediquen gran parte del tiempo de clase a la preparación-para-la-prueba, permitan la copia entre estudiantes o eviten que los de menor rendimiento sean evaluados (Koretz, Linn, Dunbar y Shepard, 1991; Heubert y Hauser, 1999; Koretz, 2005; Holcombe, Jennings y Koretz, 2012).

Características de la prueba ENLACE y condiciones de su aplicación y uso

En 2005 se elaboraron versiones de ENLACE de Lengua¹ y Matemáticas para diversos grados de educación básica (de tercero de primaria hasta tercero de secundaria) y en 2008 se ampliaron para el último grado del nivel medio superior. De acuerdo con la SEP, los exámenes para el nivel básico se basan en los planes y programas de estudio nacionales, de tipo criterial, censal, anual y de opción múltiple (SEP, 2013).

Las pruebas ENLACE/básica fueron diseñadas originalmente con propósitos pedagógicos, pero sus fines cambiaron un par de años después para rendir cuentas a la sociedad. En 2008 se empezaron a publicar rankings de escuelas de manera oficial y se asociaron los resultados de los estudiantes a la evaluación de sus docentes en el programa de Carrera Magisterial; el peso o ponderación de la calificación de los alumnos fue, en principio, de 20% y en 2011 aumentó a 50%. La aplicación de las pruebas no tuvo un control externo, por lo que se recurrió a personal de las mismas escuelas para administrarlas y, en algunos casos, a padres de familia y a personal voluntario.

Las consecuencias asociadas a los resultados de ENLACE/básica, tanto sociales (publicación de rankings) como monetarias (incentivos económicos a docentes), así como la falta de control externo de su aplicación forma-

ron un "caldo de cultivo" ideal para que la prueba se pudiera corromper. Ciertos indicios de ello (algunos documentados) lo constituyen: el robo y quema de cuadernillos, la filtración de su contenido, la copia y dictado de respuestas, la enseñanza-para-la-prueba, los "jueves de ENLACE", las tareas escolares con reactivos-tipo-ENLACE, entre otros.

Por su parte, ENLACE/MS inició en 2008 y continúa hasta la fecha. Al igual que su antecesor, estas pruebas evalúan las competencias de Lenguaje² y Matemáticas de los estudiantes que terminan el tercer grado de bachillerato, de forma censal³ y anual, con un formato de opción múltiple. Las pruebas no están alineadas a un currículo específico sino a un perfil referencial⁴ que se elaboró para evaluar las competencias disciplinares básicas. La administración de las evaluaciones se realiza por aplicadores externos a la escuela (Ceneval, 2012) y sus resultados no tienen consecuencias para los planteles, docentes y estudiantes.

Comparación de los resultados de ENLACE con los de Excale y PISA

Para conocer el grado de corrupción e inflación que sufrieron los resultados de ENLACE, tanto para educación básica como media superior, se contrastaron las tendencias de las puntuaciones medias que obtuvieron los estudiantes mexicanos a lo largo del tiempo con las que arrojaron las pruebas de Excale y PISA, que en este estudio funcionan como pruebas control.

Para lo anterior, se recurrió a la información derivada de un reciente trabajo que está por publicarse, denominado "Tendencias en el aprendizaje de la educación en México: una comparación entre ENLACE, Excale y PISA" (Contreras-Roldán y Backhoff, 2014). Este estudio comparó las tendencias en el aprendizaje de los dominios escolares que miden las pruebas en los grados en que al menos dos de las tres evaluaciones coincidieron, a saber: tercero y sexto de primaria, tercero de secundaria y tercero de bachillerato.

Hay que recordar que aunque se hable de una prueba, como ENLACE, en realidad se trata de varias, de acuerdo con el grado y dominio de que se trate. Por ejemplo, en este estudio, ENLACE/básica abarca las pruebas de Lengua y Matemáticas, de tercero, sexto y noveno grados; mientras ENLACE/MS se refiere a los mismos dominios del doceavo grado.

Por otro lado, PISA es una prueba diseñada para estudiantes de 15 años. En México, aproximadamente, 25% de quienes estudian se encuentra

cursando tercero de secundaria y 75% el primero de bachillerato.⁵ En este trabajo, los resultados de PISA/15 (que incluyen a todos los estudiantes, independientemente del grado que cursan), se desagregaron en dos grupos: PISA/sec, que comprende 25% de quienes están en secundaria, y PISA/MS, que abarca 75% de los que cursan el bachillerato.

El cuadro 1 muestra las pruebas que se utilizaron para realizar las comparaciones en cada grado escolar, en ambas asignaturas. Así, los resultados de las pruebas de ENLACE/básica de Lengua y Matemáticas de tercero y sexto grados se compararon con los de Excale de los mismos grados y dominios. Los resultados de tercero de secundaria se compararon con los de Excale, así como con los de PISA/15 y PISA/sec. Finalmente, los de ENLACE/MS se compararon con PISA/15 y PISA/MS.

CUADRO 1

Pruebas utilizadas para comparar los resultados de los estudiantes de educación obligatoria en las pruebas de ENLACE, Excale y PISA

Evaluación	3° primaria	6° primaria	3° secundaria	3° bachillerato
Enlace/básica	X	X	X	
Enlace/MS				X
Excale	X	X	X	
PISA/15			X	X
PISA/sec			X	
PISA/MS				X

El número aproximado de estudiantes que participaron en esta comparación varió considerablemente, según la prueba y grado de que se trate. El cuadro 2 muestra que en ENLACE/básica participaron cerca de 50 millones de estudiantes, en ENLACE/MS lo hicieron aproximadamente 5.5 millones de alumnos, en Excale participaron cerca de 450 mil escolares y, finalmente, en PISA lo hicieron un poco más de 280 mil alumnos.

CUADRO 2

Número aproximado de estudiantes que participaron en el estudio

Evaluación	3° primaria	6° primaria	3° secundaria	3° bachillerato
Enlace/básica	16 401 023	15 917 200	12 599 025	
Enlace/MS				5 419 724
Excale	79 645	97 806	154 331	
PISA/15			282 042	
PISA/sec			67 262	
PISA/MS				185 907

Fuente: Contreras-Roldán y Backhoff (2014).

La comparación de los resultados de las tres pruebas se basó en las tendencias de las puntuaciones obtenidas por los estudiantes en cada dominio, calculadas por el valor de las pendientes de las líneas de regresión respectivas, que se calcularon con las puntuaciones promedio (medias) de cada año en que se aplicaron las pruebas.

Los cuadros 3 y 4 muestran los resultados para la educación básica, mientras que el 5, lo correspondiente al nivel medio superior. En los tres cuadros se presentan las puntuaciones promedio (medias) de los estudiantes de acuerdo con el dominio evaluado, el grado escolar, la prueba correspondiente y el año de aplicación de cada evaluación. El último renglón de los tres muestra los valores de las pendientes de las líneas de regresión calculadas con la información disponible para cada caso.

Un aspecto que se observa en estos cuadros es la diferencia en la cantidad de evaluaciones realizadas y años en que se aplicaron las pruebas. Sin embargo, lo importante de esta información es el valor de las pendientes de cada prueba-asignatura-grado, que indica la "tendencia" en el aprendizaje de los estudiantes; es decir, la velocidad con que cambian las puntuaciones de un año a otro. Por ejemplo, los resultados en Lengua (comprensión lectora) de tercero de primaria, según ENLACE, se incre-

mentaron a una tasa promedio de casi nueve puntos (8.88) por año, lo que se traduce en una mejoría equivalente a un grado escolar completo cada cinco años.⁶ Si se compara el crecimiento de las puntuaciones de Excale para el mismo grado, se observa que apenas es de 2.76 puntos por año, lo que equivale a menos de una tercera parte del incremento que muestra ENLACE.

CUADRO 3

Comparación de las puntuaciones medias y pendientes de Español: ENLACE y Excale de educación básica, y PISA/Sec y PISA/15

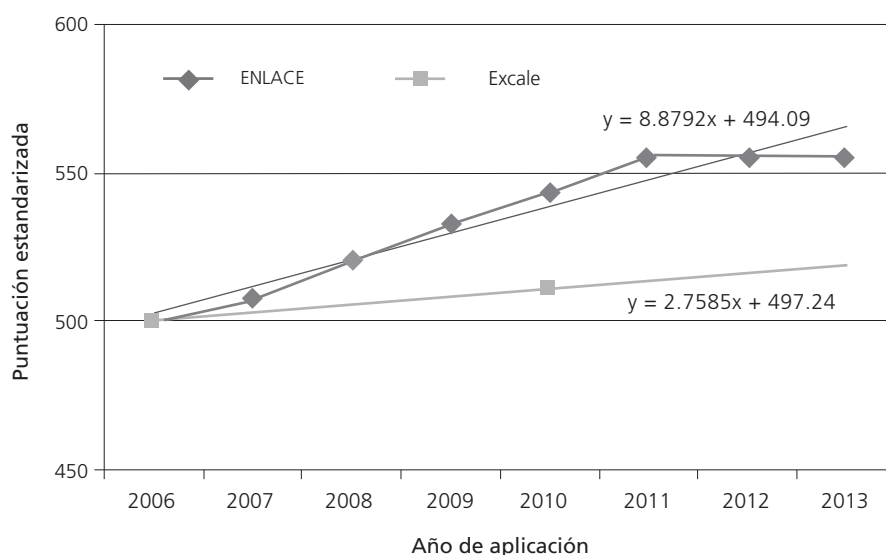
Año	3° primaria		6° primaria		3° secundaria		PISA/Sec	PISA/15
	ENLACE	Excale	ENLACE	Excale	ENLACE	Excale		
2000	-	-	-	-	-	-	375	467
2003	-	-	-	-	-	-	367	441
2005	-	-	-	500	-	500	-	-
2006	500	500	500	-	500	-	367	444
2007	508	-	509	516	514	-	-	-
2008	521	-	512	-	514	498	-	-
2009	533	-	514	518	514	-	385	465
2010	544	511	545	-	505	-	-	-
2011	556	-	549	-	501	-	-	-
2012	556	-	548	-	497	496	392	442
2013	556	-	550	-	494	-	-	-
Pendiente	8.88	2.76	8.14	4.54	-2.01	-0.57	1.73	0.97

Fuente: Contreras-Roldán y Backhoff (2014).

Para observar visualmente este fenómeno inflacionario se presenta la gráfica 1, que muestra las tendencias del aprendizaje en Lengua de ENLACE y Excale, así como las fórmulas de regresión correspondientes.

GRÁFICA 1

*Tendencias de los resultados de ENLACE
y Excale: Lengua 3º de primaria*



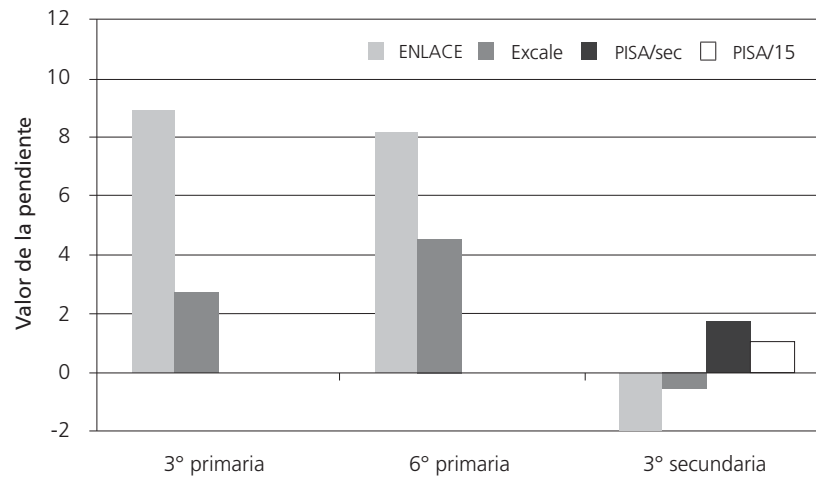
Fuente: Contreras-Roldán y Backhoff (2014).

Por su parte, la gráfica 2 presenta una comparación de los valores de las pendientes de Lengua de ENLACE, Excale, PISA/sec y PISA/15 para estudiantes de educación básica. En esta gráfica se puede apreciar que tanto en tercero como en sexto grados de primaria el crecimiento en las puntuaciones de ENLACE fue considerablemente mayor que el de Excale. Sin embargo, el fenómeno de la inflación no se observa en tercero de secundaria, cuyos resultados son un tanto heterogéneos: un pequeño decremento en las pruebas de ENLACE y Excale y un ligero incremento en PISA.

El cuadro 4 y la gráfica 3 muestran las tendencias en el aprendizaje de Matemáticas de educación básica, donde se podrá observar que la inflación de resultados de ENLACE se presenta en los tres grados; con mayor severidad en primaria (especialmente en sexto) que en secundaria.

GRÁFICA 2

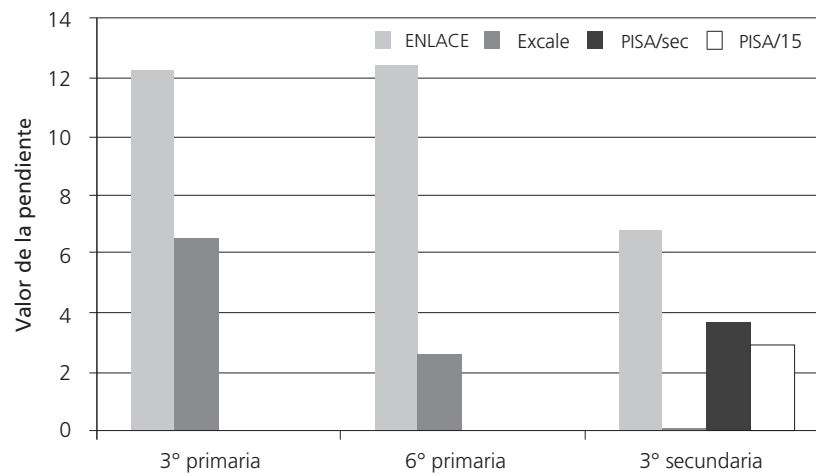
Tendencias de los resultados de ENLACE y Excale: Lengua, educación básica



Fuente: Contreras-Roldán y Backhoff (2014).

GRÁFICA 3

Tendencias de los resultados de ENLACE y Excale: Matemáticas, educación básica



Fuente: Contreras-Roldán y Backhoff (2014).

CUADRO 4

Comparación de las puntuaciones medias y pendientes de Matemáticas: ENLACE y Excale de educación básica

Año	3° primaria		6° primaria		3° secundaria		PISA/sec	PISA/15
	ENLACE	Excale	ENLACE	Excale	ENLACE	Excale		
2003	-	-	-	-	-	-	355	385
2005	-	-	-	500	-	500	-	-
2006	500	500	500	-	500	-	364	406
2007	508	-	510	512	511	-	-	-
2008	516	-	517	-	519	504	-	-
2009	530	-	527	511	519	-	385	419
2010	528	526	547	-	523	-	-	-
2011	541	-	551	-	527	-	-	-
2012	569	-	571	-	545	501	385	413
2013	592	-	588	-	554	-	-	-
Pendiente	12.24	6.52	12.44	2.63	6.79	0.08	3.67	2.87

Fuente: Contreras-Roldán y Backhoff (2014).

Finalmente, el cuadro 5 y la gráfica 4 muestran los resultados de Lengua y Matemáticas para el caso de bachillerato, donde las pruebas de ENLACE/MS se comparan con las de PISA. Es importante señalar que esta comparación no es la más pertinente, pero sí la única posible, dado que ENLACE/MS evalúa las competencias de estudiantes de tercero de bachillerato, mientras que PISA/MS, de jóvenes de 15 años que terminan el primero de bachillerato y PISA/15 los aprendizajes de alumnos de 15 años, inscritos en secundaria y en bachillerato.

Dicho lo anterior, en la gráfica 4 podemos notar que los resultados en el dominio de Lengua no presentan evidencias de haberse inflado. Por el contrario, los de ENLACE/MS en Matemáticas presentan un incremento poco creíble, que sugiere una inflación de resultados.

CUADRO 5

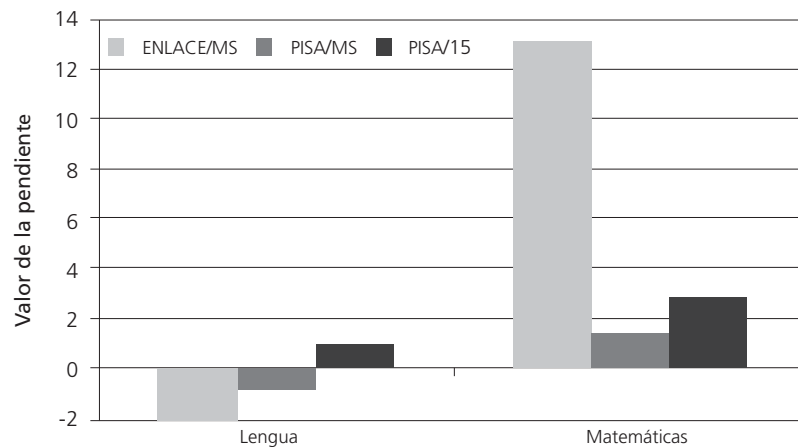
Comparación de las puntuaciones medias y pendientes en Lengua y Matemáticas de bachillerato en ENLACE/MS, PISA/MS y PISA/15

Año	Lengua			Matemáticas		
	ENLACE/MS	PISA/MS	PISA/15	ENLACE/MS	PISA/MS	PISA/15
2000	-	467	422	-	N/A	387
2003	-	441	400	-	422	385
2005	-	-	-	-	-	-
2006	-	444	410	-	437	406
2007	-	-	-	-	-	-
2008	512	-	-	510	-	-
2009	587	465	425	622	457	419
2010	523	-	-	524	-	-
2011	516	-	-	542	-	-
2012	509	442	422	558	430	413
2013	504	-	-	574	-	-
Pendiente	-2.09	-0.84	0.97	13.11	1.40	2.87

Fuente: Contreras-Roldán y Backhoff (2014).

GRÁFICA 4

Tendencias de los resultados de ENLACE y PISA: Lengua y Matemáticas, educación media superior



Fuente: Contreras-Roldán y Backhoff (2014).

Conclusiones

Richard Wolfe, un conocido investigador canadiense, afirmó que los diseñadores de los estudios de logro educativo, tanto nacionales como internacionales, insisten en que su motivación es pura, que los rankings de naciones, estados y escuelas no son relevantes, y que lo que realmente es importante es el conocimiento de los procesos de enseñanza y aprendizaje que se relacionan con los resultados de las evaluaciones de logro. Sin embargo, el hecho es que la atención pública y política se centra, por lo general, en el ordenamiento que producen las evaluaciones, y lo sustancial parece no importar (Wolfe, 1999).

El uso que se le dan a las evaluaciones estandarizadas, que se centran en la rendición de cuentas, tienden a responsabilizar a las escuelas y a los maestros de los resultados de sus estudiantes. Si además, a los resultados se les asocian consecuencias fuertes (premios y castigos), ocurre lo que Donal Campbell denominó la corrupción de la medida y que trae como consecuencia una inflación de las puntuaciones. Este fenómeno puede pasar desapercibido si no se le observa, como fue el caso de las pruebas ENLACE/básica que presentaban las condiciones ideales para que se corrompieran: rankings de escuelas, incentivos económicos a los docentes y falta de control externo en su aplicación. Los resultados aquí expuestos dan evidencia de que, en efecto, las pruebas ENLACE tanto de educación básica como del nivel medio superior presentan una inflación importante. La magnitud de la inflación que se reporta en este trabajo para el caso de México es similar, e incluso mayor, a la que han encontrado otros autores para el caso de Estados Unidos, cuando se comparan las evaluaciones estatales censales (con altas consecuencias) con las nacionales muestrales (sin consecuencias), que es del orden de tres a cinco veces (Koretz, 2005).

Mientras que la inflación en el caso de ENLACE/básica se puede deber a las condiciones de su aplicación y a las consecuencias asociadas a sus resultados, el caso de ENLACE/MS no es claro, dado que no cumple con estas dos características. Una posible explicación, en este caso, es que la dificultad de los reactivos de las nuevas versiones del examen haya cambiado, hipótesis que habría que corroborar.

Finalmente, independientemente de las razones por las cuales se corrompe una medida de la inflación, es importante tener en cuenta que, como lo anticipó Campbell, hace casi cuarenta años, el uso intensivo de

un indicador público es suficiente para debilitarlo y distorsionarlo. Habrá que tener en cuenta esta experiencia, para no repetir el mismo error, en lo que sustituya a ENLACE.

Notas:

¹ En ENLACE de educación básica a este dominio se le refiere como Español y mide, básicamente, comprensión de lectura.

² En ENLACE de educación media superior a este dominio se le refiere como Comunicación y mide, básicamente, comprensión de lectura.

³ Sin embargo, algunas instituciones de educación media superior no son evaluadas, como es el caso de los bachilleratos de la Universidad Nacional Autónoma de México.

⁴ Para conocer más sobre el marco teórico de ENLACE/MS, consultar el Manual técnico en http://www.enlace.sep.gob.mx/content/ms/docs/2012/Manual_Tecnico_ENLACE_MS.pdf

⁵ Aunque también se encuentran estudiantes de 15 años en otros grados de secundaria y bachillerato, éstos representan una proporción insignificante.

⁶ Si consideramos que las diferencias entre un grado y otro son aproximadamente de media desviación estándar.

Referencias

- Ben-Simon, A. y Cohen, Y. (2004). "International assessment: merits and pitfalls", trabajo presentado en la 30ª Conferencia Anual de la Asociación Internacional para la Evaluación Educativa, junio, Filadelfia.
- Campbell, D.T. (1975). "Assessing the impact of planned social change", en G. Lyons (ed.), *Social research and public policies: The Dartmouth/OECD Conference*, Hanover: Dartmouth College. The Public Affairs Center (pp. 3-45).
- Ceneval (2012). *Manual técnico de la prueba Evaluación Nacional de Logro Académico en Centros Escolares de Educación Media Superior (ENLACE MS)*. Ciudad de México: Centro Nacional de Evaluación para la Educación Superior.
- Contreras-Roldán, S. y Backhoff, E. (2014). "Tendencias en el aprendizaje de la educación en México: una comparación entre ENLACE, Excale y PISA", *Revista Nexos* (en revisión).
- Heubert, J. P. y Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation. A report of the National Research Council*, Washington, DC: National Academy Press (en línea). Disponible en: http://www.nap.edu/openbook.php?record_id=6336
- Holcombe, R.; Jennings, J. L. y Koretz, D. (2012). *The roots of score inflation: examination of opportunities in two states' test*, documento de trabajo, Harvard Graduate School of Education.
- Husen, T. (1967). *International study of achievement in mathematics. A comparison of twelve countries*, Estocolmo: Almqvist & Wiksell (en línea). Disponible en: <http://www.jstor.org/stable/3443305>
- INEE (2013). *Instituto Nacional para la evaluación de la Educación (INEE) Secretaría de Educación Pública (SEP)*, Boletín de prensa núm. 14, 19 de diciembre. Disponible en: http://www.inee.edu.mx/images/stories/2013/principal/Comunicado_14_Enlace-Excale_19-12-13.pdf

- Koretz, D. M. (2005). *Alignment, high stakes, and the inflation of test scores*. CSE Report 655, Los Angeles: University of California.
- Koretz, D. M.; Linn, R. L.; Dunbar, S. B. y Shepard, L. A. (1991). "The effects of high stakes testing on achievement: preliminary findings about generalization across test", trabajo presentado en el *Annual Meeting of the American Educational Research Association*, Chicago, Illinois (en línea). Disponible en: <http://nepc.colorado.edu/files/HighStakesTesting.pdf>
- Ravela, P. (2006). *Para comprender las evaluaciones educativas. Fichas didácticas*, Santiago: PREAL (en línea). Disponible en: <http://m.preal.org/detalle.asp?det=729>
- SEP (2013). *Evaluación Nacional del Logro Académico en Centros Escolares. Educación básica. Manual técnico 2013*, Ciudad de México: Secretaría de Educación Pública (en línea). Disponible en: http://www.enlace.sep.gob.mx/ba/manuales_tecnicos/
- Wolf, R. M. (2004). "The contribution of IEA to research and education", trabajo presentado en The 1st IEA International Research Conference (IRC-2004), Lefkosia, Cyprus.
- Wolfe, R. (1999). "Measurement obstacles to international comparisons and the need for regional design and analysis in mathematics surveys", en G. Kaiser, E. Luna e I. Huntley (eds.), *Internacional comparisons in mathematics education*, Londres: Falmer Press.