

Evidencias de fallas de control y fraudes en los resultados de la prueba ENLACE en el nivel medio superior

HUMBERTO GUTIÉRREZ PULIDO* | MARTHA ELENA AGUIAR BARRERA**
PORFIRIO GUTIÉRREZ GONZÁLEZ***

Se estudia el contexto y la evolución de los resultados de la prueba ENLACE, los cuales experimentaron enormes incrementos en matemáticas en todos los niveles. Contrario a lo establecido en los manuales, los niveles de dominio en educación media superior (EMS) se obtuvieron directamente de los puntajes. Del análisis de 6.4 millones de puntajes individuales en matemáticas en EMS, se demuestra que el incremento es generalizado y paulatino durante el periodo (2008-2014), lo cual se atribuye a que los ajustes hechos a la prueba o a sus resultados favorecieron dicho patrón. Además, se identifica un importante sesgo en la distribución de dichos puntajes de los últimos años, que refleja claramente las fallas de control y los fraudes en ENLACE. Esto fue lo que propició tales incrementos desproporcionados, incluidos 75 mil jóvenes que en 2013 y 2014 obtuvieron resultados perfectos. En contraste, en lectura en EMS sus resultados no se inflaron.

We examine the context and evolution of results of the ENLACE test, which have shown vast improvement in mathematics at all levels. Contrary to the approach established in the manuals, levels of mastery in secondary education (Spanish acronym EMS) were obtained directly from scores. An analysis of 6.4 million individual scores in mathematics in secondary education found that the improvement was generalized and gradual over the period (2008-2014), which is attributed to adjustments made in the test or its results favoring such a pattern. Also, a significant bias in the distribution of such scores was identified in the last years studied, which clearly reflects failure of control and fraud in ENLACE testing. This was behind such disproportionate improvements, including 75 thousand students in 2013 and 2014 who obtained perfect scores. In contrast, results for reading in secondary education were not inflated.

Palabras clave

Evaluación de pruebas
Escala de medición
Fraudes en pruebas estandarizadas
ENLACE
Distribuciones de probabilidad

Keywords

Test evaluation
Scales of measurement
Fraud in standardized tests
ENLACE
Distributions of probability

Recepción: 4 de noviembre de 2018 | Aceptación: 19 de noviembre de 2019

DOI: <https://doi.org/10.22201/iisue.24486167e.2020.169.59168>

- * Profesor titular del Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad de Guadalajara (México). Doctor en Probabilidad y Estadística. Líneas de investigación: aplicaciones de la estadística a problemas sociales y de la industria; control estadístico de procesos; diseño de experimentos. Publicaciones recientes: (2020), *Calidad y productividad*, México, McGraw-Hill Interamericana; (2019, en coautoría con C. Orozco), "Análisis de supervivencia para el tiempo de permanencia de los alumnos de la Licenciatura en Matemáticas de la Universidad de Guadalajara", *Actas del XIX Congreso Internacional de Investigación Educativa*, vol. II, pp. 156-163. CE: humberto.gutierrez@cucei.udg.mx
- ** Profesora titular del Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad de Guadalajara (México). Doctora en Diseño Curricular y Evaluación Educativa. Líneas de investigación: procesos de enseñanza y aprendizaje de las matemáticas; evaluación y diseño curricular; desarrollo y aplicación de la tecnología para el aprendizaje y la enseñanza de las matemáticas. Publicación reciente: (2018, en coautoría con H. Gutiérrez y P. Gutiérrez), "Diseño y aplicación de un cuestionario sobre la práctica docente del profesorado de matemáticas en ingeniería y ciencias", *Góndola. Enseñanza y Aprendizaje de las Ciencias*, vol. 13, núm. 1, pp. 33-54; CE: martha.aguiar@academicos.udg.mx
- *** Profesor titular del Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad de Guadalajara (México). Doctor en Ciencia y Tecnología. Línea de investigación: desarrollo de metodologías estadísticas para problemas industriales y biotecnológicos. CE: pgutierrezglez@gmail.com

INTRODUCCIÓN

Uno de los elementos clave del desarrollo de un país es la educación escolarizada, porque permite adquirir conocimientos individual y socialmente valiosos que le dan al individuo una mayor capacidad y habilidad para decidir en libertad (PNUD, 2010). El bienestar y la prosperidad de los jóvenes dependen, más que nunca, de las competencias que la educación y la capacitación les ofrecen (UNESCO, 2012). La importancia que se le ha dado a la educación escolarizada ha llevado a organismos multinacionales y a iniciativas al interior de los países a buscar evaluaciones sobre los conocimientos y habilidades que los sistemas educativos son capaces de generar en niños y jóvenes. Una organización pionera y líder en este campo es la IEA (International Association for the Evaluation of Educational Achievement), que en 1964 realizó la que se considera la primera evaluación a gran escala (Postlethwaite, 1967).

Al interior de los países se han desarrollado esfuerzos particulares; destaca Estados Unidos, donde en 1965 se aplicó la prueba EEOS (Equality of Educational Opportunity Survey), la cual dio origen al Informe Coleman (Coleman *et al.*, 1966); y en 1969 se aplicó por primera vez la prueba NAEP (National Assessment of Educational Progress), que fue evolucionando en enfoques y perspectivas (Beaton *et al.*, 2011), hasta convertirse en la principal prueba en ese país.

En México, la evaluación educativa masiva llegó tarde si se compara con los casos anteriores. La primera gran experiencia se dio con la participación de México en la prueba PISA (Programme for International Student Assessment) en el año 2000, y la implementación de EXCALE y ENLACE (Evaluación Nacional del Logro Académico en Centros Escolares) en 2004 y 2006, respectivamente (Backhoff y Contreras, 2014).

La prueba ENLACE buscó tener una cobertura censal¹ sobre el 100 por ciento de los alumnos contemplados. Se aplicó anualmente en el nivel básico (primaria y secundaria) de 2006 a 2013; y en la enseñanza media superior (EMS) de 2008 a 2014. Los primeros resultados de dicha prueba confirmaron y ampliaron el diagnóstico sobre la situación crítica del sistema educativo nacional en sus niveles básicos. Por ejemplo, en 2006, en el nivel primaria, 78.7 por ciento de los niños evaluados fueron catalogados en un nivel insuficiente o elemental en lectura, y 82.4 por ciento en matemáticas (SEP, 2013a). Estos resultados, en los que se enfatizaba la jerarquización de alumnos, escuelas, municipios y estados, se difundieron en forma masiva entre padres de familia, centros escolares y en los medios de comunicación.

Adicionalmente, la Secretaría de Educación Pública (SEP) estableció el Programa de Estímulos a la Calidad Docente, con la idea de premiar e incentivar a aquellos docentes cuyos alumnos alcanzaran los primeros lugares o que logran avances significativos en ENLACE (SEP, 2013b). Esto, aunado a la forma de utilizar esa prueba, la hicieron caer en la categoría que en el idioma inglés se conoce como “*high-stakes test*”, que son aquellas que se utilizan para tomar decisiones importantes, como reconocimientos, castigos, avances o compensación para estudiantes, educadores, escuelas o distritos.²

Existen varias investigaciones que analizan diferentes aspectos de la prueba ENLACE, como su diseño y aplicación (Vázquez, 2016; Coll, 2013), y muchos más sobre sus resultados; otros son estudios comparativos que consideran factores demográficos, de contexto escolar o relacionados con algún otro programa o intervención, por ejemplo, Campos y Santillán (2016), Gutiérrez *et al.* (2015a; 2015b), Campos y Urbina (2011), Sacona (2010),

1 Caso contrario a la prueba EXCALE, misma que se aplicó a muestras de estudiantes. EXCALE y ENLACE se sustituyeron en 2014 con la prueba PLANEA, en medio de la polémica de la reforma educativa de 2013, que incluyó, a nivel constitucional, la obligación de evaluar a los docentes. Ésta fue derogada en 2019.

2 Ver: <https://www.edglossary.org/high-stakes-testing> (consulta: 15 de octubre de 2019).

Sánchez (2009) y Cervera *et al.* (2008). En general, los estudios existentes usan datos ya agregados, en forma acrítica, y utilizan información de algunos años de ENLACE, pero ninguno se enfoca a analizar todo el periodo y a cuestionar los grandes incrementos en sus resultados. Es claro, entonces, que hacen falta más estudios que profundicen en los resultados individuales y su evolución considerando todo el periodo de aplicación, de manera que se puedan evidenciar mejor los problemas de diseño, aplicación, control y evaluación de la prueba en los diferentes niveles educativos.

Un estudio que sí cuestiona los resultados es el de Backhoff y Contreras (2014), junto con otro muy similar de los mismos autores (Contreras y Backhoff, 2014), donde se compara la evolución de los puntajes promedio de la prueba ENLACE con los de EXCALE y PISA. Los autores concluyen 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 (Backhoff y Contreras, 2014: 1267).

El análisis que respalda estas conclusiones se apoya, en buena medida, en la comparación de las tendencias de los puntajes promedio de ENLACE, EXCALE y PISA. El problema es que EXCALE y PISA sólo se aplicaron de dos a tres veces en el mismo periodo que ENLACE, lo que implica una débil estimación de las tendencias de aquéllas. Además, en dicho trabajo no se analizaron los resultados en función de los niveles de dominio, que fue la escala de mayor uso y difusión en ENLACE. Otra limitación de dicho estudio, en su parte estadística, es que sólo analiza el puntaje promedio nacional y no desagrega a diferentes niveles territoriales o de contexto escolar, ni analiza los puntajes individuales.

Lo anterior ratifica la necesidad de más estudios que profundicen sobre ENLACE en sus resultados y diferentes aspectos —dado que la gran base de datos generados durante esos años lo hace posible— para así contar con análisis más minuciosos orientados a identificar sus fallas. Con ello se aportarán elementos adicionales para tener un juicio más profundo y documentado sobre esa etapa de las evaluaciones estandarizadas en México, así como para la planeación y evaluación educativa actual y futura.

Con base en lo anterior, el objetivo de este trabajo es analizar brevemente el contexto en el que se originó la prueba ENLACE, algunas de las reacciones que provocó y, sobre todo, la evolución de sus resultados. Esto se enmarca en la revisión de algunas posiciones críticas sobre las pruebas masivas estandarizadas, incluidos algunos reportes de fraudes en su aplicación en diferentes partes del mundo.

En el análisis que se hace se cuantifican los cambios en los niveles de dominio o logro y en los puntajes promedio en las habilidades matemáticas y de lectura para los tres niveles educativos; también se hace un estudio detallado de los 6.4 millones de registros de resultados de exámenes de ENLACE en EMS en ambas habilidades. Se analiza también su evolución entre 2008 y 2014, considerando diferentes niveles de desagregación territorial y de contexto escolar. En los dos análisis se encuentran evidencias de fallas de control y fraudes en la aplicación de la prueba. En otros estudios sobre ENLACE no se han reportado análisis detallados como los indicados, junto con sus hallazgos y evidencias.

LAS PRUEBAS MASIVAS ESTANDARIZADAS Y SUS PROBLEMAS

En la introducción se señaló que existen diferentes pruebas masivas para evaluar los conocimientos y habilidades que los sistemas educativos son capaces de generar en niños y jóvenes. Con frecuencia los resultados de

estas pruebas se usan para jerarquizar, castigar o incentivar, sea por los propios gobiernos o por las diferentes instituciones que usan los datos para esos fines; pero con esto la atención pública y política se centra justamente en las jerarquizaciones que producen las evaluaciones y, en buena medida, se olvida lo sustancial de la evaluación: identificar desde las grandes fallas del sistema educativo hasta los procesos de enseñanza y aprendizaje a nivel aula (Wolfe, 1999). Además, se ha visto que cuando los incentivos para mejorar los puntajes de las pruebas son altos, la tentación de manipular los resultados se convierte en una distorsión no intencionada y concreta (Battistin, 2016).

Con relación a lo anterior, existen amplias evidencias documentadas que indican que al relacionar una evaluación cuantitativa con mejoras sociales se promueve la corrupción de los resultados de lo que se evalúa. Este fenómeno, llamado ley de Campbell,³ ha sido muy estudiado, en particular por Koretz (2019; 2017; 2005) y Nichols y Berliner (2005) en Estados Unidos, en el sector educativo. Estos autores hacen un seguimiento sobre los efectos perniciosos de las pruebas masivas estandarizadas, tanto de directivos como del profesorado, y sostienen que, a pesar de estos efectos, estas pruebas se siguen utilizando para evaluar la efectividad de las reformas promovidas (Koretz, 2017).

En México, la prueba ENLACE usó en forma profusa el indicador de nivel de logro o dominio (deficiente, elemental, bueno y excelente) que alcanzaba un niño o un joven para premiar a docentes y evidenciar o premiar escuelas. Esta forma de exponer y utilizar los resultados es muy parecida a lo que ocurre en la evaluación estatal (EE) aplicada en Estados Unidos a partir de la ley “Que ningún niño se quede atrás” (*No Child Left Behind*) (US Department of Education, 2003), en la que los resultados difieren mucho de los federales; por ejemplo, se reportan casos en donde escuelas

que obtienen puntuaciones casi de 100 en matemáticas en EE tienen resultados reprobatorios a nivel federal (Nichols y Berliner, 2005).

En su artículo, Cizek (2001) detalla varios mecanismos para identificar las trampas. Esta investigación analizó 40 escuelas (23 sospechosas y 17 de control) de 7° y 8° grado en Iowa Tests of Basic Skills, y se obtuvieron evidencias de trampas mayores a las esperadas. También en Olson y Fremer (2013) se presentan diferentes métodos para prevenir e identificar manipulaciones y trampas.

Uno de esos casos emblemáticos se dio en Atlanta en 2009, conocido como el “*Atlanta public schools cheating scandal*”, en referencia al escándalo propiciado por los resultados de pruebas sospechosas en las escuelas públicas de esa ciudad. El gobierno de Georgia informó que 178 educadores en 44 de las 100 escuelas del distrito habían instrumentado las trampas (Goldstein, 2011).

Angrist *et al.* (2015) analizaron los datos de la prueba oficial de Italia de 2009 a 2012 para el nivel básico y encontraron evidencias significativas de manipulación de los puntajes en lectura y matemáticas que atribuyeron a los propios aplicadores y evaluadores de las pruebas.

Battistin (2016) muestra que en las pruebas oficiales Key Stage 2 del Reino Unido, aplicadas a los estudiantes que se gradúan de la escuela primaria, hasta 2007 todos los exámenes que caían tres puntos o menos por debajo de la calificación mínima de aprobación eran revisados, lo que generó una tendencia a aumentar los puntajes de esos exámenes, y de esa manera sobreestimar el porcentaje de alumnos aprobados.

ENLACE: ANTECEDENTES Y SU IMPLEMENTACIÓN

En este apartado se identifican algunos aspectos característicos de la implementación de

³ Esta ley señala que “Entre más se utilice cualquier indicador social cuantitativo para la toma de decisiones sociales, el mismo estará sujeto a mayores presiones de corrupción y tendrá mayor potencial para distorsionar y corromper los procesos sociales que pretende monitorear” Campbell (1976: 49).

ENLACE y algunas de las reacciones que causó. Mientras que en Europa y Estados Unidos en las últimas cuatro décadas del siglo XX se consolidaron las pruebas masivas estandarizadas para evaluar los resultados educativos, México enfrentaba una situación política y socioeconómica crítica, de forma que el sistema educativo atendió otras problemáticas y prioridades. En el aspecto económico fue un periodo de inestabilidad macroeconómica, con devaluación e inflación, y un elevado endeudamiento externo que culminó con las crisis del periodo 1982-1995 (Hiernaux-Nicolás, 1995). En el aspecto político también hubo inestabilidad; se agotó el modelo de partido hegemónico y se dio paso al proceso de transición a la democracia (Woldenberg, 2012).

En lo social, en esas décadas se verificaron los efectos de la primera parte de la transición demográfica (Partida-Bush, 2005), caracterizada por tasas de mortalidad en rápido descenso y tasas de natalidad relativamente constantes que provocaron un incremento poblacional a tasas muy altas, principalmente en las ciudades, que experimentaron una intensa inmigración desde el campo; por ejemplo, mientras que en 1950 vivían 7.2 millones de personas en localidades urbanas, para 1990 esa cifra alcanzó los 48.7 millones (Anzaldo y Barrón, 2009). Estos fenómenos demográficos generaron una alta presión al sistema educativo del país para expandir la matrícula en todos los niveles escolares.

A la par con lo anterior, el grupo dominante del Sindicato Nacional de Trabajadores de la Educación (SNTE), de ser un importante aliado político del partido en el gobierno, tanto a nivel nacional como estatal, se convirtió en un poderoso grupo con intereses corporativos que logró incidir sobre la política educativa del país en prácticamente todos sus aspectos, no sólo los laborales (Rodríguez, 2015). Su creciente participación en la dirección educativa y en la definición de planes y programas llevó

a que las secciones sindicales del sindicato lograran controlar, y en algunos casos dirigir, la política educativa en algunos estados. La cúspide de este proceso se dio en el gobierno de Felipe Calderón (2006-2012), cuando se designó como subsecretario de Educación Básica a Fernando González Sánchez, yerno de la lidereza del SNTE (Ornelas, 2008).

Estos elementos del contexto social y político de México confluyeron en generar un sistema educativo nacional inercial, centrado en mejorar la cobertura y con fuertes derivas en su operación, mismas que se reflejaban en malos resultados educativos. Por ejemplo, en el año 2000 el promedio de escolaridad de la población de 15 años y más era de apenas 7.5 años; 9.5 por ciento de la población de 15 años y más era analfabeta y 8.7 por ciento de los niños de 6 a 14 años no asistían a la escuela (INEGI, 2001). Un resultado clave, en términos diagnósticos, fue la prueba PISA. Desde su primera aplicación, esta prueba arrojó malos resultados para México: en el año 2000 los estudiantes mexicanos ocuparon las posiciones 34 y 35 de 41 países evaluados en lectura y matemáticas, respectivamente (OCDE, 2003). En 2003 y años posteriores los resultados para México siguieron más o menos en esa tónica.⁴

Las anteriores cifras, aunadas a las prácticas sindicales y de control del SNTE más allá de lo laboral y académico (Muñoz, 2008), sintetizan la mala situación del sistema educativo nacional, por lo menos en su nivel básico, lo que llevó a poner en la agenda nacional la necesidad de cambios y mejoras en dicho sistema. En 2006, a finales del primer sexenio de la alternancia (2000-2006), la SEP aplicó la prueba ENLACE, orientada a evaluar las habilidades en lectura y matemáticas de la totalidad de los estudiantes de tercero a sexto grado de educación primaria y de los tres grados de educación secundaria. En 2008 la prueba se extendió a los estudiantes del último ciclo del bachillerato (SEP, 2014b).

⁴ Ver: <https://data.oecd.org/searchresults/?q=pisa> (consulta: 20 de julio de 2017).

Las cifras de aplicación de la prueba demuestran su masividad: en comprensión lectora, en 2006 se aplicó a 7.4 millones de niños de primaria y a 1.3 millones de estudiantes de secundaria. Estas cifras fueron al alza en los años posteriores, y en 2013 se aplicó a 13.8 millones de niños de esos mismos niveles (SEP, 2014a).

Los resultados de estas pruebas se podían consultar en el sitio web de ENLACE: a nivel de alumno se especificaba su nivel de logro o dominio (insuficiente, elemental, bueno o excelente); y un resumen diagnóstico para informar a los padres y orientarlos sobre en qué cosas debería de trabajar el estudiante para mejorar su desempeño (SEP-DGEP, 2007). Además, se señalaba en qué percentil de desempeño se ubicaba el alumno dentro de su centro escolar, y en el conjunto de escuelas ubicadas en el mismo grado de marginación en el municipio, estado y país. La jerarquización también se extendió a escuelas, municipios y entidades federativas. Adicionalmente, la SEP estableció, como ya se dijo, el Programa de Estímulos a la Calidad Docente, asociado a los resultados obtenidos en ENLACE (SEP, 2013b).

Esta jerarquización —y los estímulos— facilitaron la difusión de los resultados, lo que trajo aparentes ventajas cuando se buscaba movilizar a la sociedad y al sistema educativo en su conjunto. Como se puede constatar a partir del análisis noticioso de esos años, los medios de comunicación empezaron a dar cuenta de los malos resultados educativos, principalmente en matemáticas, y a pedir explicaciones a profesores, directivos escolares y funcionarios públicos en cada sector o escuela; en los siguientes años abundan notas de prensa y declaraciones de funcionarios comunicando los avances de ENLACE.⁵

En ese análisis noticioso las críticas a la prueba ENLACE ocupan un lugar importante.

Como reacción se empezó a instrumentar un enfoque en resultados, en lugar de en procesos; se reportaron escuelas que enseñaban a los alumnos, en el aula, a contestar las pruebas, y que encargaban tareas escolares con reactivos tipo ENLACE: los “jueves de ENLACE”, filtración de sus contenidos, entre otras (Backhoff y Contreras, 2014).

En la cobertura mediática se señalaban fallas en el control de aplicación de la prueba. Y es que la prueba no tenía control externo, por lo que se recurría a personal de las mismas escuelas para administrarlas y, en algunos casos, a padres de familia y personal voluntario (Backhoff y Contreras, 2014).

La reacción política y mediática que se generó con los resultados de ENLACE propició dar explicaciones fáciles, como lo señala Ramírez (2010); una de las más socorridas fue la de poner a los profesores en el centro de las críticas del deterioro del sistema educativo, y culparlos de los deficientes resultados de los alumnos. Sin embargo, se olvidaba que ese magisterio era parte de los resultados de las propias políticas educativas que lo habían formado y de las ataduras sindicales que los habían subordinado. Ante esto, una reacción de un sector del profesorado, principalmente el encabezado por la disidencia magisterial, fue negar la legitimidad de los resultados de la prueba ENLACE, quemar las pruebas e incluso bloquear su aplicación en los territorios en los que mantenían control (Hernández, 2012; Chacón, 2010).

Lo cierto es que, a pesar de las críticas, la SEP profundizó y amplió la aplicación de la prueba ENLACE, y sus resultados se empezaron a utilizar como referencias diagnósticas en la planeación educativa, para asignar estímulos a los profesores (SEP, 2013b) y para la toma de decisiones. Esto se puede constatar en la prensa y en los planes estatales de las entidades federativas.⁶ Hubo entonces

⁵ Esto se puede constatar consultando documentos y noticias en la web cuyo título incluía las palabras “Prueba Enlace”. Esta búsqueda se hizo el 15 de octubre de 2019 en Google Chrome, indicando *allintitle*: “Prueba Enlace”, y arrojó un total de 10 mil 300 resultados. Cuando se usó “Prueba PISA”, arrojó 3 mil 660 resultados, y con “Prueba PLANEA” se obtuvieron 3 mil documentos.

⁶ En 2015 se analizaron los planes estatales de desarrollo y, cuando estuvieron disponibles, los programas sectoriales de educación de las 32 entidades federativas de México. Se encontró que en 26 entidades se hacía referencia a datos

reacciones y acciones encaminadas a mejorar la capacitación docente y los procesos de enseñanza, sobre todo en matemáticas, que fue donde los resultados iniciales de ENLACE eran claramente más bajos.

La masividad de ENLACE proporcionó gran cantidad de información, no sólo la relacionada con los resultados en matemáticas y lectura, sino también con factores que tienen un impacto importante en el aprendizaje, como la institución, la escuela, el hogar y el individuo, lo que permitió varios estudios (Martínez *et al.*, 2013; Hoyos *et al.*, 2012) sobre la mejora de la calidad educativa.

Con la prueba ENLACE, en México se logró afianzar la evaluación e involucrar a diferentes actores del sistema educativo. Año tras año la SEP y los gobiernos estatales empezaron a reportar mejoras en los resultados educativos y en la disminución de la problemática, lo que hacía dudar más de la prueba ENLACE a sectores críticos e informados, aunque no siempre respaldados en análisis detallados.

TENDENCIAS Y CONTRADICCIONES EN LOS RESULTADOS DE LA PRUEBA ENLACE

Esta sección se divide en cuatro subsecciones: en la primera se analizan los valores promedio de las dos escalas con los que reportaron los resultados de dicha prueba, considerando los tres niveles educativos y las dos habilidades. Se ve que la mayor contradicción se localiza en EMS, por eso ahí se focalizó un análisis detallado de 5.6 millones de registros de los puntajes obtenidos de la prueba ENLACE para el periodo 2008-2014 para matemáticas (segunda subsección) y lectura (tercera). En la cuarta subsección se analiza la relación entre ambas escalas de medición.

Tendencias generales de los resultados

En la Tabla 1 se muestran los dos indicadores con los que se reportaron los resultados de la prueba ENLACE: puntaje promedio y porcentaje de estudiantes por nivel de logro o dominio (insuficiente, elemental, bueno y excelente) para el primero y el último año en que se aplicó la prueba en cada nivel. Para cada indicador se calculó su cambio relativo en términos porcentuales durante el periodo considerado.⁷ Se observó que los puntajes promedio en matemáticas tuvieron incrementos relativos de 16.7, 7.3 y 14.4 por ciento en primaria, secundaria y EMS, respectivamente. Conviene comparar estos incrementos, particularmente los de secundaria y EMS, con el cambio en el puntaje promedio de los estudiantes mexicanos en la prueba PISA en matemáticas, en la que creció 1.7 por ciento (de 406 a 413) entre 2006 y 2012.⁸ Los puntajes promedio de ENLACE, como puede verse, tuvieron incrementos bastante más altos, e implicarían una mejora educativa importante en un periodo corto, lo cual está prácticamente descartado, especialmente a la luz de otras investigaciones que han establecido la complejidad y lo difícil que es mejorar los aprendizajes (Martínez, 2018). Más bien estos incrementos en matemáticas representan algún tipo de anomalía, inflación o distorsión.

En cuanto a lectura, se ve que hay pequeños decrementos en los puntajes promedio en secundaria (-1.1 por ciento) y EMS (-3.7 por ciento); mientras que en primaria se registró un aumento de 10.1 por ciento (Tabla 1). En el caso de PISA, en lectura, el puntaje promedio entre 2006 y 2012 creció 3.4 por ciento (de 410 a 424), lo que implicaría que en ENLACE en secundaria y EMS no hubo mejora en lectura y que los fenómenos de distorsión asociados a las fallas de control de dicha prueba no se reflejaron con un incremento en los puntajes promedio en esa habilidad.

de ENLACE en el diagnóstico o para algún indicador de los planes; las excepciones fueron: Colima, Morelos, Puebla, Tabasco, Tlaxcala y Zacatecas.

⁷ Es decir $(X_{\text{final}}/X_{\text{inicial}} - 1) \times 100$.

⁸ Los datos de la prueba PISA se obtuvieron de: <https://data.oecd.org/searchresults/?q=pisa> en julio de 2017.

Tabla 1. Cambios relativos en los puntajes promedio y en el nivel de logro en lectura y matemáticas en la prueba ENLACE para los tres niveles educativos.

Nivel educativo	Asignatura	Año	Porcentaje de alumnos por nivel de logro o dominio				Logro bajo*	Logro alto*		Puntaje promedio	
			Insuficiente	Elemental	Bueno	Excelente		Valor	Cambio relativo al final del periodo	Valor	Cambio relativo al final del periodo
Primaria	Lectura	2006	20.7	58.1	19.6	1.7	78.7	21.3	101.2	500.0	10.1
		2013	13.5	43.7	33.4	9.4	57.2	42.8		550.7	
	Matemáticas	2006	21.0	61.4	16.0	1.6	82.4	17.6	176.7	500.0	16.7
		2013	12.4	38.8	29.0	19.8	51.2	48.8		583.5	
Secundaria	Lectura	2006	40.7	44.6	14.0	0.7	85.3	14.7	33.7	500.0	-1.1
		2013	37.4	42.9	18.0	1.7	80.3	19.7		494.5	
	Matemáticas	2006	61.1	34.7	3.8	0.4	95.8	4.2	422.1	500.0	7.3
		2013	46.4	31.6	14.9	7.1	78.1	21.9		536.3	
Media superior	Lectura	2008	12.4	35.3	45.5	6.7	47.7	52.3	-14.5	512.4	-3.7
		2014	19.1	36.3	40.1	4.6	55.3	44.7		493.2	
	Matemáticas	2008	46.5	37.8	12.2	3.4	84.4	15.6	151.9	509.9	14.4
		2014	26.6	26.6	26.6	26.6	60.7	39.3		583.3	

*Logro alto es la suma de los porcentajes de alumnos en el nivel de logro bueno y excelente; y logro bajo es la suma de insuficiente y elemental.

Fuente: elaboración propia a partir de datos proporcionados por la SEP para EMS y de los resultados publicados en shorturl.at/c1NQ2 y shorturl.at/noxD6 (consulta: 27 de agosto de 2017).

Por su parte, los cambios en el otro indicador contrastan considerablemente con los cambios anteriores en los puntajes de ENLACE.⁹ Los cambios relativos del porcentaje de alumnos que alcanzaron niveles de logro alto en matemáticas fueron de 176.7, 422.1 y 151.9 por ciento en primaria, secundaria y EMS, respectivamente. Mientras que, en lectura, en ese mismo orden por nivel educativo, los cambios fueron de 101.2, 33.7 y -14.5 por ciento. Los cinco crecimientos relativos, que van de 33.7 a 422.1 por ciento, son demasiado grandes para pretender que efectivamente se hayan dado en la realidad, ya que eso implicaría mejoras radicales en el sistema educativo, cosa que no se puede dar en la realidad en periodos cortos (6 y 7 años), y en un problema tan complejo.

Este gran incremento o inflación de los resultados no puede deberse sólo al desgaste por el uso de la prueba y a ciertas reacciones locales a la misma, como ha ocurrido con otras pruebas (Koretz, 2005) ya que el periodo es corto, y el cambio muy grande. Esto hace suponer que hay otros factores específicos que son responsables del crecimiento desmedido de los indicadores.

Análisis de los puntajes de matemáticas en ENLACE en EMS

Para profundizar en el anterior planteamiento nos propusimos analizar los puntajes directos de ENLACE, y no sólo sus promedios, que es lo que contiene la Tabla 1, en el entendido de que ahí se podrían encontrar evidencias

⁹ Se debe tener precaución al comparar los cambios de ambos indicadores, dada su diferencia en magnitud. Para atenuar un poco esto y comparar mejor los cambios, los cuatro niveles de logro se resumen en dos: bajo y alto (Tabla 1).

específicas de cómo se generaron los grandes incrementos expuestos en dicha tabla. Se decidió concentrar el análisis en EMS debido que en ese nivel el sentido del cambio en matemáticas y lectura es opuesto.

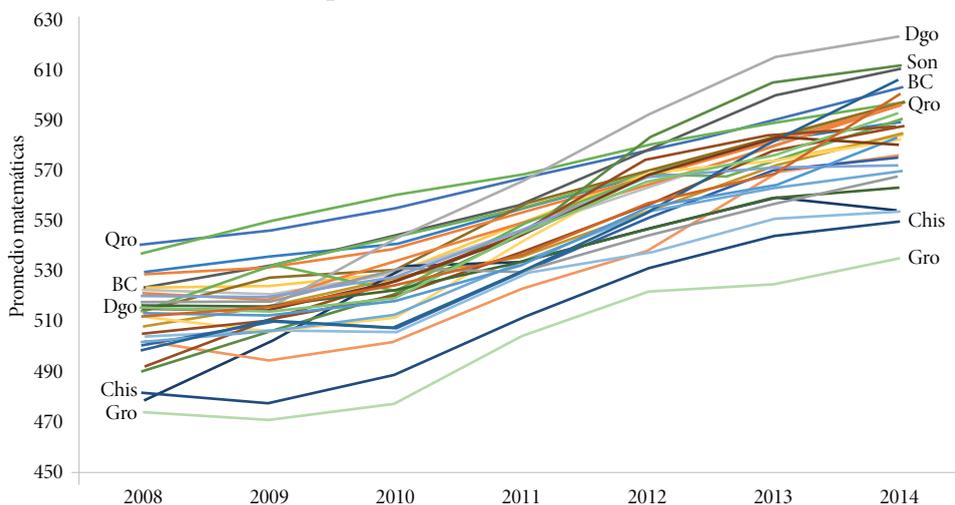
Los puntajes por alumno no se encuentran publicados en el sitio web de ENLACE. Ante ello se solicitaron por transparencia a la SEP, y ésta respondió que no contaba con dicha información. Esto se impugnó ante el Instituto Nacional de Transparencia (INAI), que obligó a la SEP a entregar la información (“INAI pide a SEP entregar puntaje...”, 2017). La SEP entregó los resultados por alumno de EMS en una escala estandarizada, lo que hace que la gran mayoría de los valores vayan de -3 a 3. Se deja asentada la duda sobre las razones de la negación inicial y sobre por qué no se entregó la información en la escala original.

La base de datos proporcionada por la SEP tiene 6.4 millones de registros de resultados de exámenes de EMS para los siete años, con 808 mil 346 para 2008; y cierra con 1 millón 027 mil 944 registros en 2014. Esta diferencia se debe a que la cobertura de la prueba fue aumentando

durante el periodo. Para analizar estos registros, con la intención de que las tendencias fueran más comprensibles, los datos estandarizados se recalcularon a la escala original.¹⁰

En la Gráfica 1 se presenta la evolución en los 32 estados del país de los puntajes promedio de los alumnos en matemáticas. En todos los casos se aprecia una clara tendencia ascendente. Para confirmarla se ajustó una línea recta con los promedios de cada estado como variable dependiente, y al año de aplicación de la prueba como variable independiente. En los 32 casos la línea recta tuvo un muy buen ajuste, acorde al coeficiente de determinación R^2 , cuyos valores estuvieron entre 89.3 y 99.6 por ciento. Dada la interpretación de este coeficiente como el porcentaje de la variación de la variable dependiente que es explicada por el modelo (línea recta en este caso), se confirma la consistente y fuerte tendencia lineal ascendente de los puntajes promedio en matemáticas en las 32 entidades. Lo anterior es una evidencia en contra de una inflación asociada sólo a la reacción de las escuelas a nivel local, porque sería extraño que se haya dado desde

Gráfica 1. Puntaje promedio en matemáticas de la prueba ENLACE en EMS para los estados de México de 2008 a 2014



Fuente: elaboración propia con base en los datos correspondientes proporcionados por la SEP.

¹⁰ Con la transformación $X_i = 500 + 100 Z_i$, donde Z_i son los valores estandarizados. Para validar esto se analizó la relación entre la media de los datos estandarizados para cada año con los puntajes promedio en la escala original que estaban reportados en Backhoff y Contreras (2014); y se encontró una relación casi perfecta con una correlación de 0.9999.

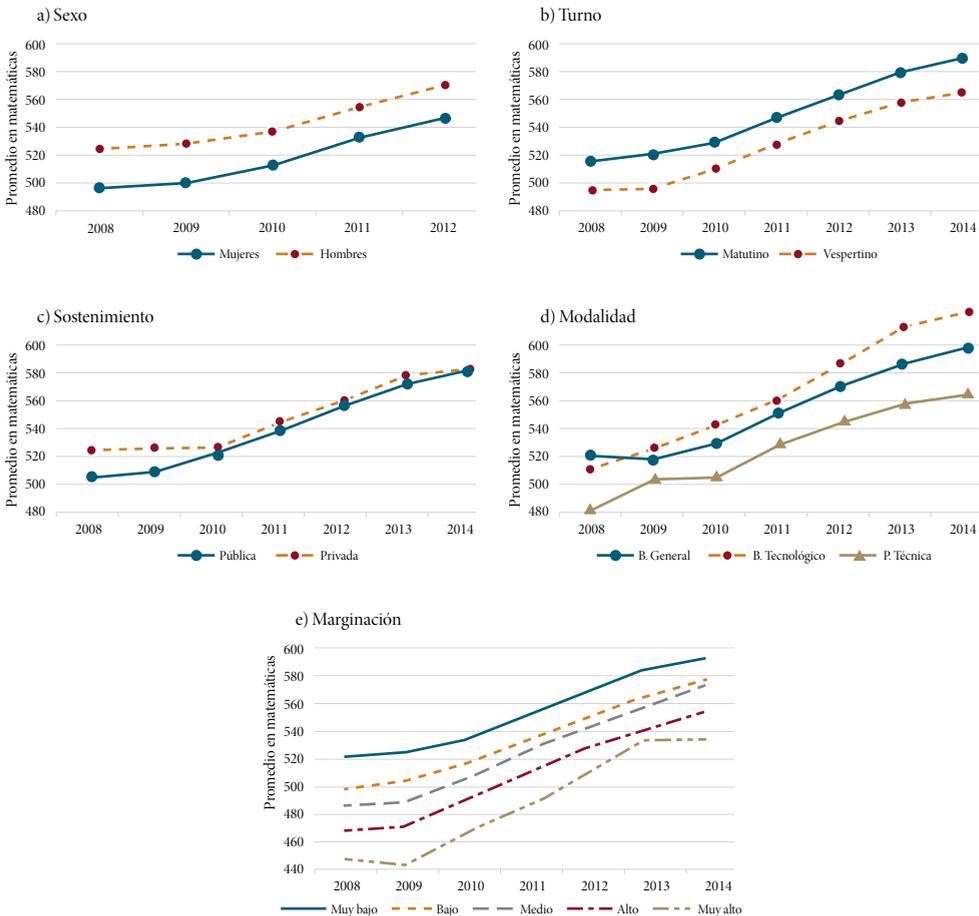
el primer año, todos los años y de manera general en las 32 entidades. Las diferencias en la magnitud de las tendencias que, como se ve en la Gráfica 1, sí las hay, pudiera ser una evidencia de las diferentes formas en que a nivel estatal se reaccionó en EMS ante ENLACE.

Para extender el análisis anterior y ver qué tan generalizado fue el incremento de los puntajes los resultados se desagregaron respecto de otros factores demográficos y del contexto escolar: sexo de los alumnos, turno del centro escolar, tipo de sostenimiento, modalidad de bachillerato y grado de marginación de la localidad donde se ubica la escuela (Gráfica 2). De ahí se confirma la generalidad que tuvo la

tendencia ascendente en los puntajes promedio de matemáticas, ya que ésta se conserva de manera similar para todos los niveles de dichos factores.

De esta manera, en el periodo 2008-2014 los puntajes promedio en matemáticas para EMS tuvieron un incremento de 14.4 por ciento, al pasar de 509.9 a 583.3, con un ritmo medio de 12.3 puntos por año. Esta mejora, que como puede verse en las Gráficas 1 y 2, es paulatina y generalizada —siguiendo la terminología establecida a nivel internacional (Koretz, 2005), en Backhoff y Contreras (2014)— se atribuye a la inflación de los resultados, y se asocia a tres causas: publicación de *rankings* de escuelas,

Gráfica 2. Puntaje promedio en matemáticas de la prueba ENLACE en EMS por diferentes factores sociodemográficos y del contexto escolar



Fuente: elaboración propia con base en los datos correspondientes proporcionados por la SEP.

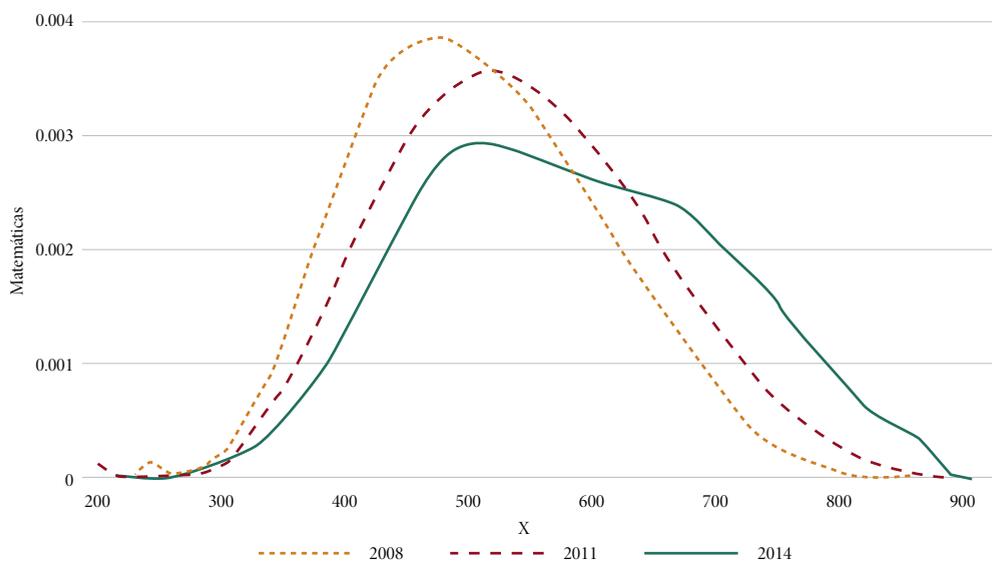
la asociación de estímulos económicos a los docentes y la falta de control externo de su aplicación. Pero creemos que, en un universo de diez mil preparatorias muy diversas, estas causas no se podían reflejar en forma tan homogénea y general en los resultados. No dudamos de que dichas causas estén presentes, pero por su naturaleza no podrían haberse reflejado en forma tan paulatina y generalizada. Planteamos que, dada la generalidad y constancia del cambio en matemáticas, en realidad su principal motor se dio por ajustes a la prueba o a sus resultados, que de forma intencionada o no llevaron a generar una tendencia de mejora artificial.

Para profundizar en esta hipótesis, en la Gráfica 3 se muestra la estimación de la densidad de probabilidades para los puntajes de los años 2008, 2011 y 2014 en matemáticas.¹¹ Se destaca que las curvas en su conjunto se van

desplazando paulatinamente a la derecha y con un fuerte sesgo la de 2014. El desplazamiento implica que todos los jóvenes de EMS empezaron a obtener mejores resultados en matemáticas año con año. Esto ya se había visto para el promedio de los puntajes por estado y por diferentes niveles de desagregación (Gráficas 1 y 2), pero el hecho de que el incremento sea para toda la distribución poblacional, y no sólo para el promedio, aporta información adicional para cuestionar buena parte de ese gran y consistente aumento de los resultados en matemáticas a través de los años.¹²

Por su parte, el fuerte sesgo y la forma de la distribución de los puntajes altos de matemáticas para el año 2014 (Gráfica 3) es claramente una anomalía que debe de tener otro origen, y se plantea como hipótesis que esto se debe a los problemas de control y fraudes específicos en la prueba. Del lado derecho de la curva de

Gráfica 3. Distribuciones de probabilidad para los puntajes de los alumnos en matemáticas para los años 2008, 2011 y 2014 en EMS de ENLACE



Fuente: elaboración propia con base en el procedimiento *density* del software R a partir de los datos correspondientes proporcionados por la SEP.

11 Se aplicó el procedimiento *density* del software R (<https://cran.r-project.org/>), que estima en forma no paramétrica la función de densidad de probabilidades de una variable a partir de un conjunto de datos (consulta: 12 de septiembre de 2019).

12 Por ejemplo, el primer decil en 2008 fue igual a 386.5 puntos, es decir, 10 por ciento de los alumnos evaluados tuvieron puntajes menores a 386.5; mientras que, para 2014, el mismo decil aumentó y llegó a 426.3 puntos, que significa un incremento de 10.3 por ciento en los puntajes de los estudiantes con peores resultados.

2014 se ve que en lugar de que ésta caiga en forma suave, tiene una alta densidad desde poco antes de 500 y hasta antes de 700. A partir de este valor la curva cae más rápido, conservando valores altos, y cae en forma brusca poco antes de 900. Esta distribución de los puntajes de 2014 es atípica, sobre todo si se contrasta con las curvas de los primeros años. Es fácil atribuir esto a un considerable aumento de la proporción de alumnos que obtuvieron resultados altos,¹³ incluida una gran cantidad con resultados perfectos con más de 800 puntos en matemáticas.

En efecto, el porcentaje de alumnos en EMS con más de 800 puntos en la prueba ENLACE en matemáticas para los años 2008, 2013 y 2014 fue de 0.17, 2.9 y 4.4 por ciento, respectivamente. Al convertir estos porcentajes de 2013 y 2014 a cantidades se obtiene que 75 mil jóvenes en EMS alcanzaron resultados perfectos en matemáticas (más de 800 puntos) en esos años. De aquí que, o se multiplicaron los jóvenes genios matemáticos o hubo problemas de control e incluso fraudes específicos que propiciaron una buena parte de estas cifras, o anomalías, como robo y filtración del contenido de la prueba, copia, etc., de lo que se dio cuenta sobre todo en los medios de comunicación (ver nota 7), pero que no se había evidenciado con un análisis como éste.

Como ocurre en todo proceso, los problemas de control se reflejan en un aumento de la variación, y éste ha sido el caso en matemáticas en EMS: la desviación estándar de los resultados se incrementó 24.3 por ciento en el periodo, al pasar de 99.1 en 2008 a 123.4 en 2014, lo que es una evidencia adicional de la presencia de fraudes y fallas de control de la aplicación de la prueba ENLACE; esta tendencia se fue agudizando al pasar de los años, y favoreció artificialmente puntajes altos y perfectos en cierta cantidad de escuelas de EMS del país. Suponemos que la

identificación de dichas escuelas es posible con los datos disponibles, aspecto que queda pendiente para posteriores investigaciones.

En otras partes del mundo el análisis de los resultados individuales de las pruebas ha sido la forma de detectar problemas en escuelas o regiones específicas (ver, por ejemplo, Battistin, 2016; Angrist *et al.*, 2015; Goldstein, 2011; Cizek, 2001). Pero en ninguno de estos casos, o de otros reportados, se ha encontrado una tendencia tan sistemática y homogénea de mejora en los diferentes niveles de desagregación.

Por otro lado, en la base de datos proporcionada por la SEP hay registros que tienen como observación que el alumno copió. Sin embargo, esos registros no fueron excluidos del reporte de resultados de la SEP, aspecto que Calderón (2015) también señala. Caso contrario ocurrió en aquéllos en los que los alumnos respondieron menos de 50 por ciento de las preguntas de la prueba, porque en los casos de previsibles puntajes bajos no se consideraron los resultados de los alumnos para reportar los niveles de logro.¹⁴

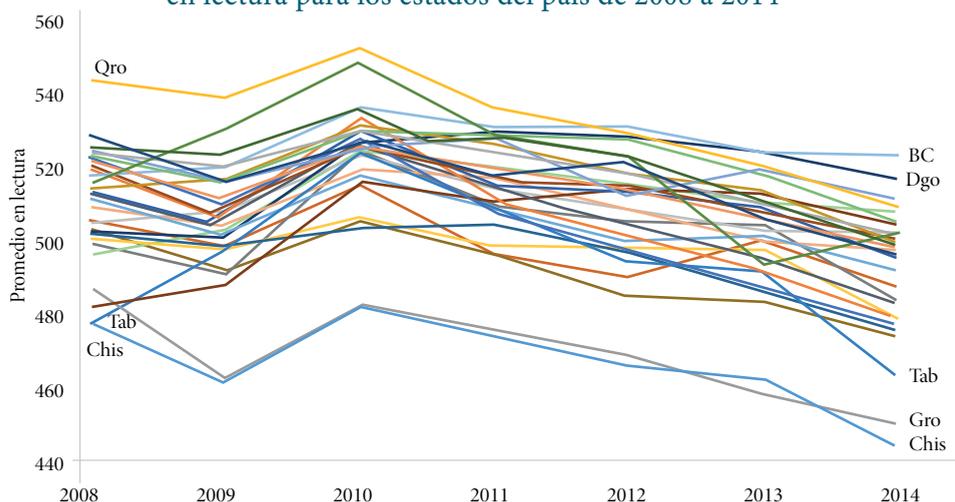
Análisis de los puntajes de lectura en ENLACE en EMS

Es clave contrastar la evidencia encontrada en matemáticas con lo que pasó en lectura en EMS, porque este caso se sale del patrón de los otros cinco de la Tabla 1. Las Gráficas 4 a 6 son las equivalentes a las previas, pero ahora para lectura. Como puede verse, aquí no se da una evolución tan homogénea entre las diferentes variables del contexto escolar, incluidos los promedios de las 32 entidades federativas. El patrón que se reproduce más en esta asignatura es que en los primeros tres años de la prueba hay pocos cambios, y a partir del 2010 en la mayoría de los casos hay un ligero descenso, con algunas excepciones. Este descenso implica una disminución de -3.7 por ciento en los

¹³ Por ejemplo, el valor del percentil 75 de los resultados en matemáticas en 2008 fue de 578.2 puntos, y para 2014 subió a 675.3; un sorprendente incremento de casi 100 puntos. Y de 16.8 por ciento en el punto de corte a partir del cual se ubicaba el 25 por ciento de los alumnos con calificaciones más altas.

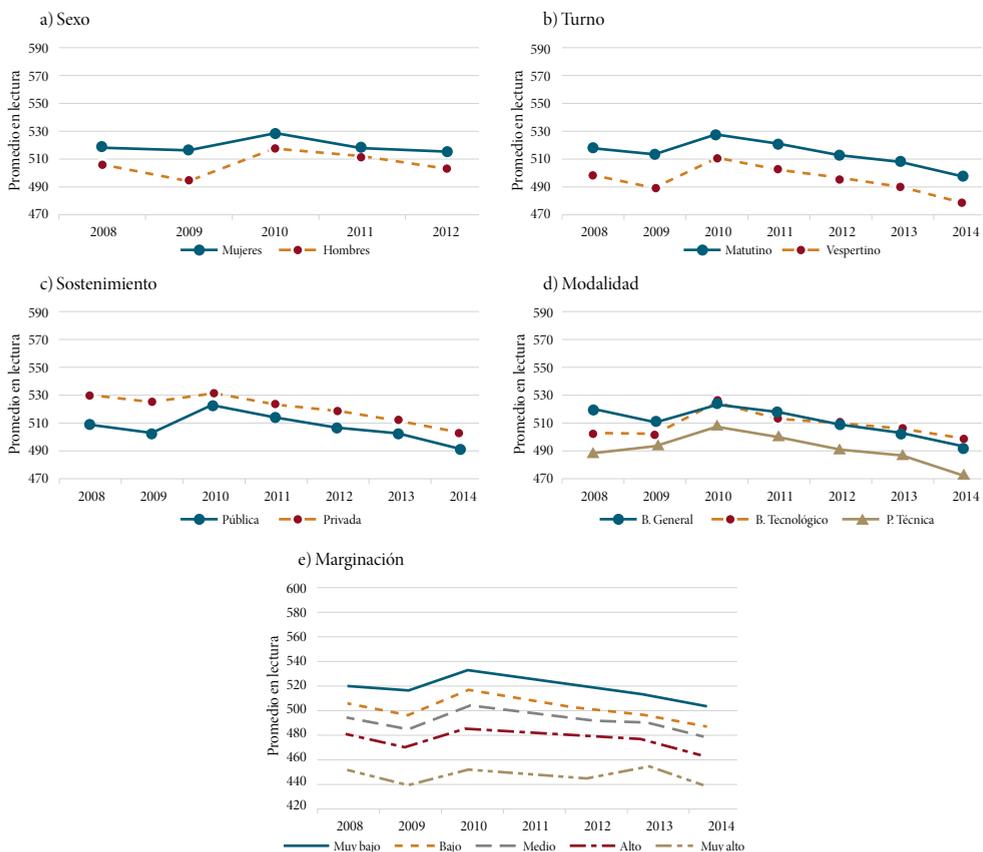
¹⁴ Ver: http://www.enlace.sep.gob.mx/ms/estadisticas_de_resultados/ (consulta: 12 de octubre de 2019).

Gráfica 4. Puntaje promedio en la prueba ENLACE en EMS en lectura para los estados del país de 2008 a 2014



Fuente: elaboración propia con base en los datos correspondientes proporcionados por la SEP.

Gráfica 5. Puntaje promedio en lectura de la prueba ENLACE en EMS por diferentes factores sociodemográficos y del contexto escolar



Fuente: elaboración propia con base en los datos correspondientes proporcionados por la SEP.

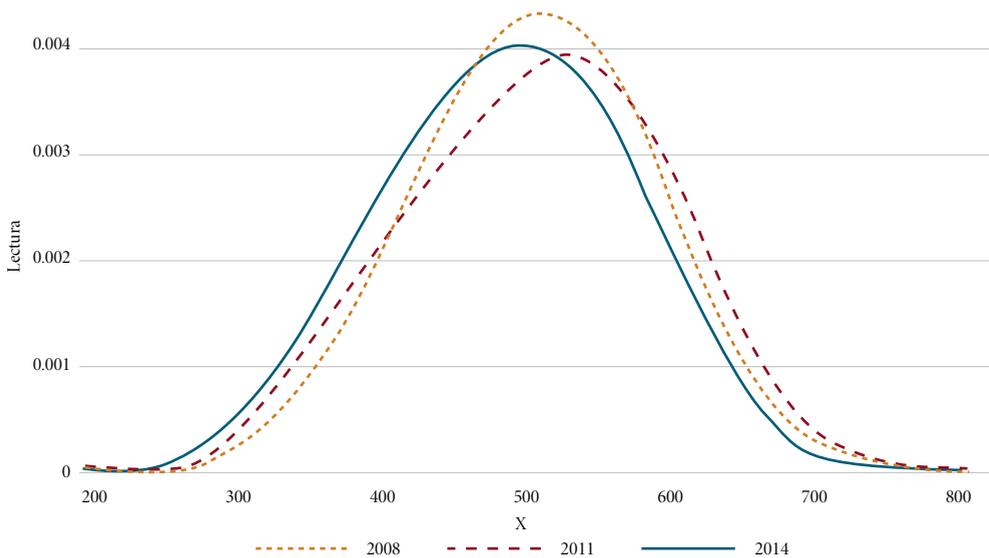
puntajes promedio, y de -14.5 por ciento en el porcentaje de alumnos con nivel de dominio alto en lectura (Tabla 1).

De esta manera, en lectura las fallas de control asociadas a ENLACE, que ya se han comentado antes, no se reflejaron con una inflación de los resultados a los se refiere Koretz (2005), ni la ley de Campbell (ver nota 3).

En los contrastes entre las tendencias en matemáticas y lectura en EMS, es de destacar también el caso de las distribuciones de

probabilidad (Gráficas 3 y 6). En lectura se ven distribuciones relativamente simétricas y acampanadas, parecidas a una distribución normal, con tendencia central y variabilidad (amplitud) similares para los diferentes años. Esto ratifica el hecho de que, en la prueba de lectura, con el análisis hecho, no se detectan problemas de control ni desgaste del uso de la medida. Tampoco la sospecha de una intervención en el diseño o evaluación de la prueba para favorecer puntajes más altos.

Gráfica 6. Estimación de la distribución de probabilidad para los puntajes de los alumnos en los años 2008, 2011 y 2014 para lectura en EMS de ENLACE



Fuente: elaboración propia con base en el procedimiento *density* del software R a partir de los datos correspondientes proporcionados por la SEP.

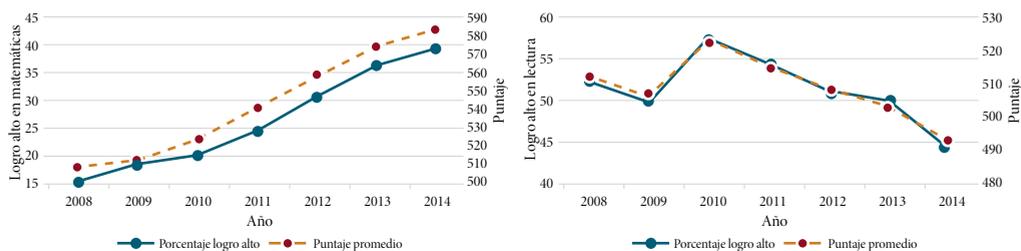
Relación entre las dos escalas de medición de ENLACE en EMS

Por otra parte, un aspecto clave en esta investigación fue averiguar el tipo de relación entre los dos indicadores para reportar los resultados de la prueba ENLACE. Los manuales de la prueba explican que éstos se originan por dos escalas diferentes de medición. En el *Manual Técnico 2013* de educación básica se señala que ENLACE es una evaluación censal que mide habilidades y competencias con dos escalas, una de ellas con valores que van de 200 a 800, donde los resultados tienen comparabilidad

en el tiempo. Además, afirman que esa comparabilidad es una de las características centrales del modelo de evaluación de la prueba, y que es objetiva para reportar los resultados (SEP-DGEP, 2013). Sobre la otra escala señalan que es subyacente respecto a los constructos:

...con la cual se puede asignar a los alumnos a niveles de logro y potenciar la comparabilidad entre años distintos. Esta escala subyacente está basada en la Teoría de Respuesta al Ítem; en ella la puntuación no depende sólo del número de respuestas correctas, sino de cuáles reactivos

Gráfica 7. Relación entre el promedio de los puntajes y el porcentaje de alumnos con logro alto en matemáticas y lectura: 2008 a 2014



Fuente: elaboración propia con datos de la Tabla 1.

se respondieron correctamente (SEP-DGEP, 2013: 71) (subrayado propio).

En el *Manual técnico ENLACE media superior 2008-2010* se describen los niveles de dominio y la forma de nombrarlos: se señala que el nivel de dominio no se determina por el número de reactivos que están contenidos en los conjuntos correspondientes a cada nivel, sino que “un ítem o tarea actúa como punto de inflexión entre dos niveles de dominio. El cambio de un nivel a otro se produce cuando se identifica un reactivo que un estudiante promedio del nivel actual no puede responder” (Ceneval, 2013: 54-58).

De acuerdo con lo que dicen los manuales se esperaría que ninguna de las escalas tuviera una relación directa, y que en cierto sentido fueran complementarias. Sin embargo, éste no fue el caso. En la Gráfica 7 se muestra la relación entre el puntaje promedio y el porcentaje de alumnos con nivel de logro alto. En el caso de matemáticas ambos indicadores tienen una tendencia ascendente prácticamente idéntica (correlación de 0.995); y en lectura, los indicadores correspondientes también tienen un patrón idéntico entre sí (correlación de 0.994). Esto lleva a sospechar que los porcentajes por nivel de logro en EMS se obtuvieron directamente de los puntajes.

Lo anterior se confirmó con el análisis de los puntajes de los 6.4 millones de registros

de resultados de exámenes de EMS, a partir de los cuales se obtuvo la coincidencia en los puntos de corte de los puntajes (X) que generan los porcentajes por nivel de dominio que se muestran en la Tabla 1 para los años 2008 y 2014 para ambas habilidades. Para el resto de los años se pueden consultar en la página web de la prueba para EMS.¹⁵

De esta manera, se confirma que por lo menos en EMS los niveles de logro se obtuvieron directamente de los puntajes. Esto es grave, porque con esto en la práctica desaparece una de las escalas de medición establecida en los manuales de la prueba (el nivel de logro), y que fue precisamente con la que se publicitaron las mejoras en los resultados de ENLACE.

Lo anterior es una evidencia adicional para respaldar la sospecha (hipótesis) de que se tomaron decisiones sobre la estructura de ENLACE en matemáticas en EMS o en sus resultados que propiciaron los incrementos comentados. Y también para sospechar que algo parecido se hizo en los niveles básicos, donde también se observan los incrementos espectaculares referidos en la Tabla 1.

El sesgo y el desplazamiento a la derecha de las distribuciones de los puntajes en matemáticas (Gráfica 3), y el hecho de que los niveles de logro se obtuvieran de dichos puntajes, explican en buena medida la falacia del gran incremento de los porcentajes de jóvenes que alcanzaron logros altos en matemáticas.

¹⁵ Ver shorturl.at/fwFL1 (consulta: 12 de octubre de 2019).

CONCLUSIONES

Con la prueba ENLACE se logró afianzar en México la evaluación del sistema educativo mediante pruebas estandarizadas, pero llegó con tres décadas de retraso respecto de Estados Unidos y de varios países de Europa. En este retraso influyó la situación política y socioeconómica crítica que enfrentó el país en las últimas tres décadas del siglo XX, incluido el fortalecimiento del poder y control del SNTE sobre la política educativa.

Entre las características distintivas de la prueba ENLACE destacan su carácter censal y la manera de difundir los resultados, mediante la jerarquización de niños, escuelas, municipios y estados, además de usar sus resultados para incentivar económicamente a profesores de escuelas públicas. Estas características lograron un rápido conocimiento de la prueba y un cierto involucramiento de diferentes actores del proceso educativo para tratar de mejorar los resultados; pero a la vez generaron rechazos, y favorecieron fallas de control, fraudes incluidos. Esto también ha ocurrido con pruebas similares en otras partes del mundo.

Del análisis de los resultados de las dos escalas utilizadas en la prueba ENLACE se encontró que en EMS los niveles de dominio se obtuvieron directamente de los puntajes, lo cual contradice lo que dicen los manuales técnicos de la prueba sobre la forma de obtener los niveles de dominio.

Con la escala de puntajes se reportan incrementos en los promedios en matemáticas en los tres niveles educativos y en lectura en primaria, que se pueden considerar anormalmente altos, mientras que, con la otra escala, la del nivel de dominio o logro, los incrementos implican mejoras relativas de 101.2 a 422.1 por ciento (Tabla 1). Evidentemente que esos niveles de mejoras no se pudieron dar en la realidad, como da cuenta la prueba PISA para los jóvenes de 15 años evaluados en el periodo.

El que la SEP no tenga publicados los resultados individuales de ENLACE no abona a la

transparencia y rendición de cuentas a las que están obligados. Esto quedó más en evidencia cuando negó la existencia de los resultados individuales. Sólo mediante la resolución del INAI entregó los valores; y gracias a ello fue posible analizar los 6.4 millones de registros para EMS, de donde se pudo ver que, de manera sospechosa, los incrementos en el puntaje de matemáticas siguieron una tendencia lineal ascendente (clara, paulatina y consistente), y que esto ocurrió de forma similar con los promedios para los 32 estados del país y en todos los niveles del entorno escolar en los que fue posible desagregar los resultados.

Esta sospecha se fortaleció al ver que toda la distribución de los puntajes en matemáticas se fue desplazando en forma paulatina para tener mejores resultados. La generalidad y sistematicidad de ese cambio no puede deberse sólo a reacciones a nivel local a la prueba y a desgastes en su uso; más bien se ha encontrado evidencia para respaldar la hipótesis de que, en parte, se debió a ajustes en la prueba o en sus resultados que generó una mejora artificial en los puntajes de matemáticas. Esto toma fuerza a partir de todos los análisis hechos, y las evidencias de la falta de rigor en la gestión de ENLACE: no aplicar lo que dicen los manuales sobre la escala principal de la prueba, el incremento en los efectos de las fallas de control, fraudes incluidos; no eliminar para el reporte de los resultados los registros sospechosos de copia y si excluir los exámenes donde los alumnos contestaron menos del 50 por ciento de los ítems de la prueba.

Del análisis de distribución de los puntajes en matemáticas de EMS para los años 2008 a 2014 se identificó un fuerte sesgo en la distribución de los puntajes altos de los años finales del periodo (2013 y 2014). Esto es evidencia de las fallas de control y fraudes en ciertas escuelas que favorecieron que más alumnos obtuvieran puntajes altos, incluidas 75 mil pruebas con resultados más que perfectos en 2013 y 2014 (más de 800 puntos).

Ambas irregularidades, y la forma de determinar los niveles de dominio, se tradujeron en el falaz aumento en la proporción de estudiantes que alcanzaron niveles de dominio alto en matemáticas en EMS, que fue lo que publicitó la SEP en esos años. Se deja abierta la hipótesis, a verificar con evidencias, de que estas anormalidades también se dieron en matemáticas en los otros niveles educativos, y quizás en lectura en el nivel primaria.

En el caso de lectura en EMS no se encontró evidencia de dichas irregularidades. Una posible razón, que se deja como hipótesis, consiste en que para las autoridades educativas en el EMS la prioridad fue matemáticas, y que esta decisión se tomó a partir de los resultados iniciales de ENLACE, donde se plasma una situación de desastre en esa disciplina, y no tanto en lectura. En la primera habilidad, 46.5 por ciento de los jóvenes evaluados obtuvo un nivel de logro deficiente en 2008, mientras

que en lectura sólo fue el 12.4 por ciento. Esto explica la razón de movilizar voluntades para atender la problemática en matemáticas, con los resultados e irregularidades ya expuestas. En cambio, en lectura, a juzgar por los resultados, no se dieron cambios significativos en la realidad, ni en el diseño o aplicación de la prueba ENLACE, por lo que en esa disciplina prevaleció la inercia que ha venido arrastrando el sistema educativo nacional; a ello se suma el impacto de los cambios globales de la sociedad de la información, los que, al parecer, afectarán negativamente las habilidades de comprensión lectora.

Sigue pendiente analizar los puntajes individuales de las pruebas de ENLACE en los dos niveles básicos, pues la Tabla 1 da pautas para esperar irregularidades similares a las encontradas en EMS. También queda pendiente identificar las escuelas y regiones donde estos problemas fueron más patentes.

REFERENCIAS

- ANGRIST, Joshua D., Erich Battistin y Daniela Vuri (2017), "In a Small Moment: Class size and moral hazard in the Italian Mezzogiorno", *American Economic Journal: Applied Economics*, vol. 9, núm. 4, pp. 216-49.
- ANZALDO, Carlos y Eric Alan Barrón (2009), "La transición urbana de México, 1900-2005", *La situación demográfica de México*, México, Consejo Nacional de Población, pp. 53-65.
- BACKHOFF, Eduardo y Sofía Contreras Roldán (2014), "'Corrupción de la medida' e inflación de los resultados de ENLACE", *Revista Mexicana de Investigación Educativa*, vol. 19, núm. 63, pp. 1267-1283.
- BATTISTIN, Erich (2016), "How Manipulating Test Scores Affects School Accountability and Student Achievement", *IZA World of Labor*, núm. 295, pp. 1-11.
- BEATON, Albert E., Alfred M. Rogers, Eugenio González, Mary Beth Hanly, Andrew Kolstad, Keith Rust, Emmanuel Sikali, Lynne Stokes y Yue Jia (2011), *The NAEP Primer (NCES 2011-463)*, Washington, DC., U.S. Department of Education-National Center for Education Statistics.
- CALDERÓN, David (2015), "Evaluación, validez y relevancia: un comentario crítico a 'Corrupción de la medida de inflación de los resultados de ENLACE', de Eduardo Backhoff y Sofía Contreras Roldán", *Revista Mexicana de Investigación Educativa*, vol. 20, núm. 64, pp. 295-308.
- CAMPBELL, Donald T. (1976), "Assessing the Impact of Planned Social Change", *Occasional Paper Series*, núm. 8, pp. 1-74.
- CAMPOS Vázquez, Raymundo y Freddy Urbina Romero (2011), "Desempeño educativo en México: la prueba ENLACE", *Estudios Económicos*, vol. 26, núm. 2, pp. 249-292.
- CAMPOS Vázquez, Raymundo M. y Alma S. Santillán Hernández (2016), "Análisis de diferencias en puntajes en la prueba ENLACE entre niños y niñas en el sistema escolar mexicano", *Estudios Económicos*, vol. 31, núm. 1, pp. 65-123.
- CENTRO Nacional de Evaluación para la Educación Superior (CENEVAL) (2013), *Manual técnico ENLACE media superior 2011-2012*, México, CENEVAL.
- CERVERA Gómez, Luis Ernesto, Gilberto Martín Lizárraga Bustamante y Claudia Paola Sánchez Guillén (2008), "Estudio georreferencial de la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE) en el municipio de Juárez, Chihuahua: análisis espacial", *Revista Electrónica de Investigación Educativa*, vol. 10, núm. 1, pp. 1-23.

- CIZEK, Gregory J. (2001), "Cheating to the Test", *Education Matters*, vol. 1, núm. 1, en: <https://www.educationnext.org/cheatingtothetest/> (consulta: 11 de junio de 2020).
- COLEMAN, James (1966), *Equality of Educational Opportunity*, Washington, DC, U.S. Government Printing Office.
- COLL, Tatiana (2013), "La reforma educativa, el poder del Estado y la evaluación", *El Cotidiano*, núm. 179, pp. 43-54.
- CONTRERAS-Roldán, Sofía y Eduardo Backhoff (2014), "Tendencias en el aprendizaje de la educación en México: una comparación entre ENLACE, Excale y PISA", *Revista Nexos*, octubre, en: <http://www.nexos.com.mx/?p=22749> (consulta: 11 de junio de 2020).
- GOLDSTEIN, Dana (2011, 21 de julio), "How High-Stakes Testing Led to the Atlanta Cheating Scandal", nota de prensa publicada en *Slate.com*, en: <https://bit.ly/2B19uHk> (consulta: 11 de junio de 2020).
- GUTIÉRREZ Pulido, Humberto, Martha Elena Aguiar-Barrera y Lizeth Díaz Caldera (2015a), "Contexto escolar y comprensión lectora en la prueba ENLACE en bachilleratos de Jalisco", *CPU-e Revista de Investigación Educativa*, núm. 20, pp. 1-24.
- GUTIÉRREZ-Pulido, Humberto, Martha Elena Aguiar-Barrera, Porfirio Gutiérrez-González y Cecilia Garibay-López (2015b), "Resultados en matemáticas en enseñanza media superior: ENLACE Jalisco y factores asociados", *Revista de Educación y Desarrollo*, núm. 35, pp. 43-52.
- HERNÁNDEZ, Lilian (2012, 2 de junio), "SEP evita trampa en la prueba ENLACE", *Excélsior*, Sección Nacional, en: <https://goo.gl/s6YiF4> (consulta: 30 de noviembre de 2018).
- HIERNAUX-Nicolás, Daniel (1995), "Reestructuración económica y cambios territoriales en México: un balance 1982-1995", *Revista de Estudios Regionales*, núm. 43, pp. 151-176.
- HOYOS, Rafael, J. Manuel Espino y Vicente García (2012), "Determinantes del logro escolar en México. Primeros resultados utilizando la prueba ENLACE media superior", *El Trimestre Económico*, vol. 79, núm. 4, pp. 783-811.
- "INAI pide a SEP entregar puntaje detallado de pruebas ENLACE" (2017, 21 de noviembre), *El Financiero*, Redacción, Sección Nacional, en: <https://goo.gl/9V8BVn> (consulta: 8 de septiembre de 2020).
- Instituto Nacional de Estadística, Geografía e Informática (INEGI) (2001), *Síntesis de resultados. XII Censo general de población y vivienda 2000*, Aguascalientes, INEGI.
- KORETZ, Daniel (2005), "Alignment, High Stakes, and the Inflation of Test Scores", *CSE Report*, núm. 655, Los Ángeles, University of California
- KORETZ, Daniel (2017), *The Testing Charade: Pretending to make schools better*, Chicago, University of Chicago.
- KORETZ, Daniel (2019), "Measuring Postsecondary Achievement: Lessons from Large-Scale Assessments in the K-12 Sector", *Higher Education Policy*, vol. 32, núm. 4, pp. 513-536.
- MARTÍNEZ Rizo, Felipe (2018), "¿Por qué es tan difícil mejorar los niveles de aprendizaje? A propósito de las nuevas reformas a la educación básica mexicana", *Perfiles Educativos*, vol. 40, núm. 159, pp. 162-176.
- MARTÍNEZ, Tomás, Erica A. Soto, Patricia Silva y Francisco Velasco (2013), "Efectos de la infraestructura básica en los resultados de la prueba ENLACE de la educación media superior tecnológica mexicana", *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, vol. 11, núm. 4, pp. 93-107.
- MUÑOZ Armenta, Aldo (2008), "Escenarios e identidades del SNTE: entre el sistema educativo y el sistema político", *Revista Mexicana de Investigación Educativa*, vol. 13, núm. 37, pp. 377-417.
- NICHOLS, Sharon y David Berliner (2005), *The Inevitable Corruption of Indicators and Educators through High-Stakes Testing*, East Lansing, Great Lakes Center for Education Research and Practice.
- OLSON, John F. y John Fremer (2013), *TILSA Test Security Guidebook: Preventing, detecting, and investigating test security irregularities*, Washington, DC, Council of Chief State School Officers.
- Organización para la Cooperación y el Desarrollo Económico (OCDE) (2003), *Literacy Skills for the World of Tomorrow. Further results from PISA 2000*, París, OCDE.
- ORNELAS, Carlos (2008), "El SNTE, Elba Esther Gordillo y el gobierno de Calderón", *Revista Mexicana de Investigación Educativa*, vol. 13, núm. 37, pp. 445-469.
- PARTIDA-Bush, Virgilio (2005), "La transición demográfica y el proceso de envejecimiento en México", *Papeles de Población*, vol. 11, núm. 45, pp. 9-27.
- POSTLETHWAITE, Neville (1967), *School Organization and Student Achievement: A study based on achievement in mathematics in twelve countries*, Estocolmo, Almqvist & Wiksell.
- Programa de las Naciones Unidas para el Desarrollo (PNUD) (2010), *La verdadera riqueza de las naciones: caminos al desarrollo humano (Informe sobre el Desarrollo Humano 2010)*, Nueva York, Colorcraft of Virginia.
- RAMÍREZ Rosales, Victoria (2010), "El normalismo: proyectos, procesos institucionales y actores", *Revista Iberoamericana de Educación Superior*, vol. 1, núm. 2, pp. 98-113.

- RODRÍGUEZ Gómez, Roberto (2015), “El proyecto educativo SEP-SNTE y la prueba ENLACE”, *Revista Mexicana de Investigación Educativa*, vol. 20, núm. 64, pp. 309-324.
- SACONA Benegas, Unai (2010), *La incidencia del programa ‘Escuelas de calidad’ en el rendimiento educativo: un estudio exploratorio en escuelas de la Ciudad de México*, Tesis de Maestría, México, FLACSO México.
- SÁNCHEZ, Consuelo (2009), “Evaluación de Enciclomedia. Algunos hallazgos relacionados con la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE)”, *Revista Latinoamericana de Estudios Educativos*, vol. 39, núm. 3-4, pp. 251-259.
- Secretaría de Educación Pública (SEP) (2013a), “ENLACE 2006-2013. Medias y niveles de logro por entidad federativa”, en: <https://goo.gl/yBXn9H> (consulta: 3 de febrero de 2018).
- Secretaría de Educación Pública (SEP) (2013b), *Programa de estímulos a la calidad docente*, México, SEP, en: <http://www.programaestimulos.sep.gob.mx/docs/Lineamientos2013.pdf> (consulta: 8 de mayo de 2018).
- Secretaría de Educación Pública (SEP) (2014a), *ENLACE 2006-2013. Medias y niveles de logro por entidad federativa*, en: http://www.enlace.sep.gob.mx/content/ba/pages/estadisticas/estadisticas_2013.html (consulta: 8 de mayo de 2018).
- Secretaría de Educación Pública (SEP) (2014b), *ENLACE. Media superior 2014: resultados nacionales*, México, SEP, en: http://www.enlace.sep.gob.mx/ms/estadisticas_de_resultados/ (consulta: 9 de octubre de 2019).
- Secretaría de Educación Pública-Dirección General de Evaluación de Políticas (SEP-DGEP) (2007), *ENLACE básica. Manual técnico 2007*, México, en: http://www.enlace.sep.gob.mx/ba/manuales_tecnicos/ (consulta: 8 de mayo de 2018).
- UNESCO (2012), *Education for All Global Monitoring Report 2012. Youth and skills putting education to work*, París, UNESCO.
- VÁZQUEZ, Gabriela (2016), “Problemas y limitaciones del uso de pruebas estandarizadas para la evaluación educativa en México”, *Temas de Educación*, vol. 22, núm. 1, pp. 97-114.
- WOLDENBERG, José (2012), *Historia mínima de la transición democrática en México*, México, El Colegio de México.
- WOLFE, Richard G. (1999), “Measurement Obstacles to International Comparisons and the Need for Regional Design and Analysis in Mathematics Survey”, en Gabriele Kaiser, Eduardo Luna e Ian Huntley (eds.), *International Comparisons in Mathematics Education*, Londres, Falmer Press, pp. 225-240.