

## EVALUACIÓN, VALIDEZ Y RELEVANCIA

*Un comentario crítico a*

*“Corrupción de la medida’ e inflación de los resultados de ENLACE”, de Eduardo Backhoff y Sofía Contreras Roldán*

DAVID CALDERÓN

### Resumen:

En este artículo comento críticamente el trabajo de Backhoff y Contreras, publicado en el número 63 de la *Revista Mexicana de Investigación Educativa*, sobre la inflación de resultados de la prueba de la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE). Desarrollo tres comentarios en diálogo con los planteamientos de los autores: *a)* la necesidad de considerar las evaluaciones educativas en el contexto de la tensión entre precisión en la medida, posibilidad de implementar el ajuste de las intervenciones y rendición de cuentas; *b)* una revisión del comparativo entre ENLACE, por un lado, y Excale y PISA por el otro, puntualizando límites y alternativas en dicho abordaje; y *c)* algunas dificultades sobre la explicación de los autores sobre la inflación de resultados por la incidencia de factores externos, y propongo considerar el posible peso del factor interno: el del diseño del instrumento. Para concluir, ofrezco algunos rasgos que pueden ser importantes para contar con una evaluación de los aprendizajes que resulte válida, pedagógicamente útil y socialmente significativa.

### Abstract:

This article represents a critique of the work by Backhoff y Contreras, published in issue 63 of *Revista Mexicana de Investigación Educativa*, with regard to grade inflation on the ENLACE test. I develop three comments in the form of a dialogue with the authors' statements: *a)* the need to consider educational evaluations in the context of tension between precision in measurement, the possibility of adjusting interventions and the rendering of accounts; *b)* a review of the comparison of ENLACE on one hand, and Excale and PISA on the other, with an emphasis on the limits and alternatives in this approach; and *c)* some difficulties regarding the authors' explanation of grade inflation due to the influence of external factors, and my proposal to consider the possible weight of an internal factor: test construction. In conclusion, I offer some factors that may be important for evaluating learning that is valid, pedagogically useful, and socially meaningful.

**Palabras clave:** prueba ENLACE, evaluación del aprendizaje, evaluación de pruebas, validación de pruebas, crítica, México.

**Keywords:** ENLACE test, evaluation of learning, evaluation of tests, validation of tests, criticism, Mexico.

---

David Calderón es director general de *Mexicanos Primero*. Insurgentes Sur 1647 Piso 12, Torre Prisma, colonia San José Insurgentes, 03900, Benito Juárez, México, DF. CE: davidc@mexicanosprimero.org. Agradezco a mi colega Pablo Velázquez el apoyo técnico.

## Introducción

El artículo de Backhoff y Contreras, publicado en el número 63 de la *Revista Mexicana de Investigación Educativa* (RMIE), destaca por proponer un abordaje técnico sobre lo ocurrido con la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) que, hasta ahora, solo incidentalmente había estado presente, y que es fundamental para el rumbo futuro de la evaluación en el país. Hace un aporte valioso a la discusión de dicha prueba y, en general, a la reflexión sobre el sentido de las evaluaciones del logro de aprendizaje.

Los autores ensayan una explicación sobre la inflación de resultados en dicha prueba censal, recurriendo al comparativo con pruebas muestrales de aplicación controlada para grados y edades semejantes: los Exámenes de la Calidad y el Logro Educativos (Excale) y el Programa Internacional para la Evaluación de los Estudiantes (PISA por sus siglas en inglés). Ofrecen, a través de regresiones referidas al incremento de los puntajes, el contraste entre el aumento acelerado para ENLACE y un incremento más modesto en los otros dos ejercicios de medición. Indican tres aspectos como posibles factores de la inflación detectada: la publicación de rankings de escuelas, la asociación de estímulos económicos a los docentes y la falta de control externo de su aplicación. Los autores previenen, al final, sobre el uso intensivo de los indicadores sociales; dejan al lector con una nota de caución, pero también de esperanza acerca de la confiabilidad de las evaluaciones a diseñar para la siguiente etapa en México.

En lo que sigue, desarrollaré tres comentarios en diálogo con los planteamientos de Backhoff y Contreras. Primero, expreso la necesidad de considerar las evaluaciones educativas en un contexto amplio, el de la inevitable pero fructífera tensión entre precisión en la medida, posibilidad de implementar el ajuste de las intervenciones y rendición de cuentas. Segundo, propongo una revisión del comparativo entre ENLACE, por un lado, y Excale y PISA por el otro, con una nota sobre algunos límites y alternativas en dicho abordaje. Tercero, planteo algunas dificultades que percibo en la explicación de los autores con respecto a la inflación por la incidencia de factores externos, y propongo considerar el posible peso del factor interno: el del diseño del instrumento. Para concluir, pongo a consideración algunos rasgos que pueden ser importantes para contar con una evaluación de los aprendizajes que resulte válida, pedagógicamente útil y socialmente significativa.

### **El gran contexto de la evaluación educativa y la “corrupción de la medida”**

Cuando el esfuerzo de evaluación educativa está dirigido a mejorar, entonces requiere: *a)* ser válida y confiable internamente, *b)* ser accesible como insumo oportuno y relevante de diagnóstico para la intervención y *c)* establecer un parámetro objetivo para identificar lo adecuado de la intervención educativa, sirviendo como verificación posterior a ella, de manera que el progreso sea explicado. Así, la serie en el tiempo de una evaluación justa y pertinente permite cumplir con el propósito pedagógico, por el punto *a*, y con el de rendición de cuentas, por el punto *c*.

El punto *a* implica que lo evaluado ha sido adecuadamente definido y que el instrumento diseñado arroja un indicador válido tras su administración. En evaluaciones de gran escala como ENLACE, Excale y PISA la realidad que se quiere indiciar es el logro de aprendizaje. Se debe reconocer, en cualquier esfuerzo de medición, que el indicador del cual se quiere establecer una medida responde a la forma de la metonimia: se toma la parte por el todo. Epistemológicamente es imposible que se presente la realidad tal como es, duplicada; las pruebas son una formalización analítica en la que algunos elementos se aíslan –la respuesta a demandas en forma de ítems de pregunta de opción múltiple, o en forma de tareas revisables– para calificar la situación presente en cuanto a incorporación de nociones o despliegue de habilidades y, con ello, calificar a su vez el logro de aprendizaje del alumno o la consideración de componentes del sistema (como puede ser la labor de los maestros apoyando el alcance de los aprendizajes esperados). Una vez establecido el constructo indicial, se definen (por prueba interna y pilotaje, regularmente con una muestra controlada) los parámetros de referencia, los puntajes y los niveles de desempeño. Además de los arduos trabajos para arribar a un constructo estable y su correspondiente traducción en un instrumento pertinente –eliminando lo más posible los sesgos– vienen los problemas de aplicación. Un instrumento que es adecuado por su diseño puede perder su carácter confiable por determinados eventos previos, simultáneos o posteriores a la aplicación, especialmente por prácticas como la copia, el dictado de respuestas o el acceso previo a los ítems.

El punto *b* implica que los resultados son entregados a quienes pueden y deben intervenir para acercarse a lo deseable. Si la medida educativa es presentada en forma excesivamente técnica y genérica, si no se hace

un gran esfuerzo de comunicación y formación para la interpretación, únicamente los técnicos podrán intervenir. En esa circunstancia, no solo se dejará fuera a los agentes fundamentales del aprendizaje –los alumnos y los maestros– lo mismo que los responsables del apoyo –los padres y la “mesoestructura”– sino, incluso, a los decisores de política pública. Si no se puede responder a la pregunta “¿Y yo qué puedo hacer?”, los resultados de mediciones subsecuentes, con validez interna, no deberían entonces mostrar cambios significativos, pues la reiteración de prácticas produce, con todas las condiciones iguales, reiteración de resultados. No basta identificar qué se quiere cambiar (en este caso, el bajo logro de aprendizaje) sino quiénes y cómo pueden desarrollar dicho cambio. Las responsabilidades del servicio público para propiciar el aprendizaje, sin una devolución que llegue hasta donde ocurre el hecho educativo, quedarán opacas, envueltas en la “tragedia de los comunes”; un efecto compuesto del que nadie acepta la paternidad, y en el cual el aporte de cada persona y grupo no se identifica, de manera que la voluntad de cambio personal se hundirá en una inercia que se experimenta inamovible. Una evaluación sin consecuencias reales de reajuste –porque no se comunicó, porque no se usa para informar la intervención– es irrelevante.

El punto *c* plantea, entonces, la importancia de una frecuencia adecuada de la medición: lo deseable es que la serie incluya a los mismos sujetos antes y después de la intervención, para confirmar lo apropiado de las estrategias elegidas para mejorar los resultados. Evaluaciones muy distanciadas entre sí en el tiempo no permiten captar qué tan cruciales resultaron diversas acciones para explicar el logro posterior, mientras que, en el otro extremo, si los cambios de práctica son frecuentes en exceso no alcanzarán a consolidarse, ni a mostrar con suficiencia sus frutos. En ambos casos, por defecto o por exceso de frecuencia, aun si en la medición se capta un progreso o avance, no sabríamos decir a ciencia cierta por qué ocurrió, y las intervenciones orientadas a la mejora no podrían juzgarse para ser rectificadas o ratificadas. Debe, adicionalmente, reconocerse que la medición del logro de aprendizaje tiene exigencias propias que no se parecen a otras acciones de cambio social planificado; por ejemplo, las evaluaciones sobre servicios de salud no implican una intervención continua, como sí ocurre con la educación, que en el formato escolar es cotidiana e intensiva. Dado que el aprendizaje es un derecho de las per-

sonas, monitorear el cumplimiento de las garantías que le corresponde al Estado ofrecer para el ejercicio de este derecho hace que la rendición de cuentas sea sumamente relevante. La verdadera rendición de cuentas no debiera construirse como simple premio o castigo –y menos circunscribirla a incentivos solamente materiales– restringido a los agentes en el aula, sino abarcar primero y especialmente a los funcionarios responsables del diseño e implementación de la política pública, como evaluación de su actuación específica.

Este gran contexto implica que las funciones pedagógica y de rendición de cuentas en las evaluaciones educativas no solo son ambas importantes, sino que están inextricablemente unidas entre sí para conservar su sentido propio. La función pedagógica de la evaluación no involucra únicamente a los diseñadores de currículum y a los altos funcionarios, ni tampoco la rendición de cuentas se agota en un cuerpo técnico o en un dispositivo social formal como el Congreso. El reto formidable de toda evaluación educativa a gran escala es mantenerse técnicamente sólida por su diseño y administración, mientras que se cumple también con su aterrizaje en prácticas de aula distintas y en el involucramiento de la sociedad con el cumplimiento de las responsabilidades sobre un derecho humano. La gran escala en el grupo de sujetos evaluados debiera corresponderse con la gran escala de las estrategias de mejora derivadas de la devolución de resultados, y con la gran escala de la difusión y debate de los resultados, con consecuencias de rectificación de prácticas y mandatos.

Puede decirse que, en el caso concreto de ENLACE, haber aparejado los resultados con incentivos burdos y no haber desarrollado una devolución significativa, con el desajuste multiplicado por su frecuencia excesiva sin cambios de práctica, es una receta para el desastre. Sin embargo, la afirmación de Backhoff y Contreras (2014:1270-1271) en el sentido de que “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” debe, a mi juicio matizarse, pues los autores, aun reconociendo que puede haber un continuo, pareciera que no dejan espacio suficiente a proponerse la correspondencia entre ambas funciones, y no su mera adición.

Supongamos por un momento que se contara en México con dos evaluaciones de gran escala en simultáneo, en la que ambas cumplieren

con los términos de su polo sin incluir ningún aspecto de la columna opuesta; incluso entonces sería un válido cuestionamiento de consistencia verificar si coinciden entre sí, y un válido cuestionamiento por parte de alumnos, padres, maestros y ciudadanos en general ponderar qué dicen los resultados con respecto al esfuerzo y responsabilidad de los involucrados. No es un decreto que la dimensión pedagógica y la de rendición de cuentas resulten, por sí mismas, incompatibles; por el contrario, no hay verdadera función pedagógica sin responsabilidad, ni hay responsabilidad válida si no es precisamente la referida al cumplimiento de la implementación pedagógica.

En el artículo que comentamos se alude a la “Ley de Campbell” (Backhoff y Contreras, 2014:1268 y 1281): “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, 1976:49). Esta frase, que proviene de un *Occasional Paper* en el que Campbell analiza el uso de evaluaciones cuantitativas para valorar el impacto del cambio social planificado, aparece en medio de multitud de otros llamados de atención igualmente atendibles, a lo largo de 71 páginas de exposición. Por ejemplo, la tendencia a preferir la no-evaluación: “Los administradores de programas y los beneficiarios tienen el poder de sabotear nuestros esfuerzos de evaluación, e intentarán hacerlo si sus propias carreras e intereses se ponen en juego” (Campbell, 1976:48). En este texto clásico se subraya la importancia de cerrar la distancia entre quien evalúa y quien implementa, so pena de que la evaluación resulte socialmente irrelevante, pues no informa los cambios deseables si no se garantiza la coordinación (Campbell, 1976:14): “Los evaluadores externos tienden a carecer del conocimiento cualitativo esencial acerca de lo que ocurrió. El conflicto crónico entre evaluadores e implementadores, que ya sería suficientemente malo bajo una dirección local unificada, tiende a ser exacerbado”. Igualmente, alude a un inquietante diferencial de poder (Campbell, 1976:30):

Nótese que en el presente el sistema escolar registra el desempeño del alumno, pero nunca registra el reporte del alumno sobre el desempeño de la escuela. Los maestros son anualmente calificados por su eficiencia pero nunca tienen la ocasión de calificar sistemáticamente las políticas que se les pide implementar.

En su pieza, Campbell no está abogando para que las evaluaciones sean desconocidas u ocasionales, sino a favor de que desde el diseño se prevea cómo evitar los abusos de excesiva concentración en un único indicador, de qué modo favorecer la complementariedad de instrumentos cuantitativos y cualitativos y, sobre todo, cómo vincular a los decisores con el ajuste de las intervenciones. Sin una teoría de cambio para guiar el proceso –hacer qué, por quiénes, genera qué– la medición, ocasional y eventual, no genera correctivos ni rediseño.

### **Comparando ENLACE con Excale y PISA**

En su texto, Backhoff y Contreras recurren a Excale y PISA como referentes de control para tratar de estimar la inflación de resultados en ENLACE. Comparan las tendencias de aumento de puntaje en los dominios escolares incluidos en las tres pruebas (Lengua/Español/Comprensión lectora, por un lado, y Matemáticas, por el otro), en los grados escolares o edades normativas en los cuales las pruebas usadas como control coinciden con ENLACE, y se revisan en la serie del tiempo.

Sin dejar de considerar las diferencias en diseño y número de estudiantes involucrados en cada prueba, la estrategia de comparación brinda un contraste que, sin duda, favorece la reflexión. El punto fuerte del argumento se ilustra con las gráficas 1 a 4 (Backhoff y Contreras, 2014:1277, 1278 y 1280), en las cuales se destaca la discrepancia de tendencias, comparativamente altas en la mayor parte de los casos para ENLACE, y más moderadas pendientes en el caso de Excale y PISA.

Tomando el dominio de Matemáticas, las puntuaciones de ENLACE marcan una tendencia en el tiempo que puede ser superior desde aproximadamente el doble (ENLACE *vs.* Excale, ambos de tercero de primaria, o también ENLACE de tercero de secundaria *vs.* PISA secundaria) hasta casi seis veces más (ENLACE *vs.* Excale, ambos de sexto de primaria). La estrategia del artículo comentado tiene numerosas limitaciones, pero es fecunda para producir preguntas; de ellas, la central es: ¿qué explica la discrepancia?

Dos gráficas del artículo comentado tienen restricciones importantes como evidencia. La gráfica 1 compara las tendencias de ENLACE y Excale para Lengua en tercero de primaria; la dificultad estriba en que para generar la tendencia de Excale únicamente se cuenta con dos observaciones en la serie, mientras que para ENLACE hay ocho. Establecer tendencia

con pocos datos es siempre riesgoso para obtener conclusiones sólidas. Por su parte, la gráfica 4 compara ENLACE de tercero de educación media superior con PISA, para los dominios de Lengua y Matemáticas. Aquí la dificultad subyace en que los grupos evaluados se refieren a edades muy distintas: los jóvenes a los cuales se les aplica ENLACE tienen típicamente 17 años, mientras que a quienes se les administró PISA tienen en su gran mayoría 15 años de edad.

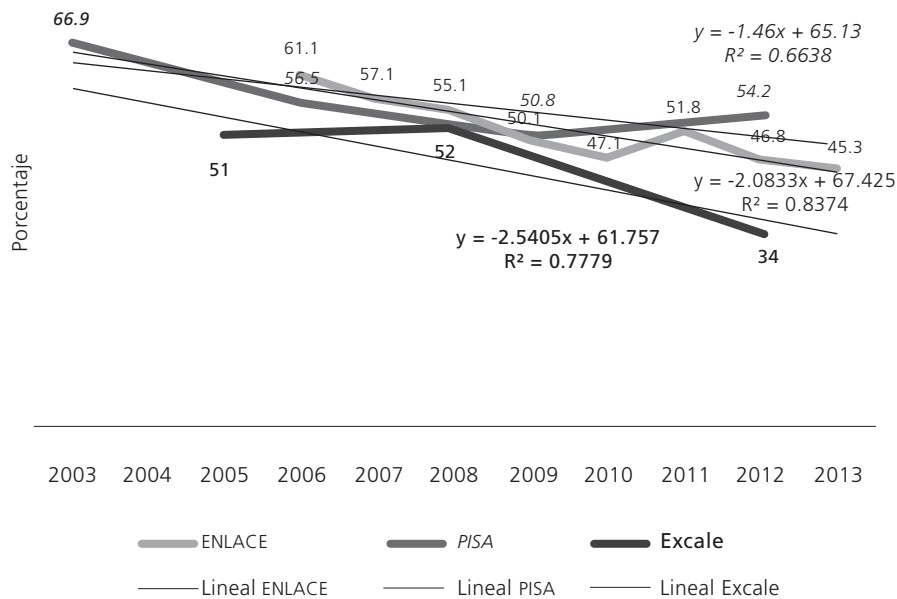
En cambio, las gráficas 2 y 3 comparan con un mejor ajuste en la población de referencia. Las barras muestran un crecimiento acelerado del lado de ENLACE, salvo para Español de tercero de secundaria. El dominio de Matemáticas, presentado en la gráfica 3, muestra crecimiento estable y significativo para los resultados de ENLACE. Una explicación plausible –alternativa a la corrupción de la medida, que es la sostenida por Backhoff y Contreras– es considerar que las tendencias son divergentes porque la demanda de ENLACE es menor, por ejemplo en el comparativo en tercero de secundaria. Siguiendo esa alternativa de explicación, el crecimiento implicaría que se trata de escala: como no miden en los mismos términos, mejorar en ENLACE en una magnitud considerable aún no “alcanza” para modificar en la misma proporción el correspondiente progreso en PISA; lo que ENLACE demanda en su campo resulta una exigencia poco compleja para las tareas solicitadas en la prueba de PISA. Para el caso de Matemáticas en primaria, la preocupación ciertamente puede ser mayor: tanto ENLACE como Excale están alineadas al currículum nacional, de manera que habiendo divergencia en la demanda de cada prueba hacia los alumnos sustentantes, ello podría implicar que el enfoque de diseño interno de ENLACE es mucho menos exigente, y favorece puntajes más altos.

Propongo una alternativa para ampliar los comparativos. Los puntajes usados en el artículo de Backhoff y Contreras son medias, las cuales fueron concebidas por los diseñadores como el despliegue de un virtual 500, que después permite tratar los resultados con franjas de desviación estándar de 100 puntos. En términos de uso pedagógico y de reajuste de política pública puede resultar más interesante plasmar los resultados en términos de niveles de logro que agregados en el gran promedio. Para ello, consideramos lo que cada prueba señala como umbral de mínimo aceptable y la distribución en los niveles de logro por arriba y por debajo de esa marca.



Así, comparando de nuevo estas distintas pruebas para el dominio de Matemáticas en tercero de secundaria, podemos ver una coincidencia en términos de captar la insuficiencia: el porcentaje de alumnos con logro insuficiente –según lo que el diseño de cada prueba maneja como nivel o niveles por debajo del mínimo aceptable– se asemeja, aunque la similitud se va perdiendo gradualmente, como puede verse en la gráfica 1.

GRÁFICA 1  
Tendencias en el porcentaje de nivel de logro insuficiente en ENLACE, PISA y Excale: Matemáticas tercero de secundaria



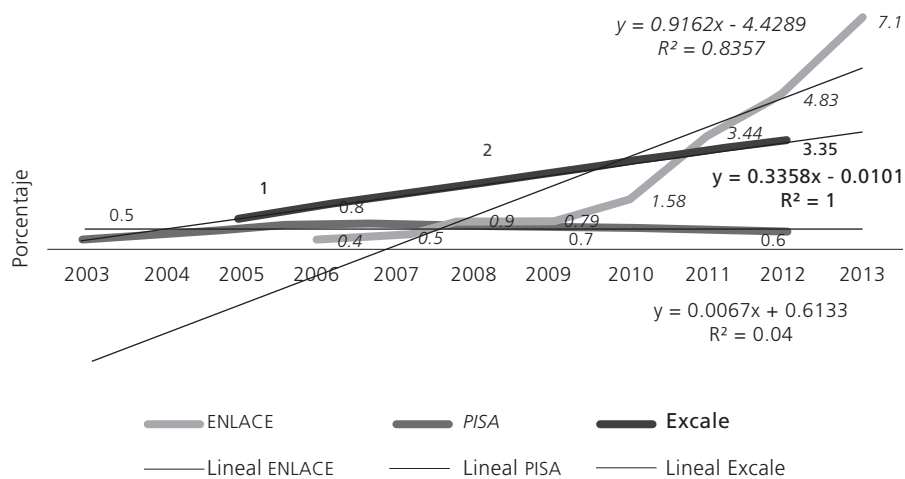
Fuente: Elaboración propia; para ENLACE, a partir de 2009 se eliminan, para mayor precisión, las escuelas donde más de 10% de los alumnos evaluados tienen resultados no confiables.

Podría decirse que ENLACE, con el control de Excale y PISA, capta en forma similar la preocupante situación de buena parte de los alumnos matriculados: casi la mitad no están teniendo el logro de aprendizaje esperado, y la mejoría alcanza a comparativamente muy pocos. En cambio, si observamos el extremo superior en el registro de niveles de logro, el crecimiento desmedido del lado de ENLACE es notable. La tendencia de

PISA es plana, la de Excale es más marcada y la de ENLACE se dispara, como puede verse en la gráfica 2.

GRÁFICA 2

*Tendencias en el porcentaje de nivel de logro sobresaliente en ENLACE, PISA y Excale: Matemáticas tercero de secundaria*



Fuente: Elaboración propia; para ENLACE, a partir de 2009 se eliminan, para mayor precisión, las escuelas donde más de 10% de los alumnos evaluados tienen resultados no confiables.

Aquí se disparan los resultados positivos de ENLACE. Una hipótesis para explicar por qué la inflación afectaría más al extremo superior puede ser que la copia y el dictado de resultados no fueron tan frecuentes y que, en cambio, la distorsión provenga especialmente de “preparar para la prueba”, situación que favorecería diferencialmente: a los estudiantes que ya suelen responder con éxito a la ejercitación reiterativa les permitiría aumentar puntajes por sobre-exposición, pero en cambio afectaría casi imperceptiblemente a los alumnos que requerirían de parte del maestro una dedicación más atenta. Si los sucesos deshonestos en el aula fueran la causa principal de distorsión, entonces la reducción de insuficiencia y el aumento de desempeño superior en ENLACE tendrían tasas de progreso semejantes (sería muy extraño dictar respuesta a los pocos de mejor desempeño, y ocultarles esa

respuesta favorable a contingentes más amplios de desempeño menor). En todo caso, queda entonces para la reflexión la discrepancia en el porcentaje de estudiantes en los niveles de desempeño superior para ENLACE como una posible evidencia adicional de la inflación de resultados.

### **Dificultades para explicar la inflación de resultados**

Recapitulando, dos abordajes distintos –comparar promedios y niveles de logro– llevan a considerar que los resultados de ENLACE/educación básica pueden distorsionar el uso de dicho instrumento como medición confiable y sesgar el juicio en su difusión y uso. La complicación que se presenta para una nueva generación de evaluaciones del logro de aprendizaje en México es identificar, con la mayor certeza posible, los factores que intervienen en ello. Las tres prácticas señaladas por Backhoff y Contreras requerirían de una verificación más concreta, para recuperar las lecciones que correspondan. La publicación del ranking de las escuelas no resulta, a mi juicio, un factor importante en este caso. Ocurrió en forma aislada y eventual, no generado por la autoridad ni los diseñadores, y su difusión tocó algunos espacios de opinión pública lo que, al máximo, se tradujo en una leve presión hacia las autoridades, más que a las escuelas mismas.

Otro factor mencionado por los autores es la falta de control externo en la aplicación. A pesar de que para la aplicación de ENLACE se emitieron directivas en el sentido de que el aplicador no fuese el maestro del grupo y que se contara con la presencia de padres en la administración, la variación ciertamente fue muy grande. Como control adicional, se incorporó un algoritmo para identificar el “factor copia”, que comparara el tren de respuesta de los alumnos de un mismo grupo, para detectar pautas de coincidencia poco plausibles. Sin embargo, en la difusión de resultados prácticamente no se corrigió el agregado de resultados, como en cambio sí hicimos en otro trabajo (Mexicanos Primero, 2013), para restar los que fueran poco confiables según dicho procedimiento. Las anulaciones por alteración durante la administración fueron pocas, y la evidencia sobre dictado de respuestas o sobre evitar que algunos alumnos –identificados como de bajo desempeño– se presentaran, no ha sido corroborada de forma que rebase la mera anécdota, ni que pudiera comprobarse una frecuencia que altere los resultados en modo sistemático.

El restante factor, en cambio, tiene un peso no desdeñable; haber entregado incentivos económicos para los maestros en un esquema de mera

adición, favoreció una amplia “preparación para la prueba”, con sus consecuentes daños no solamente para la confiabilidad, dados los aumentos acelerados de puntaje, sino por la empobrecedora reducción del currículum real a los temas que previsiblemente la prueba abordaría en sus ítems, efecto más grave aún.

Una alternativa para redimensionar el impacto de factores externos es si la inflación de resultados pudiera provenir del diseño del instrumento. Así como Backhoff y Contreras aluden a que ello sería una explicación plausible en el caso del aumento en los puntajes de ENLACE/media superior (2004:1281), propongo que otro tanto también se considere para el caso de ENLACE/básica. No hay evidencias disponibles de que así ocurrió, ni implica en sí mismo deshonestidad en los diseñadores, pero no está de más documentar cómo funcionaron el jueceo de ítems y los posibles cambios en el equipo de expertos, así como la presión de autoridades estatales y de las representaciones sindicales en búsqueda de prestigio y beneficios.

### **Conclusiones**

La inflación de resultados de ENLACE afecta a las tres dimensiones de validez, uso pedagógico y rendición de cuentas, aunque de forma diversificada. Backhoff y Contreras dedican su artículo a algunas señales de falta de confiabilidad y atribuyen la corrupción de la medida al uso inadecuado e intensivo de la prueba.

Todavía falta hacer un examen de en qué medida se diseñaron y llevaron a las aulas propuestas honestas de cambio de práctica, distintas de la ya muy mencionada “preparar para la prueba”; valdría la pena que los investigadores educativos aborden las estrategias –¿ocasionales?, ¿fallidas?– que emprendieron los equipos de las autoridades educativas estatales y federal como modificaciones más estructurales de la intervención en vista a la mejora. Igualmente, es deseable una serena recapitulación que incluya el recuento del uso comunicativo, desde los reportes individuales dirigidos a las familias hasta las notas periodísticas; las evidencias documentales de fraude y su procesamiento posterior: ¿hubo sanciones?; también revelador sería dar seguimiento al efecto y documentar la opinión de los maestros sobre los estímulos que se entregaron ligados a los resultados de la prueba. Estos insumos son urgentemente necesarios para el rediseño que se avecina en las evaluaciones nacionales.

Para no caer en la petición de principio, afirmando lo que se debiera probar, es necesario recapitular la experiencia de ENLACE, desde su diseño hasta la verificación de alteraciones acaecidas durante la administración de la prueba en el aula. La opinión de que el control de la aplicación deba hacerse necesariamente con funcionarios en el aula ha de sopesar el hecho de que –en las evaluaciones de gran escala– la mayoría de los aplicadores no son ellos mismos técnicos, sino figuras habilitadas; conviene que se considere entonces qué elementos harían posible la ciudadanización. El proceso electoral es un referente de articulación entre confiabilidad y participación, por lo que no es desdeñable tomar lecciones acerca de alternativas a un despliegue costoso y masivo de funcionarios. No por ser figuras oficiales quedan ajenas a eventuales presiones o a errores de administración de los instrumentos. Por otro lado, si se quiere contestar que no cualquier publicación de resultados ayuda a mejorar, también debe reconocerse que difícilmente el sigilo de resultados lo hace. La irrelevancia pedagógica no es opción, ni tampoco lo es una convocatoria abstracta a apoyar a las escuelas, prestigiar a los maestros o confiar en lo atinado de los funcionarios mayores.

ENLACE/básica se canceló ya, y por ello lo sensato es aprender e innovar. Ahora, como nunca, es necesario contar con una evaluación válida de los aprendizajes, pues está plasmado en el mandato constitucional que “el máximo logro de aprendizaje de los educandos” (CPEUM, 2013) sea garantizado, como responsabilidad del Estado mexicano. Sin un referente válido del logro de aprendizaje, no hay manera de monitorear el cumplimiento del derecho, y de confirmar que materiales y métodos, infraestructura, gestión y finalmente la idoneidad de docentes y directivos contribuyan a ello, tal como la Constitución establece. Contar con instrumentos tanto censales como muestrales, que compartan el mismo referente y se articulen para sumar orgánicamente las fortalezas de ambos abordajes, es una perspectiva promisoría para la siguiente etapa de la evaluación de logro de aprendizaje.

El Instituto Nacional para la Evaluación de la Educación (INEE) y las autoridades educativas tienen la tarea de concretar un sistema de evaluación con validez, sin dejar de pensar desde el momento mismo del diseño los parámetros y directrices que llevarán la propuesta de mejora a cada aula. Igualmente, para que la ciudadanía pueda mandar y demandar la responsabilidad específica sobre las decisiones de política educativa, y su implementación oportuna y efectiva, se requerirá de un esfuerzo muy ho-

nesto de transparencia y rendición de cuentas. La difusión de resultados y su uso requieren de más y mejor participación de todos los grupos sociales; el esfuerzo de una discusión crítica, con las mejores aportaciones de la comunidad académica, no solo es necesaria en el espacio de consejos de expertos, sino también en la máxima apertura, como ilustra la publicación del artículo de Backhoff y Contreras en la RMIE.

El derecho a saber de los ciudadanos acerca de la intervención de los agentes del Estado –por la pre-existencia del mandato que funda y justifica su actuación y su sueldo– no puede obviarse. Aceptar que la marcha del sistema escolar y en concreto el logro de aprendizaje de los alumnos pueda tratarse como un arcano de la Razón de Estado sería una involución al proceso democrático, pero más sencillamente sería restringir a los alumnos y maestros mismos a objetos de la medición, pero no sujetos de la acción, que recibirán –desde arriba y desde afuera– una intervención, un “tratamiento” prescriptivo posterior a la evaluación, sin participar en forma explícita y responsable.

Brindar evidencia objetiva del progreso en el logro de aprendizaje da dirección y sentido a todo el sistema educativo, ante la tendencia a considerar que la escuela es un dispositivo solo de contención, fatalmente predeterminada por el contexto, y no de desarrollo de potencialidades de los alumnos. Nos pone de nuevo ante la decisión de si nos conformamos con que el sistema escolar sea el reflejo de las limitaciones de la comunidad o si, por el contrario, queremos que sea el laboratorio de su proyecto.

## Referencias

- Backhoff, E. y Contreras Roldán, S. (2014). “Corrupción de la medida’ e inflación de resultados de ENLACE”, *Revista Mexicana de Investigación Educativa*, número 63, volumen XIX, octubre-diciembre. Disponible en: <http://www.comie.org.mx/v1/revista/portal.php?idm=es&sec=SC03&&sub=SBB&criterio=ART63011>
- Campbell, D. (1976). *Assessing the impact of planned social change*, Occasional Paper Series, Dartmouth: Public Affairs Center. Disponible en: <https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/08%20Assessing%20the%20Impact%20of%20Planned%20Social%20Change.pdf>
- CPEUM (2013). “Artículo Tercero”, *Constitución Política de los Estados Unidos Mexicanos*. Disponible en: <http://www.ordenjuridico.gob.mx/Constitucion/articulos/3.pdf>
- Mexicanos Primero (2013). *IDEI. Índice de Desempeño Educativo Incluyente. El avance en los estados de 2009 a 2012*, Ciudad de México: Mexicanos Primero. Disponible en: [http://www.mexicanosprimero.org/images/stories/idei/IDEI\\_2009-2012.pdf](http://www.mexicanosprimero.org/images/stories/idei/IDEI_2009-2012.pdf)