

## Evaluación Educativa

*Con la llegada de los nuevos operativos previstos a nivel de la Ciudad de Buenos Aires (FEPBA y TESBA) y nacional (Aprender 2019) para fines de agosto y principios de septiembre, la evaluación de la calidad educativa vuelve a estar en el centro del debate. Compartimos dos artículos que nos proponen reflexionar acerca del reduccionismo de estos dispositivos de medición de aprendizajes. La Dra. Sverdlick nos invita a pensar sobre la reiterada retórica de la evaluación y la posibilidad de nuevos aportes; y la Dra. Hebe Roig presenta las limitaciones de este tipo mediciones y plantea colocar a la evaluación en el lugar que le corresponde.*

# UNA EVALUACIÓN SIN CAMBIO: EL DISPOSITIVO NACIONAL DE EVALUACIÓN DE LOS APRENDIZAJES APRENDER

**Hebe Roig**

Profesora Adjunta Regular de la Universidad de Buenos Aires dedicada a docencia e investigación en educación a distancia, tecnología educativa, y evaluación de proyectos y sistemas educativos. Directora del Posgrado Virtual de Especialización en Evaluación Universitaria de la UBA.



En Argentina llevamos más de 25 años midiendo algunos aprendizajes de los estudiantes de primaria y secundaria. Desde entonces, muchas voces han advertido las limitaciones de este tipo mediciones para promover la mejora de las experiencias educativas en nuestras escuelas. Trataremos de refrescar algunas de estas advertencias para proponer que debemos poner a la evaluación en el lugar de humildad que le corresponde. Lugar de humildad y subsidiaridad en el arduo camino del aprendizaje personal e institucional que es tergiversado por las políticas educativas que encuentran en los números la forma de justificar decisiones.

Desde 1993 hasta la fecha, en nuestro país se han puesto en marcha diversos dispositivos de evaluación del aprendizaje de los estudiantes en el sistema de educación obligatoria. Fueron variando sucesivamente a lo largo de los diferentes periodos políticos del país y cada uno

## Evaluación Educativa

tuvo un marco normativo y económico particular. Desde entonces, podemos identificar los roles protagónicos del Consejo Federal de Educación (CFE), el Ministerio de Educación Nacional, los medios masivos de comunicación y los gremios docentes. Concertación de políticas entre las jurisdicciones, opiniones instaladas por los medios de comunicación y realidades laborales, ponen en escena diferentes expectativas sobre la educación del país y el lugar que se le asigna en este contexto a la evaluación de escuelas, docentes y estudiantes.

El dispositivo que analizamos aquí inicia su desarrollo bajo el gobierno de Cambiemos, a fines del año 2015. A principios del 2016 se presentan y hacen públicas sus características que quedan definidas inicialmente en la Res. CFE N° 280/16. Tal como se expresa en la web del Ministerio de Educación, APRENDER “es el dispositivo nacional de evaluación de los aprendizajes de los estudiantes y de sistematización de información acerca de algunas condiciones en las que ellos se desarrollan”.

La preponderancia que este gobierno da a la evaluación como política educativa es explícita en la Declaración de Purmamarca (Consejo Federal de Educación; 12/02/2016) y las posteriores resoluciones que establecerán el dispositivo APRENDER. En el nivel organizacional del Ministerio Nacional, se crea una Secretaría de Evaluación Educativa especialmente dedicada a la evaluación del sistema, en el mismo nivel jerárquico que una Secretaría de Innovación y Calidad Educativa (en cuyo seno se encuentra una Dirección de Información y Estadística Educativa). La evaluación asume jerarquía de Secretaría ya que en el conjunto de acciones del Ministerio, se revitaliza la concepción de un Estado que descentraliza la gestión educativa, pero la controla por la evaluación.

Un tercer dato que termina de dejar en claro que medir es la política educativa más importante de este gobierno, se refleja en el presupuesto educativo. El estudio de los presupuestos educativos 2016-2019 presentado por los Observatorios y Equipos de

## Evaluación Educativa

Investigación en Educación de UNIPE, FFYL-UBA, CCC, AULA ABIERTA, UNDAV, UNQUI Y UNGS1 muestra que “la partida para evaluaciones educativas es una de las pocas que exhibe un importante incremento, tanto nominal como real entre 2016 y 2019” y aclara que “esto se explica por la decisión de realizar costosas evaluaciones estandarizadas censales”.

A diferencia de los anteriores Operativos Nacionales de Evaluación (ONE), APRENDER fue planteado con dos nuevas características: ya no se buscaría una radiografía del sistema a través de una muestra de escuelas, sino un mapa detallado de todas las escuelas primarias y secundarias del país (carácter censal) y, como evaluar sería “indispensable para el aprendizaje continuo” (Declaración de Purmamarca, 2016), la medición sería anual (a partir de 2018 será bianual por los problemas económicos del país). La lógica que está en la base de estos cambios es que más evaluación y del mismo tipo es la solución para los problemas educativos.

Si bien APRENDER duplica la apuesta sobre la necesidad de la medición para llegar a cada escuela del sistema educativo, su modelo posee características similares a los antiguos ONE: se trata de pruebas objetivas para la evaluación de aprendizajes acompañadas de cuestionarios complementarios (que miden ciertos factores asociados) y se analizan los resultados agregando los datos para obtener información sobre diferentes niveles (país, jurisdicción, escuela).

El uso de pruebas objetivas de aprendizaje en conjunto con cuestionarios de factores asociados al desempeño académico proviene de la tradición de investigación sobre escuelas eficaces que se desarrolla en el mundo anglosajón desde la década de 1970.

Limitaciones del modelo

## Evaluación Educativa

Revisaremos brevemente las principales dificultades que ofrecen las pruebas objetivas de aprendizaje y el estudio estadístico de factores asociados al aprendizaje.

En las pruebas objetivas el estudiante debe escoger la respuesta correcta entre diversas opciones formuladas por quienes construyen el instrumento. Eventualmente, pueden incluirse enunciados en los que habrá que completar una palabra o resultados numéricos. Entre sus limitaciones, señaladas por diversos autores a lo largo del tiempo, se destaca el tipo de conocimiento que pueden observar y medir.

Permiten estudiar sólo aquellos aprendizajes que tienen la posibilidad de ser presentados mediante la escritura, excluyendo las múltiples formas de representación y expresión del conocimiento que van más allá de los lenguajes verbal y matemático. El estudiante hace su mayor esfuerzo en la lectura de preguntas e ítems sin poner en juego su propia escritura. La relación entre leer, escribir y aprender, que requiere ser cultivada desde la educación primaria, elude una visión integral de la alfabetización ya que la evaluación queda limitada a la comprensión lectora.

En segundo lugar, para que la prueba sea técnicamente correcta, los ítems de selección múltiple deben tener una sola respuesta correcta. Se asume que en las pruebas objetivas se deben incluir aquellos temas o contenidos sobre los que hay acuerdo indiscutido entre los especialistas de la materia que se trate. La definición de preguntas que sólo tienen una respuesta correcta es un problema si se consideran diferentes enfoques dentro de un campo disciplinar. Este aspecto limita el espectro de temáticas posibles de evaluación, particularmente en las ciencias sociales y naturales (contenidos que se incluyen en las pruebas de nivel secundario), y puede atentar contra la relevancia del contenido indagado.

Considerando incluso que las preguntas y las opciones hayan sido elaboradas técnicamente de modo correcto, este tipo de pruebas restringe las posibilidades de apreciación de los aprendizajes: “no pueden medir, por definición, la capacidad de los estudiantes para

## Evaluación Educativa

formular espontáneamente hipótesis pertinentes, para coleccionar datos de laboratorio, para planear un experimento original, para estructurar un argumento defendible o para realizar trabajos creativos” (Ausubel, 1978; 668).

En síntesis, desde el punto de vista del contenido, las pruebas objetivas miden comprensión lectora y temas sobre los que no hay debate en la materia.

Quienes defienden hoy las pruebas objetivas destacan la importancia de no reducir la evaluación del aprendizaje al uso de este tipo de instrumentos. De hecho, en nuestras escuelas, es un tipo de instrumento muy poco utilizado por los docentes.

Una segunda línea de advertencias sobre los problemas vinculados a este tipo de pruebas, es la instancia de implementación, cuando el estudiante se enfrenta a la prueba. Ausubel (1978) señala que los ítems de selección múltiple pueden castigar a estudiantes más expertos que consideran sutilezas que no quedan expresadas en las alternativas de respuesta ofrecidas, pero lo comúnmente advertido es la falta de práctica de los estudiantes con este tipo de instrumentos.

Finalmente, como un tercer aspecto que cabe considerar sobre este tipo de instrumentos, es necesario advertir sus repercusiones sobre la enseñanza y el aprendizaje. El entrenamiento en este tipo de pruebas produce mejores resultados, de allí la tentación de fomentar el uso de este tipo de instrumentos en la vida cotidiana de las escuelas para mostrar “mejores” resultados, aún a costa de empobrecer la enseñanza.

Toda evaluación tiene efectos de “retroacción”, ejerce influencia sobre lo que enseña el docente y sobre lo que los estudiantes consideran relevante estudiar. En este caso, se jerarquiza la lectura y el aprendizaje de aquellos contenidos que pueden ser catalogados como verdades indiscutibles en el marco de los núcleos prioritarios de aprendizaje (NAP).

## Evaluación Educativa

Tal como hemos mencionado, el segundo tipo de instrumento que pone en juego APRENDER es el cuestionario sobre factores asociados al desempeño estudiantil que aplica principalmente a estudiantes y directivos de las instituciones educativas. Con ellos se busca información sobre las condiciones contextuales en las que tienen lugar los aprendizajes evaluados<sup>2</sup>.

El diseño e implementación de estos cuestionarios se basa en el estudio de la correlación estadística entre el desempeño de los estudiantes en pruebas de rendimiento y variables de contexto. Las variables de contexto estudiadas desde los años 70's a la fecha, dan cuenta de que efectivamente algunas de estas variables se manifiestan en simultáneo con un mayor o menor desempeño en las pruebas objetivas, pero el abanico de variables contextuales posibles debe ser recortado por los evaluadores, ya que es imposible considerarlas todas. Cuáles son esas variables y para qué se medirán depende de las hipótesis e intereses de quienes diseñan la evaluación.

Marchesi y Martín (2000) destacan los problemas conceptuales y metodológicos de los estudios sobre factores asociados al aprendizaje. Respecto a los problemas conceptuales, en primer lugar, sostienen que no se ha elaborado una teoría comprensiva a partir de los estudios sobre las variables que inciden en la eficacia de la escuela (eficacia entendida como buenos resultados en las pruebas), se mide “lo que funciona” sin ofrecer herramientas a la comprensión. En segundo lugar, observan que este tipo de diseños desconocen características propias de cada escuela no previstas en la selección de variables realizada a priori por los evaluadores. En tercer lugar, y realmente muy grave, reducen los resultados escolares a logros académicos en pruebas estandarizadas, desvalorizando todos los otros objetivos y resultados de la actividad educativa escolar.

En cuanto a lo metodológico, estos autores observan la limitación del análisis estadístico: correlación no significa causalidad. Esto es, que dos variables se comporten de modo

## Evaluación Educativa

semejante no significa que una sea la causa de las variaciones de la otra. Sólo establece que varían en conjunto, pudiendo ser ambas resultado de otras variables no medidas.

Este tipo de diseños ofrece información descriptiva que no puede orientar la toma de decisiones. Esta advertencia fue uno de los mayores aportes de Stufflebeam y Shinkfield (1987) cuando denominaron a este tipo de estudios como “cuasievaluación”.

### Medición y toma de decisiones

En este tipo de diseños evaluativos, los resultados y la toma de decisiones no tienen relación directa entre sí. Las decisiones que se pueden tomar no surgen de la supuesta transparencia de los datos estadísticos, sino de la interpretación que se hace de ellos.

El bajo rendimiento académico, aunque seamos conscientes de la imagen limitada y sesgada que puede ofrecer este tipo de evaluaciones, sin duda presenta un problema.

Planteado el problema, se justifica que es necesario intervenir para solucionarlo, pero la forma de intervenir no deriva de la indicación de que existe un problema.

En Matemática, un problema es una cuestión que se plantea para hallar un dato desconocido a partir de otros datos conocidos. No estamos ante un problema matemático. La ecuación problema = solución no es como nos hacen creer. La solución de los problemas educativos no responde a una fórmula, sino que requiere intervenciones en las que se consideren recursos, prácticas, teorías, creencias, intereses, oportunidades, tradiciones, mitos, regulaciones institucionales... y como punto de partida, la reflexión y explicitación del sentido de la educación, sentido que marca las diferencias más profundas en nuestra comprensión de los problemas.

## Evaluación Educativa

La distancia entre medición de los aprendizajes y decisiones educativas se hace patente en los debates sobre el Plan Nacional Aprender Matemática del Ministerio de Educación (2018). Este plan se lanzó “con el propósito de innovar en la estrategia tradicional de enseñanza utilizada por los docentes”<sup>3</sup>. A partir de los resultados de APRENDER se argumentó la necesidad de mejorar la enseñanza de la Matemática, pero el modo y las estrategias para mejorar la enseñanza no podían derivar de modo directo de los resultados de las pruebas. Frente a las características de la capacitación docente propuesta y el método de enseñanza promovido por el Plan Nacional Aprender Matemática, hubo un importante debate establecido por especialistas en Didáctica de la Matemática y numerosos matemáticos. Estos señalaron que la decisión de capacitación se orienta sobre principios con los que no acuerdan: el plan supone que existe un único método que puede ser aplicado en cualquier situación y contexto, y sus fundamentos se basan en ideas propias de los años 60 y 70 que hoy estarían ampliamente superadas<sup>4</sup>. Este solo ejemplo da cuenta de la profunda distancia que hay entre medir un problema y las decisiones que pueden tomarse tratando de encontrar soluciones.

La evaluación reducida a la medición de los aprendizajes no deja de informar desde 1993, que hay una importante discrepancia entre lo que se espera que los chicos y jóvenes dominen en los campos de Lengua, Matemática, Ciencias Naturales y Ciencias Sociales, y lo que se observa en los resultados de las pruebas. Desde entonces, diferentes gestiones del Ministerio de Educación de Nación, de las jurisdicciones y de las escuelas, han llevado adelante múltiples y muy variados proyectos de mejora educativa. Tan largo recorrido debería decirnos algo sobre los cambios en la educación, pero fundamentalmente, sobre la utilidad de este tipo de pruebas para sopesar un universo tan complejo y diverso.

La alarma de incendio no dice dónde comenzó el fuego ni cuál fue la causa. No es necesario que la alarma suene cada año en el país, en cada jurisdicción y en cada escuela. Vayamos a lo relevante. La evaluación a gran escala basada en la medición de aprendizajes

## Evaluación Educativa

y sus factores asociados no es ayuda para mejorar la vida y las repercusiones sociales de la escuela. Sí ayuda, en cambio, a legitimar discursos y decisiones en las instancias jerárquicas del sistema educativo entre quienes no estén advertidos de sus limitaciones.

### Referencias bibliográficas

Ausubel, D. (1976). Psicología educativa. Un punto de vista cognoscitivo. México: Trillas.

Marchesi, A. y Martín, E. (2000). Calidad de la enseñanza en tiempos de cambio. Madrid: Alianza Editorial.

Stufflebeam, D. y Shinkfield, A. (1987). Evaluación sistemática. Barcelona: Paidós.