

Comentarios al capítulo 6

Evaluación e investigación educativa

TERESA BRACHO*

El trabajo, en el contexto del libro en que se presenta, me parece una importante aportación, pues sin duda la investigación educativa tiene mucho que decir sobre la evaluación de la educación en México.

Primero. Es realmente muy poco lo que se ha escrito sobre evaluación educativa en México y las más de las veces se trata de textos de toma de posición o de demarcación de líneas y territorios ente los enfoques cuantitativos *versus* los cualitativos, los académicos *versus* los evaluadores, e incluso entre los políticos y administradores de la educación *versus* las organizaciones gremiales. Un buen texto sobre evaluación es una deuda pendiente en la literatura académica de México y el texto presentado por Martínez Rizo y Santos es un aporte que constituirá una referencia importante en el futuro.

* Profesora e investigadora de la Facultad Latinoamericana de Ciencias Sociales (FLACSO), sede México. CE: teresa.bracho@flacso.edu.mx

En mi comentario trataré de abordar algunas cuestiones que los autores dan por sentado, y que juzgo importante explicitar o reiterar para cualquier lector. Concluyo con mi propia visión sobre el lugar de la evaluación educativa en el proceso de políticas públicas en el ámbito educativo y en el de investigación educativa.

I

¿Qué significa evaluar en educación? Aunque parece obvio, creo que es razonable acotar las definiciones que pueden encontrarse en la literatura internacional para expandir la clasificación de Ravela que utilizan los autores en el texto.

De entre las definiciones más conocidas pueden destacarse las siguientes:

- 1) Evaluación educativa como el proceso que indica el grado hasta el cual se han cumplido los objetivos de la educación. Asociada a ésta, más recientemente se pueden encontrar definiciones que incluyen la investigación sistemática sobre el valor de un objeto. A mi parecer, esta especificación es muy importante no sólo porque refiere al *reconocimiento del valor de aquello que se evalúa*, sino también porque trata de identificar la función de la evaluación como identificación del grado del alcance en los objetivos logrados en el proceso educativo.
- 2) Desde otra perspectiva, el énfasis en la definición se pone en su función: ofrecer elementos de información para la toma de decisiones sobre el proceso educativo. Desde este punto de vista, se busca eliminar el carácter de "juicio de valor" contenido en el primer conjunto de definiciones, para poner énfasis en el *contenido informativo* del proceso evaluativo y subrayar el *uso de la información en la toma de decisiones* para la mejora de los procesos evaluados.

Las definiciones anteriores van asociadas a dos tipos de evaluación, o de sus funciones, que pueden claramente distinguirse en la literatura y que los autores del trabajo comentado mencionan: la evaluación de resultados o sumativa (*summative evaluation*), que refiere a aquella que busca constatar los resultados del proceso; usualmente, las evaluaciones de logro académico se ubican en este tipo de función. Y la evaluación formativa (*formative evaluation*), que es aquella que tiene lugar en el momento en que el proceso evaluado (sea una política, un programa educativo) que está en curso y busca reorientar las acciones para un mejor resultado.

Es fácil identificar las coincidencias de algunas de estas clasificaciones con las de Ravela, quien refiere a *los usos* diferentes de la evaluación, más que a sus tipos. Probablemente la diferencia tradicional sobre evaluación formativa y sumativa corresponda a lo que los autores, junto con Ravela, identifican como “duras” y “blandas” pero que, a mi juicio, se sobrepone con la clasificación que refiere a la certificación (ésta, por definición, sería una evaluación dura, al menos desde mi perspectiva).

La razón relevante, desde el punto de vista de la investigación evaluativa, es que el *propósito buscado* a que refieren los distintos tipos de evaluación, tiene claras consecuencias técnicas sobre el diseño mismo de la evaluación, el tamaño de muestra (o, en su caso, censo), de los instrumentos de medición y los reactivos utilizados, de los sujetos a quienes se aplican los instrumentos, así como el diseño del levantamiento de la información. Dicho de otra manera, no todas las pruebas otorgan información relevante sobre distintas cosas. Una evaluación bien diseñada en todos los casos puede dar información sólo sobre los propósitos buscados en su proyecto y, por lo tanto, su interpretación y su uso deben acotarse a éstos. De igual forma, no necesariamente pueden extrapolarse resultados obtenidos de diseños generados para medir el rendimiento de los alumnos de manera automática, a los resultados institucionales, si no se conocen y evalúan al mismo tiempo las condiciones institucionales y de contexto, por ejemplo.

Buena parte de las “malas interpretaciones” o de las interpretaciones simplistas de la evaluación educativa, que señalan los autores del trabajo comentado, refieren al uso inadecuado de sus resultados. Valdría la pena discutir de quién es esta responsabilidad. Si el evaluador se tomó el trabajo de acotar sus resultados, o no, si el mal uso deriva de una intención “oculta” para sesgar la interpretación que hace algún sujeto o entidad que no es el evaluador, etc.

Pero valdría agregar que es probable que ello se asocie también a las expectativas previas que de las pruebas se tiene: hay quienes tienden a sobrevalorar las posibilidades de la evaluación, otros por el contrario, tienden a desacreditarla de entrada o a minimizar su potencial. Los primeros probablemente asumen (sin un criterio adecuado sobre la naturaleza de la medición, la investigación y de la evaluación educativa) que toda medición del aprendizaje es “científica” y/o técnicamente apropiada; que por tanto, tiene validez para sustentar *cualquier* tipo de decisión sobre el proceso educativo (sea para establecer un ranking, para diseñar incentivos a los actores, certificaciones o juicios sobre el sistema general). Los segundos, parten de un juicio contrario sobre el hecho de que la evaluación educativa, en tanto que necesariamente acota su objeto de atención en función de su propósito, es absolutamente inadecuada para emitir *cualquier juicio* sobre el proceso educativo. Ambas posiciones parten de una pre-noción sobre el potencial de los recursos evaluativos y sus usos, más que de un análisis riguroso de cada uno de estos procesos de evaluación, con el fin de poder saber cuándo una evaluación sirve a un propósito, y lo hace de manera adecuada, aunque al mismo tiempo, pueda no servir a otros usos, de suyo deseables.

II

Una diferencia de matices –o énfasis– entre lo expresado en el texto y mi posición personal: Sobre la calidad en los resultados de la evaluación y la incorporación de grupos antes excluidos. En más de una ocasión se refie-

re a que los bajos resultados en las evaluaciones mexicanas pueden explicarse por la incorporación de grupos socioeconómicos en desventaja, o a la incorporación en la segunda mitad del siglo XX de grupos antes excluidos del sistema. Mi opinión es que es válida esta afirmación, pero sólo *parcialmente*. Me explico:

Es válido insistir en tener en consideración las condiciones de contexto socioeconómico familiar en la interpretación de resultados, pero también debe aclararse que se trata de ofertas educativas diferentes. Insisto, no sólo son niños y sus familias diferentes, sino que por muy diversas razones, en las que no abundo, se trata de escuelas distintas, de maestros con diferente origen y calidades. Valga reiterar, hay que evitar la conclusión de que "los pobres siempre tienen resultados bajos, por ser pobres"; tal vez insistir en que esos niños han asistido a escuelas "pobres".

Es cierto que uno de los posibles factores de explicación de los resultados bajos puede ser el cambio de la población escolar a partir de mediados del siglo XX. Pero esto es sólo una suposición sobre la que no hay elementos de prueba rigurosa. No sabemos qué resultados obtenían los estudiantes del sistema mexicano hacia mediados del siglo pasado, ni podemos estar seguros que si se les hubiera sometido a una prueba de aprendizaje hubiesen alcanzado niveles similares a los de los países desarrollados. Es plausible suponer que probablemente en algunas dimensiones del conocimiento escolar, sus resultados hayan sido superiores a los que actualmente podrían reportarse, pero no en todas las áreas de conocimiento. En suma, sin tener información comparable y precisa sobre el desempeño de la educación en momentos anteriores, todo juicio sobre la mejora o caída de la calidad es, cuando mucho una hipótesis plausible.

Tampoco es sostenible, a mi juicio, que sea la heterogeneidad de la demanda educativa, o las condiciones de los grupos sociales menos favorecidos los que sostengan la explicación de un posible menor rendimiento. Es mi opinión, que ello es explicable por las condiciones de la oferta y

no de la demanda. La expansión de las oportunidades educativas fue producto de una gran preocupación por la cobertura, por la incorporación a la escuela y no por su calidad. Y la educación fue uno de los bienes sociales que mejor se distribuyeron. Reconozcamos que en su momento no se levantaban las voces en torno a la calidad educativa, sino a la atención escolar, a elevar los promedios de escolaridad. El supuesto, como señalan los autores, de que la evaluación de los docentes era suficiente para constatar que los resultados educativos esperados se alcanzaban, no se cuestionaba. Pero ello no se asocia a la heterogeneidad de la demanda, sino a la heterogeneidad en la composición de los docentes.

Adicionalmente, la expansión educativa no se preocupó por ajustar los modelos de enseñanza a las condiciones de demanda y la transferencia de un modelo educativo implícito en el sistema tradicional para incorporar a nuevos grupos, de suyo en desventaja, se hizo no sólo con menos recursos, lo que puede contribuir a entender por qué condujo a resultados pobres. Pero el factor de explicación no son los niños que se incorporan, ni sus familias. Es la oferta deficiente para esos grupos.

Por último, el sistema de administración de la educación tampoco cambió para ajustarse a la expansión del siglo XX, sino que simplemente creció. Así como puede pensarse que la atención a todos los niños en las escuelas hubiese requerido ajustar los modelos de educación, también requería un cambio en su administración, supervisión, etcétera. Habrá que discutir si esa administración es la adecuada para conseguir una educación de calidad para todos.

III

En otro tema, pero asociado al anterior, me parece que hay una desmesurada valoración sobre la capacidad de los maestros para hacer “buena” evaluación. Entiendo el sentido en que se pondera a favor de éste, como una forma sustantiva de “evaluación formativa”, para volver al inicio de

este comentario, pero pareciera más de una petición de principio que de una realidad sobre las formas como los docentes y los directivos evalúan y juzgan el desempeño y el progreso de sus estudiantes.

Quiero sólo reiterar, junto con las ideas expresadas por los autores, que en tanto se trata de la evaluación formativa sustantiva del sistema, pues es la que sirve a ajustar de manera inmediata las acciones de la intervención escolar o del proceso de enseñanza aprendizaje, sus funciones no sustituyen a la evaluación de resultados de procesos, a la evaluación de orden más general, que permite emitir otra clase de juicios sobre el sistema en su conjunto. Y agregaría énfasis en la idea de que debemos formar mejor a los actores escolares, maestros, directores y supervisores, para desempeñar de mejor manera esta función.

Valdría la pena insistir en que necesitamos un mejor conocimiento de las formas como se evalúa el rendimiento en el espacio escolar, así como las capacidades y recursos con que cuentan las escuelas y los docentes para realizar esta función.

IV ...

Para concluir baste sólo una pequeña reflexión sobre los usos de los resultados de evaluación en la toma de decisiones. Buena parte de las discusiones sobre la evaluación educativa en el caso mexicano se ha centrado más en el tema de la *credibilidad* de los resultados y de sus usos posibles, más que sobre la calidad de su proceso de construcción, o, en otras palabras, sobre su rigor metodológico, la calidad de su información o su "cientificidad". Pareciera que en México se ha puesto más énfasis en la discusión sobre qué puede hacerse con los resultados de la evaluación, más que sobre *qué puede ser evaluado, qué es pertinente en este momento someter a evaluación; cómo y cuáles son las interpretaciones válidas de las distintas evaluaciones y, por tanto, de sus usos legítimos.*

Debe reconocerse que la evaluación sistemática se introdujo en un contexto sociopolítico de la transición política, y en el debate público se dio prioridad al tema de la credibilidad del cambio y de la racionalidad de las políticas públicas. Fuera de los círculos de expertos y los técnicos incorporados en las nuevas instituciones de evaluación, como el INEE, pocos en el ámbito político y menos aún en el público general, se preocuparon por preguntarse qué clases de evaluación eran viables, pertinentes, y para qué propósitos.

Es por ello que en la discusión pública sobre la evaluación de la educación en nuestro país se dio más importancia a las ideas de “transparencia de los resultados”, “clasificación de escuelas” y “rendición de cuentas desde las escuelas”, que a la atención pública informada sobre qué clase de evaluaciones se necesitaban, cómo y con qué recursos, pero sobre todo, sobre la discusión sustentada de qué habría que hacer con sus resultados.

En ello el Instituto Nacional para la Evaluación de la Educación desde su inicio se ha esmerado por atender explícitamente a esos propósitos de calidad técnica, pertinencia, interpretabilidad de los resultados de las evaluaciones y difusión de los mismos con criterios claros sobre los límites de aplicación válida de sus productos.

Pero la temática sobre evaluación educativa en nuestro país recién inicia. Ya existe un Instituto que ha sido capaz desde su arranque de atender a estos problemas. Ya existen distintas pruebas (¿serán demasiadas?) que indican diferentes factores del proceso educativo y diseñadas con propósitos diversos. Avanzar para completar un panorama evaluativo de todo el complejo sistema educativo seguramente no irá por la creación de otras pruebas de aprendizaje (aunque siempre habrá espacio para mejorarlas técnicamente), sino del diseño de nuevos instrumentos que nos acerquen de mejor manera a todas las condiciones contextuales explicativas en que se producen esos resultados de aprendizaje: desde una mejor identificación de los contextos sociales, económicos y culturales, así como de las condiciones escolares, docentes y de la propia administración educativa.

Debemos de reconocer que todos los tomadores de decisión en cualquier política, toma en cuenta la información que tiene al alcance y que juzga pertinente a su objeto de decisión. Los resultados de las evaluaciones más recientes de nuestro sistema educativo, incluidos los derivados de las pruebas estandarizadas de aprendizaje, son ahora un recurso obligado en el proceso de toma de decisión. Sin embargo, los límites de pertinencia de los referentes de la evaluación, su tipo y alcances son muy diversos, y no constituyen todos los elementos que informan las decisiones en el ámbito de la política educativa. Sería simplista y arrogante pensar que así es.

Los resultados de una evaluación en educación, por naturaleza técnica, son muy complejos, delicados en su interpretación y muy lejanos en su carácter a la información utilizable en el discurso argumentativo de un tomador de decisiones. Es decir, difícilmente pueden ser usados directamente en la racionalidad de la toma de decisiones, y más bien suelen utilizarse para argumentar hacia uno u otro sentido de la decisión. Y es, a mi juicio, responsabilidad de los evaluadores el lograr que ese proceso de traducción de las evaluaciones hacia un componente plausible en la argumentación de la toma de decisiones tenga lugar de manera que no comprometa los principios técnicos, referenciales y conceptuales a que está obligada la evaluación.

V

Y la discusión apenas comienza.... También en los espacios académicos de investigación educativa, poco se ha trabajado en las investigaciones sobre la evaluación educativa. Y ahí podríamos identificar posturas predefinidas sobre la credibilidad e incluso, del valor científico, de las evaluaciones. Abrir más espacios a esta reflexión y a la discusión sistemática sobre qué hacer, cómo hacerlo y para qué juzgamos pertinente hacerlo desde las instituciones académicas, es aún una tarea pendiente.