

Reflexiones y recomendaciones sobre las evaluaciones educativas en América Latina

La realización de evaluaciones estandarizadas a gran escala para conocer la dinámica de procesos y los resultados en los sistemas educativos es cada vez más frecuente a nivel mundial, y América Latina no está al margen de esta tendencia. Si bien los países de la Región están avanzando en este ámbito, es preciso evitar el simplismo y la ingenuidad con que todavía muchos piensan en la evaluación, lo que genera sistemas mal concebidos y deficientemente implementados y, en consecuencia, produce efectos perversos para el sistema educativo, malgasto de recursos y descrédito de la evaluación externa entre los docentes.

En esta edición se reflexiona sobre la importancia de las evaluaciones educativas, los fines que persiguen y los criterios básicos para la divulgación y uso de sus resultados. Además, se formulan algunas recomendaciones para la Región en esta materia.

Extractado de “Las evaluaciones educativas que América Latina necesita”, Documento N° 40 de PREAL - Grupo de Trabajo sobre Estándares y Evaluación, marzo 2008. Disponible en www.preal.org/Publicaciones

La evaluación estandarizada en gran escala es aquella que permite producir información comparable acerca de los desempeños de estudiantes de distintos contextos culturales y regionales e, incluso, de distintos países, y que ofrece un panorama de la situación de un país o de un estado o provincia, aun cuando ello se haga a veces a través de una muestra no demasiado grande.

En América Latina, siguiendo la tendencia mundial, durante los años 90 se desarrollaron con fuerza sistemas nacionales de evaluación estandarizada. Algunos países han mantenido sus sistemas en funcionamiento con continuidad, aun cuando se produjeran cambios de enfoque o variaciones en la institucionalidad de los mismos. Otros países han tenido importantes discontinuidades y han debido comenzar prácticamente desde cero en repetidas oportunidades o tendrán que hacerlo en el futuro próximo.

En términos generales, se puede decir que las evaluaciones a gran escala en la región no solo son más frecuentes, sino también están mejorando en diversos aspectos: mayor transparencia en la difusión de los resultados; una tendencia a pasar de pruebas *normativas* (que tienen como propósito principal ordenar comparativamente a los estudiantes) a pruebas *de criterios* (que se enfocan en qué es lo que los estudiantes saben y

son capaces de hacer); una mejoría de las capacidades técnico-metodológicas para la construcción de pruebas y para el procesamiento de los datos; una creciente atención a la difusión y uso de los resultados; mayor preocupación por la investigación de los factores que inciden sobre los aprendizajes; y creciente participación de los países en evaluaciones internacionales (ver recuadro 1).

No obstante lo anterior, subsisten debilidades que necesitan ser superadas, principalmente en lo que respecta a las estrategias de difusión, el uso de los resultados y algunos aspectos claves de la calidad técnica de las evaluaciones. Ellas no pueden ser confrontadas de modo independiente, por estar fuertemente vinculadas entre sí. Debido a ello, el desarrollo futuro de sistemas de evaluación estandarizada, o la modificación de los existentes, requiere de una nueva reflexión sobre la importancia de este tipo de evaluaciones, cuáles son sus usos y propósitos, los desafíos técnicos que plantean y los criterios más apropiados para el uso y difusión de los resultados, de manera que efectivamente sirvan para mejorar la educación.

¿Por qué importan las evaluaciones educativas?

La mayoría de los países asumen que la evaluación estandarizada puede servir: (i) como base para adoptar políticas educativas mejor fundamentadas, (ii) para mejorar la gestión de los sistemas educativos y (iii) como instrumento para la colaboración y el aprendizaje continuo al interior de los mismos. Por otra parte,

crecientemente se reconoce que estas evaluaciones pueden ser un aporte desde distintas perspectivas:

- Ayudan a visualizar los resultados educativos del conjunto de estudiantes.
- Entregan información sobre el real acceso al conocimiento y sobre las capacidades que alcanzan los alumnos, más allá de la cantidad de años de estudio que tengan.
- Ayudan a hacer visibles aspectos centrales de la labor educativa: en qué medida los alumnos están aprendiendo lo que se espera de ellos al finalizar ciertos grados; cómo evolucionan a lo largo de los años los niveles de logro y la equidad en el acceso al conocimiento; cómo inciden las desigualdades sociales y culturales sobre las oportunidades de aprendizajes; cuál es la diversidad de prácticas educativas existentes en escuelas y maestros y cómo se relacionan con los aprendizajes de los estudiantes en diversos contextos sociales; cómo influyen las condiciones de la enseñanza en los progresos de los alumnos; y qué efecto tienen las inversiones en programas educativos, los cambios en la estructura del sistema, los cambios curriculares, los programas de formación y la adquisición de material educativo, entre otros.
- Aportan información importante para las decisiones que competen a las autoridades y a los encargados de formular políticas educativas, a los directivos y docentes, a los supervisores y encargados de la formación de docente, a los alumnos y sus familias, y a la ciudadanía en general.

Por otra parte, el desarrollo de un sistema nacional de evaluación estandarizada puede obligar a realizar una discusión informada sobre lo que pueden y deben haber aprendido los estudiantes al final de cada ciclo educativo, y sobre la definición de estándares y características del currículo.

Los propósitos de las evaluaciones

La implementación de un sistema de evaluación, o su reforma, requiere primero que nada definir cuál será su propósito y para qué se usarán sus re-

Recuadro 1

Participación en pruebas internacionales

Además de desarrollar sistemas nacionales de evaluación, muchos países latinoamericanos están participando en instancias internacionales de evaluación e intercambio de información sobre este tema.

- Dieciséis países participaron en el Segundo Estudio Regional en 3° y 6° de primaria de OREALC/UNESCO: Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay.
- Seis de estos países participaron en PISA 2006 y otros tres se incorporarían al ciclo PISA 2009.
- Algunos países de la región han participado además (o lo están haciendo) en los estudios de Matemática y Ciencias (TIMSS), Lectura (PIRLS) y Educación Cívica que lleva adelante la International Association for the Evaluation of Educational Achievement (IEA).
- Desde 1994 ha estado activa en la región la red de los sistemas de evaluación organizada en torno al Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), de OREALC/UNESCO, en la que participan actualmente casi todos los países latinoamericanos. Esta red se reúne dos veces al año y es un espacio de intercambio de experiencias y formación en el ámbito de la evaluación.
- Más recientemente, se ha constituido el Grupo Iberoamericano de PISA, integrado por Argentina, Brasil, Chile, Colombia, España, México, Portugal y Uruguay.

sultados. Es preciso considerar diversas opciones, que no son necesariamente excluyentes. Un sistema puede combinar varias de ellas, pero es importante tener conciencia de que cada opción necesita de un determinado diseño y tiene diferentes exigencias técnicas y de costos. Se debe analizar si el sistema nacional de evaluación será:

- *De carácter diagnóstico* (para alumnos, escuelas o sistema educativo), sin consecuencias directas para los estudiantes, cuya finalidad principal es enriquecer con información de calidad las percepciones, decisiones y acciones de diversos actores del sistema para mejorar la enseñanza y el aprendizaje.
- *De certificación de los logros educativos de los estudiantes*, cuyo propósito central es establecer quiénes han alcanzado los conocimientos y desempeños necesarios para aprobar un curso o nivel y, en consecuencia, determinar si aprueban o reprueban.

En las evaluaciones de tipo diagnóstico, los costos de aplicación de las pruebas pueden ser menores, dado que pueden hacerse en base a muestras y aplicarse solo en algunos grados claves y cada cierto número de años. Por otra parte, las pruebas pueden diseñarse de un modo en que no todos los estudiantes

responden a las mismas preguntas, sino a bloques de las mismas, lo cual permite trabajar con una cantidad muy grande de preguntas y obtener un análisis más detallado de los distintos aspectos del currículo. Estos sistemas permiten definir estándares o niveles de expectativa exigentes sin que ello produzca un fracaso generalizado. Asimismo, ayudan a construir una cultura de la evaluación y a acumular capacidad técnica para cuando se analice la posibilidad de establecer un sistema con consecuencias. El principal problema de estas evaluaciones es que pueden carecer de impacto sobre las políticas, sobre las prácticas de enseñanza y sobre la motivación de los estudiantes hacia el aprendizaje escolar, a menos que se tomen algunas acciones complementarias, tales como invertir tiempo en el análisis y discusión de los resultados por parte de diversas áreas del Ministerio de Educación y otros actores relevantes; realizar entre docentes una interpretación de los resultados desde la perspectiva de sus implicancias didácticas; y proveer a las familias y alumnos información apropiada y comprensible acerca de lo que se considera imprescindible haber aprendido en cada grado o nivel educativo y de las acciones que pueden contribuir al logro de dichos aprendizajes.

En el caso de las evaluaciones con fines de certificación, existen ventajas asociadas a la transparencia y responsabilidad por los resultados. Hoy es común que dos estudiantes que han aprobado un mismo nivel educativo en dos regiones distintas de un país, posean en realidad niveles de conocimiento muy diferentes. La evaluación para la certificación da transparencia al valor de los certificados educativos ante la sociedad y hace responsables tanto a los docentes como a los propios estudiantes por alcanzar los conocimientos y capacidades que evalúa el examen, lo cual tiene impactos positivos sobre el aprendizaje. Sin embargo, estos sistemas implican algunos requerimientos que son costosos: las pruebas deben ser de carácter censal y cada estudiante debe tener más de una oportunidad para rendirla, lo cual requiere implementar varios operativos de evaluación cada año, y deben tener una amplia cobertura curricular, lo cual requiere que sean extensas, en lo posible con preguntas de respuesta construida (que tienen un importante costo de codificación) y deben abarcar varias asignaturas o disciplinas. Adicionalmente, estos sistemas pueden generar tensiones importantes que es necesario anticipar: si las pruebas son exigentes, ello puede conducir a niveles de reprobación muy altos, que afectarían principalmente a los sectores sociales más vulnerables y podrían agudizar problemas como la deserción y el desempleo juvenil. Por ello, es preciso acompañar este tipo de evaluaciones (en realidad todo tipo de evaluaciones, pero estas en particular) con acciones pedagógicas de carácter complementario y compensatorio. Proponer estándares altos implica asumir responsabilidad sistémica por proporcionar todas las oportunidades para aprender: textos, materiales, instalaciones y preparación de maestros, entre otros.

Unidades de Evaluación

Para llevar adelante una política de evaluación de logros educativos o aprendizajes apropiada, es necesario contar con Unidades de Evaluación con las capacidades para asumir los desafíos de calidad técnica que estos implican (ver recuadro 2) y con los recursos

adecuados que requiere este complejo emprendimiento. La puesta en marcha de un sistema de evaluación serio requiere de un tiempo mínimo de entre dos y tres años, y las Unidades de Evaluación necesitan tener una cierta estabilidad en el tiempo de sus elencos técnicos, además de contar con independencia para reportar los resultados de su trabajo. Dicha independencia no debe dar lugar a su desvinculación de la política educativa. Por el contrario, la evaluación debe responder a un proyecto político-educativo con amplio respaldo y debe mantener una estrecha relación con otras áreas claves de la política educativa como la formación docente, el desarrollo curricular, el planeamiento y formulación de proyectos, la evaluación de programas y la investigación.

De más está decir que evaluar bien requiere de inversión. Más vale no hacer ninguna evaluación que hacer una mala o deficiente. Y es preferible un sistema de evaluación modesto pero con costos asequibles y sostenibles en el tiempo, que una evaluación amplia y sofisticada, que solo podrá realizarse una vez y luego no podrá replicarse.

Divulgación y uso de los resultados

Una vez realizadas las evaluaciones, es necesario tener en cuenta los siguientes criterios básicos en la difusión y uso de sus resultados:

- La responsabilidad por los resultados debe ser compartida por diversos actores: las autoridades nacionales y regionales, los cuerpos docentes, los estudiantes y sus familias. Debería evitarse todo uso de los resultados con el fin deliberado o implícito de culpabilizar o responsabilizar de manera exclusiva a ciertos actores.
- La responsabilidad por los resultados requiere de una adecuada combinación entre las exigencias y los apoyos a las escuelas y a los docentes. Por un lado, es necesario que escuelas y docentes asuman su responsabilidad respecto al objetivo de que todos los estudiantes aprendan lo que se espera de ellos. Simultáneamente, es deber de las autoridades establecer las políticas de apoyo necesarias para que escuelas y docentes puedan desarrollar su labor

de manera apropiada. Un esquema de exigencias altas sobre escuelas y docentes sin los correspondientes apoyos solo puede generar malestar y desánimo. Un esquema de apoyos sin contrapartes y exigencias puede generar autocomplacencia.

- Es inapropiado utilizar los resultados de las pruebas estandarizadas como indicador principal de la calidad del trabajo del docente o de la escuela. Esto es importante, en particular si no se ha controlado el efecto de otros factores internos y externos al sistema educativo y si no se tiene en cuenta que el aprendizaje depende también de la motivación y esfuerzo personal del estudiante. Además, dado que los resultados de pruebas estandarizadas no

Recuadro 2

Los desafíos de calidad técnica de las evaluaciones

Tras definir la política de evaluación, es necesario implementarla de acuerdo con diversos estándares apropiados de calidad técnica. Esto implica, entre otras cosas:

- Elaborar un referente o marco conceptual que defina con precisión los conocimientos y desempeños que son considerados apropiados al finalizar el grado o ciclo educativo que será evaluado, para luego traducirlo adecuadamente en especificaciones y estándares de carácter técnico.
- Incluir en las pruebas actividades con diversos grados de complejidad, que permitan informar sobre el punto en que se encuentran los estudiantes más avanzados y los menos avanzados.
- Definir los “puntos de corte” que establecen los límites entre los niveles de desempeño de una prueba.
- Diseñar muestras apropiadas a los propósitos de la evaluación.
- Definir cómo se harán comparables los resultados de pruebas aplicadas en distintos años.
- Construir datos longitudinales que permitan apreciar la evolución en el tiempo de los aprendizajes de un mismo conjunto de estudiantes.
- Establecer mecanismos de control del cumplimiento de las condiciones estandarizadas que deben regir durante la aplicación de las pruebas, para que la información sea confiable y comparable.

son el único indicador de la calidad de la educación, es preciso considerarlos en conjunto con otros aspectos relevantes como la pertinencia y relevancia de lo que se enseña, la retención, la formación de valores y hábitos o la formación ciudadana. Debe evitarse identificar el término “calidad” con resultados de pruebas estandarizadas.

- Para facilitar la comprensión y uso de los resultados, no basta con ofrecer resultados numéricos. Es fundamental ilustrar a los distintos actores acerca del tipo de tareas que los estudiantes

debieron resolver en las pruebas. Este criterio debe combinarse adecuadamente con la necesidad de mantener el carácter confidencial de un conjunto de preguntas, que puedan emplearse en futuras evaluaciones para establecer mediciones comparables en el tiempo. Si bien no se pretende que los docentes utilicen pruebas estandarizadas para evaluar a sus estudiantes, el conocimiento y la comprensión de las mismas pueden contribuir a mejorar sus modos de evaluar y a desarrollar una cultura de la evaluación.

- Al reportar las diferencias de resultados en función de la composición sociocultural del alumnado, se debe cuidar de no crear un sistema de expectativas de logro diferenciado por grupos sociales. En todo caso, las diferencias deberían producirse en las políticas para crear condiciones apropiadas para enseñar a los grupos más desfavorecidos. Se debería descartar cualquier tipo de uso de los resultados que, directa o indirectamente, propicie la selección de estudiantes por parte de las escuelas a los efectos de mejorar sus resultados. ©

Diez recomendaciones para las evaluaciones en América Latina

Luego de analizar aspectos como los propósitos de las evaluaciones estandarizadas, el uso de sus resultados y los desafíos institucionales y de calidad técnica que plantean, el Grupo de Trabajo sobre Estándares y Evaluación de PREAL formula diez recomendaciones generales:

La evaluación debe concebirse como un elemento articulado en un conjunto más amplio de acciones y políticas educativas. La evaluación por sí misma no produce mejoras. Deben existir articulaciones estables entre el área de evaluación y las áreas de desarrollo curricular, formación docente, investigación, diseño de políticas, comunicación y prensa, entre otros.

La evaluación debe contemplar un proceso de reflexión colectiva sobre el estado de la educación y los caminos para mejorarla. El primer paso es la consulta y el debate público sobre qué deben aprender los estudiantes y cuáles serán los propósitos y consecuencias de la evaluación. También es clave la discusión pública a partir de los resultados. Para ello hay que invertir en comunicación y difusión tantos o más recursos que en la propia evaluación, antes, durante y después de la misma.

La evaluación debe estar al servicio del desarrollo de un sentido de responsabilidad compartida por la educación como bien público. Debe promover el compromiso con la educación de todos los actores, cada uno según su lugar y ámbito de acción. Por este motivo, debe evitarse utilizar la evaluación para culpabilizar a actores específicos por los problemas detectados.

Los sistemas de evaluación necesitan ampliar progresivamente el abanico de fines educativos que son objeto de evaluación. Es importante en este sentido incluir a la formación ciudadana y otras disciplinas además de Lenguaje y Matemática, así como un espectro más amplio de competencias y capacidades.

Se requiere diseñar progresivamente evaluaciones de progreso de los estudiantes. Con sistemas que prevén la realización técnicamente correcta de dos o más evaluaciones a los mismos alumnos o cohortes en distintos momentos, para constatar sus avances en el aprendizaje, se puede aportar mayor información acerca del impacto de las políticas educativas, de las acciones de las escuelas y de las prácticas de enseñanza en los aprendizajes de los estudiantes.

Un sistema de evaluación es un proyecto de largo plazo, por lo que requiere de un compromiso del Estado y de una planificación cuidadosa del diseño del sistema. Ello exige tomar decisiones acerca de los fines, las consecuencias, los aspectos a evaluar, las áreas y los grados, y la periodicidad de las evaluaciones, entre otras cosas. Esto toma tiempo, por lo que no es conveniente pretender la implementación de sistemas de evaluación en plazos reducidos.

Un buen sistema de evaluación requiere inversión. Esa inversión se necesita principalmente en la conformación de equipos humanos calificados y en recursos económicos suficientes para una adecuada implementación de todos los procesos implicados.

El sistema de evaluación debe sustentarse en una actitud firme de transparencia. En particular, se requiere transparencia con relación a los resultados y la rendición de cuentas a la sociedad.

Los Ministerios de Educación deben asumir un compromiso serio con los resultados de la evaluación. Lo anterior implica promover el diálogo acerca de los problemas detectados y los modos de enfrentarlos, diseñar líneas de acción apropiadas para resolver los problemas e invertir los recursos necesarios para llevarlas adelante.

Los sistemas de evaluación deben ser objeto de evaluación periódica. Esto permite analizar la calidad técnica de la información que producen y su relevancia para diversos actores educativos y sociales.