



Programa de Promoción de la Reforma Educativa en América Latina y el Caribe
Partnership for Educational Revitalization in the Americas

DOCUMENTOS

Nº 40

Las evaluaciones educativas que América Latina necesita*

Marzo 2008

Pedro Ravela, Uruguay
Patricia Arregui, Perú
Gilbert Valverde, Estados Unidos
Richard Wolfe, Canadá
Guillermo Ferrer, Argentina
Felipe Martínez Rizo, México
Mariana Aylwin, Chile
Laurence Wolff, Estados Unidos

* Los autores de este documento son los miembros del Comité Gestor del Grupo de Trabajo sobre Estándares y Evaluación del PREAL.

Las opiniones vertidas en este trabajo son responsabilidad de los autores y no comprometen a PREAL ni a las instituciones que lo patrocinan.

© 2008 Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL)

Todos los derechos reservados

Si desea solicitar copias adicionales, contacte al PREAL y al Inter-American Dialogue.
Este estudio puede ser descargado de: www.preal.org/publicacion.asp

Inter-American Dialogue
1211 Connecticut Avenue, NW, Suite 510
Washington, DC 20036
202-822-9002
iad@thedialogue.org
www.thedialogue.org y www.preal.org

Para citar: Ravela, Pedro; Arregui, Patricia; Valverde, Gilbert; Wolfe, Richard;
Ferrer Guillermo; Martínez Rizo, Felipe; Aylwin, Mariana; Wolff, Laurence,
Las evaluaciones educativas que América Latina necesita
(PREAL, Santiago 2008).

ISBN: 0-9772271-8-9
Primera Edición
Publicado en Chile

Diseño e Impresión:
Editorial San Marino
esm.cl

ÍNDICE

Presentación	5
I. ¿Por qué son importantes las evaluaciones nacionales de logros educativos?	6
II. La situación de la evaluación estandarizada en América Latina	9
III. Los propósitos y los usos de los resultados de las evaluaciones estandarizadas	12
Evaluaciones para la certificación de los aprendizajes de los estudiantes	12
Evaluaciones de carácter diagnóstico y formativo	13
Uso de las evaluaciones para establecer incentivos	15
Criterios básicos para la divulgación y uso de los resultados	15
Advertencias sobre la comparación de resultados entre escuelas	16
Itinerario para la toma de decisiones	17
IV. Los desafíos de calidad técnica de las evaluaciones	19
V. Sobre la constitución de las Unidades de Evaluación	20
VI. Diez recomendaciones acerca de las evaluaciones que la región necesita	22

PRESENTACIÓN

Este documento aborda la importancia, propósitos y usos de las evaluaciones estandarizadas a gran escala de aprendizajes y/o logros educativos en América Latina y el Caribe, básicamente en los niveles primario y medio. Está dirigido a quienes formulan políticas educativas, docentes, académicos, empresarios, sindicatos, organizaciones sociales, agencias de financiamiento y medios de prensa, con el fin de aportar al debate y decisiones sobre la evaluación estandarizada en los sistemas educativos.

Se entiende por evaluación estandarizada en gran escala a aquella que permite producir información comparable acerca de los desempeños de estudiantes pertenecientes a distintos contextos culturales y regionales e, incluso, a 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 (por ejemplo, 5.000 estudiantes)¹.

El documento se centra en evaluaciones de *aprendizaje* –definido como el cambio ocurrido en los conocimientos y capacidades de cada estudiante a lo largo del año lectivo– y/o de *logro educativo*, entendido como la acumulación de conocimientos y capacidades a lo largo de toda la vida del estudiante. No se pretende orientar aquí acerca de otros aspectos también importantes de la evaluación del quehacer educativo,

como son las evaluaciones del desempeño docente, de las políticas educativas, de los centros educativos o aquellas que realizan los docentes en las aulas, entre otras. Tampoco se pretende incursionar en la evaluación en el nivel terciario ni en las pruebas de selección para el mismo.

Los contenidos de esta publicación ayudarán a quienes toman decisiones de política educativa a comprender y analizar las diversas opciones existentes en cuanto a los propósitos y usos de los sistemas de evaluación y las implicancias de cada una de ellas, de modo de orientarlos en cómo concebir una estrategia de evaluación.

Es preciso evitar el simplismo y la ingenuidad con que muchos piensan en la evaluación, lo que genera sistemas mal concebidos y deficientemente implementados y, como consecuencia de ello, efectos perversos para el sistema educativo, malgasto de recursos y descrédito de la evaluación externa entre los docentes. Por el contrario, se debe reconocer la gran variedad de aspectos que deben considerarse al momento de implementar un sistema de evaluación o reformar uno existente. Para que la inversión en evaluación valga la pena, debe tener propósitos claros, una filosofía orientada a construir una visión de responsabilidad compartida en relación a la educación, un diseño técnico de calidad y adecuado a los propósitos, una orientación fuerte a apoyar a los docentes en su tarea y una voluntad política manifiesta de encarar acciones dirigidas a resolver los problemas y deficiencias que la evaluación ponga de manifiesto. Asimismo, requiere de una inversión importante para construir una unidad técnica competente y

¹ El término “en gran escala” no se emplea como sinónimo de evaluaciones censales.

un plan de largo plazo, lo cual lleva tiempo y no puede ser improvisado.

I. ¿POR QUÉ SON IMPORTANTES LAS EVALUACIONES NACIONALES DE LOGROS EDUCATIVOS?

La realización de evaluaciones estandarizadas como forma de conocer mejor la dinámica de procesos y resultados en los sistemas educativos es cada vez más frecuente a nivel regional y mundial, en países de muy diversas culturas y orientaciones ideológicas de gobierno. Prueba de ello es la creciente participación de los países en las evaluaciones internacionales como PISA, TIMSS y PIRLS, y regionales como SERCE (en América Latina) y SACMEQ (en África), así como el desarrollo de diferentes tipos de sistemas nacionales y subnacionales de evaluación.

En algunos casos, la evaluación está motivada por una preocupación por la formación ciudadana y la consolidación de una sociedad democrática; en otros, por la productividad de la fuerza de trabajo y la competitividad de la economía nacional, por las oportunidades para el desarrollo integral de las personas y sus posibilidades de participación en la sociedad del conocimiento, o por la equidad y la visión de la educación como uno de los caminos para superar la pobreza. Muchos sistemas de evaluación parten de una combinación de los intereses anteriores. En casi todos los casos se asume que la evaluación puede servir:

- como base para adoptar políticas educativas mejor fundamentadas;
- para mejorar la gestión de los sistemas educativos;
- como instrumento para la colaboración y el aprendizaje continuo al interior de los mismos.

A continuación se destacan los aportes principales de este tipo de evaluaciones.

La evaluación estandarizada ayuda a visualizar los resultados educativos del conjunto de estudiantes

La educación es una actividad “opaca”, en el sentido de que sus resultados no son directa ni inmediatamente observables. Un buen docente puede apreciar si sus estudiantes están aprendiendo o no y cómo, pero no todos los docentes tienen los mismos criterios de valoración, los cuales están fuertemente vinculados a su experiencia profesional: su formación general y específica, su conocimiento de la disciplina que enseña, su capacidad para percibir los procesos y dificultades de los estudiantes, su familiarización con diversos tipos de alumnos, etc. Dada la heterogeneidad del cuerpo docente de un país, no es posible obtener una visión de la situación del conjunto por la mera agregación de los puntos de vista individuales. La evaluación estandarizada apunta a lograrlo.

Las evaluaciones entregan información sobre el real acceso al conocimiento y a las capacidades que alcanzan los alumnos, más allá de la cantidad de años de estudio que tienen

En décadas pasadas, la equivalencia entre permanencia en el sistema educativo y acceso al conocimiento y a las capacidades simbólicas se daba por descontada y los indicadores para valorar a los sistemas educativos estaban relacionados con el acceso (matrícula, cobertura, retención, etc.). En esos tiempos, las mayorías más pobres y con menor capital cultural no accedían al sistema educativo o apenas cursaban unos pocos años de educación básica. Con la progresiva universalización del acceso al sistema educativo, hoy muchos estudiantes procedentes de los sectores sociales más desfavorecidos llegan en desventaja en términos de manejo del lenguaje oral (oficial) y escrito, con lo cual la equivalencia entre años de estudio y acceso al conocimiento y a las capacidades ya no es tal. Por otra parte, el crecimiento del acceso a la profesión docente no fue acompañado por garantías de calidad en su formación, lo cual también cuestiona la mencionada equivalencia.

Así, más años en el sistema educativo no necesariamente significa que todos los niños y jóvenes estén incorporando los conocimientos, actitudes y capacidades necesarias para la vida personal y social, los que son cada vez más sofisticados y complejos. Las evaluaciones buscan dar luces sobre lo que está ocurriendo al respecto.

Las evaluaciones estandarizadas ayudan a hacer visibles un conjunto de aspectos centrales de la labor educativa

Entre otras cosas, las evaluaciones estandarizadas aportan información sobre:

- en qué medida los alumnos están aprendiendo lo que se espera de ellos al finalizar ciertos grados o niveles;
- cuál es el grado de equidad o inequidad en el logro de dichos aprendizajes;
- cómo evolucionan a lo largo de los años, tanto los niveles de logro como la equidad en el acceso al conocimiento por parte de los diversos grupos sociales;
- en qué medida y cómo las desigualdades sociales y culturales inciden sobre las oportunidades de aprendizajes de los estudiantes;
- cuál es la diversidad de prácticas educativas existentes en escuelas y maestros y cómo las mismas se relacionan con los aprendizajes de los estudiantes en diversos contextos sociales;
- cómo influyen las condiciones de la enseñanza (situación de los maestros, recursos disponibles, tiempos de estudio, etc.) en los progresos de los alumnos;
- qué efecto tienen en los logros educativos las inversiones en programas educativos, los cambios en la estructura del sistema, los cambios curriculares, los programas de formación, la adquisición de material educativo, etc.

Un sistema de evaluación de aprendizajes y/o logros educativos puede aportar información importante a diversos actores sociales

En la medida que el sistema produzca y comunique adecuadamente información sobre los aspectos antes indicados, puede ser un instrumento clave de mejora, enriqueciendo la comprensión de la situación educativa y la toma de decisiones en diversos ámbitos:

- *Las autoridades y los encargados de formular políticas educativas* pueden comprender mejor los problemas de la enseñanza y del aprendizaje; hacerse cargo de las carencias en que se desarrolla la labor docente y desarrollar políticas pertinentes para apoyar el trabajo de las escuelas. Las evaluaciones también les permiten valorar, sobre una base de evidencia empírica sólida, el impacto de las políticas y programas que han impulsado y los probables efectos de las que se proponen impulsar.
- *Los directivos y docentes* pueden, a partir de una mirada externa sobre los logros educativos en el conjunto del sistema, comprender mejor lo que están logrando y lo que no están logrando sus propios estudiantes, cómo están aprendiendo y qué dificultades tienen. Pueden aprender de las experiencias de otros docentes y escuelas que trabajan con estudiantes de características tanto similares como diferentes a los propios. Pueden tomar decisiones más apropiadas acerca de qué aspectos del currículo enfatizar y enriquecer sus propios modos de evaluar el aprendizaje de sus estudiantes.
- *Los supervisores y encargados de la formación de docentes* pueden apoyarse en la información sobre los logros y dificultades educativas en el conjunto del sistema para estudiar en profundidad las debilidades en el enfoque o en la práctica de la enseñanza que probablemente generan algunas de las insuficiencias constatadas en los aprendizajes. Desde esa nueva perspectiva, pueden

mejorar su labor tanto de orientación a los docentes como de formación de los mismos. En particular, los supervisores pueden beneficiarse de contar, como instrumento de apoyo para su labor, con un mapa de las escuelas que las caracterice tanto en términos de la composición sociocultural de su alumnado como de sus logros educativos.

- *Los padres y madres de estudiantes*, adecuadamente informados, pueden comprender mejor qué se espera que sus hijos aprendan, qué es lo que están logrando y qué pueden hacer para colaborar con la escuela y con el aprendizaje de sus hijos.
- *La ciudadanía* en general estará mejor informada sobre lo que acontece al interior del sistema educativo y, por tanto, estará más atenta a los temas y problemas de la educación. Estará, además, en mejores condiciones para exigir, tanto a los poderes públicos como a los profesionales de la docencia, la mejora continua de la educación que se brinda a los niños, niñas y jóvenes, y el uso responsable de los recursos que se destinan a la educación.

El desarrollo de un sistema nacional de evaluación estandarizada puede obligar a realizar una discusión informada sobre qué aspectos del currículo formal son exigibles a todos los estudiantes y a definir con claridad qué es lo que todos los estudiantes deberían haber aprendido al final de cada ciclo educativo

La mayoría de los currículos en la región están constituidos por largas listas de objetivos y temas, todos deseables pero no todos realizables. La elaboración de pruebas nacionales obliga a definir qué es lo que debe ser considerado como fundamental y, por tanto, lo que todos los estudiantes deben saber y ser capaces de hacer. Se pueden usar distintos términos para designar a estas definiciones: estándares, competencias fundamentales, indicadores de logro, niveles de desempeño, metas de aprendizaje, criterios de suficiencia, entre otros.

Para no confundir...

Al desarrollar estos sistemas nacionales de evaluación, es necesario explicitar tanto el rol que estos cumplen así como las funciones que no cumplen y cuáles son sus limitaciones, para evitar así los riesgos de mala interpretación respecto a lo que son y la información que recogen y difunden.

- **Una evaluación estandarizada aporta información fundamental e indispensable sobre la “calidad educativa”, aunque no sea un indicador completo de la misma.** No todos los objetivos valiosos de la educación están incluidos en este tipo de evaluaciones. Hay una gran cantidad de saberes, actitudes, valores y aprendizajes relevantes que no pueden –por la dificultad de medirlos de esta manera– o no deben –porque son propios de cada entorno local y, por tanto, no son exigibles a todos los estudiantes del país– formar parte de una evaluación estandarizada.
- **La evaluación estandarizada de aprendizajes y/o logros educativos es un componente esencial de un sistema integral de evaluación educativa, pero no es el único tipo de evaluación relevante.** También tienen importancia la evaluación realizada en el aula por los docentes, la evaluación del desempeño docente, la evaluación de los centros educativos, la evaluación de las políticas educativas, la evaluación del uso de los recursos, la evaluación de la relevancia del currículo, etc.
- **La evaluación es condición necesaria, pero no suficiente, para mejorar la educación.** Si bien existe alguna evidencia de que la mera existencia y difusión de información tiene algún impacto sobre ciertos actores, es necesario recordar que la evaluación es apenas uno de varios elementos clave de la política educativa que debe procurarse estén adecuadamente alineados.

dos: la formación docente inicial y en servicio, la carrera y condiciones para el trabajo docente, la gestión escolar, la supervisión, los diseños curriculares, los libros y materiales educativos, una inversión de recursos proporcional a las necesidades de las diversas poblaciones y una acción decidida por parte de los responsables educativos para resolver los problemas detectados, entre otros.

- **La evaluación estandarizada externa solo tendrá efectos positivos sobre la educación si es concebida, percibida y empleada como un mecanismo de responsabilización pública de todos los actores vinculados al quehacer educativo.** Existe siempre el riesgo de que la política educativa se concentre en la implementación de evaluaciones, pero que luego no se tomen acciones concretas para enfrentar y resolver los problemas que las mismas ponen de manifiesto. Muchas veces las autoridades se limitan a informar de los resultados y transferir toda la responsabilidad por la solución de los problemas detectados a las escuelas y familias. Otras veces, toda la responsabilidad se atribuye a los docentes, quienes, por su parte, tienden a transferir toda la responsabilidad a los padres o al contexto. Lo importante es evitar ingresar en una dinámica de culpabilización y, por el contrario, intentar construir una lógica de responsabilidad compartida en torno a la educación.

II. LA SITUACIÓN DE LA EVALUACIÓN ESTANDARIZADA EN AMÉRICA LATINA

Los sistemas nacionales de evaluación estandarizada se desarrollaron con fuerza en toda la región durante los años 90. 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 tenido que co-

menzar prácticamente desde cero en repetidas oportunidades o tendrán que hacerlo en el futuro próximo. Además de ello, actualmente:

- Dieciséis países están participando en el Segundo Estudio Regional en 3° y 6° de primaria que lleva adelante la 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 participan 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).
- Durante los últimos 12 años 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.

Las evaluaciones a gran escala en la región no solo son más frecuentes sino también están mejorando

Los principales cambios y mejoras que se observan en los últimos años incluyen:

- ***Mayor transparencia en la difusión de los resultados.*** Parece haberse superado la etapa en que las autoridades en varios paí-

ses interferían en la divulgación de los resultados de las evaluaciones cuando estos no les parecían favorables.

- *Tendencia creciente 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.* Asimismo, crecientemente se incluye en las pruebas de criterios una definición de cuál es el resultado que todos los estudiantes deberían alcanzar para que su desempeño pueda ser considerado satisfactorio.
- *Mejoría de las capacidades técnico-metodológicas para la construcción de pruebas y para el procesamiento de los datos.* Existen esfuerzos por desarrollar pruebas que evalúen un rango de conocimientos y capacidades más amplio y que incluyan preguntas de respuesta construida. Se han ido incorporando también metodologías más sofisticadas para el tratamiento de datos, como la Teoría de Respuesta al Ítem y el Análisis Multinivel.
- *Creciente atención a la difusión y uso de los resultados.* Se ha ido comprendiendo que no basta con implementar una evaluación y publicar un informe, sino que es necesario desarrollar una estrategia de divulgación y un conjunto de reportes apropiados a cada una de las audiencias a las que se quiere llegar. Las Unidades de Evaluación tienen una mayor conciencia acerca de los usos apropiados e inapropiados para cada tipo de evaluación, aunque esto muchas veces no está suficientemente claro entre quienes toman decisiones de política.
- *Mayor preocupación por la investigación de los factores que inciden sobre los aprendizajes.* Si bien falta mucho por hacer en términos metodológicos e interpretativos para producir buenos trabajos de investigación, en la mayoría de los países hay conciencia de que es preciso avanzar en la investiga-

ción y en la construcción de hipótesis sobre cómo las dinámicas de la gestión de los sistemas educativos, los procesos escolares, las prácticas de enseñanza y las decisiones de política educativa influyen en las oportunidades de aprendizaje de los niños y niñas.

- *Creciente participación de los países en evaluaciones internacionales.* Esto ha tenido consecuencias positivas en las Unidades de Evaluación y en la calidad de su trabajo: ha contribuido a la construcción y acumulación de capacidad técnica en evaluación, ha facilitado el intercambio y el desarrollo de un lenguaje común entre países y ha ayudado a mejorar la calidad de diversos procesos técnicos (elaboración de pruebas, incorporación de preguntas de respuesta abierta, muestreo, control de las aplicaciones, análisis de datos, modos de difusión, entre otros).

Las principales debilidades presentes en los sistemas de evaluación estandarizada en la región se observan en la estrategia de difusión, en el uso de los resultados y en la calidad técnica de las evaluaciones

A pesar de las mejoras previamente mencionadas, persiste un conjunto importante de debilidades que es preciso abordar:

- Es frecuente que desde el nivel político se demande la implementación de grandes operativos de evaluación sin propósitos claramente definidos y en tiempos y con recursos absolutamente insuficientes para hacerlo técnicamente bien.
- Aún no se tiene claro que no cualquier evaluación sirve para cualquier propósito y que es necesario diseñar cuidadosamente el sistema de evaluación para el largo plazo, en función de propósitos y usos claramente establecidos. Esto es determinante para las decisiones relativas a la utilización de censos o muestras, a los grados y disciplinas que serán evaluados, la periodicidad de las evaluaciones, el tipo de pruebas y escalas de reporte, entre otros.

- Falta una mayor discusión pública sobre qué debe ser evaluado y qué aspectos del currículo deben haber sido enseñados a todos los estudiantes al concluir determinados niveles educativos. Como resultado de ello, las metas y/o estándares de aprendizaje no son claros. Mientras no exista claridad al respecto, tampoco existirá claridad ni para la enseñanza ni para la evaluación. Esta tarea requiere de un amplio debate social y de un trabajo técnico coordinado entre las unidades de currículo y de evaluación.
- Para mejorar la investigación sobre los factores que explican los resultados, falta mejorar mucho los instrumentos complementarios que se aplican junto con las pruebas y concebir diseños de investigación más ambiciosos: estudios longitudinales, estudios de “valor agregado” y estudios cualitativos, entre otros.
- Los Ministerios de Educación han tenido, por lo general, una limitada capacidad para concebir, formular e implementar políticas que den respuesta a los problemas identificados en las evaluaciones. Falta para ello mayor articulación entre las Unidades de Evaluación y otros actores educativos relevantes, dentro y fuera de los Ministerios de Educación, y mayor articulación entre la evaluación, el desarrollo curricular, la formación inicial y el desarrollo profesional de los docentes.
- Las acciones para la divulgación y uso de los resultados de las evaluaciones aún son insuficientes, en particular para lograr que los docentes los comprendan y utilicen y para que los resultados se incorporen a la cultura escolar.
- En la mayoría de los países perdura la práctica de divulgar resultados de tipos de escuela, escuelas individuales, estados o provincias, sin una adecuada consideración de los contextos socioculturales en que operan las instituciones y subsistemas educativos. Esto conduce a interpretaciones y conclusiones erróneas sobre la eficacia educativa de dichas instituciones o subsistemas. Para ello es necesario realizar evaluaciones de “valor agregado”, que implican dos mediciones sobre una misma población en distintos momentos. Sobre este aspecto la experiencia en la región es mínima.
- En muchos países persisten importantes debilidades técnicas en el diseño de las evaluaciones: excesiva simplicidad de la mayor parte de las preguntas y dificultad para elaborar preguntas que permitan evaluar capacidades cognitivas complejas; focalización de las evaluaciones en los logros educativos correspondientes a un grado específico, lo que impide saber qué han aprendido quienes aún no logran lo esperado para dicho grado; deficiencias en la conformación de las muestras y en la forma de estimar y reportar los errores de medición; y debilidades para establecer mediciones que sean comparables en el tiempo.
- No hay suficientes profesionales calificados para diseñar y conducir este tipo de evaluaciones, lo que se agrava por la falta de continuidad de los equipos técnicos de los países, muchas veces por razones de índole político-partidaria. Esto dificulta la acumulación de conocimiento y experiencia en la región y ha dado lugar a que muchos países tengan que comenzar desde “cero” con sus sistemas de evaluación, algunos años después de haber tenido uno en funcionamiento y haberlo desmantelado.

El riesgo de las malas evaluaciones

Solo una evaluación técnicamente buena y cuyos resultados sean conocidos y utilizados de manera apropiada, puede tener algún impacto en la mejora de los aprendizajes. Como es obvio, una evaluación cuyos resultados son poco conocidos y utilizados supone un ejercicio estéril y un desperdicio de recursos.

Más graves aún son aquellas situaciones en que evaluaciones técnicamente malas son ampliamente difundidas o en que las evalua-

ciones son empleadas para fines diferentes de aquellos para los cuales fueron concebidas y diseñadas, lo cual puede tener efectos dañinos para el sistema educativo. El razonamiento según el cual lo que importa es evaluar y difundir resultados como mecanismo de rendición de cuentas, sin importar la calidad de la evaluación, es una falacia que puede hacer retroceder los sistemas de evaluación e imposibilitar una discusión seria de las ventajas y desventajas de la rendición de cuentas o responsabilización por los resultados.

III. LOS PROPÓSITOS Y LOS USOS DE LOS RESULTADOS DE LAS EVALUACIONES ESTANDARIZADAS

Sea en la implementación de un sistema de evaluación o en su reforma, es necesario considerar algunas opciones básicas de acuerdo a lo que se espera de él.

El primer paso es definir cuál será el propósito del sistema de evaluación y para qué se usarán sus resultados

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 requiere de un determinado diseño y tiene exigencias técnicas y de costos diferentes. 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 –autoridades y cuerpos técnicos, supervisores, directivos, docentes, estudiantes y sus familias– 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, aprueba o reprueba.

Es necesario también establecer un adecuado balance entre la implementación de pruebas nacionales y la participación en pruebas internacionales

Las pruebas nacionales pueden ofrecer una mirada más apropiada acerca de qué aprenden los estudiantes de aquello que se les está enseñando, mientras que las pruebas internacionales pueden servir para ubicar la situación del país en relación a lo que saben y son capaces de hacer los estudiantes en otras sociedades y enriquecer el debate sobre el currículo nacional y el enfoque de la enseñanza. Los países deberían analizar cuidadosamente en qué pruebas internacionales participar, teniendo en cuenta lo que cada una pretende evaluar y su pertinencia en relación a los objetivos nacionales. Parece recomendable participar periódicamente en al menos una evaluación regional o internacional, en el marco de la política de prioridades que cada país defina.

EVALUACIONES PARA LA CERTIFICACIÓN DE LOS APRENDIZAJES DE LOS ESTUDIANTES

Un sistema de evaluación para certificar los aprendizajes de los estudiantes a través de exámenes nacionales de alta calidad técnica tiene 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. Los sistemas de evaluación para la certificación dan transparencia al valor de los certificados educativos ante la sociedad. Además, dichos sistemas hacen 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. En todo caso, la evaluación externa con

propósitos de certificación es más apropiada para los ciclos superiores del sistema educativo, sobre todo hacia el final de la educación media.

Un sistema de evaluación para la certificación implica algunos requerimientos que son costosos

Por una parte, 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. Por otra parte, las pruebas 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.

Un sistema de evaluación para la certificación puede 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, lo cual las haría insostenibles desde el punto de vista social y político. 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. Existen diversos caminos para enfrentar este problema:

- Establecer que la evaluación estandarizada constituya solo una parte del proceso de certificación de aprendizajes de los estudiantes (por ejemplo, el 40% de su calificación final). El resto de la nota correspondería a las evaluaciones que realizan los docentes a cargo de los estudiantes. Si bien en este caso se mantiene el problema respecto a la heterogeneidad de los criterios de

evaluación empleados por los docentes, es una forma de combinar la evaluación externa y la interna, e incluso de aportar a los docentes una mirada externa que les permita reflexionar sobre sus propios criterios.

- Establecer un período de transición hacia estándares únicos y universalmente aplicables. Durante dicho período, el foco principal se centra en la mejoría o progreso de los estudiantes de cada escuela respecto a su situación anterior, y no únicamente en el cumplimiento del estándar absoluto de certificación.

La peor respuesta que podría darse a este problema sería establecer un porcentaje fijo de reprobación, por ejemplo, no debe reprobar más del 10% de los estudiantes. Ello implica utilizar pruebas más fáciles y, por lo tanto, enviar a estudiantes, familias y educadores una señal equivocada respecto a las expectativas de aprendizaje.

EVALUACIONES DE CARÁCTER DIAGNÓSTICO Y FORMATIVO

Un sistema de evaluación de tipo formativo y sin consecuencias directas para los estudiantes tiene ventajas en términos de costos, posibilidades de diseño y de establecimiento de estándares altos, y aporte a una cultura de evaluación

Los costos de aplicación de estas 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 con carácter matricial, 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, de modo que, cuando se analice la posibilidad de establecer un sistema con consecuencias, existan las

condiciones para hacerlo apropiadamente. Una de las alternativas a considerar entre los usos de carácter formativo es la de “liberar” una prueba completa para que sea aplicada en forma autónoma por parte de los maestros, con el fin de ayudarles a identificar dificultades de alumnos individuales y de enriquecer su repertorio de instrumentos de evaluación.

El principal problema de estas evaluaciones es que pueden carecer de impacto si no se toman algunas acciones complementarias

Las evaluaciones de carácter diagnóstico y formativo pueden resultar intrascendentes si no van acompañadas de una estrategia precisa y de una inversión importante para asegurar la divulgación de los resultados y su uso en acciones educativas posteriores, puesto que su efectividad se juega en que los diferentes actores reciban, comprendan y utilicen los resultados. Por esto hay que tener presente que:

- *Para que los resultados de este tipo de evaluaciones tengan impacto sobre las políticas educativas*, se debe invertir tiempo en el análisis y discusión de los mismos por parte de diversas áreas del Ministerio de Educación y otros actores relevantes, en la comprensión de los problemas y deficiencias que los resultados ponen de manifiesto y en la concepción de acciones e inversiones apropiadas para hacerles frente. Las autoridades deben estar dispuestas a someter sus políticas y decisiones al escrutinio de la ciudadanía, para lo cual es necesario también invertir en la comunicación apropiada y permanente de los resultados a la opinión pública.
- *Para que los resultados tengan impacto sobre las prácticas de enseñanza*, es imprescindible asimismo invertir tiempo en el análisis e interpretación de sus implicancias didácticas: si los estudiantes no son capaces de resolver cierto tipo de situaciones, ¿qué es lo que se está haciendo de manera insuficiente o inapropiada en las aulas y qué es lo que se debería hacer? Este tipo de análisis

debe ser realizado tanto por especialistas en la didáctica de las áreas evaluadas como por los docentes, creando para este último efecto espacios permanentes y sistemáticos de formación en servicio y de trabajo colectivo al interior de las escuelas. Es muy importante que los maestros puedan analizar la mayor cantidad posible de ítemes para poder identificar cuáles revelan un bloqueo importante para el desarrollo de nuevos conceptos o capacidades. Sin embargo, siempre será necesario conservar el carácter confidencial de una parte de los ítemes para poder realizar mediciones comparables en el tiempo.

- *Para que los resultados tengan impacto sobre la motivación de los estudiantes y sobre la actitud de las familias hacia el aprendizaje escolar*, es necesario proveerles de 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.

Evaluar con una perspectiva amplia

Es muy importante que las pruebas no se limiten a evaluar los conocimientos y competencias correspondientes a un único grado (aquel en que se aplica la evaluación), sino que incluyan una perspectiva más amplia de niveles de desempeño –desde más básicos a más complejos– a lo largo de varios grados, de modo de identificar qué han aprendido los estudiantes en grados anteriores y qué necesitan ahora. De este modo, los docentes podrán constatar y subsanar las deficiencias en los aprendizajes de cursos previos, que impiden a los estudiantes continuar avanzando y los resultados serán útiles no solo para los docentes del grado evaluado, sino también para los de grados anteriores. El énfasis no estará puesto en comunicar a los docentes que los estudiantes “aprueban” o “desaprueban”, sino en comunicar que están en diferentes puntos de un continuo de aprendizaje, en el que todos pueden y necesitan progresar.

Las pruebas de tipo diagnóstico pueden tener carácter muestral o censal, dependiendo de la estrategia de cambio educativo

El carácter censal o muestral tiene diversas implicancias, aunque también es posible combinar una aplicación por muestras controlada con una distribución censal de las pruebas para su aplicación en forma autónoma por parte de las escuelas, con fines formativos y de análisis de resultados e identificación de estudiantes que necesitan apoyos complementarios.

- Las *pruebas a base de muestras* sirven para tener un diagnóstico global del sistema. Requieren de un cuidadoso diseño de la muestra, de modo de obtener información representativa para los niveles de desagregación en los que se desee actuar y tomar decisiones (regional, provincial, municipal; urbano y rural; escuelas indígenas, etc.). El impacto de los resultados de estas pruebas depende principalmente de las medidas de política educativa que se tomen a nivel central y de una estrategia de difusión apropiada que llegue a todas las escuelas.
- Las *pruebas censales* sirven para obtener información de cada una de las escuelas e incluso de los alumnos. El impacto de sus resultados depende de la devolución de la información a cada comunidad educativa, con un enfoque y un formato adecuados para promover una mayor participación y compromiso a nivel local. La información también puede resultar muy útil para dirigir mejor las políticas hacia distritos o escuelas con mayores problemas, ya que se puede contar con un “mapa” de resultados de todas las escuelas, zonas, provincias, tipos de escuela, etc.

USO DE LAS EVALUACIONES PARA ESTABLECER INCENTIVOS

Además de las opciones examinadas, existen políticas de evaluación estandarizada dirigidas a establecer incentivos económicos en función de los resultados o a propiciar un mercado competitivo entre las escuelas

Tres son las principales modalidades en este tipo de políticas:

- Utilizar los resultados para construir *rankings* de escuelas y hacerlos públicos, como forma de: fomentar la responsabilidad de las escuelas por sus resultados; entregar a las familias información para la toma de decisiones sobre la escuela a la cual enviar a sus hijos; promover la competencia entre los centros educativos por obtener mejores resultados.
- Utilizar los resultados para entregar incentivos económicos a las escuelas que obtienen mejores resultados o mejoras en relación a sus resultados en evaluaciones anteriores.
- Utilizar los resultados como indicador de la calidad del trabajo de cada docente y como criterio para otorgar incentivos económicos.

Estos enfoques de uso de la evaluación a veces se constituyen, con o sin intención deliberada, en un mecanismo por el cual el Estado renuncia a su responsabilidad por los resultados del sistema educativo

En ocasiones, el Estado limita su accionar a la realización de evaluaciones, la entrega de resultados y el establecimiento de incentivos en función de estos resultados, transfiriendo la responsabilidad por los resultados a la relación entre escuelas y familias, como si se tratase de un asunto entre actores privados, en lugar de ocuparse de crear las condiciones para que la enseñanza resulte efectiva proveyendo los recursos necesarios, constituyendo un elenco docente adecuadamente formado y estableciendo mecanismos idóneos para la evaluación y la orientación de la labor de las escuelas. Este modo de operar no considera la complejidad de la labor educativa, en especial en medios socialmente desfavorecidos, y la necesidad de invertir en la creación de capacidades como herramienta principal para la mejora de la enseñanza y el aprendizaje.

CRITERIOS BÁSICOS PARA LA DIVULGACIÓN Y USO DE LOS RESULTADOS

La responsabilidad por los resultados debe ser compartida por diversos actores

Debería evitarse todo uso de los resultados con el fin deliberado o implícito de culpabilizar o res-

ponsabilizar de manera exclusiva a ciertos actores. La responsabilidad debe ser compartida entre las autoridades nacionales y regionales, los cuerpos docentes, los estudiantes y sus familias.

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 son el único indicador de la calidad de la educación, es preciso considerarlos en conjunto con otros aspectos relevantes como la retención, la pertinencia y relevancia de lo que se enseña, la formación de valores y hábitos o la formación ciudadana. Debe evitarse la identificación del 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 de las escuelas, 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 también 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.

ADVERTENCIAS SOBRE LA COMPARACIÓN DE RESULTADOS ENTRE ESCUELAS

Aunque los resultados de pruebas estandarizadas no dan cuenta en forma exhaustiva de la calidad educativa de una escuela, sí aportan información relevante de los niveles de desempeño alcanzados

La información sobre resultados en dos pruebas estandarizadas –en general, lenguaje y matemática– no constituye por sí misma una evaluación de la “calidad educativa” de las escuelas ni debería presentarse al público como tal. La “calidad” de una escuela incluye otros aspectos relevantes y valorados por los docentes, estudiantes y familias, tales como el desarrollo emocional, las relaciones interpersonales, la formación cívica y en valores, entre otros. Aun así, la información comparativa acerca de los niveles de desempeño alcanzados por los estudiantes en una diversidad de escuelas puede ser relevante para los equi-

pos docentes de las escuelas, en la medida en que enriquece la percepción de la propia labor, permitiéndoles ubicar los logros de sus estudiantes en el contexto de los alcanzados por los estudiantes en otras escuelas.

Para que la comparación entre los niveles de desempeño de los estudiantes sea apropiada, se deben considerar sus condiciones sociales de origen

Los resultados académicos de las escuelas deberían poder compararse con los de escuelas de similar composición social, dado que los retos y dificultades que implica enseñar a estudiantes de origen desfavorecido (o de lengua materna indígena) son muy diferentes que en el caso de estudiantes de familias con educación secundaria completa y/o terciaria. También hay que tener en cuenta las tasas de deserción de las escuelas y sus políticas de selección de estudiantes, dado que una escuela puede mejorar sus resultados a través de la exclusión de estudiantes con dificultades.

Las comparaciones entre escuelas deben considerar la diferencia entre medir el “logro educativo” y medir el “aprendizaje”

En rigor, si el propósito es entregar a las escuelas, familias y/o autoridades, información acerca de la capacidad de enseñar de los docentes y las escuelas, es necesario medir el progreso de los estudiantes a lo largo de un período lectivo (“aprendizaje”) y no únicamente el resultado final (“logro”). Las diferencias entre estos dos términos se expresan en que:

- El *aprendizaje* puede ser definido como el cambio ocurrido en los conocimientos y capacidades de cada estudiante a lo largo del año lectivo. Su medición requiere la aplicación de dos pruebas, una al inicio y otra al final. De este modo es posible constatar el progreso realizado por cada alumno.
- El “logro educativo” en cambio, se mide con una única prueba y refleja la acumulación de conocimientos y capacidades a lo largo de

toda la vida del estudiante, lo cual incluye el capital cultural familiar y sus experiencias en otras escuelas o con otros docentes.

Debe tenerse siempre presente que el aprendizaje y el logro educativo de los estudiantes dependen tanto de lo que hacen las escuelas y docentes, como del esfuerzo de los propios estudiantes, del apoyo de las familias a la tarea escolar, del contexto comunitario y cultural y de las políticas educativas. Los problemas educativos no pueden ser resueltos apelando exclusiva ni principalmente a mecanismos de mercado.

Los resultados expresados en términos de rankings de escuelas deben ser mirados con precaución

La mayoría de los *rankings* de escuelas transmiten una falsa imagen de ordenamiento. Una escuela puede ocupar el primer lugar y otra el número 40 y, sin embargo, las diferencias entre sus promedios pueden no ser estadísticamente significativos. Esto significa que no puede decirse si un promedio es realmente superior al otro, porque las diferencias entran dentro de los márgenes de error de la medición. Pero aun cuando las diferencias en los promedios sean estadísticamente significativas, pueden ser irrelevantes en términos de las proporciones de alumnos que logran los niveles de desempeño esperados.

ITINERARIO PARA LA TOMA DE DECISIONES

Considerando las alternativas existentes para una política de evaluación, antes de poner en marcha un sistema de evaluación (o de modificar uno existente) es indispensable seguir un itinerario de decisiones para definir las características del sistema que incluya al menos las siguientes preguntas básicas:

- ¿Para qué se quiere evaluar? ¿Quiénes usarán los resultados y con qué propósitos? ¿Qué se espera saber que no se sepa ya?
- ¿Cuáles serán las unidades de análisis para el reporte de resultados: estudiantes individuales, grupo de clase/maestros, escuelas,

tipos de escuela, entidades subnacionales, sistema educativo?

- De acuerdo a los propósitos definidos, ¿qué es más apropiado evaluar –aprendizaje o logro educativo– al egreso de determinados grados o ciclos?
- ¿Qué tipo de consecuencias tendrán los resultados y para quiénes?
- De acuerdo a los propósitos definidos, ¿es necesario trabajar en forma censal o son suficientes evaluaciones en base a muestras?
- ¿Qué grados y qué disciplinas es importante evaluar?
- ¿Cada cuánto tiempo es necesario y adecuado realizar las evaluaciones?

Las respuestas a estas preguntas deberían plasmarse en un plan de evaluación a corto, mediano y largo plazo, claro y explícito. En la elaboración de dicho plan es muy importante considerar los costos en dinero implicados en cada opción y los recursos humanos necesarios para su adecuada implementación. En particular, se debe balancear la inversión a realizar en la recogida de información con la inversión en la difusión y uso de los resultados. Muchos Ministerios de Educación destinan importantes sumas de dinero a recoger una gran cantidad de información todos los años, que luego casi no se analiza, ni se difunde ni se utiliza. Es absolutamente inconducente implementar un primer operativo de evaluación si no se ha elaborado un plan de trabajo de largo plazo.

“Mandatos” útiles al momento de tomar decisiones políticas respecto al sistema de evaluación

- **“No harás daño”.** Una de las primeras normas del juramento hipocrático señala que el médico se abstendrá de todo aquello que pueda generar daño o perjuicio a sus pacientes. Del mismo modo, a

la hora de pensar en un sistema de evaluación es importante considerar los riesgos de efectos no deseados y perjudiciales que la estrategia de evaluación diseñada podría tener sobre el propio sistema educativo que se desea mejorar.

- **“El consumo excesivo de este producto puede resultar perjudicial para la salud”.** La evaluación no puede estar por encima de la educación. El exceso de evaluaciones puede producir perjuicios para la salud del sistema educativo, en especial si solo hay evaluaciones pero no hay políticas para dar respuesta a los problemas que estas ponen de manifiesto.
- **“No evaluarás en vano”.** Se debe evitar recoger todos los años información que nunca será analizada ni utilizada. Para que un sistema de evaluación tenga impacto, debe tener una periodicidad que garantice que los datos puedan ser analizados, discutidos, comprendidos y utilizados. Incorporar la nueva información y traducirla en decisiones y acciones lleva tiempo. Los cambios en el sistema educativo requieren más tiempo aún.
- **“Vísteme despacio, Sancho, que estoy apurado”.** Es muy importante que quienes toman las decisiones en los Ministerios de Educación desestimen la creencia ingenua en la existencia de caminos fáciles y rápidos. No es posible montar un programa de evaluación serio en tres meses. No existe coyuntura ni “ventana de oportunidad” que lo justifique: las consecuencias de la improvisación se harán sentir tarde o temprano. La evaluación requiere reflexión cuidadosa acerca de sus fines y usos; discusión pública sobre qué es lo fundamental a evaluar; diálogo e involucramiento de diferentes actores; conformación de equipos técnicos competentes en diversos aspectos; e información previa para que los actores se involucren con la evaluación.

IV. LOS DESAFÍOS DE CALIDAD TÉCNICA DE LAS EVALUACIONES

Una vez definida la política de evaluación, es necesario implementarla de acuerdo con diversos estándares apropiados de calidad técnica. Esto implica abordar los siguientes desafíos:

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

Si bien esta construcción exige antes que nada un debate y decisiones de carácter político, las definiciones tomadas en dicho terreno deben luego ser adecuadamente traducidas en especificaciones y estándares de carácter técnico. Asimismo, el debate político-educativo debe estar alimentado por información sobre los avances conceptuales recientes en cuanto a la enseñanza y a los desempeños de los estudiantes en las disciplinas consideradas.

Incluir en las pruebas actividades con diversos grados de complejidad

Se debe contemplar actividades cuya solución requiera de competencias cognitivas suficientemente complejas, apropiadas a los desafíos de la sociedad del conocimiento, así como también actividades sencillas que permitan informar sobre el punto en que se encuentran los estudiantes menos avanzados. Las actividades deben poseer una serie de *propiedades psicométricas* que es necesario garantizar y cuyo control requiere pruebas piloto y análisis cuidadoso. También es importante, en la medida de lo posible, ampliar el uso de preguntas de respuesta construida, con el doble propósito de evaluar capacidades más complejas y de mejorar la articulación entre la evaluación externa y la cultura docente y escolar.

Conformar las pruebas a partir de la apropiada integración de las actividades en bloques y cuadernillos

Este es un desafío técnico importante y complejo, que requiere de conocimientos especializa-

dos y experiencia. Especial atención debe otorgarse a la decisión de utilizar la Teoría Clásica o la Teoría de la Respuesta al Ítem para la construcción y el análisis de las pruebas, lo cual requiere de programas de procesamiento modernos, analistas bien entrenados, análisis rigurosos y asesoramiento de alto nivel.

Definir los “puntos de corte” que establecen los límites entre niveles de desempeño en una prueba

Vinculado con esto, se debe definir la metodología a seguir para establecer cuál de los niveles debe ser considerado como aceptable para un estudiante al finalizar el grado o ciclo educativo evaluado. Un rendimiento aceptable no puede ser establecido automáticamente como equivalente al 51% o más del máximo posible en una prueba.

Diseñar muestras apropiadas a los propósitos de la evaluación

Esto tiene por objeto evitar operaciones más grandes y costosas de las estrictamente necesarias y que, al mismo tiempo, tengan un nivel de precisión adecuado. En relación a este último aspecto debe incluirse, además, la estimación y reporte de los márgenes de error de las mediciones.

Definir cómo se equiparán las evaluaciones

La equiparación de las evaluaciones se refiere a la metodología mediante la cual los resultados de pruebas aplicadas en distintos años se hacen comparables. Este es uno de los desafíos técnicos más importantes que los sistemas de evaluación deben afrontar si uno de sus propósitos es informar sobre los avances o retrocesos en los logros educativos a lo largo del tiempo. Este proceso es indispensable para poder afirmar que los eventuales cambios que se constaten responden a cambios en la realidad educativa y no meramente a cambios en el instrumento de medición. Requiere tomar recaudos tanto desde el punto de vista estadístico, como desde el punto de vista de la continuidad de los conocimientos y compe-

tencias evaluadas, y la estabilidad en la longitud y la estructura de la prueba.

Construir datos longitudinales que permitan apreciar la evolución en el tiempo de los aprendizajes de un mismo conjunto de estudiantes

Este tipo de datos, que exigen más de una medición en el tiempo, es el que evalúa el “aprendizaje” entendido como cambio y el que mejor permite establecer cuáles son los procesos propiamente escolares que tienen influencia sobre el mismo. Este enfoque es el que permitiría obtener información más relevante para la política educativa y para la investigación.

Establecer mecanismos de control del cumplimiento de las condiciones estandarizadas que deben regir durante la aplicación de las pruebas

El cumplimiento de las condiciones de aplicación de las pruebas involucra aspectos técnicos complejos y muchas veces descuidados. Sin embargo, son fundamentales para que la información sea confiable y comparable. Esto apunta a temas como:

- La calidad de la capacitación de los aplicadores.
- El establecimiento de controles de calidad durante la aplicación de las pruebas.
- La motivación a los estudiantes para realizar las pruebas.
- La logística de la distribución.
- El retorno seguro de los materiales.

Combinar las evaluaciones con estudios cualitativos

Es importante combinar las evaluaciones nacionales estandarizadas, que ofrecen una mirada sobre el conjunto del sistema, con la realización de estudios cualitativos que permitan profundi-

zar en los procesos escolares y los procesos que ocurren en las aulas. Esta combinación de enfoques es el mejor modo de aportar información rica y compleja para pensar en las políticas educativas y en las prácticas de enseñanza.

La necesaria transparencia

Sobre todos los procesos técnicos debe existir información transparente y accesible. Es necesario dar un especial énfasis a la documentación de los procedimientos técnicos seguidos en:

- La construcción de los instrumentos.
- La estimación de la precisión de las mediciones (y, por consiguiente, su margen de error).
- El diseño de las muestras y la cobertura alcanzada.
- La aplicación y control de calidad de la misma.
- La definición de niveles de desempeño y puntos de corte.
- La equiparación y comparabilidad de los resultados con mediciones anteriores.

V. SOBRE LA CONSTITUCIÓN DE LAS 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 y los recursos adecuados que implica este complejo emprendimiento.

La puesta en marcha de un sistema de evaluación serio requiere de un tiempo mínimo de entre dos y tres años

Este periodo es necesario para llevar adelante los siguientes procesos fundamentales:

- Discusión, definición y difusión pública de los propósitos del sistema de evaluación,

del tipo de consecuencias, de los usos esperados y de qué debe ser evaluado.

- Diseño del plan de evaluación a largo plazo.
- Conformación de equipos técnicos con la diversidad de capacidades requeridas (elaboración de pruebas; elaboración de cuestionarios; conocimiento de las disciplinas a evaluar y de su didáctica; currículo y estándares; muestreo; logística de la aplicación y control de calidad de la misma; captura y limpieza de datos; procesamiento y análisis; construcción de escalas; contextualización sociocultural de los resultados y análisis de factores asociados; interpretación de los datos en términos de políticas educativas, de didáctica y de prácticas de enseñanza).
- Elaboración de instrumentos, pilotaje y supervisión externa de los mismos.

Las Unidades de Evaluación necesitan tener una cierta estabilidad en el tiempo de sus elencos técnicos

El desarrollo de un sistema de evaluación requiere de una planificación a 10 ó 15 años. Si los equipos técnicos cambian con frecuencia, se desperdicia el conocimiento y la experiencia acumulada en un área compleja y se desacreditan los procesos de evaluación ante la sociedad y los educadores.

Las Unidades de Evaluación necesitan tener independencia para reportar los resultados de su trabajo

Al igual que en el caso de las Unidades de Estadísticas sociales y económicas, las unidades que realizan las acciones de evaluación y la divulgación de los correspondientes resultados no deben depender de los tiempos e intereses político-partidarios. Se ha discutido mucho respecto a si las Unidades de Evaluación deberían estar dentro o fuera de los Ministerios de Educación. El argumento principal para esta última opción es justamente la necesaria independencia y transparencia. Sin embargo, la experiencia en la región muestra casos de países con unidades

estables e independientes dentro de los Ministerios de Educación y otros con unidades inestables que no han logrado consolidar su trabajo aun estando en una institución externa. La respuesta a esta disyuntiva es que no importa tanto el lugar institucional como la cultura de continuidad y transparencia que se cree en torno a la evaluación. Esto se logra cuando existe un mandato claro y una institucionalidad sólida en relación a la evaluación, lo cual implica algún tipo de estatuto jurídico para el sistema de evaluación. Establecer esto por ley es uno de los caminos a considerar, dado que exige un acuerdo amplio, en lo posible suprapartidario, que permita establecer un plan de evaluación de la educación de largo plazo. En este terreno, al igual que en el económico, se requiere de cierta estabilidad en las reglas de juego. Si la política de evaluación cambia permanentemente, se genera desconfianza y se pierde credibilidad.

Una institucionalidad sólida requiere de órganos de gobierno y de asesoría técnica independientes y plurales, y de un presupuesto apropiado y plazas de trabajo que garanticen la operación de la unidad con la calidad técnica requerida

La independencia de la Unidad de Evaluación 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.

Para que las evaluaciones estandarizadas tengan impacto sobre las políticas y las prácticas educativas, es necesario generar nuevas interfases y modalidades de trabajo que aseguren el alineamiento de los distintos actores y ámbitos de decisión

Las Unidades de Evaluación deben asumir que su tarea va más allá de producir datos. Si su misión es producir información para otros acto-

res, se requiere de personas preparadas y con tiempo para establecer líneas de diálogo con instancias y actores responsables de:

- *Análisis de política educativa*, con el fin de mejorar tanto los diseños de las evaluaciones como los planes de procesamiento y los reportes de resultados, teniendo en cuenta preguntas relevantes para la política educativa (por ejemplo, la muestra puede ser diseñada de modo que permita evaluar el impacto de políticas o programas específicos en ciertos grupos de escuelas).
- *Análisis didáctico*, con el fin de interpretar los resultados y los problemas de aprendizaje de los estudiantes desde la perspectiva de la enseñanza y la didáctica de la disciplina evaluada, elaborar reportes con significado didáctico para los docentes, diseñar programas de formación en servicio a partir de los resultados, pensar la articulación de las evaluaciones estandarizadas con la evaluación en el aula.
- *Comunicación*, con el fin de elaborar una diversidad de reportes de resultados, con diversos formatos y lenguajes, que sean apropiados y comprensibles para diversas audiencias.

Evaluar bien requiere inversión

Más vale no hacer ninguna evaluación que hacer una evaluación mala o deficiente. Esto debe ser tenido en cuenta al tomar las decisiones acerca del Plan de Evaluación. Asimismo, más vale un sistema de evaluación modesto pero con costos asequibles y sostenibles en el tiempo, que una gran evaluación, amplia y sofisticada, que solo podrá realizarse una vez y que luego no podrá replicarse.

La inversión en evaluación debe ser valorada en función del uso que se hace de sus resultados, más que a partir de otros indicadores como el costo por alumno

Los costos de las evaluaciones son relativamente bajos cuando se los compara con los presu-

puestos nacionales y con las inversiones alternativas. Pero cualquier inversión en evaluación, alta o baja, es inútil si no se hace uso de los resultados.

VI. DIEZ RECOMENDACIONES ACERCA DE LAS EVALUACIONES QUE LA REGIÓN NECESITA

- 1. 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.
- 2. La evaluación debe contemplar un proceso de reflexión colectiva sobre el estado de la educación y los caminos para mejorarla.** El primer paso para ello es la consulta y el debate públicos acerca de qué deben aprender los estudiantes y acerca de los propósitos y consecuencias de la evaluación. Del mismo modo, es fundamental la discusión pública constructiva a partir de los resultados, con el fin de encarar las insuficiencias e inequidades en el acceso al conocimiento por parte de los y las estudiantes. Para ello es necesario invertir en comunicación y difusión tantos o más recursos que en la propia evaluación, antes, durante y después de la misma.
- 3. 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.
- 4. Los sistemas de evaluación de la región necesitan ampliar progresivamente el**

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

5. Los sistemas de evaluación de la región deberían progresivamente diseñar evaluaciones de progreso de los estudiantes, dado que son las que pueden 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 y las estudiantes.

6. 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. Todo ello exige tiempo, por lo que no es conveniente pretender la implementación de sistemas de evaluación en plazos reducidos.

7. Un buen sistema de evaluación requiere inversión, principalmente en la conformación de equipos humanos calificados, así como recursos económicos suficientes para una adecuada implementación de todos los procesos implicados.

8. El sistema de evaluación debe sustentarse en una actitud firme de transparencia en relación a los resultados y de rendición de cuentas a la sociedad.

9. Los ministerios de educación deben asumir un compromiso serio y consistente con los resultados de la evaluación, lo cual 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.

10. Los sistemas de evaluación deben ser objeto de evaluación periódica, con el fin de analizar la calidad técnica de la información que producen y su relevancia para diversos actores educativos y sociales.



Programa de Promoción de la Reforma Educativa en América Latina y el Caribe
Partnership for Educational Revitalization in the Americas

El Programa de Promoción de la Reforma Educativa en América Latina y el Caribe es un proyecto conjunto del Diálogo Interamericano, con sede en Washington, y la Corporación de Investigaciones para el Desarrollo, con sede en Santiago de Chile.

Desde su creación en 1995, el PREAL ha tenido como objetivo central contribuir a mejorar la calidad y equidad de la educación en la región mediante la promoción de debates informados sobre temas de política educacional y reforma educativa, la identificación y difusión de buenas prácticas y la evaluación y monitoreo del progreso educativo.

La ejecución de las actividades se realiza a través de Centros Asociados de Investigación y Políticas Públicas en diversos países de la región y comprenden la realización de estudios, la organización de debates y la promoción de diálogos públicos sobre temas de política educacional y reforma educativa.

Las actividades del PREAL son posibles gracias al apoyo de la United States Agency for International Development (USAID), el Banco Interamericano de Desarrollo (BID), la International Association for the Evaluation of Educational Achievement (IEA), GE Foundation, The Tinker Foundation, entre otros.



Inter-American Dialogue • 1211 Connecticut Ave. N.W. Suite 510
Washington, D.C. 20036 U.S.A. • Tel: (202) 822-9002
Fax: (202) 822-9553 • E-mail: iad@thedialogue.org
Internet: www.thedialogue.org & www.preal.org

CINDE • Santa Magdalena 75, Piso 10 • Oficina 1002 • Providencia
Santiago, Chile • Tel: (56-2) 334-4302
Fax: (56-2) 334-4303 • E-mail: infopreal@preal.org
Internet: www.preal.org

