

DICIEMBRE 2002 - Año 3 / N° 13

Directora responsable: Marcela Gajardo

E-mail: mgajardo@preal.org

Tel: (562) 3344302. Fax: (562) 3344303

Santa Magdalena 75, of. 1002.

Santiago - CHILE

Internet: www.preal.org

## La divulgación de los resultados de las evaluaciones nacionales de aprendizajes

¿Cómo presentan sus resultados los sistemas de evaluación en América Latina? Atendiendo a esta pregunta, se presentan aquí algunas conclusiones de un análisis realizado a partir de más de 70 informes nacionales producidos en la región en el periodo 1995-2000, el cual se centró en las siguientes preguntas:

• **¿Qué tipo de información producen y difunden los sistemas de evaluación?**

• **¿Cómo la presentan?**

• **¿A quiénes está dirigida?**

• **¿Con qué propósito?**

• **¿Cómo se espera que la información sea empleada?**

• **¿Sirve la información efectivamente para esos propósitos?**

**Una aproximación a estos temas y los dilemas que se enfrentan tiene por objetivo ver cómo mejorar los modos de reportar los resultados, respecto a lo cual se presentan al final algunas sugerencias.**

Se resumen a continuación algunas partes del Documento de Trabajo PREAL N° 22 “**¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?**”, de Pedro Ravela (febrero 2002), el cual se puede consultar en [www.preal.org](http://www.preal.org) sección publicaciones, o en su versión original más amplia en [www.grade.org.pe/gtee-preal/docs/Resultados.pdf](http://www.grade.org.pe/gtee-preal/docs/Resultados.pdf)

*Todos los sistemas nacionales de evaluación buscan dar cuenta a la opinión pública sobre el grado en que el sistema educativo está logrando los objetivos en ciertas áreas del currículo y en ciertos grados o niveles. Sin embargo, existen importantes diferencias en las formas de hacerlo, en las métricas que utilizan, en el tipo de información que producen, en las maneras de difundir los resultados y en el uso que finalmente dan a la información.*

### ¿Evaluación formativa o con consecuencias fuertes?

En cuanto a los propósitos que persiguen las evaluaciones, la literatura sobre este tema suele distinguir entre:

• *Evaluaciones “de alto riesgo” cuyos resultados tendrán consecuencias directas para individuos e instituciones y que, por tanto, implican un cierto nivel de riesgo o amenaza porque conllevan algún tipo de sanción positiva o negativa. Se usan, por ejemplo, para la certificación al finalizar cierto nivel educativo o para seleccionar a los alumnos que desean ingresar al nivel terciario; para el establecimiento de incentivos económicos para las escuelas o los docentes a partir de los resultados de sus alumnos en las pruebas y de otros indicadores; o para que las familias, sobre la base de los resultados de cada establecimiento publicados en la prensa, puedan juzgar la calidad de las escuelas y elegir entre ellas.*

• *Evaluaciones “de bajo riesgo” cuyos resultados tienen una función únicamente*

*informativa y formativa, pero que no entrañan consecuencias. Se usan para informar al público periódicamente sobre el estado del sistema educativo, para ayudar a la definición de políticas educativas o para aportar al mejoramiento de las escuelas y docentes a través de la devolución de los resultados a las escuelas y de la producción de materiales didácticos.*

De este último tipo son la mayoría de los sistemas nacionales de evaluación desarrollados en América Latina durante los '90, pero existen en la región algunas experiencias importantes de uso de los resultados con consecuencias fuertes, o directas. Es el caso de Chile —donde se utilizan los resultados de las evaluaciones nacionales como mecanismos para establecer incentivos y sanciones a las escuelas y para establecer incentivos económicos para los docentes— y el de México, donde durante los '90 los esfuerzos de evaluación nacional se realizaron principalmente en el marco del programa *Carrera Magisterial*, el cual establece incentivos económicos individuales para los maestros, considerando los resultados de las pruebas estandarizadas de logros del alumnado como un factor en la determinación del desempeño profesional.

La discusión de cuál debe ser la apuesta principal de los sistemas de evaluación —“consecuencias fuertes” o evaluación “formativa”— es de enorme trascendencia para avanzar en su futuro desarrollo. Desde algunos sectores, principalmente políticos y económicos, parece existir una creciente presión por ir transformando los sistemas nacionales de evaluación en sistemas de rendición de cuentas públicas y de establecimiento de incentivos. Algunos investigadores indican que el problema central de usar pruebas estandarizadas con

una estrategia de consecuencias fuertes es que exacerban la preocupación de directores y docentes por “enseñar para la prueba” y el tipo de actividades que ella incluye. Los proponentes de este tipo de evaluaciones argumentan que las pruebas miden objetivos que es importante que los estudiantes logren y que es deseable que los maestros focalicen su atención en ellos, mientras los críticos advierten que esto puede significar una distorsión de la educación y puede dejar desatendidos objetivos que son importantes aun cuando no estén incluidos en las pruebas.

La evaluación “sin consecuencias” minimiza estos riesgos, pero puede hacer que nadie en las escuelas esté obligado a hacerse cargo de los resultados y, muchas veces, ni siquiera a enterarse de ellos o a considerarlos.

La disyuntiva de fondo es en qué medida los objetivos formativos, por un lado, y de establecimiento de incentivos, por otro, pueden resultar compatibles entre sí. Esto requiere analizarse a través de una investigación dirigida a ver cómo las escuelas utilizan los resultados de las evaluaciones y de qué manera el contexto y la estrategia global del sistema de evaluación incide en la aceptación y uso de los resultados por parte de los establecimientos educacionales.

### El público de los informes de las evaluaciones

Como ya se señaló, el primer objetivo de los sistemas nacionales de evaluación es proveer información en forma sistemática a la ciudadanía. En los últimos años, los países de la región han dado un paso relevante al estar informando periódicamente a la sociedad acerca de los aprendizajes de los alumnos; sin embargo, las formas de reportar los resultados que se han desarrollado hasta el momento, en muchos casos, pueden dar lugar a serios equívocos, y se suele olvidar el hecho de que no existe aún la acumulación suficiente de capacidad técnica respecto a las formas de evaluar, a los indicadores a utilizar y a su correcta interpretación. Por esto, es necesario realizar un trabajo sistemático de difusión para que el público en general pueda atribuir significado a los datos que los informes de evaluaciones les ofrecen.

Aparte de informar al público

**Figura 1**  
**Tipo de informes de difusión de resultados producidos en países de la región**

Países	Informe general resultados	Informes específicos sobre factores asociados	Informes o publicaciones específicas dirigidas a los docentes	Informes o publicaciones específicas dirigidas a las familias	Informe de resultados por establecimiento educativo (censales)*
Argentina					
Bolivia					
Brasil					
C. Rica					
Chile					
Ecuador					
Paraguay					
Perú					
Uruguay					

\*Algunos países, como por ejemplo Ecuador, devuelven los resultados por establecimiento en el marco de operativos de carácter muestral, a los que integraron la muestra.  
Nota: Los casilleros más oscuros indican el tipo de informes de difusión de resultados.

general, los sistemas de evaluación esperan aportar información relevante a tres públicos específicos que, supuestamente, la usarán para mejorar la calidad de la educación: las autoridades y decisores de los sistemas educativos, los directivos y docentes de los establecimientos educacionales y las familias de los estudiantes.

El primer problema que aparece en algunos países es que se proponen informar a los tres destinatarios con un informe único. Asimismo, la mayoría de los reportes se limita a describir datos incluidos en los cuadros o gráficos, con escaso análisis e interpretación de los mismos. Además, en muchos reportes no se advierte acerca de los usos válidos de la información y sus limitaciones.

La elaboración de reportes de resultados para uso de los equipos docentes ha sido uno de los aspectos más trabajados de la difusión de resultados en los países de la región y ha incluido estrategias como:

- La producción y distribución a todas las escuelas de materiales y documentos didácticos derivados de los resultados generales de la evaluación, independientemente de que la misma haya sido censal o muestral.
- La entrega a cada escuela de los resultados de sus propios alumnos, en el marco de operaciones censales.
- La realización de operaciones de evaluación muestrales, las que posteriormente se aplican en el universo de las escuelas voluntariamente y con corrección autónoma, para que luego cada escuela pueda comparar los propios resultados con los de la muestra nacional.
- La implementación de talleres de

difusión de resultados y programas de capacitación en servicio a partir de los mismos, dirigidos a docentes, directivos y/o supervisores.

Aun así, para aprovechar más las evaluaciones en el mejoramiento de las prácticas de enseñanza se requiere, entre otras cosas: que los maestros conozcan al detalle no solo la definición formal de las competencias evaluadas, sino también el tipo de actividades que los alumnos debieron saber realizar adecuadamente; avanzar hacia reportes más elaborados en términos de interpretaciones didácticas de los resultados; y aportar a los docentes instrumentos de evaluación sistemáticamente elaborados y validados, que ellos puedan emplear de manera autónoma de acuerdo a sus necesidades y criterio profesional.

En el caso de la información orientada a los actores involucrados en políticas educativas, se identifica en los reportes de las evaluaciones dos tipos de debilidades:

- El carácter excesivamente descriptivo y escasamente conceptualizado de los informes.
- La visión demasiado simplificada sobre el estudio de los factores escolares asociados con los aprendizajes y su incidencia sobre la toma de decisiones.

En relación a las familias, hay que señalar que los reportes nacionales difícilmente son comprensibles, aun para la familia de nivel cultural medio.

### Las métricas empleadas para reportar los resultados

Un aspecto que requiere de atención al analizar los informes nacionales de evaluación de aprendizajes es el tipo de

métrica utilizada. En los reportes latinoamericanos siete han sido las métricas empleadas:

- Porcentaje promedio de respuestas correctas.
- Promedio de los puntajes “brutos”.
- Puntajes estandarizados.
- Puntajes de Teoría de Respuesta al Ítem.
- Porcentaje de alumnos que logran cierto nivel de puntaje o de desempeño.
- Porcentaje de alumnos que logran el dominio de objetivos específicos.
- Una transformación de alguna de las métricas anteriores a una calificación numérica expresada en la misma escala que se emplea en el sistema para calificar a los alumnos.

Detrás de cada una de estas métricas hay implicancias y significados diversos en los que es preciso formar a la opinión pública y a los docentes, de modo que puedan hacer usos e interpretaciones adecuadas de los datos. Comprender qué significan las diferentes métricas es relevante porque está en juego (i) la comprensión de la información por parte de los destinatarios y (ii) la utilidad de la información que se entrega.

La disyuntiva principal es la siguiente: limitarse a **describir lo que los alumnos logran hacer en función de tareas propuestas y a establecer comparaciones entre diferentes sectores del sistema educativo o entre años**, para obtener conclusiones del tipo “los logros en lenguaje son mejores en las escuelas urbanas que en las rurales” o “en tal provincia respecto a tal otra”, o “los resultados en matemática han mejorado —o han empeorado— con respecto a dos años atrás”; o **establecer estándares o criterios acerca de cuáles son los desempeños esperables y exigibles a todos los alumnos cuando culminan cierto grado o nivel de escolaridad**, lo que permite, además del tipo de conclusiones anteriores, otras del tipo “tal proporción de estudiantes no alcanzan los niveles de desempeño esperados en lectura en 3<sup>er</sup> grado” o “tal proporción de los estudiantes domina el objetivo o contenido X de matemática en 6<sup>o</sup> grado”.

Las conclusiones del primer tipo son útiles en términos de obtener una visión de las inequidades al interior del sistema educativo y de la evolución de los resultados a lo largo del tiempo, pero las expectativas de la opinión pública y las

familias parecen estar demandando conclusiones del segundo tipo. Estas también tienen un efecto más fuerte sobre los docentes, pues les pueden dar una señal acerca de lo que se espera que todos los alumnos sean capaces de hacer en un nivel determinado.

### El contexto sociocultural

Una de las debilidades principales de los reportes de las evaluaciones nacionales es la falta de conceptualización sociocultural de los resultados. En prácticamente todos los países se reporta en función de:

- Resultados según las divisiones políticas y/o geográficas del país.
- Resultados de escuelas urbanas y rurales.
- Resultados según forma de administración (escuelas públicas, privadas, subvencionadas).

Estos niveles de desagregación son relevantes en la medida en que permiten ver las diferencias en el logro de los alumnos entre dichos subsistemas y presentan a la luz ciertas inequidades en la distribución de los conocimientos. Sin embargo, implican una sobresimplificación de la realidad: algunos de estos agregados incluyen en su interior enormes heterogeneidades (por ejemplo, en ‘escuelas urbanas’ hay toda una gama de situaciones socioculturales); en otros casos, los diferentes tipos de escuelas tienen un alumnado perteneciente mayoritariamente a ciertos sectores sociales y no una población representativa del conjunto (*por ejemplo, la mayoría de las escuelas rurales trabajan con alumnos provenientes de familias con mínima trayectoria en el sistema educativo, mientras que la mayoría de las escuelas privadas reclutan sus alumnos de sectores medios y altos*).

En general, la investigación empírica demuestra que las diferencias de resultados obedecen principalmente a las desigualdades socioculturales con las que trabajan los distintos sectores o instituciones del sistema educativo, antes que a diferencias en su eficacia pedagógica. Lo anterior no implica afirmar que todas las discrepancia de resultados se expliquen de esta forma, pero para hacer afirmaciones válidas respecto de la eficacia pedagógica es imprescindible haber controlado el efecto del contexto sociocultural.

### Vacíos técnicos

Finalmente, es ineludible señalar tres vacíos o debilidades técnicas verificables en la mayoría de los reportes de evaluaciones que son relevantes desde el punto de vista del lector especializado y de la apertura al escrutinio público de la calidad de la información generada:

- No se reportan las tasas de respuesta a los distintos instrumentos aplicados, es decir, qué proporción de los alumnos incluidos en el diseño muestral o registrados en las escuelas efectivamente participó, lo cual es crucial dado que en muchos países se registran altos niveles de inasistencia el día de las pruebas o de falta de respuesta a los instrumentos complementarios.
- No se incluye la estimación de los márgenes de error muestral, básicos para analizar el significado de la información que se brinda.
- Falta información acerca de los procedimientos seguidos para la equiparación de los instrumentos de medición en los casos en que se aplican evaluaciones sucesivas en diferentes años en los mismos grados y disciplinas.

**Figura 2**  
Métricas empleadas por los distintos países para el reporte de resultados

Países	% promedio respuestas correctas	Promedio de puntajes brutos	Puntajes estandarizados	Puntajes «teoría de respuestas al ítem»	% alumnos por niveles de desempeño	% alumnos que dominan objetivos	Calificaciones
Argentina							
Bolivia			Puntaje T				
Brasil							
C. Rica							
Chile	Hasta 1997			Desde 1998			
Ecuador							
Paraguay							
Perú							
Uruguay							

Nota: Los casilleros más oscuros indican cuál es la métrica central utilizada en los reportes de cada país.

## Informe de evaluaciones educacionales:

## Diez pistas para avanzar en el futuro

1

• Antes de adosar a las evaluaciones algún tipo de consecuencia directa para las escuelas (como incentivos económicos, publicación de rankings u otros), sería conveniente realizar alguna investigación cualitativa acerca del impacto que realmente tiene sobre las escuelas una estrategia de este tipo. Este enfoque genera presiones sobre las escuelas, en el sentido de que pueden destinar mucho tiempo a preparar a los alumnos para responder las pruebas, lo cual puede dar lugar a una reducción del currículo cuando se trabaja con pruebas de selección múltiple o de respuesta sencilla.

2

• Si la apuesta principal es que los sistemas de evaluación aporten a los docentes información relevante para revisar y mejorar su trabajo, habría que generar conocimiento específico sobre las siguientes preguntas: ¿están los maestros recibiendo informes y publicaciones derivadas de las evaluaciones nacionales?, ¿en qué proporción efectivamente las leen?, ¿en qué medida comprenden el contenido de esas publicaciones?, ¿en qué grado la información les resulta pertinente y enriquecedora para comprender los procesos de aprendizaje de los alumnos y para analizar sus propias prácticas de enseñanza?, ¿cuáles de las distintas estrategias de divulgación (envío de publicaciones, talleres de capacitación, aplicación y corrección autónoma de pruebas, etc.) resultan más pertinentes para que las evaluaciones sirvan para mejorar la enseñanza?, ¿qué otras expectativas o demandas tienen los maestros hacia las unidades de evaluación?

3

• Aun cuando no se desee adoptar el camino de la publicación de resultados por escuela en la prensa, queda en pie el desafío de cómo informar a las familias acerca de los resultados de las evaluaciones: qué tipo de información brindarles y de qué modo hacerlo. En principio, debería tratarse, más que de información numérica, de información ilustrativa acerca de qué es lo que se espera que sus hijos sean capaces de hacer, en qué grado lo están logrando y qué pueden hacer las familias para apoyar la labor de la escuela.

4

• Se requiere conocer más acerca de cómo otros actores relevantes (autoridades educativas, profesionales de unidades de currículo o formación de docentes, políticos, dirigentes sociales) comprenden e interpretan los resultados, y qué demandas de información tendrían hacia los sistemas de evaluación. Estudios de casos nacionales mediante encuestas y grupos de discusión podrían indagar al respecto.

5

• Habría que efectuar una revisión sistemática de cómo la prensa ha venido divulgando los resultados de las evaluaciones en los países de la región, el tipo de visión que se deriva hacia la opinión pública y el tipo de errores o interpretaciones inapropiadas más comunes. Eventualmente, debería pensarse en instancias formativas dirigidas a periodistas dedicados a temas educativos.

6

• Para dar respuesta a lo que los diferentes actores esperan de los sistemas de evaluación, probablemente será necesario avanzar hacia formas de diseño de las pruebas y de los reportes de resultados de carácter “criterial”; es decir, que no se limiten a describir promedios de respuestas correctas y a comparar entre tipos de escuelas o jurisdicciones en que el país está dividido, sino que establezcan niveles de desempeño exigibles a todos los alumnos en cierto grado o nivel, y que reporten el grado en que ello ha sido alcanzado. Asimismo, los reportes numéricos deberían ilustrar ampliamente lo que los puntajes significan en términos de conocimientos y competencias de los alumnos.

7

• En el análisis y presentación de los resultados de las evaluaciones, es preciso incorporar de alguna manera la consideración de la composición sociocultural del alumnado de los diferentes tipos de escuelas o sectores del sistema educativo.

8

• Habría que dar “valor agregado” a los informes, en términos de reflexión sobre los datos y la relación de estos con lo que otras investigaciones muestran. Asimismo, es preciso reportar sistemáticamente acerca de otras variables sociales e institucionales por sí mismas, de modo de enriquecer la comprensión sobre cómo es el sistema educativo y sus heterogeneidades internas.

9

• Un desafío relevante es asumir el monitoreo en el tiempo de los resultados de nuestros sistemas educativos. Esta tarea tiene fuertes implicancias de tipo técnico para garantizar la equivalencia de los instrumentos a lo largo del tiempo.

10

• Las unidades de evaluación de la región tendrían que trabajar en forma conjunta en la definición de estándares de calidad técnica que deberían satisfacer las evaluaciones y los reportes de resultados.