

ABRIL 2001 - Año 2 / N°8

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

## ¿Cómo avanzar en la evaluación de aprendizajes en América Latina?

Al cabo de una primera etapa de instalación de sus sistemas nacionales de evaluación del aprendizaje —durante la década de los '90—, muchos países de América Latina se encuentran ingresando en una fase de revisión de lo realizado hasta el momento y de consideración de nuevas alternativas en este campo. Es éste, por lo tanto, un momento propicio para examinar por lo menos tres aspectos fundamentales:

- el papel de los sistemas de evaluación en la política educativa;
- la necesidad de mejorar la calidad técnica de los diversos aspectos constitutivos de los sistemas de evaluación; y
- las estrategias de uso y difusión de los resultados de las evaluaciones.

Ello requiere detenerse previamente a analizar los distintos fines para los que pueden servir diversos sistemas de evaluación, los paradigmas que los sustentan, las técnicas e instrumentos más adecuados según los fines particulares que persigue un país con sus pruebas, y la información complementaria que puede incorporarse en las mediciones para interpretar mejor los resultados del proceso educativo y los factores que los determinan.

*Se resumen en este número algunas partes del documento “Los próximos pasos: ¿Hacia dónde y cómo avanzar en la Evaluación de aprendizajes en América Latina?” —de los autores Pedro Ravela (editor), Richard Wolfe, Gilbert Valverde y Juan Manuel Esquivel— producido en el marco del Grupo de Trabajo sobre Estándares y Evaluación GRADE/PREAL en marzo del año 2000. El documento completo está disponible en <http://www.grade.org.pe/gtee-preal>.*

La incorporación de sistemas nacionales de evaluación de los aprendizajes en América Latina empezó en la década de los '80, cuando algunos países de la Región desarrollaron algunas experiencias en este campo, sin haber llegado entonces a institucionalizar tales prácticas de manera sistemática. En los '90, sin embargo, se asiste a una fase de instalación formal de estos sistemas en toda la Región, realizándose también una primera experiencia de evaluación a nivel regional, conducida por UNESCO/OREALC.

La multiplicación de estas actividades constituye una clara manifestación de la preocupación de los gobiernos por producir información sobre los aprendizajes que se logran al interior de los sistemas educativos. En un contexto internacional en el que el conocimiento y las capacidades de los individuos serán cada vez más importantes para el desarrollo y competitividad de los países, es previsible que en los próximos años esos esfuerzos evaluativos se mantengan. Asimismo, en la medida en que exista una mayor conciencia respecto a la necesi-

dad de incrementar y utilizar más eficientemente los recursos siempre escasos destinados al sector educación, resultará imprescindible contar con información que permita evaluar el impacto de una inversión adicional y monitorear en forma permanente y adecuada los avances y retrocesos en los resultados del sistema educativo.

El panorama de la primera década de experiencia extendida en materia de evaluación de aprendizajes a nivel nacional, muestra una importante diversidad de: enfoques, conocimientos y competencias que son evaluados; periodicidad de las evaluaciones; grados y áreas curriculares en los cuales se realizan las mediciones; el tipo de variables contextuales sobre las que se recoge y reporta información; y los análisis y formatos de devolución de la información, entre otros.

Más allá de cuán útiles hayan sido los resultados de los distintos sistemas, sí se puede decir que en la pasada década se dio un paso de enorme trascendencia para la Región: los sistemas de evaluación se han instalado en los países, se ha generado una cierta capacidad para la implementación de objetivos nacionales de evaluación a gran escala, y la sociedad y los cuerpos docentes comienzan a valorar y comprender la necesidad de este tipo de evaluaciones. Sin embargo, simultáneamente, al cabo de esta primera etapa muchos países se encuentran ingresando en una fase de revisión de lo hecho hasta el momento y de consideración de nuevas alternativas en este campo. Ello obedece a la constatación de tres tipos de debilidades:

- Insuficiente aprovechamiento de la información producida por los sistemas de evaluación, lo que tiene como consecuencia el insuficiente impacto de éstos en el sistema educativo.
- Insuficiente calidad y capacidad de evaluación de aprendizajes complejos en las pruebas que están siendo aplicadas.

- Debilidades técnicas en los procesos de desarrollo y validación de los distintos instrumentos de medición.

En la actualidad, parecen estar dadas las condiciones para dar un “salto cualitativo” en materia de evaluación, el cual exige desarrollar un proceso reflexivo en dos grandes planos: las opciones de política en materia de evaluación nacional, y las opciones técnicas para llevar a cabo las evaluaciones en función de las opciones de política escogidas. Para esto es necesario replantearse tanto los fines que pueden cumplir los distintos sistemas de evaluación (ver recuadro) y contrastarlos con lo que se desea en cada caso particular, así como considerar en el diseño de los sistemas las implicancias derivadas de los cuatro temas que se presentan a continuación.

## 1. El dilema de la granularidad: cobertura curricular versus cobertura poblacional

Un asunto crítico en el diseño de los sistemas nacionales de evaluación es el de la “granularidad”, entendida ésta como *la cantidad de detalle con que el sistema recoge y luego reporta los datos*. Por ejemplo, puede haber enormes diferencias en el costo y en el modo de utilización entre sistemas de evaluación que sólo proporcionan resultados nacionales y aquéllos que suministran resultados de todos los estudiantes o escuelas individualmente. De igual manera, pueden haber grandes diferencias entre las evaluaciones que dan información general sobre temas amplios, tales como los logros en matemáticas o lenguaje, y aquéllas que brindan información detallada sobre lo que los estudiantes saben y pueden o no hacer en áreas específicas de esas asignaturas.

Así, los sistemas se diferencian según la forma en que combinen dos dimensiones:

### A. ¿Quién es evaluado?

Si bien los estudiantes —y a menudo los padres, los profesores, los directores escolares y otros— son las fuentes primarias de datos de un sistema de evaluación educacional, no suelen ser la principal unidad para la cual se calculan los resultados y se hacen los reportes. En general se encuentran los siguientes niveles de análisis y de reporte:

- Poblaciones nacionales o internacionales (ej.: la población de escolares de 3er grado de primaria).

### Recuadro

## Objetivos posibles de los sistemas de evaluación

Un sistema nacional de evaluación de aprendizajes puede proponerse alguna o varias de las siguientes finalidades, e incluso otras adicionales, pero un sólo tipo de medición no puede cumplir con todas ellas:

- Construir un “mapa de situación” del sistema educativo, con el fin de identificar áreas prioritarias de intervención y tipos de intervenciones necesarias.
- Evaluar el impacto de políticas, innovaciones o programas específicos.
- Conocer la dirección y magnitud de los cambios que se están dando en las condiciones y logros de la educación.
- Evaluar la productividad de los maestros para efectos de establecer un sistema de incentivos.
- Brindar a los padres de familia información que les permita evaluar la calidad de las escuelas.
- Devolver información a las escuelas y maestros para que éstos examinen los resultados de su trabajo.
- Establecer la acreditación de los alumnos que finalizan un determinado nivel de enseñanza.
- Seleccionar u ordenar a los estudiantes.
- Informar a la opinión pública y generar una cultura de la evaluación.
- Contribuir a establecer estándares de calidad para el sistema educativo.
- Realizar estudios tipo costo-beneficio.
- Contribuir a la generación del conocimiento (para efectos de investigación sobre funcionamiento del sistema educativo, prácticas de enseñanza, impacto de variables sociales sobre aprendizaje, entre otros).

La experiencia indica que en muchos países de la Región ha sido insuficiente la reflexión acerca de los fines específicos que se espera que cumplan sus sistemas de evaluación y acerca de las definiciones técnicas más adecuadas para cada fin.

- Los principales estratos definidos educacional, política y socialmente (ej.: los estudiantes en escuelas públicas, en escuelas rurales, en programas bilingües, etc.).
- Principales divisiones regionales (provincias, estados...).
- Jurisdicciones menores (ciudades, municipios).
- Escuelas.
- Salones de clase (profesores).
- Estudiantes.

### B. ¿Qué es lo que se evalúa?

En este ámbito, hay también varios niveles, tales como:

- Resultados globales, incluyendo matrícula, participación en la evaluación, sin verdadera evaluación de contenidos.
- Éxito o fracaso general, culminación del plan de estudios, graduación, certificación, tal vez basadas en diferentes asignaturas y otra información.
- Puntajes en asignaturas, tales como el logro general en matemáticas o en lenguaje.
- Puntajes en áreas de asignatura (ej., solución de algoritmos, álgebra, comprensión

de lectura).

- Logro en niveles particulares de desempeño en diferentes estándares en una asignatura.
- Estadísticas de respuestas para ítemes específicos.
- Registro detallado de las respuestas a una prueba.

La granularidad del contenido determina la capacidad de interpretar y comprender la calidad del logro educacional y de tomar medidas para mejorarla, así como la capacidad para usar información de la evaluación para diseñar ajustes al currículum y a la enseñanza.

El gran dilema de los sistemas de evaluación es que atender simultáneamente estas dos dimensiones de granularidad genera usualmente un conflicto o “*trade-off*”. Suele haber una relación inversa entre las posibilidades de maximizar cobertura curricular y cobertura poblacional. Cuanto más detalladamente se desee conocer qué aprenden los alumnos en cierto nivel del sistema educativo, menos factible será contar con información desagregada a nivel de

distrito y establecimientos educativos. Por el contrario, si el propósito es generar información con estos últimos niveles de desagregación, entonces sólo será posible obtener mediciones más globales y menos detalladas de lo que los alumnos aprenden. Un desafío central para el diseño de una evaluación es desarrollar metodologías que permitan combinar diferentes propósitos de la mejor manera.

## 2. Validez de las mediciones

¿Qué significan los resultados que obtienen los estudiantes en nuestras pruebas nacionales de rendimiento? ¿Estamos realmente sacando conclusiones apropiadas, significativas y útiles a partir de esos resultados? ¿En qué medida podemos justificar la manera en que interpretamos el resultado de una evaluación?

Estas son preguntas centrales en el tema de las evaluaciones educacionales, puesto que interpretar correctamente y usar apropiadamente la información que nos dan las pruebas es algo crucial. A esto apunta el concepto de “validez”, definido actualmente como *el grado en que la evidencia empírica y la teoría dan sustento a las interpretaciones de los resultados de una medición*. La validez se refiere al ámbito del uso legítimo de esas interpretaciones y también al grado en que el uso de la prueba no produce un impacto negativo no deseado sobre el sistema educativo. La validez no es, como se pensaba antes, “que se está midiendo lo que se pretende medir”, no es meramente una cualidad “técnica” de la medición que se garantiza con una u otra prueba estadística, ni tampoco es algo muy complejo o sofisticado que necesariamente escapa a la comprensión del gran público. La validez se garantiza con la aplicación de rigor lógico y de sensibilidad en la construcción y perfeccionamiento de los instrumentos de medición, con la consideración sistemática del tipo de decisiones que podrían llegar a tomarse usando como sustentación o evidencia los resultados de esa medición.

En América Latina, donde los recursos son tan escasos, a veces se pretende que una misma prueba sirva para más de un propósito. Por ejemplo, que permita distinguir entre estudiantes que logran o no los objetivos académicos de un nivel y que, al mismo tiempo, sirva para juzgar la eficacia de distintas escuelas y la eficacia de diversos programas en los cuales participan dichas escuelas. Sin embargo, como consecuencia de lo que se dijo líneas arriba, la validación debe ser específica, según el tipo de interpretación y uso

que se va a hacer de los resultados. Es decir, validar una prueba que persigue un determinado propósito no significa que queda automáticamente validada para otro. La validez es también específica a las poblaciones. Una prueba validada para un país o provincia determinada no puede ser considerada como validada para otras poblaciones. Por último, es necesario tomar en cuenta que el tiempo cambia las características de los fenómenos y que, por lo tanto, la validación es una tarea continua y una forma de asegurar que nuevos factores que puedan aparecer con el tiempo no atenuen la validez de las conclusiones que se deriven de las evaluaciones.

Es necesario, por lo tanto, realizar un análisis más cuidadoso de la información que producen los sistemas de evaluación, en el sentido de validar las conclusiones e interpretaciones que se derivan de su análisis. Ello implica una labor de formación permanente de los usuarios (autoridades, opinión pública, medios de comunicación, maestros) respecto de los usos válidos de los distintos tipos de información y respecto al tipo de interpretaciones que no es posible extraer válidamente de los resultados de las evaluaciones, así como mecanismos continuos de consulta para garantizar la validez de los instrumentos para los propósitos viejos y nuevos que deben servir.

## 3. Enfoques de las mediciones

Los encargados de diseñar los sistemas de medición de logro en América Latina se enfrentan a la disyuntiva de desarrollar pruebas de rendimiento para comparar el logro de grupos de estudiantes con otros grupos o para descubrir qué aspectos, conocimientos u objetivos específicos logran los estudiantes. Responder a este dilema implica desarrollar pruebas en base a paradigmas con fundamentaciones teóricas diferentes. En el primer caso, se trabaja con el paradigma referido a *normas* y, en el segundo, con el paradigma referido a *criterios*.

En la mayoría de los países de la Región, el diseño de las pruebas ha estado marcado fuertemente por el primer paradigma. Las pruebas referidas a normas permiten la comparación del desempeño relativo en un área amplia de contenidos (por ejemplo, «matemáticas» o «lenguaje») de distintos grupos de estudiantes. Bajo ciertas condiciones, también permiten hacer un seguimiento a lo largo del tiempo de ese desempeño relativo general. Sin embargo, ese enfoque no permite obtener información específica y válida sobre el logro de deter-

minados conocimientos, habilidades y destrezas, como sí lo pueden hacer las mediciones basadas en criterios. Estas últimas requieren el establecimiento previo, a partir de consultas y opciones expertas, de los niveles de desempeño o estándares que son exigibles a los alumnos en el grado evaluado. Permiten luego reportar qué porción de alumnos ha alcanzado esos estándares en áreas temáticas más circunscritas, y aportan mejor información para promover su mejoría.

Una nueva alternativa en materia de evaluación que ha cobrado fuerza en los países desarrollados en la última década, es la llamada “evaluación de desempeño”, según la cual los estudiantes deben producir sus respuestas o ejecutar tareas, en lugar de simplemente seleccionar la respuesta correcta entre varias alternativas. Para esto se emplean preguntas de respuesta abierta, requerimientos de resolver problemas, producir un ensayo, materiales o discursos públicos, artefactos o documentos, y elaborar portafolios o muestras de trabajos realizados a lo largo de un periodo. Lo novedoso del enfoque es el énfasis en la medición de conocimientos y habilidades complejas tal y como se le presentan al estudiante en la vida real. Es, necesario, sin embargo, tomar en cuenta dos características de este enfoque cuando se aplica a muestras masivas de estudiantes: el tiempo que requiere un alumno para completar una tarea específica y la limitación técnica de depender del juicio humano para juzgar la calidad de la respuesta. Por otra parte, el factor costo en este caso es una seria limitante.

No obstante, es indudable que es necesario medir el conocimiento y las habilidades complejas y resulta atractiva la idea de aplicar algunos de los medios de la evaluación del desempeño a las pruebas de la Región.

## 4. Medición de aspectos contextuales

Un último aspecto que se requiere tener presente en la revisión de los sistemas de evaluación de la Región, es que no sólo es necesario mejorar sustancialmente la calidad de los instrumentos de medición de aprendizaje, sino también los instrumentos de medición de aspectos relevantes del contexto social y escolar en que los aprendizajes ocurren. La mayor parte de los países recoge información sobre variables sociales y escolares, pero en muchos casos la calidad de la misma no es suficiente y menos aún el aprovechamiento que de ella se hace para el análisis de los resultados de aprendi-

zaje y la investigación. Al respecto, es necesario:

- Mejorar la medición de las variables sociofamiliares, con el fin de contextualizar socialmente el análisis y reporte de los resultados y evitar la falacia de atribuir a los establecimientos educativos el mérito o la culpa por resulta-

dos que, en realidad, obedecen a la selección social del alumnado.

- Mejorar la medición de variables de tipo institucional y pedagógico, con el fin de desarrollar trabajos de investigación que permitan comprender la compleja trama de factores que intervienen en el logro de los aprendizajes y, de este modo, en-

riquecer el horizonte conceptual y la base empírica de la de decisiones de política educativa.

- Asegurar la recolección de medidas a lo largo del tiempo, de manera de poder el “valor agregado” que puede atribuirse a variables que el sistema educativo sí puede contribuir a modificar. ©

## Los próximos pasos

Lo presentado anteriormente permite afirmar que, si bien al cabo de esta primera década de instalación de los sistemas nacionales de evaluación en la Región se han dado pasos importantes, aún queda mucho por hacer y mejorar en este campo. En ese sentido, una posible agenda de trabajo para dar los próximos pasos estaría centrada en tres ejes principales.

### 1. El papel de los sistemas de evaluación en la política educativa.

Se trata de definir con qué estrategia se espera que un sistema de evaluación nacional de aprendizajes tenga algún impacto en la mejora de los aprendizajes. Esta reflexión debería involucrar preguntas como:

- ¿Conviene que la evaluación tenga consecuencias “fuertes” para las escuelas y maestros —ya sea bajo la forma de incentivos explícitos o bajo la forma de la publicación de un ranking de resultados— o es preferible que cumpla una función fundamentalmente informativa?
- ¿De qué modo articular los esfuerzos de reforma y actualización de las currícula? ¿De qué modo pueden las evaluaciones contribuir a mejorar la definición de las metas e indicadores de logro curriculares?
- ¿Se desea contar con información exhaustiva acerca de las competencias y conocimientos de los alumnos a nivel nacional o se prefiere producir información menos detallada al respecto, pero a nivel de cada establecimiento?
- ¿Es conveniente desarrollar pruebas nacionales de acreditación —es decir, que determinen la aprobación o reprobación de los alumnos— al cabo de algún nivel de la enseñanza?
- ¿Se espera que el sistema de evaluación permita constatar avances o retrocesos a lo largo de los años? ¿En qué áreas curriculares y en qué niveles del sistema educativo?
- ¿Con qué frecuencia realizar operaciones nacionales de evaluación?

### 2. La necesidad de mejorar la calidad técnica de los diversos aspectos constitutivos de los sistemas de evaluación.

Esto, principalmente en lo que se refiere al diseño de los instrumentos de recolección de información y a los modos de procesar y reportar los resultados. Lo anterior exige intensificar los esfuerzos de capacitación de cuadros técnicos y la acumulación de conocimiento y experiencia en una materia que aún es nueva en la Región y sobre la cual existe escasa “masa crítica”. Para estos efectos se puede, entre otras cosas:

- Facilitar el contacto de profesionales de la Región vinculados al área de evaluación con especialistas de la comunidad internacional a través, por ejemplo, de seminarios donde representantes de los distintos países sometan sus instrumentos y procedimientos al escrutinio de otros.
- Impulsar la formulación de un conjunto de estándares técnicos y condiciones metodológicas que deberán cumplir las pruebas, los procedimientos de implementación de los operativos de evaluación, los procesos de conformación y procesamiento de las bases de datos y los reportes de resultados.

### 3. Las estrategias de uso y difusión de los resultados de las evaluaciones.

Se requiere propiciar el desarrollo de trabajos de investigación que permitan recoger evidencia empírica acerca de los efectos que los distintos tipos de informes de resultados de las evaluaciones nacionales tienen en diversos públicos. Por otra parte, sería muy útil organizar eventos que permitan escuchar la opinión de los destinatarios respecto a:

- ¿Qué tipo de información esperan recibir del sistema de evaluación de aprendizajes las diversas audiencias (periodistas, padres, maestros, políticos, autoridades y técnicos de los Ministerios de Educación, etc.)?
- ¿Cómo perciben la información que actualmente se les está entregando? ¿Han podido comprenderla? ¿La han utilizado de algún modo?
- ¿Qué visión general tienen acerca de los sistemas de evaluación de aprendizajes a nivel nacional?
- ¿Cuáles son sus expectativas y prejuicios acerca de los mismos?

Lo anterior permitiría analizar la demanda potencial de información, aprender acerca de los modos de informar a los distintos usuarios potenciales, además de aportar pistas para desarrollar estrategias de formación de dicha demanda, ya que a través de estos eventos se estaría informando mejor sobre un tema relativamente nuevo en la Región y para los diversos actores.