

SISTEMAS DE EVALUACIÓN DE APRENDIZAJES EN AMÉRICA LATINA BALANCE Y DESAFÍOS

Guillermo Ferrer



**Sistemas de Evaluación de
Aprendizajes en América Latina
Balance y Desafíos**

© Preal

Autor:
Guillermo Ferrer

Primera edición:
1.000 ejemplares
Marzo 2006

Registro de Propiedad Intelectual:
N° 153.429

I.S.B.N. 956-8109-01-3

Impresión:
Alfabetas Artes Gráficas
Carmen 1985
Fono Fax: 551 5657

Índice

Introducción.....	7
Aspectos metodológicos de este trabajo	10

PRIMERA PARTE

I. ALGUNAS CONDICIONES Y PROCESOS DESEABLES PARA EL FORTALECIMIENTO Y LEGITIMACIÓN DE LOS SISTEMAS DE MEDICIÓN EN AMÉRICA LATINA	13
---	-----------

II. PRINCIPALES AVANCES Y DIFICULTADES EN EL DESARROLLO DE LOS SISTEMAS NACIONALES DE EVALUACIÓN Y MEDICIÓN DE LA CALIDAD EDUCATIVA DE AMÉRICA LATINA	19
--	-----------

Fines y propósitos del sistema	20
Marco institucional de las agencias o unidades de medición	20
La medición de aprendizajes y su relación con el desarrollo curricular y de estándares	24
Tipos de instrumentos de medición: Características y validación.....	28
– Cobertura general de las evaluaciones	28
– Diseño de instrumentos	29
– Validación técnica y adecuación lingüística	33
Formas de reporte de los resultados	35
– Reportes a autoridades de gobierno y alta dirección sectorial	35
– Reportes a actores educativos locales	37
– Informes específicos para padres de familia	41
– Seguimientos cualitativos	42

Difusión y uso de los resultados de las mediciones	43
– Uso para toma de decisiones de política o para el diseño de programas de mejoramiento	45
– Uso pedagógico local	46
– Uso de altas implicancias	48
– Responsabilización política	49
Pruebas internacionales comparadas de aprendizaje	50

III. DISCUSIÓN Y SÍNTESIS 51

SEGUNDA PARTE

Descripción de casos nacionales y subnacionales	59
Argentina	60
Bolivia	65
Brasil	69
Chile	74
Colombia	80
Costa Rica	85
Cuba	90
Ecuador	94
El Salvador	98
Guatemala	103
Honduras	107
México	112
Nicaragua	118
Panamá	121
Paraguay	125
Perú	129
República Dominicana	133
Uruguay	138
Venezuela	144
Minas Gerais, Brasil	147
Paraná, Brasil	152
São Paulo, Brasil	157
Bogotá, Colombia	162
Aguascalientes, México	166
Referencias bibliográficas	172
Entrevistas	172



Introducción

El Grupo de Trabajo sobre Evaluación y Estándares (GTEE) del PREAL –Programa para la Promoción de la Reforma Educativa en América Latina y el Caribe– viene realizando desde el año 2000 varios estudios sobre el estado de situación de la medición y evaluación de la calidad educativa en América Latina, así como del avance de los países de la Región en la propuesta y definición de expectativas claras de aprendizaje, las que en este documento serán denominadas “estándares”. Tres de estos estudios sientan las bases conceptuales de este informe sobre la situación actual de los sistemas nacionales de evaluación y desarrollo de estándares en la Región, y serán citados frecuentemente a lo largo del texto. Esos informes son:

“¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?” (PREAL, 2001);

“Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?” (PREAL, 2001); y

“Aspectos del currículum prescrito en América Latina: Revisión de tendencias contemporáneas en currículum, indicadores de logro, estándares y otros instrumentos” (PREAL, <http://www.preal.cl/GTEE/pdf/Ferrer.pdf>, 1999).

El presente informe del GTEE apunta a dar un paso más allá en la documentación de los avances de América Latina y el Caribe en cuanto al



establecimiento, consolidación, perfeccionamiento y utilidad de los sistemas nacionales y subnacionales de medición de aprendizajes, así como de los esfuerzos realizados en la Región para el desarrollo de estándares de logro académico en la educación básica.

En la **primera parte** de este documento se discute el progreso de los países latinoamericanos en sus esfuerzos por establecer expectativas claras de aprendizaje, desarrollar herramientas de medición y evaluación adecuadas para informar sobre el logro efectivo de esas expectativas, comunicar a los actores educativos y a la sociedad civil sobre la calidad de los sistemas públicos de educación en términos de aprendizajes logrados, y diseñar y ajustar políticas educativas para el mejoramiento de esos aprendizajes sobre la base de las informaciones recogidas en las evaluaciones.

Los esfuerzos iniciales de indagación estuvieron orientados a sistematizar la información disponible sobre los sistemas nacionales de medición, concretamente en relación a las agencias responsables de la evaluación en cada país, el tipo de instrumentos desarrollados, las poblaciones estudiantiles evaluadas y el tipo de uso que se da a los datos recogidos por esos sistemas. Para esa tarea se recurrió a los datos recogidos por otros investigadores¹ en los últimos años, y sobre esa base se amplió y actualizó la información mediante entrevistas y análisis documentales realizados en casi todos los países de la Región.

El informe busca aportar una visión crítica y conceptual sustentada sobre esas acciones. Para facilitar la comprensión de esa visión, se propone en la primera sección una síntesis del tipo de procesos y condiciones que, según este GTEE, y tal como se ha avanzado en estudios anteriores, contribuirían al fortalecimiento de una cultura evaluativa sólida y legitimada en los países de América Latina, y a un mayor impacto en la toma de decisiones de política y en la práctica pedagógica para el mejoramiento del logro académico de los estudiantes.

Se presenta también una ampliación y profundización de cada una de las categorías de análisis seleccionadas para este estudio de los sistemas nacionales de evaluación, y un resumen narrativo de algunos de los casos nacionales y subnacionales que pueden servir como ejemplo ilustrativo de las diferentes modalidades técnicas y organizaciona-

¹ Rojas, C. y Esquivel, J.M. (1998); Wolff, L. (1998).

les de evaluación desarrolladas en la Región. En general, se optó por presentar en esta sección los casos más institucionalizados, o los más innovadores, aunque ocasionalmente se hace referencia a sistemas nacionales que sirven de contraejemplo para ilustrar las problemáticas que suelen enfrentarse en cada uno de los aspectos de la evaluación aquí analizados.

Luego se ofrece un apartado de discusión, a modo de resumen, de los avances y dificultades más importantes observados en los sistemas de evaluación regionales.

En la **segunda parte** del informe se presenta un resumen individual para cada caso nacional y subnacional de evaluación de aprendizajes, siguiendo las mismas categorías de análisis utilizadas en la primera parte del documento. A estos resúmenes individuales podrá referirse el lector para acceder a informaciones más detalladas sobre los casos mencionados a lo largo del documento, o para comparar las características de los sistemas entre países.

Es conveniente remarcar que a lo largo del proceso de elaboración de este informe, los sistemas nacionales de evaluación han sufrido cambios de tipo institucional y metodológico, coincidentes a menudo con los recambios frecuentes de autoridades políticas sectoriales. En algunos casos se trata de ajustes parciales en la metodología o cobertura de las evaluaciones, mientras que en otros se han realizado cambios estructurales e institucionales de considerable magnitud. **En todos los casos se ha intentado ofrecer la información más actualizada disponible al momento de esta publicación, aunque es posible que algunos cambios recientes no estén debidamente registrados.** De todas maneras, se invita al lector a considerar este análisis como un “retrato dinámico” que abarca principalmente el desarrollo de los sistemas nacionales de evaluación a lo largo de los años noventa, y algunos de los cambios más importantes introducidos recientemente como respuesta a las problemáticas enfrentadas en los diferentes países a lo largo de ese período.

Claramente, la descripción presentada podría ser actualizada a diario, considerando especialmente la inestabilidad institucional que experimentan muchos de los sistemas de evaluación y ministerios de educación de la Región, pero esperamos que este trabajo sirva al propósito principal de mostrar las principales tendencias, avances y problemáticas que se ha podido observar en los últimos diez años de evaluación estandarizada de aprendizajes.

ASPECTOS METODOLÓGICOS DE ESTE TRABAJO

Todas las informaciones presentadas en este informe, tanto en la primera como en la segunda parte, se basan en fuentes bibliográficas y documentales consultadas, así como en casi 60 entrevistas realizadas a representantes de las unidades de medición y expertos en evaluación de 19 países de la Región y cinco sistemas subnacionales. Otras informaciones adicionales fueron recogidas por medio de comunicaciones electrónicas con esos representantes nacionales, quienes agregaron, modificaron o actualizaron los datos que el autor les hizo llegar para su revisión.

Cada tema o categoría de análisis considerado fue desglosado en un conjunto de subtemas, tal como se describe en el siguiente cuadro:

Categorías de análisis	Subtemas
Marco institucional	Adecuación según marco político y capacidades técnicas: <ul style="list-style-type: none"> – Estabilidad – Financiamiento y autonomía administrativa – Recursos humanos – Autonomía y capacidad para la difusión de resultados – Transparencia
Currículo y estándares	<ul style="list-style-type: none"> – Disponibilidad, adecuación y utilización de marcos curriculares nacionales para el diseño de instrumentos de medición. – Desarrollo y validación de estándares alineados con el currículo vigente.
Instrumentos	<ul style="list-style-type: none"> – Marco conceptual sólido y explícito para la elaboración de matrices de referencia. – Validación de matrices de referencia e instrumentos. – Tipos de ítems o reactivos. – Adecuación cultural y lingüística (EBI). – Cobertura muestral o censal coherente con objetivos de la evaluación. – Estudio de contexto escolar y extraescolar para análisis de factores asociados al rendimiento.
Reportes	<ul style="list-style-type: none"> – Coherencia entre tipos de reporte y usos esperados (desarrollo curricular, pedagogía, focalización de apoyos, capacitación docente, selección de alumnos, etc.). – Adecuación de reportes a distintos tipos de destina-

	rios (claridad, guías de interpretación, sensibilización, etc.).
	– Informaciones sobre factores asociados y modelos de valor agregado.
Difusión y usos	Devolución: <ul style="list-style-type: none"> – Tiempos – Alcance – Periodicidad Impacto: <ul style="list-style-type: none"> – Uso escolar – Toma de decisiones de política/diseño de programas – Responsabilización política Altas implicancias: <ul style="list-style-type: none"> – Escuelas – Docentes – Alumnos
Sistemas subnacionales	Diferencias principales con sistemas nacionales: <ul style="list-style-type: none"> – Estándares – Cobertura censal – Participación de actores locales – Tiempos y formas de devolución de resultados – Uso centrado en el mejoramiento pedagógico
Pruebas internacionales	<ul style="list-style-type: none"> – Desarrollo de capacidades técnicas – Difusión de resultados e impacto en la opinión pública – Usos específicos de los resultados

Por tratarse también de un análisis crítico, y no meramente descriptivo, de los sistemas nacionales de evaluación, este informe contiene apreciaciones personales del autor sobre las características de esos sistemas y sobre el grado de desarrollo e impacto que han logrado en los últimos años. Sin embargo, las afirmaciones realizadas están basadas en el análisis de todas las fuentes disponibles para cada caso nacional y subnacional, e intentan incorporar las voces de diferentes actores del proceso evaluativo, tanto dentro como fuera de los órganos oficiales de gestión.

También fue necesario realizar comparaciones entre las versiones de actores gubernamentales pertenecientes a diferentes períodos de ad-

ministración sectorial, a menudo correspondientes a gobiernos nacionales de diferentes orientaciones políticas y estilos de gestión. Esta multiplicidad de voces implicó una tarea de síntesis e interpretación de la realidad que, seguramente, sería materia de debate en cada caso analizado. De todas maneras, resulta relevante recordar que todos los casos presentados fueron revisados por las autoridades técnicas en cada país, quienes tuvieron la oportunidad de corregir, moderar o actualizar las informaciones e interpretaciones ofrecidas por el autor.

Primera parte

I. ALGUNAS CONDICIONES Y PROCESOS DESEABLES PARA EL FORTALECIMIENTO Y LEGITIMACIÓN DE LOS SISTEMAS DE MEDICIÓN EN AMÉRICA LATINA

Es claro que los modelos y sistemas² de medición y evaluación del desempeño académico de los estudiantes deben diseñarse sobre la base de las condiciones políticas y de las capacidades técnicas y operativas, actuales o potenciales, de cada país. Sin embargo, la experiencia en evaluación acumulada hasta el presente en la Región, así como la experiencia ganada en países más desarrollados, permite avanzar un conjunto de condiciones que pueden considerarse deseables o recomendables para cualquier país que esté planificando o ajustando su sistema de medición y evaluación de resultados de aprendizaje. Entre ellas, merecen mencionarse las siguientes:

² En teoría, el término “sistema” da cuenta de una organización y articulación relativamente complejas entre las partes de un todo. En este trabajo mantendremos la denominación de “sistema” para todos los programas nacionales de evaluación de aprendizajes, aunque el lector observará que en algunos casos esos programas no lograron articularse con otros esfuerzos de política sectorial, o bien lo hicieron pero durante períodos de tiempo acotados, por lo cual podría considerárseles “proyectos de medición” y no estrictamente sistemas.

- Los propósitos de las evaluaciones de logros de aprendizaje y los usos pretendidos de sus resultados deben ser materia de amplia consideración y debate desde el inicio del proceso de establecimiento de un programa de pruebas o de un sistema de evaluación educacional más amplio. Las bondades del enfoque, metodología, instrumentos e impacto de las pruebas y evaluaciones tienen que establecerse en función de cuán bien sirven a esos propósitos y cuán efectivo ha sido el uso de los resultados para lograrlos.
- Los países deben considerar cuál es el marco o “estatus” institucional más adecuado para la agencia evaluadora responsable. Específicamente, conviene discutir y definir con claridad, de acuerdo a las condiciones políticas y técnicas de cada país, si la unidad de medición debe ser una dependencia oficial, un ente contratado externo, o si debe establecerse un sistema mixto. En cualquier caso, lo que resulta imprescindible es definir con claridad el grado de autonomía funcional que tendría la unidad evaluadora. El marco institucional puede tener fuertes implicancias para la credibilidad y legitimidad de sus acciones, para la sostenibilidad técnica-operativa y financiera del sistema, o para las posibilidades de difusión, uso e impacto de los resultados de las evaluaciones.
- Los instrumentos de medición deben estar referidos a las expectativas o metas de aprendizaje que cada país o región se haya propuesto mediante los marcos curriculares nacionales. Para esto se requiere que esas metas estén expresadas claramente, lo suficiente como para que las competencias cognitivas disciplinares que van siendo alcanzadas por los estudiantes puedan ser reconocidas de la manera más objetiva posible por todos los evaluadores del sistema: docentes, directores, supervisores y agentes profesionales externos a la escuela.
- Esas metas deben estar descritas operacionalmente³ y ejemplificadas para facilitar la comprensión de todos los actores del sistema sobre

³ Una definición operacional de las metas de aprendizaje consiste en la descripción clara y lo suficientemente detallada del tipo de saberes conceptuales y procedimentales que el estudiante debe demostrar haber desarrollado. Esto no implica un grado de operacionalidad tal que limite la observación y restrinja el juicio evaluativo a una forma única de expresión del conocimiento, pero sí debe ser lo suficientemente descriptivo como para guiar la construcción de instrumentos de medición válidos, asegurar la unidad de criterios al momento de la corrección de respuestas, y producir informaciones comprensibles y substantivas sobre qué tipos de conocimientos han adquirido los estudiantes.

cuáles son las expectativas de logro académico en las diferentes áreas curriculares. Adicionalmente, la descripción debe diferenciar y ejemplificar los distintos niveles de logro que es posible se obtengan en la práctica para cada competencia a desarrollar en las diferentes áreas curriculares evaluadas. Puede, por ejemplo, describirse un nivel de logro insuficiente, un nivel mínimo, uno medio y uno avanzado, para cada área curricular y cada nivel (grado o ciclo) de escolaridad evaluados. Esta clasificación, con su descripción operacional y su ejemplificación, constituye el marco de interpretación básico que permitirá a los profesionales de la educación y a los ciudadanos comprender la significancia de los datos arrojados por las evaluaciones, y permitirá también generar estrategias pedagógicas y curriculares focalizadas para mejorar los resultados de aprendizaje.

- Donde no existan definiciones claras y operacionales de las competencias complejas a desarrollarse, sino unas definiciones generales y orientadoras para el desarrollo curricular, como es el caso en casi todos los países de América Latina, es preciso reforzar el vínculo de trabajo y la comunicación entre las unidades técnicas de medición y los profesionales a cargo del diseño curricular oficial para desarrollar, en conjunto, indicadores claros de desempeño académico e instrumentos de evaluación debidamente alineados con esos indicadores.
- La metodología y cobertura de las mediciones deben ser consistentes con los propósitos y usos de las evaluaciones y deben validarse técnica y socialmente⁴. Esto implica tomar decisiones racionales y discutidas en profundidad sobre los siguientes aspectos:
 - El objetivo y los sujetos de las mediciones, es decir, si se busca evaluar el sistema, a las instituciones, o a los actores individualmente (docentes y alumnos, en especial). La decisión determinará, por ejemplo, si las pruebas deben ser censales o muestrales, si estarán referidas a normas o a criterios, o si los resultados tendrán altas o bajas implicancias para los actores.
 - Si los datos recogidos por las mediciones van a ser utilizados para el mejoramiento pedagógico o curricular, será necesario diseñar instru-

⁴ Por validación social entendemos los mecanismos de discusión y decisión compartida entre actores clave de la comunidad educativa sobre la pertinencia, relevancia y calidad de los contenidos e instrumentos de evaluación. Considérese también como parte de la comunidad educativa a los representantes de la sociedad civil que se busca involucrar y comprometer en el proceso de mejoramiento de la calidad educativa.

mentos adecuados para esos fines, lo que requerirá tomar decisiones sobre, por ejemplo, el modelo de medición utilizado para medir el logro de competencias de complejidad creciente, si los reactivos (ítemes) de las pruebas exigirán respuestas de opción múltiple, respuestas abiertas, demostraciones prácticas de laboratorio, o una combinación entre estas alternativas u otras que se propongan, el grado de cobertura del currículo que tendrán las pruebas, etc.

- Todos los instrumentos de las evaluaciones requieren ser validados en forma experimental, es decir, mediante aplicaciones piloto en el campo, antes de elaborar su versión definitiva.
- En algunos países será necesario realizar validaciones específicas cuando se decida evaluar poblaciones cuya lengua materna sea diferente de la lengua oficial. En estos casos, la validación técnica de los instrumentos involucrará aspectos lingüísticos (traducciones adecuadas o elaboración de reactivos en lenguas nativas), así como aspectos culturales generales para asegurar que los ítemes no contengan sesgos que dificulten las respuestas de manera discriminatoria.
- Si además de medir los aprendizajes se considera necesario o conveniente medir los factores escolares y extraescolares –“asociables” al rendimiento académico de los estudiantes– esto es, medir variables que pueden estar incidiendo o estar vinculadas con el desempeño escolar y analizar sus relaciones con este, es necesario utilizar procedimientos técnicos sólidos para la recolección de datos y su posterior análisis estadístico. Asimismo, pueden diseñarse instrumentos estandarizados (guías de entrevistas u observación, análisis de videograbaciones de aula, etc.) para la realización de seguimientos cualitativos sistemáticos sobre los procesos pedagógicos e institucionales que puedan contribuir a explicar los diferentes resultados de aprendizaje obtenidos en las mediciones cuantitativas.
- La forma en que se reportan los resultados de las evaluaciones debe guardar coherencia con los objetivos de las mediciones y debe tener en cuenta las necesidades concretas de información de los usuarios. Por ejemplo, si lo que se desea es brindar información sobre el logro de los estudiantes en algunas capacidades específicas de áreas curriculares recientemente implementadas, de manera tal que las escuelas puedan realizar ajustes allí donde se detecten dificultades, es poco útil reportar los resultados en forma agregada con un porcentaje bruto de logro para el total de cada área curricular o disciplina. De la misma manera, si lo que se busca es mayor involucramiento de los

padres en los procesos de aprendizaje de sus hijos, será necesario brindarles información no solamente sobre el nivel de logro de sus hijos en ciertas capacidades académicas, sino también una descripción y ejemplos del tipo de desempeño que se espera de ellos en el nivel educativo en que se encuentran.

- Si lo que se desea es emitir juicios sobre la calidad de las escuelas en términos de los logros académicos de sus estudiantes, y no solo comparar o establecer un *ranking* de resultados por institución o ámbitos geográficos, es necesario reportar los resultados de las evaluaciones distinguiendo el nivel económico y sociocultural de la población estudiantil con que esas escuelas trabajan, así como las condiciones de infraestructura, equipamiento y recursos humanos con que cuentan. De esa manera, será posible comparar los niveles de logro de cada escuela con el conjunto de escuelas que trabajan en condiciones similares y así evitar conclusiones injustas sobre la calidad de cada establecimiento⁵. Si las mediciones se realizan siguiendo un modelo de “valor agregado”, es decir, calculando el progreso académico de los estudiantes respecto de sí mismos en el tiempo, y en comparación con alumnos de niveles socioeconómicos similares, también podrán reportarse datos y realizarse inferencias más justas sobre la capacidad de cada centro educativo para ofrecer oportunidades adecuadas de aprendizaje.
- Las mediciones orientadas a recabar información que permita una intervención remedial para elevar los resultados de aprendizaje deben ser acompañadas con programas fuertes, focalizados o integrales, que apoyen el mejoramiento de la gestión y los procesos pedagógicos en las escuelas. En última instancia, los sistemas de evaluación solo se consolidarán y legitimarán socialmente en la medida en que los actores perciban que la medición es, en efecto, una herramienta para el mejoramiento de la calidad de la educación y no un ejercicio solo diagnóstico o eventualmente punitivo.
- Si los esfuerzos de evaluación están orientados a movilizar la opinión pública y las fuerzas de la sociedad civil en relación al mejoramiento de la calidad educativa, es necesario diseñar estrategias comunicacionales efectivas. La eficacia de esas estrategias depende no solamente de la capacidad del sistema para llegar a diferentes au-

⁵ Para un tratamiento más detallado, y adecuadamente ejemplificado, sobre este aspecto de la medición de aprendizajes se recomienda la lectura de Ravela, 2002.

diencias por medio de documentos y publicaciones masivas, sino también del grado de involucramiento de actores y organizaciones clave de la comunidad educativa y de la sociedad civil en el proceso evaluativo. En la medida en que esos actores, como por ejemplo representantes de los sindicatos docentes, empresarios o profesionales de los medios de prensa, sean partícipes de los procesos y del diseño de las estrategias comunicacionales, se logrará un mayor alcance e impacto de la información derivada de las mediciones. Aun cuando la difusión de los resultados quiera hacerse solo en forma local y focalizada, como por ejemplo a escuelas o municipios, será también necesario consultar ampliamente a representantes de esos sectores en particular para determinar con mayor precisión cuál será la mejor estrategia comunicativa a utilizarse.

- En cualquier caso, independientemente de cuáles audiencias se desee alcanzar con prioridad, es muy importante realizar la devolución de los datos en forma oportuna, periódica, y asegurando la llegada a la mayor cantidad posible de individuos e instituciones.
- Los esfuerzos de evaluación en cada país requieren de la iniciativa y voluntad política central (nacional) para establecerse, consolidarse y lograr impactos significativos, pero también es conveniente facilitar y promover las iniciativas subnacionales (zonales, jurisdiccionales) de evaluación externa. En los países en vías de desarrollo, la cooperación entre unidades técnico-políticas centrales y locales es una estrategia clave para ganar espacios de experimentación, legitimación y de mayor impacto de las prácticas evaluativas.
- Donde se decida participar en pruebas internacionales comparadas de rendimiento académico, y al igual que con las pruebas nacionales, es necesario definir claramente cuáles serán los objetivos nacionales específicos de la evaluación para poder diseñar estrategias de comunicación y líneas de intervención que fortalezcan el impacto de las informaciones recabadas mediante la medición⁶.

⁶ Un tratamiento en profundidad de este tema puede encontrarse en Ferrer y Arregui (2002).

II. PRINCIPALES AVANCES Y DIFICULTADES EN EL DESARROLLO DE LOS SISTEMAS NACIONALES DE EVALUACIÓN Y MEDICIÓN DE LA CALIDAD EDUCATIVA DE AMÉRICA LATINA

Durante las últimas dos décadas la mayor parte de los países de la Región ha adoptado algún modelo de evaluación nacional de los resultados de aprendizajes. En algunos casos la experiencia de evaluación fue ganándose con lentitud, con aplicaciones para fines muy específicos, como el de regular la admisión a la educación superior o evaluar el impacto de un proyecto focalizado de mejoramiento educativo auspiciado por un organismo de cooperación internacional. En otros casos, la evaluación mediante pruebas estandarizadas se adoptó en forma relativamente abrupta, como parte de los programas de reforma educativa más o menos integral, muy comunes en los años noventa, y dentro de marcos institucionales y organizacionales creados específicamente para ese fin. Existe en la actualidad una variedad considerable en los niveles de desarrollo alcanzados por los diferentes sistemas nacionales de evaluación, dependiendo de las capacidades técnicas, operativas y financieras de las unidades de medición, así como de los contextos políticos dentro de los cuales se enmarcan.

Esta sección amplía la discusión anterior sobre las características deseables de los sistemas nacionales de evaluación, según indica la experiencia y la literatura especializada internacional, y presenta ejemplos ilustrativos extraídos de casos nacionales y subnacionales de medición y evaluación de aprendizajes recogidos en el marco de este trabajo de investigación. El tratamiento de las características de los sistemas de medición se organiza en torno a seis temas principales:

- Fines y propósitos de los sistemas.
- Marco institucional de las agencias o unidades de medición.
- La medición de aprendizajes y su relación con el desarrollo curricular y de estándares en cada contexto nacional.
- Tipos de instrumentos de medición.
- Formas de reporte de los resultados.
- Difusión, uso e impacto de los resultados.

Finalmente, se ofrece un resumen breve sobre la participación de países latinoamericanos en pruebas internacionales comparadas de

aprendizaje, según las informaciones obtenidas en un trabajo de investigación anterior a cargo de este autor.

Fines y propósitos del sistema

Como se afirmó desde los inicios de las actividades del Grupo de Trabajo sobre Estándares y Evaluación del PREAL⁷, en muchos países de la Región fue insuficiente la reflexión acerca de los fines específicos que se esperaba o deseaba que persiguieran sus sistemas de evaluación, con la consecuencia de que algunas de las definiciones técnicas que se tomaron inicialmente no fueron las más adecuadas para los propósitos o usos diferentes que luego se quiso dar a los instrumentos y resultados de las mediciones. Con dificultad un solo diseño serviría para satisfacer todos los objetivos –a veces incluso incongruentes entre sí– perseguidos a lo largo del tiempo por los hacedores de política, pero los costos de diseñar instrumentos complementarios y las urgencias de contar con información sustentatoria para ciertas decisiones a veces inducen a usos equivocados de los resultados de las pruebas. Esto se ve agravado por la ausencia, en muchos casos, de una visión de largo plazo sobre el desarrollo del sistema y a la necesidad de improvisar medidas “de emergencia” que, a veces, acaba por debilitar la legitimidad y consolidación de los sistemas.

Afortunadamente, la acumulación de experiencia y el relativo fortalecimiento de capacidades evaluativas obtenidos por muchos países de la Región permitirían en la actualidad una revisión de esos fines y objetivos, con la ventaja de que el debate podría estar iluminado por experiencia vivida, y no solo por concepciones teóricas o modelos importados de otros contextos. Es posible, y lo demuestran en estos días países como República Dominicana, Chile y otros, realizar ajustes profundos sin necesidad de desmontar totalmente los sistemas y las capacidades institucionales ya instaladas.

Marco institucional de las agencias o unidades de medición

La experiencia de evaluación de los resultados de aprendizaje en América Latina data de al menos dos décadas, aunque la mayor parte de los sistemas oficiales de evaluación se crearon durante los años

⁷ Ver la introducción a Ravela *et al.* (2001).

noventa. Algunos países, como Chile y Colombia, vienen realizando desde hace muchos años aplicaciones continuas de pruebas de rendimiento académico y las unidades de medición a cargo gozan de altos grados de legitimidad y continuidad técnica a pesar de los “vaivenes” políticos que se producen periódicamente por los cambios de gobierno nacional. Otros sistemas, como los de Ecuador, Guatemala, Bolivia o Perú han sido más vulnerables a las voluntades políticas de turno, lo cual ha determinado que las informaciones del sistema no siempre sean difundidas de modo público, o que los equipos técnicos se recambien de manera repentina, con el consecuente impacto sobre la calidad de los instrumentos, de los operativos, o del diseño e implementación de estrategias apropiadas para la difusión de los datos.

Durante estas dos últimas décadas los países han ensayado diferentes marcos institucionales para la administración de los sistemas de evaluación educativa. El avance radica, precisamente, en ese ensayo, en la búsqueda de la forma que mejor acomode a los intereses y metas de las políticas de Estado en educación. En general, puede afirmarse que a pesar de los cambios y de algunos casos de fuerte inestabilidad política, ha habido una tendencia a sostener los sistemas de evaluación nacionales, y a reinstaurarlos o recomponerlos cuando sus actividades fueron discontinuadas en algún momento.

En varios países, los arreglos institucionales más estables resultaron ser los que se instalaron fuera de la estructura orgánica de los ministerios de educación. Estos sistemas, por lo general denominados “institutos”, poseen mayor autonomía administrativa y técnica que otros órganos de línea y suelen llevar a cabo sus tareas de medición y devolución de resultados con más agilidad y consistencia que los sistemas dependientes de los ministerios.

Dos ejemplos de este tipo de forma institucional son los del ICFES en Colombia y el INEP en Brasil. En Colombia, si bien existe una unidad de evaluación dentro del Ministerio de Educación Nacional, el protagonismo técnico y operativo en la realización de diferentes evaluaciones nacionales lo ha tenido el ICFES, Instituto Colombiano de Fomento a la Educación Superior. Este organismo, si bien no es totalmente independiente del Ministerio, ya que este último lo financia y preside, goza de amplia discreción y autonomía en la toma de decisiones técnicas con respecto a algunas de las pruebas nacionales, como por ejemplo el Examen de Estado para el ingreso a la educación superior. En Brasil, por su parte, las evaluaciones nacionales de aprendizajes son realizadas por el

INEP, Instituto Nacional de Estudios e Investigaciones Educativas, que también se administra independientemente del Ministerio de Educación. En ambos casos, los operativos de evaluación se vienen realizando con periodicidad desde hace varios años y los equipos técnicos han tenido la posibilidad de consolidarse y desarrollar poco a poco nuevas tecnologías de medición de los aprendizajes.

México, luego de haber contado con un sistema de medición dependiente de la Secretaría de Educación Pública durante muchos años, creó recientemente el Instituto Nacional de Evaluación de la Educación (INEE), un organismo semiautónomo que contará con mayor latitud que la antigua Dirección General de Evaluación para la toma de decisiones técnicas y administrativas sobre los operativos de evaluación. También Argentina intentó en el año 2001 la creación de un instituto semiautónomo de evaluación, el IDECE, aunque los inesperados cambios de gobierno nacional en ese mismo año determinaron su disolución y la reinstalación del sistema nacional, en la actualidad denominado DINIECE, dentro del organigrama regular del Ministerio de Educación.

Las agencias de evaluación autónomas, o institutos, si bien pueden adquirir mayores grados de autonomía funcional y legitimidad técnica, también pueden presentar problemas. Quizás el mayor riesgo es que se desvinculen de las necesidades de información de los ministerios, convirtiéndose así en programas de medición de alta calidad técnica pero con escaso impacto sobre la toma de decisiones de política para el mejoramiento de la calidad educativa. En algunos años más podrá hacerse una evaluación de situación sobre el funcionamiento e impacto de los institutos nacionales de reciente formación en la Región.

En otros casos se ha dado una evolución inversa, en que sistemas o proyectos de medición anteriormente administrados por instituciones independientes se encuentran hoy administrados desde los ministerios de educación. Tal es el caso de Chile, donde la organización e implementación iniciales del SIMCE estuvieron delegadas por el Ministerio de Educación en universidades públicas o privadas con la suficiente capacidad técnica para llevar a cabo esas tareas. Hoy, sin embargo, el sistema de evaluación se administra enteramente desde la Unidad de Currículum y Evaluación del Ministerio de Educación. En este caso particular, dada la capacidad técnica y la legitimidad ganada por el SIMCE a lo largo de las dos últimas décadas, el funcionamien-

to dentro del Ministerio de Educación no parece haberle restado al sistema autonomía o estabilidad en sus funciones⁸.

Pero no todos los casos de transferencia de responsabilidades a los ministerios de educación resultan exitosos o convenientes. En Ecuador, por ejemplo, la terminación del programa de evaluación APRENDO, que fuera financiado enteramente con fondos de cooperación internacional, significó el final de los operativos. Si bien se planificó la transición del sistema a la administración pública ministerial, la ausencia de recursos materiales y profesionales especializados en el Ministerio impidió la continuidad de las evaluaciones tal como venían realizándose antes, y tampoco se han diseñado nuevos programas de medición.

Este fenómeno es común a casi todos los países que iniciaron sus sistemas de evaluación con financiamiento y apoyo técnico de los organismos de cooperación internacional, principalmente el BID y el Banco Mundial. Sin embargo, existen excepciones, como Uruguay y Argentina, que hoy financian con fondos del presupuesto ordinario sus sistemas de medición dentro de la estructura orgánica ministerial y la continuidad de los operativos no se ha visto amenazada de manera importante⁹.

Otro factor que parece contribuir a la estabilidad y permanencia de los sistemas de medición es la existencia de leyes nacionales que ordenen su creación y sostenimiento. Valga como ejemplo el sistema de Pruebas Nacionales en República Dominicana, que si bien funciona en condiciones organizacionales relativamente precarias, los operativos anuales de evaluación siguen realizándose en forma regular, tal como indica la Ley General de Educación y un conjunto de ordenanzas nacionales específicas. Esto mismo podría ocurrir en Perú, por ejemplo, donde la nueva Ley General de Educación prescribe la crea-

⁸ La legitimidad ganada por el SIMCE, vale notar, podría atribuirse a que los resultados de las pruebas no han dado indicios de mejoras sustanciales en la calidad de los aprendizajes durante la última década. Desde el punto de vista político, esto hace del SIMCE una herramienta “crítica” de los resultados de las políticas educativas del gobierno nacional. Cabría preguntarse si esa legitimidad pudiera verse afectada si los resultados fueran mejorando y el SIMCE comenzara a ser percibido como un instrumento de “validación” de las políticas educativas en curso.

⁹ Está pendiente la revisión sobre la cantidad exacta de países que discontinuaron programas de evaluación debido a problemas de financiamiento externo o interno.

ción y sostenimiento permanente de un Instituto Peruano de Evaluación, Acreditación y Certificación Educativa.

Los sistemas subnacionales de medición de aprendizajes analizados en este trabajo, por su parte, se administran enteramente desde las secretarías estatales o distritales de educación, y este arreglo institucional parece funcionar adecuadamente en esos contextos. Solo el caso del SARESP, São Paulo, muestra un arreglo institucional diferente, donde el protagonismo técnico lo tiene una institución independiente del gobierno. Quizás por la escala de estos sistemas educativos, más pequeña y más fácilmente administrable, los sistemas de evaluación logran mayores grados de comunicación y trabajo orgánico con otras oficinas sectoriales, facilitando así el flujo de datos y magnificando las posibilidades de impacto en la toma de decisiones de política educativa – fenómeno que, en muchos países, no se da en los niveles centrales de la administración del sector–. Son sistemas, también, que logran mayores grados de participación de los actores locales, en especial docentes, padres de familia y organizaciones académicas vinculadas a la formación docente¹⁰.

La medición de aprendizajes y su relación con el desarrollo curricular y de estándares

Una de las preocupaciones mayores de quienes tienen a su cargo el diseño de instrumentos de medición es que cuentan con currículos nacionales que ofrecen definiciones poco claras, y apenas operacionales, de lo que se espera que los alumnos puedan realizar con los conocimientos conceptuales contenidos en esos currículos. Casi siempre la matriz curricular para el desarrollo de los instrumentos de evaluación se realiza mediante el diseño de una “tabla de especificaciones” que permite construir los ítemes de las pruebas cubriendo un número determinado de objetivos prioritarios, aunque estos no representen la totalidad o complejidad del currículo oficial. La implicancia negativa de esta situación es, naturalmente, que el marco de interpretación de los resultados es débil puesto que no refleja un acuerdo profesional y social sobre qué deben saber y saber hacer con sus conocimientos los estudiantes del sistema. A esto se suma, también, el problema de que a menudo se prioriza los contenidos que pueden

¹⁰ La descripción de varios de estos casos subnacionales ha sido incluida en la segunda parte de este documento.

ser medidos con facilidad, y no necesariamente los más importantes según los supuestos acuerdos entre actores del sistema.

Esta problemática se agrava por la débil comunicación que con frecuencia existe entre las unidades de evaluación y las dedicadas al diseño y desarrollo curricular a nivel nacional, por razones que incluyen, entre otras, el diferente origen profesional de sus integrantes. Esto, junto con la poca claridad de las metas de aprendizaje propuestas oficialmente, suele ser uno de los mayores obstáculos para que las actividades de evaluación sean coherentes con esas metas y puedan de hecho facilitar su consecución.

Sin embargo, en algunos países se está avanzando en dirección del establecimiento de algunas especificaciones, consensuadas entre los mismos evaluadores, sobre esas expectativas de aprendizaje, aun cuando los procesos de desarrollo curricular del país no estén alineados con ese trabajo. Visto en forma optimista, puede decirse que muchos de los países de la Región cuentan con equipos de profesionales de la evaluación dedicados al establecimiento o especificación de metas curriculares claras que les permitan diseñar pruebas más enfocadas, a la vez que avanzan el trabajo, y a veces el debate, sobre qué resultados de aprendizaje concretos se espera prioritariamente de los estudiantes. Ante la falta de movimientos resueltos a favor del desarrollo de estándares de contenido y desempeño académico, el trabajo de las unidades de medición se constituye en una instancia de avance importante en esa dirección. Algunos casos representativos de esta tendencia son Uruguay, Colombia¹¹ y Ecuador, así como los sistemas subnacionales de medición como de Aguascalientes y Bogotá.

En Colombia, donde el currículo oficial se plantea como lineamientos generales y orientadores, pero con especificaciones poco claras de desempeño, se ha adelantado este tipo de trabajo en al menos tres frentes. En primer lugar están las competencias en matemática y lenguaje definidas por los equipos de evaluación del Ministerio de Educación Nacional, y que supuestamente habrían de orientar las mediciones de un programa integral de seguimiento a realizarse anualmente entre 1996 y 2005. Este programa, sin embargo, fue discontinuado luego de una fuerte

¹¹ Tanto el ICFES como la unidad de evaluación del MEN, así como la Secretaría de Educación de Bogotá, han realizado esfuerzos considerables por definir competencias complejas y operacionales que permitan una mejor evaluación e interpretación de los resultados educativos.

reestructuración del Ministerio de Educación. En otro frente, el ICFES definió un conjunto importante de competencias complejas de aprendizaje en varias áreas curriculares comprendidas en el Examen de Estado. Estas competencias, que permiten evaluar aprendizajes en diferentes y claramente definidos niveles de logro, ofrecen un marco de interpretación de los resultados mucho más sólido del que resultaría de evaluaciones diseñadas sobre la base de una típica tabla de especificaciones.

Adicionalmente, existe un esfuerzo reciente del Ministerio de Educación Nacional por establecer un conjunto de estándares nacionales en tres áreas curriculares que fueron presentados públicamente a la comunidad educativa y a la población en general en el año 2003. Si bien es prematuro juzgar la validez e impacto potencial de esos estándares, vale informar que su sola presentación ya ha generado cierto nivel de debate entre los círculos profesionales disciplinares y la movilización de algunas opiniones en los medios de prensa.

En Ecuador se vivió una experiencia similar a la de Colombia, aunque en menor escala. Dada la incertidumbre sobre el currículo que finalmente habría de oficializarse a mediados de los años noventa, el equipo técnico de APRENDO decidió llevar a cabo la definición de un conjunto de destrezas académicas básicas que luego permitieron el diseño de instrumentos de medición referidos a criterios. Las pruebas plantean cuatro ítems por destreza, y se considera que al menos tres deben ser respondidos correctamente para considerar esa destreza alcanzada. A mediados de 2001, el Ministro de Educación de Ecuador planeaba utilizar esta definición y evaluación de destrezas para establecer un sistema público de comunicación sobre los alcances de los alumnos y del sistema, mediante la decisión sobre una “línea de corte” que actuara como horizonte de logro y que pudiera ir elevándose año a año en la medida en que los resultados nacionales fueran mejorando. Sin embargo, las evaluaciones han sido discontinuadas luego del último cambio de autoridades en la administración sectorial y no se han realizado declaraciones sobre el uso que pueda darse en el futuro a las matrices curriculares definidas por los equipos de medición de APRENDO.

En Uruguay, donde el currículo nacional largamente vigente presenta obstáculos técnicos importantes para la definición de matrices de referencia, el proceso de elaboración y aplicación de pruebas ha resultado una instancia interesante de debate curricular. Concretamente, los ítems de evaluación requieren la demostración de conocimientos y capacidades cognitivas que resultan pertinentes y deseables, pero que están ausentes

en forma explícita en el currículo nacional. La validación de los ítemes de evaluación, y la valoración que los docentes hacen de ellos, ha promovido grados importantes de reflexión sobre el currículo prescrito y sobre su práctica en el ámbito escolar.

Para las evaluaciones distritales de aprendizajes que realiza la Secretaría de Educación de la Ciudad de Bogotá, se estableció también un conjunto de competencias alineadas con los lineamientos curriculares nacionales. Las pruebas, al igual que las del Examen de Estado nacional, ofrecen un marco conceptual explícito y definiciones operacionales para cada una de las competencias evaluadas, lo cual permite reportar y ejemplificar diferentes niveles de desempeño estudiantil de manera más clara que lo que permitiría el uso de los lineamientos curriculares nacionales para ese fin. Desde los puntos de vista técnico y político, las competencias básicas propuestas pueden ser consideradas como *estándares* de currículo, en tanto proporcionan una herramienta de comunicación efectiva entre todos los actores sectoriales sobre las expectativas de aprendizaje para todos los estudiantes del sistema¹².

Interesa también el caso del estado de Aguascalientes en México donde, si bien la matriz de referencia para la evaluación no involucra una propuesta curricular distinta a la nacional, los objetivos de aprendizaje seleccionados para la evaluación constituyen un esfuerzo decidido por priorizar un conjunto de contenidos que se consideran básicos y que debieran ser aprehendidos por todos los estudiantes del sistema. El valor de esta selección curricular radica en que todas las acciones de mejoramiento propuestas desde las escuelas y desde la Secretaría de Educación sobre la base de los resultados de las evaluaciones se realizan, con prioridad, en función del alcance de ese conjunto de saberes. Procesos similares han ocurrido en los sistemas subnacionales de Paraná, Minas Gerais y São Paulo, aunque en esos casos los estados cuentan con marcos curriculares propios desarrollados a partir de los Parámetros Curriculares Nacionales.

En la mayor parte de los casos, sin embargo, convendría avanzar más enérgicamente en la propuesta de ejemplificaciones que muestren qué deberían saber y cómo deberían desempeñarse académicamente los alumnos al término de algunos niveles educativos, de manera tal que los estándares no solo prescriban los contenidos de aprendizaje sino que

¹² Para más detalles sobre los procesos de desarrollo curricular y evaluación en Bogotá, consultar Ferrer, 2004.

además arrojen luz, mediante ejemplos seleccionados de las respuestas de los estudiantes a las pruebas, sobre el tipo y niveles de desempeño que se espera de ellos. De esa manera, podría orientarse mejor no solo la actividad pedagógica y el desarrollo de mejores prácticas evaluativas en el ámbito escolar, sino también estimular la participación y demanda de los beneficiarios del sistema por mejores resultados de aprendizaje. En esa dirección están dando ya pasos iniciales Argentina, El Salvador, Perú y varios otros países de la Región.

Chile está avanzando quizás con más claridad e ímpetu en esta dirección. Con el asesoramiento técnico de Australia, el Ministerio de Educación está elaborando estándares de contenido y desempeño y mapas de progreso en aprendizajes para estudiantes desde primer grado hasta el cuarto año de enseñanza media en las áreas de Lenguaje, Matemáticas, Historia y Ciencias Sociales, Ciencias (Biología, Química y Física) e Inglés como segunda lengua. También ha encargado la elaboración de estándares para la educación infantil. Se espera, a partir de la elaboración de estos instrumentos, que las pruebas nacionales puedan reportar puntajes de rendimiento referidos a logros de aprendizaje esperados, establecidos claramente y con antelación a la aplicación de las mediciones y, por supuesto, a tiempo de influir en la enseñanza. Hasta el momento, el SIMCE no ha podido reportar resultados según criterios claros sobre cuál es el nivel de rendimiento aceptable o suficiente, ni sobre cuántos alumnos lo logran. Si la medición se vincula a logros esperados, el puntaje podrá adquirir mayor significado, toda vez que profesores, padres y alumnos podrán conocer cuánto sabe el estudiante y qué debe enseñar el docente para lograr las metas. La primera medición sobre la base de logros esperados se aplicará en el año 2006.

Tipos de instrumentos de medición: Características y validación

Cobertura general de las evaluaciones

La mayor parte de los sistemas de evaluación nacionales miden resultados en las áreas curriculares de matemáticas, lenguaje y comunicación, ciencias sociales y ciencias naturales. En algunos casos se evalúan logros en lenguas extranjeras, en particular en inglés y francés. Complementariamente, en países multilingües como Perú, Bolivia y Guatemala, se aplican pruebas traducidas en lenguas vernáculas. En educación media es común que el área de ciencias naturales sea evaluada diferencialmente

en biología, física y química, mientras que el área de ciencias sociales se evalúa con pruebas diferentes para historia, geografía y educación cívica¹³.

Casi todos los países que tienen sistemas nacionales de evaluación miden logros de aprendizaje en educación primaria y secundaria. En ambos casos, las pruebas tienden a ser aplicadas al término de ciclos educativos de dos o tres años, bajo la lógica de que algunos aprendizajes no pueden lograrse al cabo de un año académico, sino luego de períodos más largos que coinciden con etapas etarias y de desarrollo cognitivo, o de secuenciación curricular multianual. Adicionalmente, algunos países, como Chile, República Dominicana o Costa Rica, aplican pruebas “de salida” del nivel de educación media o secundaria (algunas de ellas denominadas “pruebas de Bachillerato”), como requisito de graduación o como criterio de selección para el ingreso a la educación superior. México, por su parte, aplica este tipo de pruebas también como instancia de graduación de la educación primaria y acceso a la educación media superior, aunque el uso que se da a los datos es por lo general diagnóstico y no selectivo.

Diseño de instrumentos

Una de las primeras instancias de validación técnica durante el diseño de los sistemas de evaluación estandarizada consiste en la toma de decisiones sobre el tipo de cobertura poblacional y curricular de los instrumentos de medición. Tales decisiones, que refieren en especial a las poblaciones y a los contenidos de aprendizaje evaluados, deben estar ligadas coherentemente al tipo de uso que se hará con posterioridad de los resultados¹⁴.

En cuanto a la cobertura poblacional, es necesario distinguir entre los países que aplican pruebas muestrales y los que realizan mediciones censales. Ambos tipos de pruebas sirven propósitos diferentes. Como se explicó anteriormente, las pruebas censales, aplicadas al universo de estudiantes en ciertos niveles de escolaridad, arrojan resultados por establecimiento, o por alumnos, que permite establecer sistemas de incenti-

¹³ La educación ciudadana, o educación para la ciudadanía y la democracia, también va cobrando mayor protagonismo dentro de los currículos nacionales y algunos sistemas han diseñado instrumentos específicos para su evaluación.

¹⁴ Un análisis exhaustivo de estas instancias de validación para el diseño puede hallarse en Ravela (Ed.), 2001.

vos o sanciones de acuerdo a los logros obtenidos (altas implicancias), o bien pueden facilitar la focalización de apoyos centrales a aquellos centros educativos que presentan mayores necesidades y dificultades de rendimiento académico. Las aplicaciones muestrales, por su parte, apuntan a recabar informaciones sobre diferentes grupos o poblaciones del sistema, para así poder detectar problemas generales de aprendizaje y, supuestamente, diseñar estrategias integrales de apoyo a todas las escuelas y estudiantes de un determinado grupo social, o una determinada región del país, en concreto, aquellas que más lo necesitan.

La cobertura y análisis de logro curricular que se pretenda realizar determinará, en especial, si las pruebas estarán *referidas a normas* o *referidas a criterios*. La selección de uno u otro tipo de pruebas se realiza, o debiera realizarse, de acuerdo a los fines y usos que se pretende dar a los resultados. Las pruebas referidas a normas son aquellas que buscan comparar los logros entre diferentes grupos de estudiantes, sin pretender un análisis muy profundo del tipo y profundidad de conocimientos adquiridos por los estudiantes. Las pruebas referidas a criterios, por su parte, suponen el establecimiento de acuerdos sobre cuáles competencias o capacidades académicas básicas debieran lograr TODOS los estudiantes, luego definen una “línea de corte” o marca de logro mínimo o suficiente, y miden si los alumnos del sistema educativo alcanzan o no esa marca, y en qué medida lo hacen, como resultado de su experiencia escolar.

No se supone deba existir una correspondencia automática entre tipos de cobertura poblacional y curricular. Las pruebas referidas a criterios pueden ser aplicadas muestral o censalmente, y lo mismo ocurre con las pruebas referidas a normas. Sin embargo, y tal como se explica con minuciosidad en trabajos anteriores de este Grupo de Trabajo (Ravela, ed., *op. cit.*) una mayor cobertura poblacional puede restringir la capacidad de los instrumentos para recabar datos de aprendizaje con mayor profundidad, y viceversa. La decisión sobre qué tipo de cobertura priorizar depende, nuevamente, del tipo de usos que quiera darse a la información.

Los sistemas nacionales que aplican pruebas referidas a criterios ofrecen, en general, fundamentaciones conceptuales y descripciones operacionales más sólidas sobre los aprendizajes y niveles de desempeño esperados. Las pruebas referidas a normas, por el contrario, se diseñan con frecuencia mediante un modelo de medición que selecciona los ítems de evaluación de acuerdo al *índice de discriminación* que mostraron en los pilotajes de campo, pero que no refleja necesariamente la relevan-

cia de los contenidos evaluados ni asegura que estos representen el universo curricular de referencia.

En el caso de Uruguay, las pruebas aplicadas son referidas a criterios. La selección de los ítems no se realiza de acuerdo al índice de dificultad y de discriminación, sino que se hace por el “valor” de los ítems, es decir, buscando aquellos que reflejan, según criterios profesionales, los tipos de capacidades que el sistema educativo considera deben ser desarrolladas antes del término de ciertos niveles de escolaridad. Una vez diseñados los instrumentos se establece, previa consulta a expertos curriculares y pedagogos de campo, una “línea de corte”, que para el caso de Uruguay es el 60% de respuestas correctas del total de los ítems. Al reportarse los resultados, se describe el porcentaje de alumnos que alcanzó esa línea de corte, así como el porcentaje de alumnos que lograron niveles inferiores (insuficientes) de desempeño.

Los Exámenes de Estado que diseña el ICFES en Colombia, originalmente referidos a normas, se elaboran al presente según un modelo criterial. La iniciativa para esta transición se debió a un renovado equipo de técnicos que aportaron nuevas visiones, y a las demandas de la población estudiantil y de la comunidad educativa en general por un tipo de pruebas que se centrara menos en capacidades simples de memorización de datos y más en el desarrollo de habilidades cognitivas complejas, a la vez que reflejaran conocimientos conceptuales disciplinares más actualizados. Esa actualización suponía un mejor alineamiento con los contenidos de los lineamientos curriculares oficiales, con estándares internacionales de “alfabetización científica”¹⁵, y con concepciones formativas que ponen de relieve la necesidad de educar ciudadanos democráticos y participativos. Los exámenes evalúan competencias de tipo interpretativo, argumentativo y propositivo, las cuales son consideradas “modos fundamentales de participación y construcción de lo social” (MEN-ICFES, 2000).

También en Colombia, la Secretaría de Educación de Bogotá aplica pruebas criteriosales y “casi censales” (80 por ciento de cobertura en los grados evaluados). Los criterios consisten en una serie de competencias definidas localmente sobre la base de los lineamientos curriculares nacionales. Los resultados se reportan en forma de porcentaje promedio de respuestas correctas por tipos de competencia, más específicamente por niveles de desempeño para cada una de ellas. Las respuestas abier-

¹⁵ Del inglés, *scientific literacy*.

tas, por ejemplo de producción de texto, se califican y reportan según niveles de competencia preestablecidos que guían la corrección y la interpretación que hacen los docentes de los resultados al recibir los informes escolares.

Varios de los países que comenzaron aplicando pruebas referidas a normas han optado por realizar un reporte de resultados según criterios definidos a posteriori. Tales son los casos de Chile, Argentina, Costa Rica y Brasil. Esto no significa que las pruebas estén estrictamente referidas a criterios, pues el modelo de medición es normativo, pero se establecen criterios y líneas de corte para el reporte con la intención de informar mejor a los usuarios del sistema. Esta estrategia, sin embargo, alberga ciertos riesgos para la validez de las interpretaciones puesto que los ítemes seleccionados según su índice de discriminación pueden no representar fielmente el logro de las competencias o saberes evaluados, a menos que el espectro de ítemes utilizados sea lo suficientemente amplio.

La cobertura poblacional de las pruebas varía de país en país. Cerca de nueve países aplican o han aplicado pruebas en forma solo muestral, otros dos en forma exclusivamente censal, y ocho han experimentado con ambos tipos de pruebas. Todos los sistemas subnacionales analizados en este trabajo aplican pruebas censales. En general, se observa un importante grado de coherencia entre la cobertura poblacional realizada y el tipo de uso que se da a los datos. Es decir, se evita hacer inferencias de resultados sobre instrumentos que no lo permiten, aunque también es cierto que, en general, se evidencia una alarmante subutilización de los datos para el diseño de estrategias de mejoramiento de la calidad de los aprendizajes y, en oportunidades, interpretaciones insuficientemente sustentadas o generalizaciones no del todo válidas.

En cuanto al formato de las pruebas de aprendizaje, todos los países utilizan en especial preguntas cerradas de opción múltiple, y varios de ellos complementan los instrumentos con preguntas abiertas para la producción de textos, en particular en el área de lenguaje y comunicación. La corrección de respuestas abiertas ha presentado problemas técnicos para algunos países al momento de definir o utilizar criterios homogéneos de evaluación entre correctores. En otros países el proceso resultó más exitoso y las correcciones arrojan datos pedagógicamente muy útiles para los docentes. Es digno de destacar que en muchos países se está recurriendo más y más al uso de respuestas abiertas, pese a su mayor costo y dificultad de revisión y codificación, por ser estas de mayor utilidad para la identificación de los procesos correctos o equívocos que

siguen los alumnos en sus esfuerzos por responder a las preguntas y problemas que les plantean las pruebas y, por consiguiente, para desarrollar propuestas para el mejoramiento de la enseñanza y el aprendizaje de ciertos contenidos y el desarrollo de competencias.

Validación técnica

Por último, resta mencionar la validación técnica que se supone debe hacerse de los ítemes o reactivos de las pruebas. Este es un elemento crítico y técnicamente delicado en el proceso de construcción de instrumentos, y sobre el cual no existen descripciones suficientemente detalladas en los informes técnicos nacionales. La teoría de medición internacional ofrece un conjunto importante de lineamientos y estándares de calidad para la validación de los reactivos de pruebas, pero no se dispone de información suficiente para juzgar en qué medida esos estándares se respetan en la construcción de pruebas nacionales en América Latina. Los informes técnicos mencionan, en muchos casos, que la validación se realizó con la participación de docentes y expertos disciplinares, pero no se indica, por ejemplo, qué criterios específicos se utilizaron para juzgar la validez de los ítemes (Ej. sesgos culturales, “ruido” o superposición de contenidos entre disciplinas, pertinencia curricular, comparabilidad interanual, etc.) La calidad de los instrumentos es un tema que amerita estudios en profundidad y que será materia de análisis en trabajos futuros de este Grupo de Trabajo.

Adecuación lingüística

Tres países de la Región aplican pruebas en lenguas maternas originarias. Perú lo hace en quechua y aimara, Bolivia en quechua, aimara y guaraní, mientras que Guatemala ha evaluado estudiantes en las cuatro lenguas mayas predominantes, que representan al 84% de la población guatemalteca. En este último caso, sin embargo, no fue posible materializar el operativo en su totalidad pues existe un alto grado de deserción escolar en escuelas de población mayahablante.

Factores asociados al rendimiento académico

Otro aspecto destacable de casi todos los programas de medición en la Región es que las pruebas de logro académico suelen ir acompañadas

de la aplicación de cuestionarios a diversos actores de la comunidad educativa, normalmente alumnos, docentes, directores y padres de familia, para recoger información sobre los factores escolares y extraescolares que pueden “asociarse” estadísticamente al rendimiento o desempeño de los estudiantes. Los factores escolares suelen ser indicadores sobre las condiciones de infraestructura, equipamiento didáctico, recursos profesionales del establecimiento escolar, actitudes de los alumnos hacia las disciplinas evaluadas, tipos de prácticas pedagógicas y grados de implementación del currículo oficial en aula¹⁶. Los factores extraescolares estudiados, por su parte, dicen relación en particular con el nivel económico y sociocultural de los alumnos, según la información que las familias aportan sobre sus ingresos, grados de escolaridad de los padres y hermanos del estudiante, servicios básicos de energía y agua potable en el hogar, y ciertos recursos culturales y de información en el hogar, tales como TV, radio, libros, periódicos y revistas.

El valor de la información contextual extraescolar radica en que los sistemas de medición pueden reportar los resultados de las pruebas de acuerdo al nivel socioeconómico y cultural de los estudiantes, y de esa manera hacer afirmaciones más justas sobre la calidad de las escuelas, de acuerdo al grado de vulnerabilidad social de los estudiantes con que trabajan. Dentro de este modelo de análisis, destacan los esfuerzos experimentales por medir “valor agregado” que han realizado algunos sistemas subnacionales, como la provincia de Buenos Aires en Argentina. Tal modelo de análisis permite, mediante pruebas de rendimiento aplicadas en dos momentos en el tiempo, medir el progreso en aprendizajes que una cohorte de estudiantes del mismo contexto socioeconómico logra respecto de sí misma al cabo de un lapso determinado. Lamentablemente, este operativo en Argentina ha sido discontinuado y no se han realizado publicaciones de los resultados obtenidos en las mediciones.

¹⁶ A pesar de los esfuerzos técnicos en este sentido, los datos recogidos mediante estos instrumentos no siempre resultan útiles para el análisis estadístico de factores asociados, en parte debido a que la información se recolecta a modo de encuestas pero no como parte de un modelo estadístico lo suficientemente sólido como para permitir análisis rigurosos posteriores, y también porque la “medición” de prácticas pedagógicas y otras variables relevantes es de difícil realización a gran escala.

Formas de reporte de los resultados

Se distinguen tres grupos principales de audiencias o destinatarios de los datos e informaciones producidos por los sistemas nacionales y subnacionales de medición de aprendizajes. El primer grupo está conformado por las autoridades políticas de gobierno, en los poderes Ejecutivo y Legislativo, y por la alta dirección del sector educativo de cada país. El segundo grupo de destinatarios es el público general, y organizaciones académicas y no gubernamentales. El tercer grupo está integrado por los actores educativos locales, principalmente docentes, directores y supervisores de escuelas, alumnos, padres y madres de familia.

A lo largo de la última década los sistemas nacionales de medición fueron desarrollando formatos diferentes para el reporte de los datos obtenidos con la intención de hacerlos más útiles según los tipos de destinatarios que se busca informar. En general, se observa una evolución hacia formas de reporte que apuntan crecientemente a la devolución de resultados en los niveles escolares, y con particular énfasis en la presentación de datos curricular y pedagógicamente útiles. Esto contrasta con la tendencia inicial de algunos países, sobre todo los que iniciaron sus sistemas de evaluación a principios de los años noventa, de producir informes generales que presentaban datos muy “gruesos”, en términos de porcentajes globales de rendimiento por asignatura y comparativos entre las diferentes subpoblaciones estudiantiles evaluadas, cuya utilidad estaba en especial limitada a la selección de grupos o sectores en los cuales podría focalizarse algún tipo de intervención o, lo que ha sido más común, para movilizar a la opinión pública sobre el deficiente nivel de logros educacionales.

Reportes a autoridades de gobierno y alta dirección sectorial

Todos los países de la Región producen informes generales, a modo de resumen ejecutivo, sobre el rendimiento académico de los estudiantes evaluados. Estos informes tienen como finalidad principal la de informar a las autoridades políticas y técnicas superiores del sector, y así facilitar la toma de decisiones de política educativa sobre la base de indicadores relacionados con los *procesos* y los *resultados* de la educación, que actúan como complemento de los indicadores tradicionales sobre *insumos* educativos tales como salarios, infraestructura o equipa-

miento, u otros *productos* del sistema, como por ejemplo tasas de ingreso, repitencia, deserción y graduación.

Los informes a autoridades superiores ofrecen en general una descripción comparativa de resultados entre las diferentes poblaciones evaluadas. Para tal fin, se presentan los datos de rendimiento académico en forma agregada, por asignatura y grado de escolaridad al nivel nacional, y luego en forma desagregada por regiones geográficas o administrativas y por tipos de escuelas (públicas-privadas, urbanas-rurales, unidocentes-polidocentes, bilingües-monolingües, etc.). En algunos casos, los informes desagregan los datos según el nivel económico o sociocultural de la población evaluada. La presentación de los datos suele hacerse en forma gráfica (barras, tablas) y forma narrativa muy resumida. En casi todos los casos se ofrece algún tipo de conclusiones generales, de tipo comparativo entre subpoblaciones pero evitando cifras específicas. En conclusión, aunque no en todos los casos, se presenta un conjunto de recomendaciones sobre líneas de acción posibles para el mejoramiento de los procesos educativos o de la equidad en el acceso al conocimiento.

Tales conclusiones y recomendaciones suelen basarse en el análisis de variables escolares y extraescolares asociadas estadísticamente al rendimiento académico. Sin embargo, existen limitaciones respecto a la validez de las interpretaciones que puedan hacerse sobre la base de estos análisis. Como bien nota Ravela (2001) en su trabajo sobre informes nacionales de resultados, las interpretaciones sobre los factores asociados al rendimiento pierden validez si se realizan solo en función de análisis bivariados, puesto que estos no toman en consideración otros factores que afectan simultáneamente el rendimiento y moderan el efecto de las demás variables.

Respecto a las informaciones de difusión pública (sociedad civil), el escenario es menos consistente pues parece no haberse invertido tiempo y recursos suficientes para el desarrollo de formatos específicos para este fin. Los sistemas de medición de altas implicancias, por su responsabilidad de informar de modo sistemático a los usuarios directos de los datos, hacen públicos los resultados por escuela o por alumnos. Tales son los casos del SIMCE en Chile o del ICFES en Colombia. En los países donde las pruebas no tienen consecuencias directas para los actores locales, los esfuerzos por presentar los resultados a la sociedad civil han sido esporádicos y el acceso a los datos depende mayormente de la voluntad y capacidad de los medios de prensa por hacer una cobertura más o menos periódica de los resultados.

Reportes a actores educativos locales

Aunque resta mucho por mejorar en el diseño y difusión de resultados para los actores locales, no son pocos los esfuerzos realizados en esa dirección. Esta afirmación se basa en la variedad de formatos que han producido varios países, muchos de ellos ofreciendo datos muy relevantes, así como tratamientos sustantivos de las respuestas de los estudiantes y recomendaciones pedagógicas y curriculares para docentes y alumnos. A continuación se presentan algunos casos nacionales y subnacionales a modo de ejemplo.

- *Pruebas de salida de bachillerato e ingreso a educación superior (altas implicancias)*

En los últimos años los países que aplican pruebas de egreso de educación secundaria o ingreso a educación superior han desarrollado nuevos modelos conceptuales de evaluación y formatos de reporte que ofrecen informaciones de interés y utilidad para docentes y alumnos, más allá de la mera transcripción de puntajes y posiciones obtenidos para la acreditación o selección de estudiantes. Los modelos conceptuales mencionados implican en algunos casos un mayor alineamiento con los contenidos curriculares de la educación secundaria, en contraste con la medición clásica de habilidades cognitivas, y el establecimiento de criterios de logro que permiten evaluar a los alumnos en relación a metas de aprendizaje acordadas en vez de obtener simples medidas de tendencia central (pruebas normativas).

En República Dominicana, los informes institucionales devueltos a las escuelas presentan los resultados desagregados al nivel de curso o sección y por área de conocimiento. Se reportan los porcentajes de estudiantes promovidos y aplazados, el grado de correlación entre la calificación escolar final (interna) y la calificación en prueba estandarizada, una comparación de resultados con centros educativos semejantes, y con todos los centros educativos en los ámbitos nacional, regional y distrital. Los alumnos, por su parte, reciben un informe individual con el puntaje obtenido por dominios o competencias, y por nivel de desempeño, en cada área de conocimiento; se les ofrece también un análisis narrativo de sus logros y dificultades principales, y una tabla comparativa de sus resultados con el promedio nacional y con el promedio de las escuelas del mismo nivel socioeconómico o zona geográfica. Finalmente, reciben

un conjunto de recomendaciones específicas para mejorar su rendimiento académico, lo cual resulta en particular útil para aquellos alumnos que no aprobaron el examen y deben volver a presentarse.

Los informes de la prueba PAES en El Salvador también devuelven los resultados a los estudiantes uno a uno. En esos informes, cada estudiante obtiene información sobre las competencias evaluadas en cada asignatura, el nivel de logro alcanzado en dichas competencias, el promedio global de rendimiento en cada una de las áreas y una explicación conceptual de los niveles de logro alcanzados.

Los resultados de los Exámenes de Estado colombianos se devuelven a todos los centros educativos y a los estudiantes individualmente. Todas las escuelas medias reciben un reporte institucional donde se detalla exhaustivamente cuál fue el nivel de logro de cada uno de sus estudiantes en las diferentes áreas del currículo, y cuáles objetivos específicos de cada área muestran mayores avances y dificultades. Los estudiantes, por su parte, reciben un reporte individual que les permite conocer detalladamente cuáles han sido sus logros y dificultades más significativos¹⁷.

Los casos anteriores corresponden a pruebas de altas implicancias para los alumnos, pero no para las escuelas, al menos con formalidad. Según se desprende de las informaciones recabadas para este estudio, solo en Chile se informa públicamente y de manera individual, por escuela, los resultados de la prueba de ingreso a la educación superior¹⁸. En ese caso, y dado que algunos de los padres y alumnos tienen la posibilidad de escoger el colegio público, privado o subvencionado de su preferencia, los resultados institucionales pueden acarrear consecuencias directas para las escuelas secundarias en términos de captación de matrícula.

- *Devolución con fines de mejoramiento curricular y pedagógico (bajas implicancias)*

La mayor parte de los sistemas de medición de aprendizajes que realizan devoluciones de resultados sin consecuencias directas para las

¹⁷ Este tipo de reportes puede considerarse como una instancia importante de acercamiento al tipo de instrumentos y formas de reporte que propone el movimiento de estándares.

¹⁸ Hubo intentos de hacer públicos los resultados comparados en algunos países, pero no se consolidaron como estrategias permanentes de difusión de resultados, en gran parte debido a la reacción de la comunidad educativa en contra de la publicación de *rankings* institucionales.

instituciones o individuos, lo hace con la intención de ofrecer insumos de información que permitan a los actores locales reflexionar sobre el desempeño de los estudiantes y elaborar estrategias de mejoramiento ante las dificultades observadas. Algunos países y sistemas subnacionales han desarrollado formatos elaborados con diligencia para estos propósitos, entre los cuales merecen mención los de Uruguay, Bolivia, Aguascalientes, Minas Gerais, Bogotá, Paraná y São Paulo.

En Uruguay, la información desagregada por escuelas es entregada confidencialmente a cada establecimiento, presentando los resultados de sus estudiantes y los de estudiantes en otras escuelas de contexto socioeconómico semejante. La presentación de datos sobre niveles de suficiencia en el desempeño académico se basa en una línea de corte (nivel *satisfactorio*) equivalente a la respuesta correcta al 60% del total de los ítems, y en el porcentaje de alumnos que alcanza ese nivel de rendimiento. Los informes sobre 6° grado de primaria presentan también comparaciones interanuales, cubriendo períodos de tres años, sobre el progreso de la población estudiantil en el logro de las competencias evaluadas. Tales datos se expresan en forma de puntos porcentuales de diferencia entre los resultados de las aplicaciones comparadas (año 1 y año 3) para el total de la población evaluada, y en forma desagregada por tipo de contexto socioeconómico. Quizás el valor más importante de estas devoluciones radica en la presentación de una selección de ítems de la prueba para que puedan ser reaplicados en forma autónoma por los docentes durante el año escolar, siguiendo las pautas de corrección explicadas en los informes. De esa manera, las escuelas no participantes de la muestra nacional pueden obtener una medida más objetiva del nivel de desempeño de sus alumnos, a la vez que acceden a nuevas formas de evaluación, a la posibilidad de realizar reflexiones más sistemáticas sobre el currículo y sobre los procesos de enseñanza-aprendizaje y, asimismo, como se verá más adelante, a oportunidades de desarrollar iniciativas y recibir apoyo para mejorar sus planes y programas institucionales.

El SIMECAL de Bolivia, tanto para el caso de las pruebas censales como las muestrales, devuelve los resultados a las escuelas por medio de un informe institucional sobre el rendimiento de sus alumnos que incluye el puntaje promedio institucional, una descripción operacional de los niveles de logro definidos para cada área y el porcentaje de alumnos de la escuela ubicados en cada nivel. Luego presenta una descripción de las fortalezas y debilidades del conjunto de alumnos por cada eje temático en cada área curricular. También distribuye un folleto con análisis de

ítemes seleccionados y propuestas metodológicas para mejorar los resultados en relación a esos ítemes. Para el caso de los bachilleres egresantes, se ofrece un informe individual por alumno con su desempeño en cada eje temático evaluado por área curricular. Para el conjunto de escuelas y de equipos técnicos locales, se distribuye un folleto con indicadores de eficacia escolar según los datos recogidos con las pruebas y cuestionarios de contexto. Los factores asociados al rendimiento que se publican refieren a las características del alumnado y sus familias (educación previa, NES familiar y actitudes, vivienda), procesos pedagógicos y recursos materiales escolares. Si bien se cuenta con información sobre las características socioeconómicas y culturales de las familias, no se han definido estratos específicos que permitan desagregar los resultados según esta variable.

En el estado de Aguascalientes, en México, los informes de las pruebas estadales se reportan por niveles de desempeño logrados en cada área. Los niveles definidos se corresponden con una partición en cuartos de una escala de 1 a 100 y se clasifican en *crítico e insuficiente* (no dominio), y en *aceptable y deseable* (dominio). Los resultados hacia el fin de cada año académico se presentan en términos de “ganancia educativa”, en una escala de -4 a 4, dependiendo de los avances o retrocesos que muestra la comparación entre pruebas inicial y final. Por ejemplo, si un alumno comenzó el año en un nivel *insuficiente* y culminó el año en un nivel *deseable*, la ganancia, según la fórmula utilizada, se calcula en 3 puntos sobre 4. Si el resultado fue inverso, es decir, que comenzó en un nivel *deseable* y decreció a un nivel *insuficiente*, la “ganancia” se expresa en forma negativa, esto es, como -3 puntos en la escala. Estos datos se presentan agregados al nivel estadual, y desagregados a los niveles de zona educativa, escuela, grupo (curso) y alumno en forma individual.

En los informes que distribuye la Secretaría de Educación de Bogotá a cada escuela, los resultados para cada asignatura se reportan desagregados por competencias, por niveles de desempeño alcanzados para cada una de ellas, y se expresa el valor esperado (expectativa oficial) y el valor promedio logrado por los alumnos del plantel en una escala de 0 a 306. Se ofrece además un análisis de patrones de error de una selección de ítemes (para el conjunto de escuelas), y una serie de materiales con recomendaciones pedagógicas sustentadas en los datos presentados.

En Minas Gerais, también cada escuela recibe un informe individual. Los informes presentan gráficos separados para cada grado y área curricular evaluada, y cada uno de ellos muestra el porcentaje de alumnos que

se ubica en cada uno de los tres niveles de desempeño (proficiencia) predeterminados (crítico, intermedio y suficiente). Los porcentajes se desagregan al nivel de las diferentes capacidades que conforman cada una de las competencias evaluadas en cada área curricular. Además, se ofrece un gráfico de barras comparando el desempeño promedio de la escuela respecto del municipio correspondiente, la superintendencia regional y el estado. Los informes se complementan con una revista explicativa para orientar a los profesores en la lectura e interpretación de los datos.

En los estados de São Paulo y Paraná, además de presentar datos y análisis de resultados comparables a los de los casos anteriores, se promueve la participación de las escuelas en la elaboración de los informes. Para esto, cada centro educativo se responsabiliza por el procesamiento primario de los datos (conteo de respuestas, estadísticas descriptivas básicas, etc.) según el formato y las pautas indicadas desde la unidad de evaluación de la Secretaría. Posteriormente, esos datos son elevados a la administración central donde son procesados de nuevo y verificados. Por último, son devueltos a las escuelas con un nuevo formato y con guías metodológicas para la interpretación de los resultados. En el caso de Paraná en particular, cada escuela tiene a su cargo la tarea de elaborar un “boletín escolar” que incluye, además de los datos de las evaluaciones de aprendizajes, un conjunto de datos institucionales complementarios, tales como tasas de aprobación, reprobación y deserción, y el perfil profesional de los docentes. Todos esos datos también son reprocesados por la unidad de medición y devueltos a las escuelas en un nuevo formato que permite la comparación de los resultados y del perfil institucional con los de las demás escuelas del municipio y del estado en su totalidad.

Informes específicos para padres de familia

Este grupo de destinatarios de la información sobre aprendizajes es, en casi todos los países de la Región, el menos atendido por los sistemas de medición y evaluación. Quizás la razón principal que justifica esta debilidad es la creencia por parte de las autoridades políticas sectoriales que este tipo de informaciones puede conducir a las comparaciones y a la competencia entre escuelas por captar matrícula. Otras de las razones que a menudo se esgrimen es que los padres de familia no están lo suficientemente preparados para comprender las informaciones que se podrían ofrecer, y que por lo tanto no se justifica el esfuerzo de difusión

de los resultados entre estos actores del sistema. Ambas razones son muy debatibles, por supuesto, pero hasta el momento ese debate no parece tener lugar en la agenda sobre evaluación y usos de la información sobre aprendizajes.

En la mayor parte de los casos nacionales que tienen pruebas censales se solicita a las escuelas que compartan la información con los padres de familia, pero se lo hace en forma de recomendación, sin ningún tipo de prescripción normativa. El único país que facilita sistemáticamente a las familias el acceso a los resultados de su escuela es Chile, y lo hace en el marco de un esquema de libre selección de centros. Fuera de este tipo de esquema, destacan los casos subnacionales de Paraná y Minas Gerais. En el primero, grupos de representantes de padres de familia son invitados regularmente a los encuentros estatales de capacitación docente, y allí son informados sobre los resultados de las pruebas anuales y sus implicancias en términos de las acciones de política educativa propuestas. En el caso de Minas Gerais, no todos los padres de familia reciben informes de los resultados, pero sí los representantes de este grupo en los gobiernos escolares de cada institución. Estos representantes tienen acceso a los informes de resultados, y de su voluntad depende el que sean o no socializados entre los demás padres.

Seguimientos cualitativos

Además de los informes regulares mencionados anteriormente, algunos países realizan seguimientos cualitativos de los procesos institucionales y pedagógicos que pueden explicar los resultados obtenidos por los estudiantes. La Unidad de Curriculum y Evaluación del Ministerio de Educación de Chile, donde funciona el SIMCE, cuenta con un equipo especial de seguimiento que aplica encuestas y realiza observaciones escolares para llevar un mejor registro sobre la evolución de la implementación y alcance de los nuevos marcos curriculares, así como sobre el uso de textos escolares de distribución masiva y estilos pedagógicos de los docentes. Los resultados de esas indagaciones se presentan con preferencia en forma de porcentajes de respuestas a las encuestas realizadas. Dado que la muestra de escuelas participantes en esos seguimientos no es estadísticamente representativa, los datos no se presentan como asociaciones o correlaciones entre variables de proceso y resultado, sino como tendencias que deben ser estudiadas en mayor profundidad en próximos operativos nacionales. Hasta donde sabemos, estas informacio-

nes son de uso interno para el diseño de políticas curriculares desde el Ministerio, pero no se han divulgado entre miembros o instituciones de la comunidad académica.

En México, la Dirección General de Evaluación de la SEP también realizó en años anteriores seguimientos cualitativos sobre las escuelas participantes en la muestra nacional. Los resultados de esos seguimientos fueron devueltos a las escuelas mediante informes que describen las características organizacionales y pedagógicas de las escuelas que muestran mejoras de sus resultados en el tiempo, así como de aquellas cuyo rendimiento ha sido decreciente. También se ofrecen orientaciones metodológicas para replicar esas investigaciones en cada centro, a modo de autoevaluación, y se presenta un conjunto de recomendaciones pedagógicas y organizacionales para elevar la calidad de los aprendizajes.

Difusión y uso de los resultados de las mediciones

Este aspecto constituye uno de los problemas centrales de la evaluación de aprendizajes en la Región. Dado que los sistemas nacionales de medición se han creado con la explícita intención de hacer más transparentes los resultados de la gestión y de la práctica educativa en cada país, y proveer a los profesionales del sistema herramientas útiles para el mejoramiento de sus prácticas, las administraciones nacionales de educación han asumido para sí una responsabilidad de gran envergadura que no siempre se cumple en tiempo y forma adecuados.

Como se explicó antes, el uso de los resultados de las pruebas es variado y puede clasificarse, a grandes rasgos, en usos de altas o bajas implicancias, dependiendo del tipo de consecuencias o grado de responsabilización directa que implique para los actores locales o centrales del sistema. El tipo de difusión que se planea dar a los datos, así como las estrategias para lograrlo, depende básicamente del tipo de impacto que se espera esos datos tengan. Idealmente, se espera que, cualesquiera sean las audiencias y las estrategias para llegar a ellas, los resultados de las evaluaciones dejen de ser simplemente *datos* para transformarse en *información* sustantiva sobre los logros de aprendizaje de los alumnos. Esto significa que quienes tienen a su cargo la difusión de los datos deben asumir la responsabilidad de interpretarlos, reorganizarlos y presentarlos de una manera clara y diferenciada de acuerdo al uso que se supone diferentes audiencias darán a la información. En otras palabras, los datos se transforman en información cuando *comunican* a los actores

del sistema un mensaje intencional sobre los aprendizajes logrados, idealmente de acuerdo a un marco conceptual que permita interpretar los resultados en relación a un conjunto de criterios claros sobre los aprendizajes y desempeños esperados de los estudiantes.

Otro factor que afecta la posibilidad de impacto de los datos es el tiempo de devolución de los resultados, en términos de lapsos y periodicidad de entrega. En otras palabras, el uso que los actores educativos puedan dar a los datos se verá más estimulado si los datos de las evaluaciones son devueltos poco tiempo después de las aplicaciones, y si las entregas se realizan con periodicidad. Estos factores pueden contribuir al afianzamiento del sistema de evaluación, en tanto se construya una demanda creciente de los datos en forma oportuna, es decir durante los momentos clave para el diseño de políticas de mejoramiento educativo, o de planificación de la cobertura y entrega curricular en el ámbito escolar. Varios de los sistemas de evaluación de la Región han logrado realizar las aplicaciones y devolver los resultados en ciclos constantes de ejecución, generalmente anuales o bianuales. En otros casos, las aplicaciones y los esfuerzos de devolución han sido esporádicos, o bien constantes durante un tiempo y luego discontinuados.

De todas maneras, no es posible afirmar que una mayor periodicidad y constancia impliquen de manera inevitable mayor impacto, pues este depende de otros factores también importantes, tales como la participación de los actores locales en el proceso evaluativo, la calidad de los informes, o la voluntad política por tomar decisiones informadas y no meramente intuitivas. Asimismo, es importante distinguir entre periodicidad y frecuencia de evaluación y entrega de resultados. Por ejemplo, la evaluación y entrega **anual** de resultados no garantizan por sí mismas un mayor impacto en el sistema educativo, pues este ritmo de trabajo suele dejar menos margen de tiempo a los equipos técnicos de medición para realizar análisis sustantivos de los datos y elaborar mejores informes. Es así que varios sistemas de evaluación regionales han decidido disminuir la frecuencia de aplicaciones, pero manteniendo la periodicidad, para poder ofrecer mejores insumos de información a sus audiencias. Un ejemplo de ello sería el caso de Perú, que modificó su plan original de realizar operativos nacionales en varios grados y disciplinas cada dos años, ampliando el ciclo a tres. También Chile ha modificado la periodicidad de sus pruebas, intercalando grados y materias de manera de permitir un mayor tiempo y recursos para el análisis de datos y el mejoramiento del sistema en general.

Uso para toma de decisiones de política o para el diseño de programas de mejoramiento¹⁹

En el nivel nacional, son pocos los países que han realizado un uso específico de los datos de las evaluaciones de aprendizaje para la toma de decisiones de política educativa. Entre ellos puede mencionarse a Chile, México y Uruguay.

En Chile, el uso de las evaluaciones SIMCE para la focalización de políticas de mejoramiento se materializa en al menos tres frentes de acción: 1) El uso de los resultados como uno de los dos indicadores más importantes en la selección de las escuelas que ingresan al Programa de las 900 escuelas (P-900, actualmente atendiendo alrededor de 1.500 instituciones); 2) La utilización de los datos para la asignación de fondos concursables para los Proyectos de Mejoramiento Educativo en las escuelas; y 3) El uso de los datos como uno de los indicadores principales para otorgar incentivos docentes mediante el SNED, Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos.

En México, recientemente se ha impulsado un programa de mejoramiento escolar en centros educativos urbano-marginales denominado PEC, Programas de Escuelas de Calidad²⁰, y varios de los estados han realizado la selección de las escuelas participantes en el Programa en función del contexto socioeconómico en el que laboran, así como de sus resultados en pruebas de aprendizaje nacionales o estatales. Esas mismas pruebas son utilizadas, entre otros indicadores y evaluaciones de aprendizajes diseñadas *ad hoc*, para monitorear el progreso de las escuelas participantes.

Las evaluaciones en Uruguay, por su parte, fueron utilizadas hace poco para fundamentar la continuación y ampliación de un programa de equidad para Escuelas de Tiempo Completo. Por el tipo de modelo de evaluación que se realiza en este país, los datos sirvieron para demostrar que este tipo de escuelas logran mejores resultados que otras en contextos socioeconómicos semejantes.

¹⁹ Quedan excluidos de esta sección los usos de la evaluación para la acreditación o selección de estudiantes, o para la selección de escuelas por parte de estudiantes y familias.

²⁰ Este programa se coordina federalmente con la participación de equipos de investigación en cada uno de los estados, y consiste en la indagación cualitativa en escuelas sobre los aspectos institucionales que favorecen o dificultan los resultados académicos de los estudiantes. El Programa requiere que las escuelas realicen una propuesta de mejoramiento, a la cual se le da seguimiento en forma de análisis cualitativo durante su ejecución.

En el nivel subnacional, la toma de decisiones de política en función de los resultados de las evaluaciones es un fenómeno más extendido, especialmente para el diseño de programas de capacitación pedagógica y de gestión escolar. Destacan en este sentido los sistemas de evaluación de Bogotá, São Paulo y Paraná, que en todos los casos han puesto en marcha programas integrales y continuos de capacitación docente y de directivos escolares para elevar el rendimiento académico demostrado por los alumnos en las pruebas estandarizadas locales.

En Bogotá, adicionalmente, se reporta que muchas de las decisiones sobre inversión en infraestructura escolar, programas alimentarios escolares y otros insumos se realizan teniendo en cuenta los resultados de las evaluaciones de aprendizaje. De la misma manera, los datos son utilizados para seleccionar y monitorear a las escuelas privadas que albergan matrícula pública mediante subvenciones per cápita.

Uso pedagógico local

Los sistemas de medición de aprendizajes en América Latina que buscan ofrecer herramientas pedagógicas útiles a las escuelas para el mejoramiento del desempeño académico de los estudiantes lo hacen, principalmente, mediante la elaboración de informes analíticos de los resultados de cada aplicación de pruebas y cuestionarios de contexto, y mediante la devolución de resultados en formas de talleres informativos y de discusión.

En la sección anterior se explicó con cierto grado de detalle algunos de los casos nacionales y subnacionales donde se ha puesto especial énfasis al diseño y distribución de ese tipo de informes escolares. En cuanto a la devolución presencial de los datos y la discusión sobre sus implicancias pedagógicas, valgan como ejemplos los casos de Uruguay, Colombia (Examen de Estado), y Bogotá.

En Uruguay los resultados se entregan con frecuencia en el contexto de talleres de presentación en los que se procura generar discusión entre los profesionales en torno a los resultados sobre diferentes ítemes de la prueba, e incluso en torno a la validez de esos ítemes. El trabajo de devolución se ha orientado no solo al mejoramiento de los puntajes sino a que esos puntajes sean el reflejo de una concepción pedagógica diferente y renovada. Luego de las evaluaciones censales o muestrales se distribuyen manuales de interpretación junto con una copia de la prueba completa que se aplicó. De esta manera se permite que los docentes

reapliquen la prueba administrada centralmente y que practiquen con individualidad el ejercicio de corrección. Tener posesión sobre esas pruebas les permite también diseñar instrumentos similares en forma independiente. Para el equipo de medición es importante que los docentes reciban la devolución de los datos poco tiempo después de la aplicación, lo cual por lo común demora unas dos semanas.

Para la devolución de los resultados del Examen de Estado, en Colombia, se han realizado talleres de capacitación en interpretación de resultados para 1.400 colegios. Según informaciones recogidas en 2001, se planeaba una extensión del programa a 6.600 colegios más de todo el país. Los talleres se centran en la capacitación para la interpretación de los informes institucionales y para la identificación de fortalezas y debilidades de cada centro educativo. Los talleres duran dos horas y se ofrecen a representantes de 80 colegios por vez. Antes de la aplicación del nuevo modelo de evaluación referido a criterios, se realizaron 12 talleres con representantes de 11.000 colegios para explicar a los docentes el nuevo marco conceptual y la estructura de las pruebas.

En el sistema de medición subnacional de Bogotá, la difusión de los resultados comienza antes de la aplicación de las pruebas, en forma de explicación y sensibilización a los docentes. En principio, antes de la primera aplicación general, se realizó una primera aplicación piloto para motivar y generar interés. Una vez obtenidos los resultados de las primeras pruebas generales, se realizaron talleres y seminarios de devolución e interpretación de resultados para 6.000 docentes (por cada escuela del distrito participa el director, 2 ó 3 docentes y el coordinador académico)²¹. Para presentar los resultados a la comunidad educativa y civil se utilizan videos y equipos de DVD.

En cuanto al uso específico de los datos obtenidos de las pruebas, la Secretaría de Educación de Bogotá ha diseñado dos programas de intervención. Uno se denomina “Nivelación para la Excelencia”, mediante el cual se brinda apoyo en la gestión, en la infraestructura y en los recursos pedagógicos a las escuelas de más bajos resultados. Para estas intervenciones, especialmente en el componente pedagógico, se

²¹ El alcance y resultado de los seminarios se evalúa con encuestas sobre la satisfacción del público. Según encuestas recientes, los beneficiarios de los cursos demuestran altos niveles de conformidad pues logran comprender el contenido de los reportes.

contratan universidades como agencias oferentes de la capacitación, quienes trabajan “in situ” con los docentes y directores. La capacitación se realiza mediante visitas periódicas a las escuelas durante un año completo. En cuanto al efecto de estas capacitaciones, y según las mediciones últimas, las escuelas participantes crecieron un 35% en rendimiento académico entre 1998 y 2000, cuando el promedio para todo Bogotá solo creció en 11 puntos porcentuales. El otro programa de intervención es el de “Acción para la Excelencia”, a través del cual las escuelas públicas con resultados más altos (decil superior) comparten experiencias entre ellas, se recoge información sobre sus prácticas y luego se realiza y distribuye a las escuelas publicaciones sobre esas experiencias.

Uso de altas implicancias

Algunos de los casos de evaluación de altas implicancias fueron explicados en la sección anterior, principalmente los que regulan el ingreso a la educación superior o la graduación de los alumnos de la educación básica secundaria. Otros sistemas de evaluación son utilizados también para el establecimiento de incentivos docentes, para la adjudicación de recursos financieros extraordinarios, o para monitorear y regular la promoción de los estudiantes a niveles superiores dentro de la escolaridad básica.

Tanto Chile como México hacen uso de los resultados de las evaluaciones a estudiantes como indicadores complementarios para la asignación de incentivos docentes. En el caso de Chile, se hace mediante el SNED, Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales Subvencionados, que otorga una bonificación anual al equipo docente de las escuelas que logran, entre otros desempeños profesionales, elevar el rendimiento académico de sus estudiantes respecto de mediciones anteriores. Los resultados del SIMCE también se utilizan para seleccionar y monitorear los logros de las escuelas solicitan y obtienen financiamiento para los PME, Proyectos de Mejoramiento Educativo, un sistema de fondos concursables para desarrollar y ejecutar propuestas de mejoramiento institucional. En México, los docentes pueden postularse voluntariamente al programa de Carrera Magisterial, un sistema de incentivos salariales que también utiliza los resultados de las pruebas de aprendizaje nacionales como uno de los indicadores de desempeño profesional.

En cuanto al uso de los resultados para promoción de estudiantes, se registra el caso del estado de São Paulo, donde los datos censales del SARESP constituyen un indicador determinante para acceder al ciclo siguiente de escolaridad básica. Un desempeño insuficiente en las pruebas del SARESP (menos de 50% de acierto) puede ser revisado por las autoridades si el rendimiento del estudiante ha sido muy bueno durante el ciclo y se presume que por alguna razón especial no pudo rendir al nivel de su potencial el día de la aplicación. En ese caso, el alumno tiene una segunda oportunidad para presentarse a la prueba²².

Responsabilización política

Con frecuencia se cita en los documentos oficiales de evaluación educativa el propósito de hacer de los resultados de las mediciones una instancia de responsabilización política y administrativa por los avances y logros del sistema. Sin embargo, en la práctica son escasos los ejemplos de auténtica rendición de cuentas del Estado frente a la sociedad y ante la comunidad educativa por los resultados de la educación en términos de desempeño académico. Existen algunas razones legítimas para no difundir masivamente los resultados de aprendizaje en forma de indicador crítico sobre el éxito de la administración sectorial. Entre esas razones, vale mencionar que los resultados se ven condicionados por factores que no pueden ser intervenidos por el sistema educativo, así como el hecho de que los datos obtenidos en un momento en el tiempo reflejan los resultados de un proceso educacional ocurrido durante ciclos más extensos de lo que dura por lo normal un período presidencial o de gestión ministerial. Aun así, los modelos evaluativos más modernos tienen la capacidad de medir el progreso de los aprendizajes de año en año, controlando por las variables extraescolares. Sin embargo, son varios los países de la Región en que los resultados raramente se ponen a disposición en formatos o medios que faciliten el escrutinio público.

Al menos dos casos escapan a esta tendencia. Uno es el de Chile, donde los resultados publicados cada año por el SIMCE suelen tener una alta repercusión política, la cual se manifiesta principalmente en forma de artículos académicos y de opinión que utilizan los resultados y las

²² Al momento de esta publicación, el uso de las evaluaciones del SARESP con fines de bonificación salarial y promoción de estudiantes estaba siendo reconsiderado por las autoridades de la Secretaría de Educación.

bases de datos de las evaluaciones para cuestionar los logros de la reforma educacional en términos de calidad y equidad.

También destaca el caso de la Secretaría de Educación de Bogotá, que hizo públicas sus metas de lograr, para el año 2004, que el conjunto de escuelas distritales llegara a los 180 puntos de rendimiento promedio (en una escala de 0 a 306), y con un rendimiento global mínimo de 100 puntos en cada centro educativo. Si bien aun no se dispone de información sobre el logro de estas metas o sobre sus repercusiones políticas, vale mencionar el caso por ser el único sistema que explicitó sus objetivos de política educativa en términos de logros de aprendizajes.

Pruebas internacionales comparadas de aprendizaje

Otra instancia importante en el desarrollo de los sistemas nacionales de medición en América Latina ha sido, en los últimos años, la participación en pruebas internacionales de logros de aprendizaje.

Las pruebas internacionales comparadas son instrumentos de medición estandarizados que se aplican en varios países simultáneamente, en grados o grupos etarios seleccionados, y que recogen información sobre logros de aprendizaje conceptual y procedimental en diferentes disciplinas, y sobre los factores contextuales que se presume influyen en el logro académico de los estudiantes. Las pruebas suelen incluir ítemes de opción múltiple, preguntas abiertas y/o demostraciones prácticas, como por ejemplo de laboratorio en el caso de las ciencias naturales. Entre los factores contextuales más importantes que suelen estudiarse aparecen las características familiares y socioculturales de los alumnos, la infraestructura escolar, la preparación académica de los docentes, la cobertura curricular efectuada en clase en las áreas disciplinares evaluadas y las actitudes y valores de los alumnos con relación a esas áreas (Ferrer y Arregui, 2002).

Las pruebas son diseñadas y coordinadas por agencias internacionales de evaluación, y con cierto grado de participación por parte de las unidades de medición de la calidad educativa de los países. Algunas de estas agencias son privadas e independientes, de un perfil básicamente académico, tales como la IEA (Asociación Internacional para la Evaluación del Logro Académico), o ETS (Servicios de Evaluación Educacional). Otras agencias son intergubernamentales y en ellas participan las autoridades políticas o técnicas de administración y evaluación educativa de los países miembro. Ejemplo de estas últimas son la OCDE (Organi-

zación para la Cooperación y el Desarrollo Económico), el Laboratorio Latinoamericano para la Evaluación de la Educación (LLECE), dependiente de UNESCO-OREALC, o SACMEQ (Consortio de África del Sur para la Evaluación de la Calidad Educacional). Las pruebas de la IEA o de la OCDE son de cobertura mundial, mientras que otras, como las del LLECE o SACMEQ, son de cobertura regional. En la actualidad existen no menos de 10 pruebas internacionales que miden el logro estudiantil en diferentes áreas de conocimiento. Los países que ya han participado, o están participando, en una o más pruebas internacionales de aprendizaje son Argentina, Bolivia, Brasil, Chile, Colombia, Cuba, Honduras, México, Paraguay, Uruguay, República Dominicana, Perú y Venezuela.

La experiencia de participación ha sido variada, y los países se han beneficiado en mayor o menor grado de esa participación dependiendo de las capacidades técnicas con que ingresaron al proyecto, y también de las condiciones políticas que acompañaron el proceso. Esas condiciones, concretamente la voluntad política de hacer públicos los resultados y la estabilidad de las autoridades políticas y técnicas del sector, han determinado con vehemencia el tipo de impacto que tienen los resultados sobre las decisiones de política y las prácticas educativas en las escuelas. En general, ese impacto en América Latina ha sido muy débil, debido también a la falta de estrategias adecuadas de comunicación y utilización de los resultados, así como a la insuficiente definición de los propósitos y objetivos de las evaluaciones, como se vio anteriormente.

III. DISCUSIÓN Y SÍNTESIS

Teniendo en cuenta las informaciones presentadas en este informe sobre casos nacionales y subnacionales, pueden adelantarse algunas conclusiones sobre los avances y dificultades más significativos en materia de medición y evaluación de la calidad educativa en América Latina.

Un primer aspecto considerado en este estudio ha sido el de la **determinación clara y consensuada de los objetivos de los sistemas nacionales de medición**. Concretamente, a quiénes se evalúa, con qué fines y qué tipos de pruebas se aplican para lograrlos. En general, y de acuerdo a los ejemplos revisados, pareciera haberse logrado bastante coherencia entre las decisiones sobre las poblaciones a ser evaluadas y los tipos de decisiones de política y pedagógicas que se busca informar por medio de las evaluaciones. Por ejemplo, se aprecia que los países que aplican

pruebas muestrales solo pretenden entender algunos aspectos del rendimiento del sistema, tanto como para diseñar algunas medidas generales de política, o bien promover la responsabilización política por el mejoramiento de los aprendizajes (bajas implicancias). Por otro lado, los sistemas que aplican evaluaciones censales apuntan a dos tipos de objetivos principales. Por un lado, promover la responsabilización escolar, profesional y estudiantil mediante sistemas de incentivos o desincentivos, o mediante mecanismos de acreditación, que afectan directamente a los actores locales del sistema educativo (altas implicancias). Por el otro, ofrecer insumos de información en extremo desagregados sobre desempeño estudiantil para facilitar la planificación de estrategias puntuales de intervención curricular y pedagógica. Sin embargo, persiste aun la necesidad de revisar más profundamente en cada país si las decisiones que se han tomado en esta materia hasta el momento de verdad reflejan las necesidades pedagógicas y de desarrollo curricular del sistema en toda su intensidad y diversidad. Específicamente, convendría revisar en cada caso si se prefiere lograr con cada evaluación una mayor cobertura curricular o una más amplia cobertura poblacional, asumiendo que el aumento en un tipo de cobertura restringe las posibilidades de amplitud en la otra²³, y que ambas alternativas servirían propósitos diferentes.

En segundo lugar, merece destacarse el ensayo sobre diferentes **alternativas institucionales para la organización y administración de las unidades de medición** en los diferentes países. Los tipos de organización más frecuentes son las unidades dependientes directamente de los Ministerios de Educación, los organismos semiautónomos (institutos), y las universidades contratadas por el gobierno. Como se explicó antes, no es posible afirmar que una forma de organización sea más adecuada que otras. La elección depende del contexto político y de las capacidades técnicas y financieras para la evaluación a gran escala en cada país. El avance radica, en especial, en la experimentación de formas alternativas de gestión de la evaluación, lo cual indica que existe una fuerte voluntad de sostener y perfeccionar los sistemas de medición, aun cuando sus impactos en algunos países no hayan sido de gran magnitud hasta el presente. Vale destacar que la administración de los sistemas de medición desde los órganos propios del Estado, es decir sin mediación de institutos u organizaciones semiautónomas, parece ser más viable y efectiva en los sistemas educativos subnacionales (estados, provincias). Sin

²³ Ver capítulo 1 de Ravela *et al.* (2000).

embargo, algunos sistemas nacionales han logrado altos grados de institucionalización siendo administrados directamente por los ministerios de educación.

En tercer lugar, se aprecia un esfuerzo considerable por establecer **expectativas más claras y operacionales sobre el tipo de competencias académicas que han de desarrollar los estudiantes, y que serán luego evaluadas**. Los mayores esfuerzos en este sentido han estado en manos de los equipos mismos de evaluación, los que ante la falta de referentes curriculares claros (Ferrer, 1999; 2004) han optado por proponer marcos conceptuales e indicadores de logro más específicos que faciliten una interpretación sustantiva de los resultados de las mediciones²⁴. El aspecto más débil observado en este terreno es el de la escasa vinculación que aun existe entre los equipos de evaluación y los equipos de desarrollo curricular en las sedes ministeriales del sector. Es de esperar que, en la medida en que otros grupos de profesionales, como los de medición, continúen proponiendo alternativas más elaboradas en cuanto al tipo de expectativas de aprendizaje a ser evaluadas, mayor será la motivación para que los equipos técnicos de currículo participen con sus propias propuestas de actualización y especificación curricular.

En cuarto lugar, se observa que **la forma en que se reportan los datos recogidos durante las mediciones** no siempre resultan totalmente útiles a las audiencias para quienes están destinados. En particular, deben mencionarse los casos de los países que informan sobre los logros de diferentes poblaciones sin tener en cuenta sus características económicas y socioculturales. En otras palabras, resulta injusto comparar el rendimiento académico de las escuelas o localidades sin distinguir que algunas de ellas trabajan con poblaciones altamente desfavorecidas y que por lo tanto deben recorrer un camino más arduo para obtener resultados comparables al de las escuelas más favorecidas. Es necesario, entonces, rescatar el esfuerzo que ya están haciendo varios países por reportar los datos de manera tal que las escuelas puedan compararse con instituciones semejantes y así obtener una medida más justa sobre la eficiencia de su labor pedagógica. Debe recordarse también que, aunque muy inci-

²⁴ Vale recordar que, desde la perspectiva del movimiento de estándares, la definición de expectativas de logro por parte de los evaluadores no es suficiente. Esas expectativas deberían ser conocidas ampliamente por todos los actores, y los niveles esperables o anticipados de desempeño deberían ser fijados dentro de un marco de política curricular más amplio, no solamente para el diseño de un modelo de medición de aprendizajes.

pientemente aun, existen algunas iniciativas para medir esas diferencias de forma estadística mediante el uso de modelos de valor agregado. Por último, es necesario alertar sobre la necesidad de mejorar las estrategias para comunicar a los profesionales y beneficiarios del sistema no solo sobre los niveles de logro alcanzados por los estudiantes sino también sobre los tipos de desempeños concretos, debidamente ejemplificados, que se espera de los alumnos. Si bien los padres suelen tener dificultad en expresar las demandas de aprendizaje que tienen con respecto al sistema, hay cierta evidencia de que ellos están en realidad interesados en saber más sobre las expectativas curriculares de las escuelas.

En cuanto a las estrategias de **difusión y uso de los resultados**, tanto en el caso de las pruebas nacionales como de las internacionales, se aprecia que si bien existen diferentes formas de reporte, es un aun muy débil el “contrato” de información entre las agencias evaluadoras y los usuarios del sistema. Es decir, aún falta elaborar estrategias de comunicación que permitan involucrar a los usuarios de la información –léase decisores de política, docentes y directores de centro, padres de familia, medios de comunicación, asociaciones profesionales y representantes de la sociedad civil, entre otros– en las instancias de elaboración de las pruebas, de diseño de los reportes y de definición de las estrategias de difusión a utilizarse. Esto permitiría que los datos se transformen en información sustantiva y de interés para diferentes audiencias, quienes a su vez responderían mejor a la información si fueran debidamente sensibilizadas desde los primeros momentos del proceso evaluativo.

A pesar de esta debilidad, es importante destacar que los países cuyos sistemas de evaluación son de altas implicancias poseen estrategias de difusión más sólidas y hacen esfuerzos mayores por entregar información a una mayor cantidad de actores en forma desagregada por escuela, e inclusive por curso y por estudiante. Si bien esas estrategias suelen ser parte de la lógica de la evaluación de altas implicancias, pues se apunta a una responsabilización institucional (escolar) e individual (docentes y estudiantes), hay sistemas que devuelven la información de modo sistemático a los actores escolares aun cuando el uso de esa información sea para motivar, no necesariamente obligar vía sanciones o incentivos, la responsabilización profesional para el mejoramiento de los procesos pedagógicos y de gestión.

En sexto y último lugar, y en relación al **uso mismo de los datos generados por las mediciones**, es necesario distinguir entre los sistemas

que apuntan con mayor énfasis relativo a lograr mayor responsabilización política y administrativa por los resultados de aprendizaje, los que promueven la responsabilización institucional, profesional y estudiantil (altas implicancias), y los que buscan un impacto pedagógico y curricular al nivel local. Los primeros han tenido hasta el presente menos éxito en sus intentos, pues la escasa participación de la sociedad civil en los procesos evaluativos no ha favorecido movimientos importantes en demanda por mayores niveles de calidad. Es decir, que si bien los resultados son a menudo publicados en medios masivos de comunicación, no han logrado generar corrientes fuertes de opinión por parte de algunos sectores clave de la sociedad civil, y por lo tanto el impacto y presión sobre las decisiones de política no han sido tan fuertes como se esperaba. Aun así, conviene destacar que en algunos países se aprecia un mayor impacto de los resultados en la opinión pública –al menos al nivel del discurso, cuando no de las acciones– de lo que ocurría unos pocos años atrás.

Los países que intentan promover la responsabilización institucional, profesional y estudiantil cuentan con marcos normativos y mecanismos de gestión más específicos que aseguran un uso más extendido y sistemático de los resultados de las mediciones. Esto, por supuesto, no garantiza mayores niveles de calidad necesariamente, y es por eso que los países que aplican pruebas de altas implicancias han desarrollado en los últimos años formatos de entrega de los resultados más elaborados y pedagógica y curricularmente útiles para docentes y alumnos. Tal como fue descrito en la sección anterior, son los sistemas de acreditación de estudiantes para la salida de bachillerato o ingreso a educación superior los que más avances han realizado en este sentido.

Si bien desde la teoría, o desde la experiencia internacional, suele advertirse que las pruebas de altas implicancias pueden introducir desviaciones en la labor pedagógica, en tanto los esfuerzos se centren en mejorar los puntajes de las mediciones en perjuicio de otros valores formativos de la educación, esto no aparece como un problema mayor en América Latina, al menos hasta el presente. Para este trabajo no se recogieron opiniones de actores locales, pero otras investigaciones del autor en países de la Región que cuentan con este tipo de pruebas indican que los mecanismos de incentivo o acreditación no implican necesariamente una reducción del trabajo pedagógico al “entrenamiento” para los exámenes exclusivamente. De hecho, muchas de las escuelas que logran altos puntajes en las pruebas estandarizadas cuentan con propuestas edu-

cativas y curriculares propias que ponen fuerte énfasis en el rol formativo integral de sus alumnos (Ferrer, 2004).

Algunas conclusiones adicionales

Además de las generalizaciones arriba planteadas, es preciso hacer referencia más explícita a dos aspectos de la evaluación que han sido mencionados anteriormente, en concreto, el de los sistemas subnacionales de evaluación y el de la participación de los países de la Región en pruebas internacionales.

Con relación a las evaluaciones subnacionales, se observan algunas iniciativas fuertes tendientes a desarrollar sistemas de medición que acomoden mejor las necesidades locales o regionales de información. Dada la potencialidad de esos sistemas –por la escala con la que trabajan y por su circunscripción geográfica– para elaborar instrumentos más alineados a las definiciones curriculares locales –asumiendo que las hubiera– y para diseñar estrategias más efectivas de comunicación de los resultados, convendría en una segunda fase de este estudio indagar más profundamente en el tema. Hasta el presente, por lo pronto, puede adelantarse que las iniciativas han logrado grados importantes de consolidación interna y legitimación frente a la comunidad educativa, y que las posibilidades de impacto en la práctica curricular y pedagógica escolar se vislumbran como considerablemente más importantes que las de los sistemas nacionales.

En cuanto a las pruebas internacionales, importa destacar que ya es una veintena de países los que han participado en al menos una evaluación internacional comparada, y que varios de ellos han participado en varias pruebas durante el último lustro. Las experiencias de participación han sido muy diferentes en cada país, dependiendo de las capacidades técnicas y de los contextos políticos en que se enmarcaron, y por lo tanto los beneficios y problemas de esa participación han sido variados. Si bien la experiencia ha significado para los equipos técnicos de evaluación en varios de los países una excelente oportunidad de aprendizaje y capacitación, el impacto de los datos recogidos ha sido bajo. La difusión masiva de los resultados, vía la prensa escrita principalmente, ha generado algunas opiniones críticas debido a las grandes diferencias de rendimiento entre América Latina y los países más desarrollados, pero los resultados no han llegado a circular entre los medios profesionales de la educación de manera tal de promover decisiones y cambios en las prácti-

cas pedagógicas, curriculares y de gestión²⁵. Tal como se planteó para el caso de las pruebas nacionales, es necesario reforzar el vínculo de comunicación entre evaluadores internacionales y los usuarios potenciales de la información, antes, durante y después de la aplicación de las pruebas. Solo de esa manera podrá lograrse, en el corto o mediano plazo, una mayor difusión e impacto de los resultados de las pruebas en el mejoramiento de la calidad educativa.

²⁵ En los últimos dos años, la prueba PISA de la OCDE ha generado niveles importantes de debate educativo en la Región, en buena parte debido al tipo de especificación de contenidos académicos que propone, así como las definiciones que ofrece sobre aprendizajes socialmente significativos y los diferentes niveles de desempeño posibles.



Segunda parte

DESCRIPCIÓN DE CASOS NACIONALES Y SUBNACIONALES DE SISTEMAS DE EVALUACIÓN DE APRENDIZAJES EN AMÉRICA LATINA

Nota: Los datos sobre publicidad y uso de resultados presentados en estas tablas fueron provistos por autoridades técnicas o asesores de cada país o distrito regional. Esos datos pueden no resultar congruentes con indagaciones posteriores realizadas por el autor, y transcritas en los párrafos descriptivos sobre cada sistema. Se optó por incluir ambos tipos de informaciones para resaltar las diferencias que existen entre los objetivos oficiales sobre uso de los datos, y lo que efectivamente ocurre en la práctica.



ARGENTINA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
ARGENTINA	1993- 1994	7, 5M, 6M	L, M	MN	G, U, P	CD, CG	OREALC 1997, TIMSS-R
DiNIECE	1995	3, 9, 5m, 6M 7	L, M L, M, CS, CN	MN MN	G, U, P G, U, P	CD, CG CD, CG	(2000, 2003), PISA-Plus 2001, PIRLS 2001,
	1996-1998	(3, 9) (5M, 6M) 6, 7	L, M L, M, CS, CN	(MN) (CN) MN	G, U, P G, U, P	CD, CG CD, CG	IEA-Civic Education 2001, PISA 2003
	1999	3, 7, 9 (6) (5M, 6M)	L, M L, M, CS, CN	MN (MN) (CN)	G, U, P G, U, P	CD, CG CD, CG	
	2000	(3) (6, 5M, 6M) 9	L, M, CS, CN L, M	(MN) (CN) MN	G, P		
	2002						
	2003	3, 6, 9, 5M	L, M	MN			
		3, 6, 9, 3 Polimodal (antiguo 5M)	L, M	MN			
	2005	Pruebas de pilotaje 3, 6 9, 3 Polimodal (antiguo 3M)	CS CN				

[2]

Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

La medición de la calidad educativa ha estado a cargo del Ministerio de Educación desde 1993. Originalmente, y hasta el año 2001, la oficina evaluadora fue el SINEC, Sistema Nacional de Evaluación de la Calidad Educativa, dependiente de la Secretaría de Programación y Gestión Educativa y de la Subsecretaría de Evaluación de la Calidad Educativa. En el año 2000 se constituyó el IDECE, Instituto de Calidad Educativa, entidad semiautónoma del Ministerio de Educación. El Instituto fue disuelto poco tiempo después y hoy lo reemplaza la DiNIECE, Dirección Nacional de Información y Evaluación de la Calidad Educativa, órgano de línea del Ministerio de Educación, Ciencia y Tecnología.

Currículo y estándares

Argentina aprobó en 1994 los Contenidos Básicos Comunes (CBC), dentro de un marco de reforma y de un proceso de descentralización en el cual se transfirió la administración de todas las escuelas nacionales a las provincias (estados federados). Cada provincia, así como la Ciudad Autónoma de Buenos Aires, son responsables por el diseño de sus currículos, compatibles con los CBC. Actualmente, se están elaborando los Núcleos de Aprendizajes Prioritarios para todos los años de escolaridad. Estos Núcleos, basados en los CBC na-

cionales y aprobados por consenso federal, priorizan algunos contenidos para la enseñanza en matemática, lengua, ciencias sociales y ciencias naturales.

Tipos de pruebas

El sistema mide anualmente muestras representativas de alumnos y en el último año de educación media realiza mediciones censales. Los instrumentos son normativos, se basan en preguntas de opción múltiple y son acompañadas por cuestionarios a alumnos, directores y docentes para el relevamiento de factores asociados al rendimiento. Actualmente se está diseñando un nuevo modelo de medición referido a criterios, que será aplicado por primera vez en 2006.

Tipos de reporte

Los resultados de las mediciones se publican desagregados por área de conocimiento, por género, por provincia, por ámbito geográfico (urbano y rural), y por tipo de gestión (pública y privada). Además, se clasifica a la población de alumnos de acuerdo a algunos indicadores sobre su vulnerabilidad socioeconómica, y a las escuelas según sus recursos materiales y pedagógicos. Para las mediciones censales se reportan los datos desagregados por escuelas. Los resultados se reportan como porcentaje de respuestas correctas para el total de estudiantes y para los diferentes estratos considerados.

Uso y difusión de resultados

Las mediciones en Argentina son de bajas implicancias, en tanto los resultados no conllevan consecuencias directas para los actores del sistema (la excepción se dio en el año 2000, cuando los resultados de secundaria por establecimiento se pusieron a disposición del público). Hasta el año 2000 se publicaron cuadernos con recomendaciones metodológicas para docentes, en los cuales se discutían algunos ítemes de las pruebas y se explicaban las dificultades que estos habían presentado para los alumnos. Actualmente esas publicaciones han sido discontinuadas y los resultados de las mediciones no parecen haberse instalado en la agenda de debate público ni en la del sistema educacional. Solamente los resultados de las pruebas internacionales parecen tener cierto grado de impacto en los medios de comunicación y en la opinión pública.

Sistemas subnacionales de evaluación	La provincia de Buenos Aires ha desarrollado, con la asesoría técnica de UNESCO-OREALC, un sistema de medición basado en el modelo de “valor agregado” que permite medir el progreso académico de los alumnos teniendo en cuenta los factores socioeconómicos de cada población escolar. La Ciudad Autónoma de Buenos Aires también posee un sistema propio de evaluación, pero los resultados no se hacen públicos sino que se devuelven directamente a las escuelas para fines de investigación y mejoramiento pedagógico e institucional. También realizan, o han realizado, mediciones independientes las provincias de Catamarca, Chaco, Córdoba, Entre Ríos, La Pampa, La Rioja, Mendoza, Misiones, Neuquén, Río Negro, Salta, San Juan, Santa Cruz, Santa Fe, Tierra del Fuego y Tucumán. Los grados de cobertura y las metodologías de evaluación varían de provincia en provincia, así como el grado de apoyo que han solicitado al Ministerio de Educación para el desarrollo de instrumentos propios. Además, casi todas las provincias han desarrollado algún tipo de estrategia propia para la difusión de los resultados del operativo nacional de evaluación.
Pruebas internacionales de aprendizaje	Argentina ha participado en las pruebas internacionales del LLECE (1997), TIMSS-R (2000), PISA-Plus (2001), PIRLS (2001), IEA-Civic Education Study (2001), PISA (2003).
Avances observados	Vale mencionar la legitimidad política que ha ido ganando el sistema nacional de evaluación, la continuidad en la ejecución de los operativos anuales y la creciente percepción sobre la necesidad de mejorar la utilidad de las evaluaciones para fines pedagógicos.
Dificultades observadas	Entre las dificultades más importantes que enfrenta el sistema en la actualidad, cabe mencionar la falta de comunicación entre la unidad de medición y las oficinas de desarrollo curricular y formación y capacitación docente, así como la resistencia hacia las mediciones cuantitativas que ofrecen algunos académicos y técnicos dentro del Ministerio. Simultáneamente, es necesario que la unidad de medición haga análisis más complejos de los datos obtenidos y que comience a reportar los datos de

rendimiento en forma diferencial para las escuelas de diferentes estratos socioeconómicos. No se observa un marco conceptual definido para la interpretación de los resultados ni de los factores asociados al rendimiento, pero a pesar de que las pruebas son referidas a normas, y de que el vínculo entre las oficinas de evaluación y de desarrollo curricular es sumamente débil, se ha ensayado con la presentación de datos en forma más sustantiva, en relación a las capacidades específicas que el conjunto de los estudiantes demuestran de acuerdo al porcentaje de logro mostrado en las evaluaciones.

El Plan de Evaluación de la Calidad Educativa 2003-2007, recientemente aprobado por el Consejo Federal de Educación tiene como objetivos principales los de desarrollar nuevas metodologías cuanti-cualitativas, profundizar los trabajos de análisis e interpretación de los resultados, brindar asistencia para el fortalecimiento de los sistemas provinciales de evaluación, y lograr mayores grados de articulación con otros programas y direcciones del Ministerio para desarrollar planes más eficaces de mejoramiento de la calidad educativa. Se plantea también la necesidad de hacer mejor uso de los datos para diseñar políticas tendientes al mejoramiento de la equidad en el acceso al conocimiento, así como de elaborar informes diferenciados para distintos tipos de audiencias. Se continuará participando en pruebas internacionales.

BOLIVIA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
BOLIVIA (SIMECAL)	1996-2000	1, 3, 6, 8, 4M	L, M	MN, CN, CR (La Paz)	G, P, U		OREALC 1997 ALL (2001)

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)], U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia], P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional Por Ley del Poder Ejecutivo, el SIMECAL depende directamente del despacho del Ministro de Educación. En la práctica, sin embargo, se encuentra ubicado dentro de la Dirección General de Estrategia Sectorial. Esto determina que las informaciones que se recogen mediante las evaluaciones no lleguen rápidamente a las autoridades políticas con mayor poder de decisión. Hasta el año 2002, esta situación aun no había sido resuelta.

Currículo y estándares La Reforma Educativa boliviana introdujo, entre los años 1997 y 1998, un nuevo marco curricular nacional

para educación primaria organizado en torno a competencias, indicadores de desempeño y estrategias de evaluación por áreas disciplinares. En educación secundaria aún continúan vigentes los antiguos programas introducidos en los años setenta. Las nuevas propuestas curriculares, y muchas de las acciones efectivas de reforma, reivindican fuertemente la educación en lenguas y culturas originarias. Los marcos, organizados y difundidos en forma de módulos, no pretenden una cobertura curricular exhaustiva, sino que proponen tipos de actividades que responden a una nueva “lógica curricular” y pedagógica. Oficialmente, las pruebas del SIMECAL están referidas a los marcos curriculares nacionales. Sin embargo, existen críticas por parte de los equipos técnicos de desarrollo curricular en cuanto a la forma relativamente aislada en que el SIMECAL desarrolló las tablas de especificaciones y los instrumentos de evaluación.

Tipos de pruebas

Originalmente, las pruebas desarrolladas por el SIMECAL fueron estrictamente normativas y referidas al antiguo currículo nacional. Después, se desarrollaron pruebas referidas a criterios sobre la base de las nuevas competencias curriculares y de tres niveles de logro (A, B y C) establecidos desde la unidad de evaluación misma. Los instrumentos se basan en preguntas de opción múltiple. Se aplican pruebas en las tres lenguas originarias principales, y todas se acompañan con cuestionarios a docentes, directores, padres y alumnos para el análisis posterior de factores asociados al rendimiento. El SIMECAL fue creado básicamente para la evaluación de aprendizajes escolares, pero de manera lenta fue asumiendo nuevas responsabilidades: Prueba de Aptitud Académica (bachilleres egresantes), evaluación de maestros interinos, bachillerato pedagógico, asesores pedagógicos, profesores (sistema de bonos y ascensos), preselección de candidatos a directores distritales, directores de escuela e ingreso a institutos normales.

Tipos de reporte

Los informes de resultados elaborados por el SIMECAL indican los porcentajes de alumnos que alcanzan cada nivel de desempeño en las áreas evaluadas. A su vez, los datos son desagregados principalmente por departamento, por lengua vernácula, por tipo de dependencia, y por género. Tanto en el caso de las pruebas censales como en las

muestrales, las escuelas reciben un informe institucional del rendimiento de sus alumnos que incluye el puntaje promedio institucional, una descripción operacional de los niveles de logro establecidos por área, y el porcentaje de alumnos de la escuela ubicados en cada nivel. Luego se presenta una descripción estandarizada de las fortalezas y debilidades del conjunto de alumnos por cada eje temático en cada área curricular. También se distribuye un folleto con análisis de ítemes seleccionados y propuestas metodológicas para mejorar los resultados en relación a esos ítemes. Para el caso de los bachilleres egresantes, se ofrece un informe individual por alumno con su desempeño en cada eje temático evaluado por área. Para el conjunto de escuelas y de equipos técnicos locales, se distribuye un folleto con indicadores de eficacia escolar según los datos recogidos con las pruebas y cuestionarios de contexto. Los factores asociados al rendimiento que se publican refieren a las características del alumnado y sus familias (educación previa, NES familiar y actitudes, vivienda), procesos pedagógicos, y recursos materiales escolares. Si bien se cuenta con información sobre las características socioeconómicas y culturales de las familias, no se han definido estratos específicos que permitan desagregar los resultados según esta variable.

Uso y difusión de resultados

Los datos de las evaluaciones son presentados oficialmente en diversas instancias: presentación pública por parte del Ministro de Educación, presentación a los diferentes equipos de reforma educativa y curricular, entrega de trípticos y folletos a escuelas, realización de talleres informativos en ciudades capitales con participación de la sociedad civil y de equipos técnicos locales, y publicación de resultados globales en prensa escrita. Se realizan también estudios secundarios que asocian los datos de rendimiento a variables de nutrición infantil. Se ignora qué uso efectivo se da a los resultados en las escuelas. Los datos no han sido utilizados en la medida en que se esperaba para el diseño de políticas educativas. Problemas de comunicación al interior del Ministerio, y probablemente la falta de acuerdos sectoriales sobre la validez e importancia del logro de los aprendizajes evaluados hace que los resultados tengan bajo impacto en la agenda de política. Diversas opiniones al interior de la unidad de evaluación, así como de agencias

	<p>de cooperación internacional, coinciden en que los datos del SIMECAL deberían utilizarse como uno de los indicadores principales para la definición de mejores políticas sectoriales (no solamente curriculares y pedagógicas) e intersectoriales.</p>
Sistemas subnacionales de evaluación	<p>No se han desarrollado sistemas subnacionales.</p>
Pruebas internacionales de aprendizaje	<p>Bolivia ha participado en la prueba de LLECE-UNESCO (1997) y ALL (2001).</p>
Avances observados	<p>Se observa continuidad en los programas de evaluación y relativa estabilidad de los cuadros técnicos. Se han realizado esfuerzos por mejorar de manera gradual los instrumentos de medición, desarrollar pruebas criteriosales y producir reportes de calidad que permiten, potencialmente, un buen uso pedagógico y curricular de los datos.</p>
Dificultades observadas	<p>Entre las dificultades más importantes se encuentra la ausencia de políticas sectoriales que enfatizan el seguimiento a la calidad de los aprendizajes y que para ello utilicen sistemáticamente los datos recogidos por el SIMECAL. Existe la percepción y el convencimiento de que los datos podrían utilizarse como indicadores clave para la formulación de políticas intersectoriales, pero la falta de comunicación y articulación entre órganos de gobierno hace que esa formulación no se realice. Tampoco se realiza un seguimiento adecuado sobre el uso efectivo que las escuelas hacen de los datos proporcionados por el SIMECAL.</p>
Proyección a futuro	<p>Las autoridades del SIMECAL consideran que el sistema ha asumido más responsabilidades de evaluación que las que su capacidad operativa permite. Por ello, es menester concentrar los esfuerzos sobre todo en la evaluación de aprendizajes escolares y dedicar todos los recursos restantes a difundir más ampliamente los datos y a “educar” a audiencias clave para el uso de las informaciones en la formulación de políticas educativas.</p>

BRASIL

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
BRASIL	1990-1993	1, 3, 5, 7	L, M, CN	MN	G, P		ETS-
INEP/ MEC(SAEB)	1995	4, 8, 11	L, M	MN	G, P	CD, DC	Matemáticas (1991;
	1997	4, 8, 11	L, M, CN	MN	G, P	CD, DC	OREALC (1997); PISA (2000,
	1999	4, 8, 11	L, M, CN, CS	MN	G, P	CD, DC	2003), TIMSS 2003
(ENEM)	2001-2005	4, 8, 11	L, M	MN	G, U, P	CD	
	1998-2005	Término de Educación básica	5 competencias [7]	V	G, U	SES	

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental, V=Voluntario

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos, SES=Selección educación superior

[7]

I. Dominar linguagens – Dominar a norma culta da língua portuguesa e fazer uso da linguagem matemática, artística e científica. II. Compreender fenômenos – Construir e aplicar conceitos

das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas. III. Enfrentar situações-problema – Selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema. IV. Construir argumentação – Relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente. V. Elaborar proposta – Recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural.

Marco institucional

La agencia evaluadora al nivel nacional en Brasil es el SAEB– Sistema de Evaluación de la Educación Básica y está administrada por el Instituto Nacional de Investigaciones Educativas (INEP), un órgano dependiente del Ministerio de Educación pero con cierto grado de autonomía administrativa. Un subsistema de evaluación del INEP es el ENEM, Examen Nacional de Enseñanza Media, que guarda relación técnica con el SAEB pero se administra y aplica en forma independiente.

Currículo y estándares

Las primeras aplicaciones del SAEB se realizaron antes de la elaboración de los Parámetros Curriculares Nacionales, aprobados en el año 1997, cuando cada estado contaba con un marco curricular propio. Por esta razón, fue necesario entonces construir y validar con los estados una matriz de referencia que contuviera un 60% de contenidos comunes a todos los currículos estatales. Posteriormente, se fue actualizando esa matriz para incorporar los contenidos de los Parámetros Curriculares Nacionales, de manera que ya no es necesario validar permanentemente la matriz con las Secretarías de Educación de cada estado. En la actualidad, participan en la definición de contenidos especialistas contratados para tal fin, pero no se trata de un trabajo conjunto con los equipos técnicos estables de la administración nacional del currículo. Los docentes son consultados durante la elaboración de las pruebas, pero no puede considerarse un proceso altamente participativo de estos actores. La matriz de referencia busca lograr un equilibrio entre el currículo prescrito y el implementado en las escuelas, pero el diseño de las pruebas y cuestionarios no permite realizar análisis posteriores sobre cuánto del rendimien-

to se debe a la implementación efectiva del currículo nacional (Oportunidad de Aprendizaje).

Tipos de pruebas

Los instrumentos de medición son referidos a normas y se construyen de acuerdo al modelo psicométrico de la Teoría Clásica y de TRI. La matriz de referencia se basa en descriptores que involucran dos niveles de especificación: competencias y habilidades. Se mantienen los mismos tópicos de evaluación en todos los grados evaluados, pero se va ajustando las prioridades o nivel de complejidad para los grados superiores. El modelo es fuertemente cognitivista, en tanto busca detectar en qué punto del proceso de aprendizaje se produce una ruptura o detenimiento en la adquisición de saberes más complejos. Para esto, luego de la aplicación de las pruebas, se realizan estudios de seguimiento cualitativo en aula. Las pruebas van acompañadas de cuestionarios que relevan datos escolares y extraescolares para estudios posteriores de factores asociados al rendimiento. Para tal fin, se diseñó una caracterización sociodemográfica de las escuelas y sus entornos, así como una caracterización profesional de los directores y docentes a cargo de los grupos evaluados. Los cuestionarios se aplican a alumnos, docentes y directores. Estas informaciones también se cruzan estadísticamente con los datos del censo nacional de educación y del Instituto Brasileño de Geografía y Estadística. El ENEM, Examen nacional de Enseñanza Media, se aplica a los egresantes de ese nivel de escolaridad, es voluntario y requiere de una inscripción o registro previo por parte de los alumnos. La matriz de referencia para la prueba fue construida ad hoc pero se sirve del banco de ítems del SAEB y refiere por lo tanto a los contenidos de los Parámetros Curriculares Nacionales. La prueba apunta a ofrecer indicadores relevantes para la selección de individuos en el mercado laboral y, eventualmente, para el ingreso a la educación superior (estos indicadores son utilizados opcionalmente por las universidades).

Tipos de reporte

Los informes de resultados reportan, principalmente, el porcentaje de alumnos por grado que logra diferentes niveles de desempeño en una escala predeterminada para las diferentes áreas del currículo. Complementariamen-

te, se presenta una descripción del tipo de capacidades que los alumnos demuestran haber logrado en cada uno de esos niveles de desempeño. Los datos se presentan agregados al nivel nacional, y luego se desagregan por región, estado y municipio. El ENEM entrega un boletín individual de resultados. Si alguna persona desea tener acceso al mismo, debe pedir autorización del estudiante. A solicitud de las escuelas se puede hacer un boletín de resultados de las mismas.

Uso y difusión de resultados

El rediseño de los instrumentos de evaluación del SAEB a partir del año 2001 ha determinado que se revean también las estrategias de difusión y uso de los datos recogidos. Hasta el momento, el énfasis ha estado puesto principalmente en la sensibilización de audiencias clave para asegurar un mayor impacto posterior de las informaciones. Para tal fin, se han realizado visitas informativas y asambleas a todos los estados con la participación de equipos técnicos estatales y locales, docentes, directores de escuela, universidades y medios de prensa. Por demanda de los docentes, las asambleas ponen especial acento en las características técnicas de las pruebas y de la matriz de referencia. Los resultados son de conocimiento público pero su uso está destinado básicamente a la toma de decisiones de política al interior del Ministerio de Educación. Por tratarse de pruebas muestrales, los resultados no se devuelven a las escuelas individualmente. Sin embargo, el INEP ha realizado informes específicos de uso pedagógico que muestra una selección de ítemes, presenta los desempeños promedio de los alumnos para cada uno de ellos, y propone un conjunto de explicaciones sobre las posibles causas de desacierto en las respuestas. Adicionalmente, se han encargado investigaciones cualitativas que buscan comprender algunos de los factores pedagógicos y de prácticas sociales en aula que expliquen los resultados en el ámbito nacional. Los resultados no tienen altas implicancias para las escuelas o los alumnos, y no se observa un alto impacto en la responsabilización política por la calidad de los aprendizajes. Los bancos de datos y de ítemes están disponibles para investigaciones externas al INEP, pero solo pueden accederse por convenio y mediante arreglos legales previos. Los resultados del ENEM son utilizados

para la selección de inscriptos a la ProUni, para la obtención de becas de estudio integrales o parciales (para estudiantes de bajo nivel socioeconómico). Además, cientos de instituciones de educación superior los usan para sus procesos selectivos.

Sistemas subnacionales de evaluación Ver separadamente los casos estatales de Paraná, São Paulo y Minas Gerais.

Pruebas internacionales de aprendizaje LLECE (1997) y PISA (2000)

Avances observados Se aprecia un importante grado de institucionalización interna, en términos de estabilidad de los equipos y de continuidad en las evaluaciones. Externamente, el SAEB ha logrado instalarse sin oposición por parte de los sindicatos u otros actores de la comunidad educativa, lo cual se atribuye a los esfuerzos realizados por mantener a los actores locales de forma adecuada informados sobre los objetivos y fines de la evaluación. Vale destacar también el tránsito hacia un modelo de evaluación más centrado en el uso pedagógico y curricular de los datos recogidos por los operativos.

Dificultades observadas A pesar del interés y esfuerzos realizados por maximizar el uso pedagógico de los datos, se cree que llevará un tiempo considerable, y mucho trabajo de difusión y capacitación, lograr un impacto efectivo en las escuelas. No se dispone de informaciones que den cuenta del uso dado a los datos para la toma de decisiones de política educativa a nivel nacional.

Proyección a futuro Se espera poder diversificar en mayor grado los tipos de reporte de resultados para impactar con eficacia en diferentes tipos de audiencias.

CHILE

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
CHILE (PER Prueba de Evaluación del Rendimiento Escolar)	1982-1984	4, 8	L, M, CN, CS	MN, CN	G, U, P		OREALC 1997, TIMSS-R 1998, TIMSS 2003, IALS 1998, IEA-Civic Education (2000), PISA Plus
SIMCE Sistema de Medición de la Calidad de la Educación	1988	4	L, M, CS, CN, AA	CN, ME	G, U, P		
	1989	8	L, M, CS, CN, AA	CN, ME	G, U, P		
	1990	4	L, M, CS, CN, AA	CN, ME	G, U, P		
	1991	8	L, M, CS, CN, AA	CN, ME	G, U, P	CD, DC, FA	
	1992	4	L, M, CS, CN, AA	CN, ME	G, U, P	CD, DC, FA	
	1993	2M	AA	CN, ME	G, U, P	CD, DC, FA	
	1994	8	L, M, CS, CN, AA,	CN, ME	G, U, P	CD, DC, FA	
		4	L, M, CS, CN, AA	CN, ME	G, U, P	CD, DC, FA	
	1995	2M 8	L, M, AA L, M, CS, CN	CN, ME CN, ME	G, U, P G, U, P	CD, DC, FA CD, DC, FA	
	1996	4	L, M, CS, CN, AA	CN, ME	G, U, P	CD, DC, FA	
	1997	8	L, M, CS, CN, AA	CN, ME	G, U, P	CD, DC, FA	
	1998	2M	L, M	MN, ME	G, U, P	CD, DC, FA	

Continuación

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
	1999	4	L, M, CN, CS	CN, ME	G, U, P	CD, DC, FA	
	2000	8	L, M, CN, CS	CN, ME	G, U, P	CD, DC, FA	
	2001	2M	L, M	CN, ME	G, U, P	CD, DC, FA	
	2002	4	L, M, CN, CS	CN, ME	G, U, P	CD, DC, FA	
	2003	2M	L, M	CN, ME	G, U, P	CD, DC, FA	
	2004	8	L, M, CN, CS	CN	G, U, P	CD, DC, FA, I	

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos, I= Incentivos docentes.

Marco institucional	El SIMCE, Sistema de Medición de la Calidad Educativa, es administrado desde la Unidad de Currículum y Evaluación del Ministerio de Educación. Este sistema, originalmente creado fuera del Ministerio bajo otra denominación (PER) y la dirección técnica de la Universidad Católica de Chile, aplica pruebas de logro académico desde el año 1981. A partir del año 1988 comenzó a funcionar dentro de la estructura del Ministerio de Educación, por lo cual no es un organismo autónomo aunque su legitimidad técnica le otorga grados importantes de liderazgo y capacidad de decisión de política evaluativa al nivel nacional.
Currículo y estándares	En 1996 y 1998, en una de las últimas fases de la amplia reforma educacional emprendida a principios de la década pasada, fueron aprobados por ley los Objetivos Fundamentales y los Contenidos Mínimos Obligatorios para la educación básica primaria y secundaria, y que constituyen el marco curricular oficial en Chile. Los instrumentos de medición de aprendizajes del SIMCE fueron incorporando los contenidos de esos referentes curriculares gradualmente, en la medida en que los nuevos programas se implementaban en las escuelas. La LOCE–Ley Orgánica Constitucional de Enseñanza de 1990, ordena también la elaboración de estándares nacionales de currículo, mandato que está siendo cumplido en la actualidad con la elaboración de estándares de contenido y desempeño para todos los grados de escolaridad en las áreas de Lenguaje, Matemáticas, Historia y Ciencias Sociales, Ciencias (Biología, Química y Física) e Inglés como segunda lengua.
Tipos de pruebas	Las pruebas son censales en todos los grados de aplicación. Históricamente han sido diseñadas con referencia a normas, aunque recientemente se ha redefinido el modelo para poder reportar los resultados con referencia a criterios. Los instrumentos incluyen preguntas cerradas de opción múltiple, así como preguntas para el desarrollo de respuestas abiertas. Todos los operativos son acompañados con cuestionarios de contexto para el estudio posterior de factores escolares y extraescolares asociados al rendimiento académico. Complementariamen-

te, un equipo técnico especial realiza un seguimiento cuanti-cualitativo de implementación curricular a nivel nacional mediante encuestas y observaciones a docentes y directores de escuelas. Ese seguimiento permite indagar más profundamente sobre la incidencia que las diferentes fuentes de prescripción curricular utilizadas en la escuela (marcos curriculares, programas de estudio, libros de texto, etc.) tienen sobre el logro académico de los alumnos.

Tipos de reporte

Los informes generales de difusión presentan los resultados agregados al nivel nacional, y desagregados por regiones geográficas, tipo de escuela (urbana o rural) y tipos de administración (municipal, privada o subvencionada). A lo largo de los últimos años se han utilizado diferentes formas de reporte, dependiendo de los objetivos de información para diferentes audiencias. Algunos informes presentan los resultados en forma de promedios de logro y desviación estándar para las diferentes asignaturas y estratos evaluados, mientras que otros reportan porcentajes de respuestas correctas para cada signatura, o más detalladamente por objetivos al interior de cada una de ellas. Varios de los informes enfatizan las diferencias de resultados entre tipos de escuelas, o bien la evolución de resultados a lo largo del tiempo. La comparabilidad interanual de resultados ha sufrido cuestionamientos técnicos, por lo que recientemente se ha recurrido a la metodología de *equating*, lo cual permite inferencias más confiables sobre el progreso en los aprendizajes a lo largo de la última década. Complementariamente, el equipo técnico del SIMCE produce informes analíticos-interpretativos de los resultados y sus implicancias para la comprensión de la evolución de la implementación y alcance de los nuevos marcos curriculares. Todas las escuelas evaluadas reciben un informe institucional que resume los niveles de logro académico de los grupos (cursos) participantes en la evaluación de cada una de las asignaturas.

Uso y difusión de resultados

Chile es probablemente el país latinoamericano donde los resultados de aprendizaje medidos estandarizadamente tienen la mayor repercusión política interna. Los esfuerzos realizados por el gobierno para impulsar y sostener altos niveles de inversión en la reforma educa-

cional han requerido mostrar evidencias sobre el éxito de los programas implementados. Muchos de los informes oficiales de resultados académicos, así como las investigaciones independientes que hacen uso de los datos del SIMCE, ponen especial énfasis en las variaciones de logro educativo a lo largo de los años noventa y en el tamaño de las brechas entre escuelas de diferentes dependencias y estratos socioeconómicos. Estos informes son frecuentemente balanceados, o desafiados, por investigaciones independientes de la comunidad académica nacional, que cuestiona el impacto de la reforma y de los mecanismos de financiamiento en términos de mejoramiento de la calidad educativa. En cuanto al uso de las evaluaciones SIMCE para la focalización de políticas de mejoramiento, se destacan al menos tres estrategias fuertemente consolidadas: 1) El uso de los resultados como uno de los dos indicadores más importantes en la selección de las escuelas que ingresan al Programa de las 900 escuelas (P-900- en este momento atendiendo alrededor de 1.500 instituciones); 2) La utilización de los datos para la asignación de fondos concursables para los Proyectos de Mejoramiento Educativo; 3) El uso de los datos como uno de los indicadores principales para otorgar incentivos docentes mediante el SNED, Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos. El uso pedagógico y curricular de las evaluaciones centrales se da, básicamente, mediante la difusión entre docentes de nuevas y más sofisticadas formas de evaluación, y a través del seguimiento cualitativo que hace la UMC de la implementación curricular en las escuelas. Ese seguimiento permite observar el grado de apropiación formal y efectivo de los nuevos marcos curriculares, y los resultados (tanto del seguimiento como de las pruebas nacionales e internacionales) permiten identificar cuáles aspectos del currículo necesitan de mayor estímulo para su implementación y logro.

Sistemas subnacionales de evaluación	No se han desarrollado sistemas subnacionales de medición y evaluación de aprendizajes.
Pruebas internacionales de aprendizaje	Chile ha participado en las pruebas internacionales comparadas OREALC 1997, TIMSS-R 1998, IALS 1998, IEA-Civic Education Study 2000, y PISA Plus 2001.

Avances observados	Se observa una transición importante hacia el uso de los datos de las evaluaciones para la definición y diseño de políticas de mejoramiento educativo, en contraste con el uso original que se dio al sistema para informar al público para la selección de escuelas. Si bien los resultados individuales por escuela aun se difunden públicamente, el mayor énfasis está puesto sobre el análisis sustantivo de los datos y sus implicancias para el mejoramiento de la calidad educativa. Destaca la continuidad de los operativos a lo largo de los años, la consolidación y estabilidad de los equipos técnicos, y la institucionalización externa lograda por el sistema.
Dificultades observada	A pesar de los informes analíticos y las recomendaciones pedagógicas elaborados por el SIMCE, aun resulta difícil impactar sobre los procesos de desarrollo curricular y de enseñanza-aprendizaje en las escuelas.
Proyección a futuro	El proceso actual de elaboración de estándares requerirá una redefinición o ajuste de los instrumentos de medición de aprendizajes.

COLOMBIA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
COLOMBIA ICFES- Ministerio de Educación	1991/1994	3, 5, 7, 9	L, M	MN, MR	G, U, P		TIMSS (1995); OREALC (1997); IEA-Civic Education (2000); PIRLS (2001)
Pruebas SABER	1997/1998 1998/1999 2002-2003	3, 5, 7, 9 7, 9 5, 9 [7]	L, M, L, M, CN L, M, CN	MN MN CN	G, U, P G, U, P		
ICFES: Exámenes de Estado	1980-2005	11	L, M, CS, CN, LE	CN	G, U	DC, SES	

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental, C=Ciudadanía

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos, SE=Selección educación superior

[7]
En algunas entidades territoriales se realizó, adicionalmente, la evaluación de los grados 3 y 7.

Marco institucional

Todos los operativos de medición y evaluación de aprendizajes se realizan a través del ICFES, Instituto Colombiano para el Fomento de la Educación Superior, un órgano semiautónomo del Ministerio de Educación Nacional. El ICFES depende políticamente del Ministerio pero goza de altos grados de autonomía administrativa, lo cual le ha permitido lograr considerable estabilidad técnica y desarrollo institucional a lo largo de las tres últimas décadas. Dentro del ICFES funciona el Servicio Nacional de Pruebas, la unidad técnica encargada del diseño, aplicación y procesamiento de los operativos nacionales de evaluación. El operativo principal, de mayor estabilidad en el tiempo y reglamentado por ley nacional, es el correspondiente a los Exámenes de Estado, que regulan el acceso de los estudiantes a la educación superior. Estas pruebas son administradas por el ICFES en forma independiente y no guardan relación técnica ni política con la administración ministerial de la educación básica (primaria y secundaria), aunque ofrecen datos de alto valor curricular y para las escuelas medias de todo el país. Por otro lado, la Dirección de Calidad del Ministerio de Educación y el ICFES han desarrollado en forma conjunta un Sistema Nacional de Evaluación de aprendizajes en educación básica, denominado SABER. Este proyecto fue concebido como un plan de monitoreo de la calidad educativa que habría de llevarse a cabo entre los años 1997 y 2005, en forma de seguimiento longitudinal de aprendizajes. Sin embargo, la rotación de autoridades políticas e importantes reestructuraciones en el Ministerio de Educación del sector en los últimos años, y los consecuentes cambios en las prioridades y líneas de acción de política, han impedido que ese proyecto se complete y las aplicaciones anuales han sido discontinuadas.

Currículo y estándares

En Colombia conviven tres referentes curriculares oficiales que se han ido sumando a lo largo de la última década. Estos son los Lineamientos Curriculares, los Indicadores de Logros Curriculares Generales, y los nuevos Estándares nacionales aprobados en 2002 para tres asignaturas. Las evaluaciones de SABER aplicadas hasta 1999 tomaron como referente curricular los dos primeros documen-

tos. Los Exámenes de Estado, por su parte, fueron diseñados originalmente sobre la base de un conjunto de habilidades cognitivas universales, pero recientemente fueron reformulados en base a criterios que incorporan los contenidos de los marcos curriculares nacionales. El diseño de la matriz de referencia de estos exámenes puede considerarse, desde el punto de vista técnico, como una propuesta de estándares, en tanto explican clara y exhaustivamente qué tipos y niveles de desempeños académicos se espera de los estudiantes, a la vez que constituyen herramientas útiles de información para las escuelas secundarias, los alumnos y sus familias.

Tipos de pruebas

Las pruebas SABER son muestrales, están referidas a criterios y evalúan el rendimiento académico de acuerdo a una escala predefinida de tres niveles de desempeño de complejidad creciente para cada asignatura y año evaluados. Las pruebas se complementan con cuestionarios de contexto que permiten el análisis de factores escolares y extraescolares asociados al rendimiento académico (características de escuela, maestro, alumno y familia). Los Exámenes de Estado se aplican censalmente a todos los aspirantes a la educación superior. Las pruebas miden el alcance de conocimientos en las áreas curriculares de formación general, y también permiten la evaluación de conocimientos en profundidad para algunas disciplinas, según los requisitos de ingreso de los diferentes programas de educación superior. Se evalúan también conocimientos en lenguas extranjeras y formación interdisciplinar (específicamente, *medio ambiente y violencia y sociedad*).

Tipos de reporte

Los informes de las pruebas SABER presentan los resultados de aprendizaje agregados al nivel nacional, y desagregados al nivel departamental, por tipos de escuela (urbano o rural) y tipo de gestión (privada u oficial). Los datos de rendimiento de cada subgrupo de estudiantes se reportan en forma de porcentaje de respuestas correctas para cada uno de los niveles de desempeño predefinidos (B, C, y D – desde el más básico al más complejo). En cuanto a los Exámenes de Estado, los informes principales son los que llegan a los centros educativos. Todas las escuelas medias reciben un reporte

institucional donde se detalla exhaustivamente cuál fue el nivel de logro de cada uno de sus estudiantes en las diferentes áreas del currículo, y cuáles objetivos específicos de cada área muestran mayores avances y dificultades. Adicionalmente, los estudiantes reciben un reporte individual que les permite conocer en detalle cuáles han sido sus logros y dificultades más significativos. Los informes de difusión general reportan los porcentajes de alumnos que lograron cada uno de los niveles de desempeño en las diferentes áreas evaluadas. También se reportan datos desagregados por departamento geográfico, comparando el desempeño relativo de cada uno respecto del promedio nacional.

Uso y difusión de resultados

Los informes que se produjeron sobre los resultados de las pruebas SABER se difundieron entre las autoridades políticas y técnicas del sector, y también fueron presentados a diversos actores locales mediante folletos explicativos, foros, talleres de interpretación de resultados y otros eventos públicos. Los datos de contexto levantados por los cuestionarios han sido analizados sobre todo por organizaciones e investigadores independientes. Los resultados de los Exámenes de Estado, por su parte, tienen como destinatario principal las escuelas medias y sus alumnos. Dado el grado de desagregación de los datos que reciben las escuelas, especialmente en cuanto a los conocimientos específicos demostrados por los estudiantes, los informes constituyen herramientas clave para el mejoramiento en el diseño y la implementación curricular. Para los alumnos, los datos devueltos son indicadores precisos sobre el tipo de conocimientos que han logrado, y sobre cuáles deben trabajar en mayor profundidad si necesitan mejorar los resultados y sus posibilidades de ingreso a la educación superior. No hay evidencia de que los datos hayan impactado en las decisiones de política en el ámbito de la educación básica, aunque sí se utilizan por lo común como materia de estudio por parte de investigadores del ICFES y de otras instituciones académicas.

Sistemas subnacionales de evaluación

El sistema de evaluación subnacional que mayor desarrollo ha logrado es el diseñado e implementado por la Secretaría de Educación de Bogotá (ver informe aparte).

Pruebas internacionales de aprendizaje	Colombia ha participado en las pruebas internacionales comparadas TIMSS (1995); OREALC (1997); IEA-Civic Education (2000); y PIRLS (2001).
Avances observados	La larga experiencia en medición de aprendizajes acumulada por el ICFES ha permitido el desarrollo de importantes capacidades técnicas y operativas a nivel nacional, aunque esas capacidades no siempre se transfieren al Ministerio de Educación ni son utilizadas por este para orientar más racionalmente las decisiones de política educativa. Cabe destacar los esfuerzos por desarrollar modelos psicométricos e instrumentos referidos a criterios, lo cual garantiza mayor articulación curricular entre la educación media y los requisitos de ingreso a la universidad. El ICFES goza de un grado importante de institucionalización interna y externa, lo cual se debe tanto al respaldo normativo que lo regula como a la voluntad de los equipos técnicos por actualizar los instrumentos y hacerlos más útiles como herramientas de gestión escolar.
Dificultades observadas	Si bien se reconoce oficialmente el valor de los datos recogidos por las pruebas SABER y los Exámenes de Estado, los resultados han tenido muy bajo impacto en la toma de decisiones de política educativa. Esto se debe principalmente a la alta rotación de autoridades políticas del sector, lo cual ha impedido que se mantengan líneas de acción a largo plazo, y que se utilice los datos de las evaluaciones para sustentarlas.
Proyección a futuro	La administración ministerial actual ha hecho explícita su voluntad de seguir ampliando y socializando los estándares elaborados en 2002, y de reforzar los mecanismos de medición y evaluación de aprendizajes en función de esos estándares.

COSTA RICA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
COSTA RICA IIMEC,	1986	3, 6, 9, 11/ 12	L, M	CN	G, P		
Universidad de Costa Rica y Departamen- to de Pruebas Nacionales (Ministerio de Educación)	1987 1989-1990 1996 1997	6, 9, 11/12 3,6,9 3, 6, 9 3, 6	L, M, CS, CN L, M, CS, CN L, M, CS, CN	CN MN MN MN	G, P G, U, P		OREALC 1997 (No concluyó proceso de evaluación)
Pruebas Nacionales de Bachillerato o de Conclusión del II Ciclo de EGB	1988-2003 [7]	Bachillerato	L, M, CS, CN, LE	CN	G, U, P		
Primer ingreso	1994-1996	1					
Destrezas cognoscitivas	1996	1					

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)], U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia], P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

[7]
Existieron en décadas anteriores

Marco institucional

Las primeras acciones de evaluación estandarizada en Costa Rica comenzaron en 1986 en manos del IIMEC, Instituto de Investigación para el Mejoramiento de la Educación Costarricense, en la Universidad de Costa Rica, y por solicitud del Ministerio de Educación para realizar una evaluación formativa de la educación básica. Luego, a partir del año 1988 y por encargo del Consejo Superior de Educación, el IIMEC diseñó e implementó las pruebas de graduación del Bachillerato. En 1996 venció el convenio entre el Ministerio de Educación y el IIMEC y en 1997 se estableció el CENE-EDU, Centro Nacional de Evaluación de la Educación, dependiente del Consejo Superior de Educación. El CENE-EDU fue disuelto poco tiempo después, en 1998, y actualmente la medición nacional de aprendizajes se administra desde la División de Control de Calidad y Macroevaluación del Sistema Educativo (Departamento de Pruebas Nacionales), órgano funcional del Ministerio de Educación que aplica pruebas tanto para el sistema de educación formal como para el Subsistema de Educación Abierta (Programa de Alfabetización por Suficiencia, Bachillerato a Distancia y Bachillerato por Madurez). Otra dependencia del Ministerio de Educación, el Departamento de Investigación Educativa, realiza estudios secundarios sobre la base de los datos de las pruebas nacionales, aunque su relación con la División de Control de Calidad es aparentemente muy débil.

- Currículo y estándares** Las pruebas de conocimiento de todos los niveles, incluido el Bachillerato, están referidas al currículo nacional vigente, sobre el cual se construyó una serie de Temarios de Referencia para la elaboración de las matrices. La elaboración de esos Temarios, sin embargo, fue realizada específicamente por los equipos de evaluación, sin la participación de los equipos de desarrollo curricular del Ministerio. Tampoco se han utilizado como referentes los estándares elaborados por la CECC. Los temarios son aprobados por el Consejo Superior de Educación (órgano de autoridad máxima en el sistema educativo costarricense), que a su vez consulta con curriculistas de las diversas asignaturas antes de su aprobación anual.
- Tipos de pruebas** Las pruebas de destrezas cognoscitivas fueron muestrales y referidas a normas. Las pruebas de Bachillerato, así como las Pruebas Nacionales de Conclusión del II y III ciclo de EGB, también son normativas pero de cobertura universal en esos niveles de escolaridad. Algunas aplicaciones de estas últimas (1996-1997) se realizaron muestralmente y mediante pruebas referidas a criterios. Todas las pruebas censales sumativas (promoción y graduación) constituyen el 60% de la calificación final del alumno en las asignaturas evaluadas. En todos los casos, las pruebas se basan en un modelo de respuestas de opción múltiple únicamente, excepto en la prueba de Bachillerato en la cual se evalúa también la redacción de ensayos. Se intentó la evaluación de producción escrita en las demás pruebas, para ser corregida por profesores designados en cada escuela, pero investigaciones del IIMEC demostraron luego que las correcciones eran inapropiadas y la experiencia fue discontinuada a partir del año 1991.
- Tipos de reporte** Algunos informes reportan los resultados nacionales y regionales en forma criterial, es decir por porcentajes de alumnos que respondieron correctamente a los ítems de los diferentes objetivos de cada área curricular. Tal es el caso del informe comparativo 1996-1997, que buscaba lograr una medición de impacto de las medidas correctivas tomadas desde el Ministerio de Educación en esos años. Otros informes, como el referente a un estudio muestral de 1998, presentan las medias aritméticas de logro por objetivo en cada área curricular evaluada a nivel nacional, y

luego desagregadas por género, región, tipos de escuela (tamaño), y tipos de administración (público o privada). Se ofrecen también análisis descriptivos de los resultados, descripciones de las principales dificultades que presentaron los alumnos y un conjunto de recomendaciones metodológicas. Los informes oficiales sobre las pruebas de Bachillerato presentan los resultados agregados al nivel nacional, y desagregados al nivel regional por asignatura y modalidad, y en forma de porcentajes de alumnos que promocionaron en cada asignatura. Se presenta también una discusión de los resultados y un conjunto de conclusiones y recomendaciones para el mejoramiento del nivel de aprendizajes. Estudios específicos realizados por la División de Control de Calidad del Ministerio ofrecen también datos sobre correlaciones entre resultados de la prueba y el rendimiento de los alumnos en Educación Diversificada, así como informaciones sobre el valor predictivo de las pruebas de Bachillerato respecto al desempeño de los alumnos graduados en algunas carreras de educación superior.

Uso y difusión de resultados

La División de Control de Calidad produce informes diferenciados de resultados para autoridades superiores del MEP, autoridades regionales y escuelas, incluyendo padres y alumnos. Asimismo, se producen informes técnicos para investigadores y profesionales académicos, así como para la prensa. Además de los informes formales, se realizan con frecuencia estudios de profundización sobre los resultados, tales como los mencionados en el apartado anterior. Las pruebas de Bachillerato y las de Conclusión del II y III ciclo de la EGB son de altas implicancias para los alumnos, representando el 60% de la calificación individual. El otro 40% se deriva del promedio general de los últimos tres años de estudio en cada asignatura. En el caso del bachillerato, ese 40% se deriva del promedio de cinco asignaturas en los dos últimos años de estudio. Los resultados de las pruebas han tenido hasta el momento escaso impacto en las decisiones de política educativa al nivel nacional, con excepción de las modificaciones específicas realizadas sobre los temarios de evaluación o sobre las pruebas mismas.

Sistemas subnacionales de evaluación

No se han desarrollado sistemas subnacionales de evaluación.

- Pruebas internacionales de aprendizaje** Costa Rica participó en la prueba del LLECE- 1997, pero por razones de organización no concluyó el proceso de evaluación y por lo tanto no publicó los resultados finales.
- Avances observados** La larga trayectoria en materia de medición estandarizada de aprendizajes puede considerarse un avance en términos de la capacidad técnica desarrollada a lo largo de los años. En general, y a pesar de que persiste el debate sobre la utilidad de las pruebas nacionales, la opinión pública parece favorecer el sostenimiento de los operativos anuales de medición.
- Dificultades observadas** Los datos recogidos por las pruebas anuales no han sido utilizados mayormente para la toma de decisiones de política o para el diseño y mejoramiento de programas orientados a elevar la calidad de los aprendizajes. El uso de los datos parece estar restringido casi exclusivamente a las decisiones sobre promoción y graduación de los estudiantes, lo cual cuestiona el valor evaluativo-formativo potencial de las pruebas.
-

CUBA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
CUBA	1975	3, 4, 6, 9, 12	L, M	MR			OREALC 1997
	1996-1998, 2000	6, 9, 12	L, M	MN	G, U	CD	
	2002	6, 9, 12	L, M	CN			

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)], U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia], P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

La medición y evaluación de aprendizajes en Cuba se realiza a través del SECE, Sistema de Evaluación de la Calidad de la Educación, que es una unidad técnica del Instituto Central de Ciencias Pedagógicas, organismo adscrito al Ministerio de Educación. El equipo de evaluación está integrado por investigadores del Instituto y representantes electos de las direcciones de Supervisión, de Educación Primaria, de Educación Básica Secunda-

ria, de Educación Preuniversitaria, y de la Dirección de Formación y Capacitación Docente (20 personas en total). Este grupo, a su vez, se apoya logísticamente en grupos provinciales de evaluación, y en tareas de coordinación que realiza la Universidad Pedagógica Nacional. Las acciones evaluativas del SECE son de carácter en especial investigativo, a diferencia de otros programas anteriores de evaluación nacional que tenían funciones de acreditación (promoción y graduación) de estudiantes de educación básica y preuniversitaria.

Currículo y estándares En Cuba existe un currículo nacional para todos los niveles de escolaridad. Este currículo sufrió modificaciones importantes en los años 1975 y luego en 1987, aunque no se considera que hayan atravesado “reformas” propiamente dichas. Más bien, las actualizaciones curriculares se van realizando periódicamente a partir de las informaciones recogidas a través del sistema nacional de asesoramiento y supervisión, que en Cuba está fuertemente afianzado. No se cuenta con estándares nacionales de currículo, y no existen por el momento iniciativas para su desarrollo en el corto o mediano plazo.

Tipos de pruebas Las pruebas de aprendizaje que viene aplicando el SECE desde 1996 son referidas a normas y se basan en un modelo de ítems de opción múltiple y también de preguntas abiertas. La cobertura es censal al nivel de las escuelas y muestral al interior de cada institución. Esto implica que los resultados son representativos al nivel municipal, provincial y nacional, pero no al nivel escolar debido al tamaño de la muestra en cada centro educativo. Los instrumentos, además de medir el desempeño estudiantil en lenguaje y matemáticas, levantan información sobre aspectos axiológicos y actitudinales de los estudiantes. A través de cuestionarios de contexto aplicados a docentes, alumnos y familias, se estudia también el efecto sobre los aprendizajes de algunas variables extraescolares, tales como el nivel sociocultural de los alumnos, género y grupo racial de pertenencia. Se realizan también comparaciones entre el desempeño académico de los alumnos en las pruebas estandarizadas, y el desempeño alcanzado en las mismas áreas de conoci-

	miento según la calificación que realizan los docentes en las escuelas.
Tipos de reporte	Información no disponible.
Uso y difusión de resultados	Los resultados de las pruebas han sido utilizados principalmente en eventos masivos de capacitación docente, vía televisión nacional, y a través del asesoramiento continuo que reciben las escuelas por parte de equipos técnicos locales y provinciales. A los Consejos de Dirección de Educación de las provincias se les entrega un informe con los resultados y una guía explicativa para su interpretación. Los datos se utilizan también para estudios de casos que profundizan la indagación sobre las razones que explican los resultados insuficientes en las escuelas.
Sistemas subnacionales de evaluación	No se han desarrollado sistemas subnacionales de evaluación.
Pruebas internacionales de aprendizaje	Cuba participó en la prueba internacional comparada del LLECE- 1997.
Avances observados	Las autoridades del Instituto Central de Ciencias Pedagógicas sostienen que existe una política de integración entre las diferentes áreas del Ministerio de Educación, lo cual ha permitido que se difundan los datos de las evaluaciones de rendimiento académico, y que se comprenda su importancia para el diseño de políticas educativas más eficaces. Además, se considera que el uso extensivo que se ha dado a los datos para la capacitación docente ha contribuido positivamente a la legitimación del sistema de evaluación estandarizada que se realiza desde el Instituto.
Dificultades observadas	Se considera que es necesario superar las diferencias de percepción que sobre la evaluación, y las formas de realizarla, tienen los docentes y los investigadores-evaluadores participantes del SECE. Es decir, a pesar de que el sistema de evaluación externa ha ganado legitimidad, tal como se explicó en el apartado anterior, los docentes no siempre perciben que los resultados de las pruebas reflejan fielmente el desempeño de

sus estudiantes. Respecto al rol de la evaluación para el diseño de políticas, y pese a que los resultados se entregan formalmente a las Comisiones Centrales de Asignatura (comisiones de especialistas encargados de revisar el currículo), se considera que aun es necesario reforzar el uso de los datos para el mejoramiento curricular.

Proyección a futuro

Se espera que los datos de las evaluaciones tengan mayor impacto sobre los procesos de desarrollo y actualización curricular. Asimismo, se espera que el sistema de evaluación externa gane mayores grados de legitimidad entre los docentes.

ECUADOR

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
ECUADOR APRENDO- Sistema Nacional de Medición de Logros Académicos 1994-2000	1996-2000	3, 7, 10	L, M	MN	G, U, P		

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

El Sistema Nacional de Medición de Logros Académicos “APRENDO” se gestó y fue implementado entre los años 1994 y 2000 dentro del marco del Proyecto EB/ PRODEC (Proyecto de Desarrollo, Eficiencia y Calidad de la Educación Básica), financiado por el Banco Mundial y con recursos del Ministerio de Educación.

APRENDO no perteneció a la estructura orgánica del Ministerio, y sus actividades cesaron una vez concluido el convenio financiero con el Banco Mundial. Todos los operativos de evaluación fueron diseñados e implementados con el asesoramiento técnico de la Universidad Católica. Las responsabilidades por la medición y evaluación de aprendizajes fueron transferidas a los equipos técnicos del Ministerio de Educación en el año 2001, aunque hasta el momento no se ha desarrollado un sistema de evaluación que reemplace o continúe con las acciones de APRENDO.

- Currículo y estándares** Ecuador llevó a cabo una reforma curricular y aprobó los nuevos marcos curriculares nacionales en 1996. El sistema de evaluación APRENDO comenzó la tarea de definición de “destrezas” cognitivas para su evaluación con anterioridad a la aprobación del nuevo currículo oficial.
- Tipos de pruebas** Las pruebas de rendimiento aplicadas por APRENDO hasta el año 2000 fueron muestrales y diseñadas con referencia a criterios. Cada alumno respondía cuatro preguntas por destreza, todas las cuales se presentaban en formato de opción múltiple. Las pruebas fueron acompañadas por cuestionarios de contexto escolar y extraescolar a alumnos, padres, docentes y directores. No se dispone de información sobre las variables específicas estudiadas.
- Tipos de reporte** Los informes globales de resultados muestran, en forma de gráficos, los porcentajes de alumnos que lograron cada una de las destrezas evaluadas y, más detalladamente, cada uno de los niveles de dominio logrado por destreza. Esos niveles se clasifican en *inicio* (0 ó 1 respuestas correctas por destreza), *avance* (2 respuestas correctas), y *dominio* (3 ó 4 respuestas correctas). Los valores se presentan agregados al nivel nacional, y desagregados por región geográfica (sierra o costa, y agregados provinciales dentro de cada región), tipo de escuela (fiscal o privada), y ámbito geográfico (urbano o rural). Se ofrece también una descripción narrativa de los resultados de diferentes subpoblaciones en forma comparada. Se presenta finalmente un conjunto de con-

	<p>clusiones generales y las implicancias de los resultados para el diseño de políticas de mejoramiento de la calidad de los aprendizajes. Los informes oficiales no presentan resultados de las encuestas de contexto, ni análisis de factores asociados al rendimiento académico.</p>
<p>Uso y difusión de resultados</p>	<p>Los informes generales sobre resultados fueron difundidos mediante medios masivos de comunicación, y se distribuyeron a universidades, sindicatos e institutos pedagógicos. También se realizaron presentaciones públicas, las cuales fueron comentadas posteriormente por medios de prensa local. El equipo de APRENDO elaboró y presentó al Ministerio de Educación una propuesta de capacitación docente a partir de los resultados y de acuerdo a los contenidos de la reforma curricular, pero el proyecto no logró concretarse. También se propuso a las facultades de educación de diferentes universidades un cambio curricular de formación docente basado en los resultados de APRENDO, pero esto tampoco fue llevado a cabo. Los informes de resultados también fueron entregados a las escuelas participantes de la muestra, aunque de desconoce qué tipo de uso se dio a los datos en ese ámbito. Solo se dispone de información sobre escuelas privadas con buenos resultados que utilizaron los datos como publicidad institucional. Hasta el momento en que cesaron los operativos de APRENDO, los resultados no habían sido utilizados para informar decisiones de política educativa al nivel nacional.</p>
<p>Sistemas subnacionales de evaluación</p>	<p>No se han desarrollado sistemas subnacionales de medición de aprendizajes.</p>
<p>Avances observados</p>	<p>Ecuador no ha participado en pruebas internacionales comparadas de aprendizaje.</p>
<p>Pruebas internacionales de aprendizaje</p>	<p>El proyecto APRENDO contribuyó significativamente al desarrollo de capacidades locales en medición de aprendizajes y proporcionó, mientras se mantuvieron los operativos, datos sobre rendimiento académico que no habían sido estudiados anteriormente en Ecuador.</p>

Dificultades observadas El hecho de que el sistema de evaluación dependiera, no de la estructura orgánica del Ministerio de Educación sino de un proyecto especial de cooperación internacional, significó que las capacidades técnicas desarrolladas y los datos recogidos de las evaluaciones nunca fueran adecuadamente transferidas a los equipos de planta ministeriales.

Proyección a futuro En un documento de noviembre de 2002 titulado “Contrato Social por la Educación en el Ecuador”, y publicado electrónicamente por el Ministerio de Educación y Culturas, se anuncia la intención de crear un sistema integral de evaluación que abarque no solamente los aprendizajes de los alumnos sino todas las instancias de gestión y práctica educativa en el sistema. Hasta el momento no se dispone de informaciones sobre programas diseñados o implementados para llevar a cabo esas acciones.

EL SALVADOR

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
EL SALVADOR Ministerio de Educación	1993	Ed. Parvularia (6 años), 1, 2					
SINEA - Sistema Nacional de Evaluación de los Aprendizajes SABA	1993	3, 6, 9, 2M	L, M, CS, CN, ES	MN	G	DC, CG	
	1994	3, 4, 5, 6					
	1995	3, 5					
	1996	3, 4, 5					
	1998	3, 6, 9, 2M	L, M, CS, CN, ES	MN	G	DC, CG	
	2001	3, 6, 9	L, M, CS, CN, ES	MN	G, U		
PAES- Prueba de Aprendizaje y Aptitudes para Egresados de Educación Media	1997-2004	2M, 3M (Técnica)	L, M, CS, CN	CN	U	DC, SES	

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos, SES=Selección educación superior.

Marco institucional

El proyecto SABE de evaluación de aprendizajes en educación básica depende de la Dirección Nacional de Monitoreo y Evaluación del Ministerio de Educación. Los objetivos de evaluación del proyecto han ido mutando a lo largo de la última década, dependiendo de los cambios en la administración sectorial, de los requisitos de monitoreo de las agencias internacionales de financiamiento (AID y Banco Mundial, principalmente), y de programas de evaluación contratados para organizaciones privadas (IN-CRE, Intercultural Center for Research in Education). Por su parte, las pruebas para egresantes de la educación media (PAES), si bien tienen carácter oficial y están normadas por la Ley General de Educación, son diseñadas y administradas descentralizadamente por la Universidad Centro Americana de El Salvador.

Currículo y estándares

Los marcos curriculares nacionales (educación parvularia y grados 1° a 9°) fueron aprobados inicialmente en el año 1991, pero sufrieron modificaciones hasta el año 1997. Adicionalmente, El Salvador participó del proyecto de elaboración de estándares promovido por la CECC, para lo cual se realizó una amplia consulta, a modo de validación, con representantes del medio docente, de la sociedad civil, de diferentes cuadros técnicos del sector educativo y del sector productivo. Los procesos de implementación de esos referentes curriculares se vieron interrumpidos por causa de los terremotos sufridos en el último quinquenio, y por lo tanto las escuelas no han tenido acceso universal a los documentos prescriptivos. Existe cierto grado de consenso sobre la necesidad de resolver este vacío normativo pero aún no se logran acuerdos firmes sobre cómo llevarlo a la práctica. No se descarta una nueva ronda de validación de ambos referentes curriculares. El proyecto SABE de evaluación se desarrolló en forma paralela a la elaboración de los marcos curriculares nacionales. Por tanto, las

matrices de referencia fueron incorporando gradualmente los contenidos prescritos por esos marcos, con cierto grado de participación de los equipos de desarrollo curricular en el proceso de validación de ítemes. Para las pruebas PAES, sin embargo, no se han utilizado los contenidos curriculares ni los estándares centroamericanos. Estas pruebas son construidas sobre la base de un conjunto de habilidades cognitivas universales (Taxonomía de Bloom) pero no están validadas curricularmente desde los marcos prescriptivos vigentes.

Tipos de pruebas

Las pruebas de SABE son muestrales y diseñadas sobre un modelo de preguntas de opción múltiple. Adicionalmente, se aplican pruebas de producción de texto escrito y oral a una submuestra de alumnos. Los instrumentos están referidos a criterios de evaluación consistentes en 10 objetivos de dominio básico que incluyen contenidos específicos para cada área curricular evaluada. Recientemente, el proyecto SABE ha experimentado con la metodología de valor agregado, con aplicaciones diferenciales en tres municipios correspondientes a las tres regiones geográficas principales del país. Sin embargo, las aplicaciones de seguimiento de cohorte fueron luego discontinuadas y no se llegó a concretar el análisis longitudinal. También se han realizado aplicaciones a submuestras de centros educativos pertenecientes a programas especiales, tales como el de “Escuela 10”. Las aplicaciones de SABE se acompañan con cuestionarios de contexto que son respondidos por docentes, directores de escuela y padres de familia. Se relevan datos sobre variables personales, escolares y de contexto socio-económico. Originalmente las pruebas y los cuestionarios eran administrados por equipos diferentes, lo cual dificultaba la tarea de análisis de factores asociados al rendimiento, pero actualmente se está coordinando más racionalmente la aplicación y análisis de los datos derivados de ambos instrumentos. La prueba PAES de egreso de bachillerato es censal y, aunque fue concebida originalmente como una prueba referida a normas, actualmente es elaborada con referencia a criterios.

Tipos de reporte

Los informes de SABE presentan los resultados en forma de promedios porcentuales de los objetivos alcanza-

dos por los estudiantes a nivel nacional, y luego desagregados por departamento geográfico, zona, género y tipo de escuela. También se reporta los porcentajes de respuestas para cada una de las opciones de los ítemes, así como el porcentaje de alumnos que alcanzó los diferentes objetivos evaluados en cada área. La prueba PAES de egreso de bachillerato devuelve los resultados a los estudiantes individualmente. En el informe cada estudiante obtiene información sobre las competencias evaluadas en cada asignatura, el nivel de logro alcanzado en dichas competencias, el promedio global de rendimiento en cada una de las áreas, y una explicación conceptual de los niveles de logro alcanzados.

Uso y difusión de resultados

En general, los resultados de las pruebas SABE no han tenido impacto en la toma de decisiones de política. Esto se atribuye principalmente a que el sistema de evaluación funcionó durante muchos años en forma aislada de otros equipos técnicos del Ministerio, y tampoco se hicieron campañas de sensibilización y estímulo a la participación de la comunidad educativa en los procesos evaluativos. Consecuentemente, no ha existido una real demanda de información por parte de los usuarios centrales y locales. A partir del año 1998 se introdujeron cambios de organización del sistema para permitir mayor participación de los equipos de currículo en la validación de los instrumentos, así como la participación de los equipos docentes en la aplicación de las pruebas. Los resultados se devuelven a todas las escuelas de la muestra y se realiza en forma de talleres orientados al desarrollo de capacidades de análisis y comparación de los datos. La respuesta de los docentes a las informaciones se considera altamente positiva, pero no se ha podido dar a estas devoluciones la continuidad que se esperaba. Los resultados de la prueba PAES se entregan a los estudiantes individualmente. En el año 1999, además, fueron devueltos a cada colegio indicando el resultado de cada estudiante, el promedio de cada sección (grupo) y de la institución, en forma desagregada por asignatura y por modalidad (bachillerato general y técnico), y en forma comparada con otros departamentos y con el promedio nacional. Los porcentajes de aciertos obtenidos por la institución se presentaron desagregados por cada

una de las áreas temáticas evaluadas, y mostrando los niveles de dominio cognitivo alcanzados: conocimiento, comprensión y análisis. La prueba PAES es obligatoria para los estudiantes egresantes del bachillerato pero no implica consecuencias para su graduación. Sin embargo, los resultados ya están comenzando a utilizarse por las universidades para la selección de postulantes. Junto con los talleres de devolución, se realizó una encuesta de cobertura nacional donde directores y docentes expresaban sus inquietudes respecto a las causas que pueden explicar los bajos resultados: condiciones escolares, características de los alumnos y su entorno, características de la prueba misma, adecuación de las políticas de apoyo (capacitación docente, materiales), currículo, etc. Adicionalmente, se recogieron las propuestas de mejoramiento expresadas por los equipos docentes para cada uno de esos factores. Los resultados de la encuesta fueron sistematizados y publicados en forma de libro por el Ministerio de Educación.

Sistemas subnacionales de evaluación	No se han desarrollado sistemas subnacionales.
Pruebas internacionales de aprendizaje	No se ha participado en este tipo de pruebas.
Avances observados	Se han dado pasos importantes para reforzar la validez curricular de las evaluaciones y por comenzar a difundir más ampliamente los resultados dentro del Ministerio de Educación y entre las escuelas evaluadas.
Dificultades observadas	Hasta el momento, no se conocen medidas de política ni repercusiones en la opinión pública derivadas de los resultados de las evaluaciones.
Proyección a futuro	Se espera cambiar la estrategia de capacitación mediante el uso de los resultados de las pruebas SABE. Una de las estrategias pensadas es la del uso de los datos para capacitar a asesores pedagógicos, quienes a su vez ofrecerían asesoramiento técnico a las escuelas que lo solicitaran.

GUATEMALA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinarias [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
GUATEMALA SINMELA- Sistema Nacional de Medición del Logro Académico/	1992-1996	3		ME			
PRONERE- Programa Nacional de Evaluación del Rendimiento Escolar	1998-2000	3, 6	L, M	MN [7]	G, U, P	G	
PRONERE + Juárez Asociados	2004	1, 3		MN (Escuelas públicas, únicamente)			
DIGEBI	2003			MN		DC	
Ministerio de Educación/ Universidad de San Carlos	2004	Último año de Secundaria		MN			

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales,
ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales,
LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

[7]
La muestra incluye una sobrerrepresentación intencional de escuelas de programas especiales

Marco institucional

Las primeras evaluaciones estandarizadas de aprendizajes se realizaron a principios de la década pasada en el marco del Proyecto BEST para el Fortalecimiento de la Educación Básica (USAID), bajo la coordinación del Centro Nacional de Pruebas de Rendimiento Escolar. Esas evaluaciones, más tarde discontinuadas, se utilizaron para evaluar la implementación del Proyecto en las escuelas participantes. En 1997, como parte del Proyecto de Reforma Educativa y con apoyo financiero del Banco Mundial se creó el SINMELA, luego denominado PRONERE, en convenio entre el Ministerio de Educación y la Universidad del Valle de Guatemala. La Universidad diseñó y ejecutó los operativos de evaluación hasta el 2001, año en que caducó el convenio financiero con el Banco Mundial. El PRONERE fue desmantelado y no ha sido reemplazado con otros programas de evaluación.

Currículo y estándares

Guatemala no cuenta con un currículo nacional unificado, ni se han avanzado propuestas de estándares en los últimos años. Las pruebas del PRONERE se elaboraron originalmente utilizando un banco de ítems facilitado por una institución educativa privada, y más adelante se construyó una matriz de referencia sobre la base de observaciones de clase y reportes de docentes sobre cobertura curricular en aula. Sin embargo, no se construyó una matriz de referencia que reflejara acuerdos

nacionales sobre prioridades de aprendizaje para la evaluación.

Tipos de pruebas

Las pruebas son muestrales, representativas al nivel nacional, fueron construidas con referencia a normas y contienen 40 preguntas en cada área evaluada. Los instrumentos fueron elaborados y aplicados en español y en cuatro lenguas originarias mayoritarias. En muchas escuelas bilingües, sin embargo, no se pudo aplicar pruebas pues los estudiantes abandonan antes de llegar el 3^{er} grado de escolaridad. Las pruebas se acompañan con cuestionarios de contexto a estudiantes, maestros y directores de centros educativos. Las pruebas de DIGEBI para escuelas de educación bilingüe, son referidas a criterios

Tipos de reporte

Los informes elaborados por PRONERE consisten en descripciones narrativas y comparativas entre subpoblaciones de alumnos respecto al porcentaje de respuestas correctas que lograron en promedio para cada área evaluada. Los datos se presentan agregados al nivel nacional, y desagregados al nivel departamental, por ámbito (rural, urbano) y por grupo lingüístico.

Uso y difusión de resultados

Con la información obtenida se realizan resúmenes ejecutivos y presentaciones gráficas para autoridades nacionales y regionales, y también para universidades formadoras de docentes. Las descripciones narrativas y las recomendaciones pedagógicas presentadas se basan en una selección de los ítems de la prueba que demostraron ser más difíciles de responder. Se realiza una descripción de los centros educativos evaluados y del nivel socioeconómico de las familias de la muestra, y se presentan algunos de los factores de contexto que muestran mayor incidencia en el rendimiento académico de los alumnos. Estos datos se han utilizado para algunos eventos de capacitación docente, especialmente de docentes bilingües por medio del PRONEBI, Programa Nacional de Educación Bilingüe. En general, los resultados han tenido escaso impacto en la opinión pública y en la toma de decisiones de política educativa. Los resultados de la prueba del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación de UNES-

	CO OREALC, aplicada por el Banco Mundial, no han sido difundidos hasta la fecha.
Sistemas subnacionales de evaluación	No se han desarrollado sistemas subnacionales de evaluación.
Pruebas internacionales de aprendizaje	En el año 2002, el Banco Mundial aplicó las pruebas del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación de UNESCO OREALC a una muestra representativa de estudiantes de 3º a 4º grado de primaria.
Avances observados	La experiencia del PRONERE permitió desarrollar capacidades técnicas locales para la medición de aprendizajes, lo cual constituye un avance en la medida en que esas capacidades sean utilizadas para nuevos emprendimientos de evaluación. Se observa también cierto grado de movilización por parte de algunos sectores de la sociedad civil en demanda por la continuidad y uso de las evaluaciones.
Dificultades observadas	La cesación de actividades de evaluación al término del convenio con el Banco Mundial, y la ausencia de voluntad del gobierno por mantener el sistema en funcionamiento, dan cuenta de los escasos niveles de institucionalización que logró el PRONERE a pesar de sus esfuerzos por continuar las aplicaciones durante cuatro años consecutivos. Una mayor divulgación de los datos, así como mayores grados de participación de actores sectoriales clave parecieran ser imprescindibles para recuperar y mantener el sistema de evaluación.
Proyección a futuro	No existen actualmente propuestas desde el Ministerio de Educación para la reactivación del PRONERE u otra agencia similar.

HONDURAS

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
HONDURAS Secretaría de Educación. Componente V del Proyecto PEEP (Eficiencia de la Educación Primaria)	1990-1994	1, 2, 3, 4, 5	L, M, CN, CS	ME (10 municipios)	Secretaría de Educación y USAID	Usado para evaluar proyectos apoyados por USAID	OREALC 1997
UMCE	1997 1998 1999 2000 2002-2004	3, 6 2, 3, 4, 6 2, 3, 4, 5 3, 6 3,6	L, M L, M L, M L, M L, M, CN	MN, CN MN MN MN MN	G, U G, U G, U G, U G, U	CD, FA	

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional	<p>El sistema de evaluación actual tiene origen en un acuerdo entre la Secretaría de Educación Pública (SEP) y el Banco Mundial para el monitoreo del PROMEB, Programa de Mejoramiento de la Educación Básica. Anteriormente (1990-1994) existieron esfuerzos por realizar evaluaciones estandarizadas en el marco del Programa de Eficiencia de la Educación Primaria (PEEP), financiado por USAID. Sin embargo, esas evaluaciones nunca llegaron a sistematizarse ni institucionalizarse. En 1995 se crea la UMCE, Unidad de Medición de la Calidad Educativa, administrada por la Universidad Pedagógica Nacional Francisco Morazán. Esta Unidad cuenta con altos grados de autonomía administrativa y operativa pese a que, formalmente, reporta directamente a la Dirección General de Evaluación de la Calidad de la SEP. Los operativos nacionales se organizan mediante una Red Desconcentrada de Evaluación, que funciona en las 18 direcciones departamentales, y que participan directamente en el proceso de aplicación de las pruebas y recolección de datos.</p>
Currículo y estándares	<p>Hasta 2003 no se contaba en Honduras con un marco curricular nacional, sino con los Rendimientos Básicos e Indicadores de Evaluación del Nivel Primario, que fueron elaborados ad hoc para el PEEP, pero no todas las escuelas tuvieron acceso a este documento. Dado que este referente curricular no era apropiado para el diseño de una matriz de referencia de evaluación, la UMCE desarrolló un conjunto de objetivos e indicadores de desempeño para las diferentes áreas disciplinares evaluadas. En el año 2003, y luego de varios años de debate sobre diferentes propuestas curriculares, fue aprobado oficialmente un marco curricular nacional que incorpora, entre otros referentes, los contenidos y objetivos que la UMCE venía evaluando desde 1997. Queda pendiente la culminación del proyecto de estándares de la CECC, aunque no es claro aun cómo ambos referentes curriculares podrán ser integrados y posteriormente evaluados.</p>
Tipos de pruebas	<p>Las pruebas aplicadas por la UMCE son muestrales a nivel nacional, referidas a criterios y basadas princi-</p>

palmente en un modelo de respuesta de opción múltiple. En el área de lenguaje, además, se evalúa redacción escrita y lectura oral. Las pruebas son complementadas con cuestionarios a alumnos, padres de familia, docentes y directores de centro para el estudio posterior de factores asociados al rendimiento. Se estudian variables de NES y otras variables intraescolares (incluyendo un modelo de medición de eficacia escolar). Los equipos técnicos han encontrado dificultades para establecer correlaciones confiables entre esas variables y el rendimiento académico, aunque la incorporación de un modelo de análisis jerárquico lineal (HLM) a partir de 2002 ha permitido aumentar la confiabilidad y validez de las correlaciones. Además del programa regular de evaluación en los grados y áreas mencionadas, la UMCE presta sus servicios para la medición de rendimiento académico como indicador de impacto de algunos programas específicos financiados por agencias de cooperación internacional.

Tipos de reporte

La UMCE produjo una serie de documentos de resultados entre los años 1998 y 2001. Tales informes presentan los datos de rendimiento en forma de porcentajes del alumnado a nivel nacional que lograron los objetivos de aprendizaje evaluados. Esos objetivos se organizan por dominios cognitivos, áreas de dominio y competencias específicas. Los datos se presentan también desagregados por departamento, por tipos de escuela, por tipos de gestión (pública o privada) y por situación geográfica. Se ofrece un análisis descriptivo de los resultados nacionales y por departamento, un análisis de ítemes seleccionados, y un conjunto de recomendaciones pedagógicas para docentes y técnicos locales. La UMCE realiza también informes sobre estudios longitudinales, análisis multivariados, y análisis comparativos de resultados entre grados por área de aprendizaje, así como de factores asociados al rendimiento.

Uso y difusión de resultados

Los informes oficiales son elevados a las autoridades políticas de la SEP y a autoridades educativas regionales. Se espera llegar también a las escuelas, a los padres de familia y al público en general, aunque di-

versos obstáculos, operativos y presupuestarios, han limitado este tipo de difusión. Sin embargo, se reporta que alrededor de 2.000 maestros, directores y técnicos locales han tenido acceso a los datos desde 1998. Esos datos llegan, en muchos casos, de la mano de los equipos técnicos encargados de las capacitaciones docentes. Actualmente se está realizando una campaña de divulgación de los resultados de años anteriores con la asistencia de organizaciones internacionales especializadas. Además, a partir de 2003, los resultados de las evaluaciones se encuentran disponibles en la página web de la UMCE.

Sistemas subnacionales de evaluación	No se han desarrollado sistemas subnacionales de evaluación.
Pruebas internacionales de aprendizaje	Honduras participó en el estudio comparado del LLECE-1997.
Avances observados	La UMCE ha logrado un alto grado de consolidación interna, en términos de constitución y capacitación de equipos técnico-profesionales, y de continuidad en los operativos de evaluación a través de los años. Asimismo, vale destacar los esfuerzos por desarrollar un modelo psicométrico de alto valor pedagógico, especialmente ante la ausencia de referentes curriculares sólidos y actualizados en el ámbito nacional.
Dificultades observadas	Destaca principalmente la escasa comunicación y débiles acuerdos entre la UMCE y la SEP sobre las metas y usos de la evaluación, lo cual resulta en la subutilización de los datos para la toma de decisiones de política. Desde el punto de vista técnico, las limitaciones presupuestarias no han permitido expandir la muestra y así favorecer análisis jerárquicos multivariados sobre los resultados de rendimiento y las variables que lo afectan. Esas dificultades, sin embargo, comenzaron a ser subsanadas a partir del año 2002 con la introducción del modelo de análisis jerárquico lineal.
Proyección a futuro	Se espera, principalmente, lograr un mayor impacto de los resultados en la opinión pública, en la toma de

de decisiones de política educativa y en el uso pedagógico y curricular de los datos en el nivel escolar. También se proyecta la creación de un sistema nacional de evaluación que integre las informaciones de la evaluación externa realizada por la UMCE, con las de la evaluación interna realizada por la misma Secretaría de Educación. Se espera, por lo tanto, que el contrato de asociación entre Secretaría de Educación y UMCE se extienda más allá del plazo actualmente vigente (2005).

MÉXICO

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
MÉXICO SEP y SNEE Sistema Nacional de Evaluación Educativa Evaluación de la Educación Primaria (EVEP)	1996-2000	1, 2, 3, 4, 5, 6	L, M, CS, CN	MN	G, U	FA	TIMSS 1995 OREALC 1997; PISA 2000, 2003
Evaluación de Estándares Nacionales	1997-2003	2, 3, 4, 5, 6	L, M	MN	G, U		
Evaluación del factor "Aprovechamiento escolar" del programa Carrera Magisterial	1995-2005	3, 4, 5, 6, 1M, 2M, 3M	L, M, CN, CS, LE	CR	G, U	I	
IDANIS (Instrumento de Diagnóstico de Alumnos de Nuevo Ingreso a Secundaria)	1995-2003	6	Habilidades cognitivas	CR	G, U		

[2]
Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos, I=Incentivos Docentes.

Marco institucional

Hasta el año 2002, la evaluación de la educación en México fue responsabilidad de la Dirección General de Evaluación de la Secretaría de Educación Pública (SEP) de la nación. Actualmente, las acciones de evaluación son coordinadas entre esa Dirección y el recientemente creado Instituto Nacional de Evaluación de la Educación (INEE), un órgano gubernamental semiautónomo que, si bien pertenece a la estructura funcional de la SEP, goza de mayores grados de autonomía administrativa y política que otros órganos de gobierno en el sector. La creación del Instituto responde a una necesidad largamente debatida de contar con un sistema de evaluación integral más objetivo y eficiente, y menos vulnerable a las decisiones coyunturales de la administración sectorial. Simultáneamente, las Secretarías de Educación de todos los estados de la Federación cuentan con direcciones de evaluación propias, aunque no todas constituyen “sistemas de evaluación” propiamente dichos. Los estados que han optado por desarrollar un sistema de evaluación propio, lo han hecho siguiendo una de dos modalidades principales: la utilización de las pruebas nacionales mediante muestras estatales ampliadas o adaptadas, o bien el desarrollo de instrumentos propios y sistemas originales de difusión y uso de los resultados. En ambos casos, los estados cuentan con el apoyo técnico de la

Dirección General de Evaluación, y actualmente del Instituto Nacional de Evaluación Educativa.

Currículo y estándares

El currículo nacional vigente para educación primaria y secundaria de México fue aprobado en el año 1993 dentro de un marco general de reforma. La prescripción y actualización curricular es responsabilidad casi exclusiva de la administración federal, con un bajo nivel de injerencia y participación permitido a las administraciones estatales. El currículo fue elaborado cuando ya se contaba con una larga experiencia en medición de aprendizajes y, por lo tanto, se disponía de gran cantidad de información sobre los logros y dificultades de aprendizajes de los estudiantes. Sin embargo, la participación y capacidad de influir en las decisiones curriculares por parte de los equipos de evaluación fue marginal, y sigue siendo débil en la actualidad. Las pruebas nacionales de aprendizaje (excepto el IDANIS y el EXANI, que miden habilidades cognitivas generales) están referidas al currículo nacional, aunque para la elaboración de las matrices de referencia la Dirección General de Evaluación realizó una especificación de contenidos y competencias académicas en forma paralela o complementaria a la prescripción curricular oficial. Ese proyecto de especificación curricular, así como de los instrumentos de medición elaborados sobre esa base, se denominó “estándares nacionales” de educación primaria y de educación secundaria. Estos estándares no pueden considerarse propiamente tales desde una perspectiva de política curricular, aunque han sido útiles para definir más acabadamente las expectativas de aprendizaje que se evalúan anualmente en las áreas de matemáticas y lenguaje.

Tipos de pruebas

Todas las pruebas nacionales diagnósticas son muestrales. Las pruebas que se aplican para el programa de Carrera Magisterial son censales para la población de docentes registrados en el programa. La prueba IDANIS es censal en el Distrito Federal, y en los estados se aplica de acuerdo a la demanda puntual de cada entidad. Todas las pruebas están referidas a normas, aunque los resultados se reportan por niveles de desempeño determinados a posteriori. Las pruebas contienen ítemes de opción de respuesta múltiple y, ocasionalmente, requieren el desa-

rollo de respuestas abiertas. Las pruebas aplicadas en el programa de Carrera Magisterial, que responden los alumnos de los docentes participantes, constituyen un indicador más, entre otros, de desempeño profesional considerados en el esquema de incentivos del programa. También se ha desarrollado un programa de autoevaluación de centros educativos. El programa se basa en una adaptación del modelo oficial escocés de autoevaluación y contiene 33 indicadores sobre gestión escolar, práctica pedagógica e infraestructura. Las escuelas evalúan su propia gestión sobre la base de esos indicadores y elevan los resultados a la Dirección General de Evaluación. Esta, a su vez, procesa los datos y elabora los informes correspondientes. El modelo está siendo utilizado por las 15.000 escuelas que participan del Programa de Escuelas de Calidad. Adicionalmente, se aplican en México exámenes de egreso de educación secundaria y bachillerato para la selección de alumnos en transición hacia ciclos superiores de educación (bachillerato y educación superior, respectivamente.) Estos exámenes se denominan EXANI I y II, y evalúan habilidades cognitivas así como conocimientos específicos en ocho disciplinas diferentes. En cuanto al estudio de factores asociados al rendimiento, se han realizado análisis bi y multivariados considerando variables intraescolares (Ej. conocimientos y experiencia docente) y extraescolares (Ej. nivel de estudios de los padres, asistencia del alumno a educación preescolar). Estos estudios han sido complementados con seguimientos cualitativos sobre los resultados de aprendizaje. Según los últimos lineamientos de política evaluativa difundidos por el INEE, las evaluaciones nacionales de aprendizaje que se apliquen en los próximos años serán censales en algunos grados clave de escolaridad, y serán complementadas con evaluaciones a docentes, directivos y centros educativos en su conjunto.

Tipos de reporte

Se elaboran informes ejecutivos para autoridades centrales donde se presentan los resultados globales desagregados al nivel estadual, por tipo de dependencia geográfica (urbana o rural), y por porcentajes de escuelas que alcanzan cada uno de los niveles de desempeño establecidos (muy bajo, bajo, medio, alto, muy alto). Supervisores locales y escuelas reciben informes orientados al

mejoramiento pedagógico, aunque esto ocurre en forma irregular y no existen formatos estandarizados de reporte para todo el país. En los informes IDANIS se muestran los promedios de rendimiento al nivel nacional y estadual, en forma de gráficos que reportan la media aritmética y desviación estándar para cada caso. Los resultados de las evaluaciones cualitativas y las investigaciones realizadas por la Dirección de Evaluación son devueltos a las escuelas. Los informes describen las características institucionales de las escuelas que muestran mejoras de sus resultados en el tiempo, así como de aquellas cuyo rendimiento ha sido decreciente. También se ofrecen recomendaciones para replicar esas investigaciones en cada centro, a modo de autoevaluación, y se presenta un conjunto de recomendaciones pedagógicas y organizacionales para elevar la calidad de los aprendizajes.

Uso y difusión de resultados

Los datos recogidos por las evaluaciones nacionales comenzaron a hacerse públicos en el año 2000. Anteriormente, solo se conocían al interior de la Secretaría de Educación. En los últimos dos años se han publicado algunos informes oficiales sobre los operativos de evaluación realizados en los últimos años, resúmenes de los resultados a nivel nacional y estadual e informes para las escuelas con los datos recogidos por los instrumentos de autoevaluación escolar y las investigaciones cualitativas llevadas a cabo por la Dirección General de Evaluación (ver apartado anterior). El uso más estable que se ha dado a los datos hasta el momento es el de evaluación para incentivos de Carrera Magisterial, así como de graduación y selección de alumnos mediante las pruebas EXANI I y II de egreso de secundaria y bachillerato. El nuevo Instituto Nacional de Evaluación Educativa ha hecho explícita su intención de introducir gradualmente pruebas de aprendizaje censales en grados clave de la escolaridad básica, las cuales contribuirán a decidir, junto con la evaluación que realizan los docentes, la promoción y graduación de los alumnos. Los nuevos lineamientos de política evaluativa del INEE remarcan la necesidad y el compromiso por reforzar la transparencia técnica de los operativos, la publicidad de los resultados (inclusive a alumnos y familias), así como de elaborar informes diferenciados para distintos tipos de audiencias. En los primeros meses

de trabajo del INEE ya se produjo un primer informe nacional de resultados, que fue presentado al Presidente de la República, a los medios de prensa y a las administraciones estatales, además de ser publicado en la nueva página web del Instituto.

- Sistemas subnacionales de evaluación** Varios estados de la Federación han desarrollado sistemas independientes de medición y evaluación de la calidad educativa. En este trabajo se presenta detalladamente el caso del estado de Aguascalientes.
- Pruebas internacionales de aprendizaje** México ha participado en las pruebas comparativas TIMSS-1995, LLECE-1997, y PISA-2000.
- Avances observados** Destaca la experiencia técnica en medición de aprendizajes desarrollada en las últimas dos décadas, así como la relativa estabilidad de los equipos profesionales a cargo de los operativos. La creación del INEE, y especialmente el largo debate que precedió su constitución y aprobación oficial, da muestras de la importancia que ha cobrado la evaluación estandarizada dentro del ámbito de las políticas de educación pública.
- Dificultades observadas** La dificultad central hasta el momento ha radicado en que las múltiples mediciones realizadas no se traducen en una evaluación sustantiva e integral de la calidad de la educación. La falta de comunicación entre unidades de la SEP, así como los todavía insuficientes mecanismos de divulgación de los resultados, determina que los datos disponibles no se analicen y que no tengan mayor impacto sobre la toma de decisiones de política.
- Proyección a futuro** El desafío principal para el nuevo Instituto de Evaluación es el de lograr que los datos recogidos en los últimos años, y los que se obtengan a partir de 2004, sean adecuadamente analizados y utilizados para el mejoramiento de la calidad educativa. Desde el punto de vista institucional, el INEE deberá consolidar gradualmente su legitimidad dentro del sector, y ante la sociedad, proveyendo informaciones útiles para diferentes actores y asegurando la organicidad de acciones con otras agencias gubernamentales.
-

NICARAGUA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinarias [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
NICARA- GUA	1996	3,4	L, M	ME	U	DC	
	1997	5	L, M	ME	U	DC	
Evaluación del currículo transformado de 1º a 5º de Educación Primaria							
Dirección Evaluación PROYECTO APRENDES- NE- Sistema Nacional de Evaluación de la Educación Básica y Media	1996-1997 2002	4, 3M 3, 6	L, M L, M	ME MN	G, U G, U, P	ED	

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos, ED=Evaluación descentralizada.

Marco institucional	La evaluación estandarizada de aprendizajes se realiza desde la Dirección de Evaluación del Ministerio de Educación, Cultura y Deportes, y dentro del marco del Segundo Proyecto de Educación Básica (APRENDE). Los esfuerzos de evaluación realizados por esa Dirección hasta 1999 fueron relativamente desarticulados y carentes de un objetivo claro respecto al uso potencial de las informaciones recabadas. A partir del año 1999 se crea el SNE, Sistema Nacional de Evaluación de la Educación Básica y Media, que cuenta con el apoyo de APRENDE, BASE (USAID) y SIMEN (UNESCO), y que ha replanteado la estrategia de evaluación y sus fines específicos. El sistema de evaluación se inscribe en un contexto de reformas importantes para la descentralización de la administración de las escuelas públicas, especialmente a partir del Programa de Autonomía Escolar (1992) y de la erogación de la Ley de Participación Escolar (2002).
Currículo y estándares	Información no disponible.
Tipos de pruebas	Las pruebas de aprendizaje son muestrales, referidas a criterios y basadas en un modelo de respuestas de opción múltiple. Se acompañan con cuestionarios de contexto escolar y extraescolar a docentes, directores, alumnos y padres de familia. Adicionalmente, se realizan operativos para la autoevaluación de centros educativos en una pequeña muestra nacional no representativa.
Tipos de reporte	Los resultados de aprendizaje se reportan en forma de porcentajes de alumnos que se ubican en los niveles básico, intermedio y proficiente de la escala de desempeño. Tales porcentajes se presentan en forma agregada al nivel nacional, y luego desagregada por regiones, tipos de dependencia (centros autónomos, no autónomos, y

	<p>privados con y sin subvención estatal), ubicación geográfica (urbana o rural) y modalidad (regular o multi-grado). Adicionalmente, se presentan los resultados de los análisis de factores asociados al rendimiento, indicando cuáles factores escolares y extraescolares muestran el mayor impacto sobre los aprendizajes.</p>
Uso y difusión de resultados	<p>Desde el Ministerio de Educación se producen informes técnicos, resúmenes ejecutivos, cuadros y gráficas, y boletines informativos. Se realizan también talleres, presentaciones y debates para docentes, directores de escuela, técnicos municipales y departamentales, delegados municipales y departamentales y autoridades superiores en el nivel central. Se ha anunciado la publicación de los resultados en la página web del Ministerio de Educación pero aún no se ha concretado.</p>
Sistemas subnacionales de evaluación	<p>No se han desarrollado sistemas subnacionales de evaluación.</p>
Pruebas internacionales de aprendizaje	<p>Nicaragua no ha participado en este tipo de pruebas.</p>
Avances observados	<p>Se observan esfuerzos, al menos en la propuesta formal de la Dirección de Evaluación, por dotar al SNE de mayor significancia y utilidad para la toma de decisiones de política educativa y para el desarrollo curricular y pedagógico escolar.</p>
Dificultades observadas	<p>La difusión de los datos aún no es suficiente para impactar en la opinión pública y para movilizar la demanda de la sociedad civil por mayores niveles de calidad educativa.</p>

PANAMÁ

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
Programa de Evaluación de la OEA	1981-1985			MR	G		
Programa de Pruebas de Diagnóstico	1985-1988 1992	6, 6M 3, 6M	L, M, CS, CN	MN MN	G G		
CECE	1995	1M, 2M, 3M, 4M, 5M, 6M	L, M	MR	G		
SINECE Sistema Nacional de Evaluación de la Calidad de la Educación	1999-2000 2000 2001	3,6,9 3,6,9 3,6,9	L, M, CS, CN L, M, CS, CN L, M	MN	G G G		

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional	Durante las dos últimas décadas se ensayaron diversos proyectos de evaluación, a cargo de diferentes agencias, pero ninguno logró institucionalizarse, ni quedan registros de los datos recabados en esos operativos. El último proyecto de evaluación estandarizada fue llevado a cabo por la Dirección de Evaluación del Ministerio de Educación en los años 2000 y 2001. Sin embargo, se trata de una evaluación piloto con una muestra reducida. En el año 2002, como resultado de un extenso Diálogo Intersectorial, se firmó un decreto presidencial para la creación de una nueva Dirección de Evaluación con funciones renovadas y lineamientos más explícitos para el desarrollo de nuevos operativos y procesos evaluativos integrales.
Currículo y estándares	Las pruebas de los años 2000 y 2001 fueron construidas mientras se actualizaba el currículo nacional. La matriz de referencia fue elaborada por especialistas disciplinares y con escaso grado de participación por parte de los equipos de currículo del Ministerio de Educación. Un grupo de especialistas del Ministerio de Educación y de centros educativos convocados participó en el proceso de elaboración de los estándares promovidos por la CECC, Coordinación Educativa y Cultural Centroamericana. Los estándares de ciencias naturales, ciencias sociales, matemática y español han sido incorporados a los marcos curriculares nacionales de educación primaria pero, según opiniones extragubernamentales, no han sido aún lo suficientemente difundidos ni acompañados con programas concretos de implementación en las escuelas. Los estándares de educación premedia y media para las áreas de español, matemática, física, ciencias naturales, inglés, ciencias sociales biología y química están siendo revisados en la actualidad y próximamente serán utilizados para realizar ajustes curriculares en 7° y 12° grado.
Tipos de pruebas	La prueba piloto aplicada en 6° grado en el año 2000 tenía como objetivo principal relevar las capacidades básicas de los estudiantes en la solución de problemas, comprensión, análisis, reflexión y síntesis. Se elaboraron 18 ítemes de opción múltiple y dos ítemes de pro-

ducción escrita para cada asignatura evaluada. Las pruebas fueron acompañadas por una encuesta sobre el perfil socioeconómico de los alumnos pero las limitaciones de tiempo no permitieron procesar los datos para correlacionarlos con los resultados finales de las pruebas. Con la experiencia adquirida en esa prueba piloto, se diseñó en 2001 la Prueba Piloto Nacional “Comprender” en las asignaturas de español y matemática, para los grados 3°, 6° y 9° de los Centros de Educación Básica General. El propósito de la misma fue medir las habilidades y actitudes verbales y numéricas en los estudiantes. Estas pruebas también fueron elaboradas mediante ítemes de opción múltiple y de producción escrita, con un total de 25 reactivos para cada una de las asignaturas evaluadas.

Tipos de reporte

El informe sobre la experiencia del año 2000 consiste en una descripción del porcentaje de alumnos que respondió correctamente a cada destreza evaluada, en forma agregada a nivel nacional y desagregada por provincias y por regiones educativas. Se presentan también algunas conclusiones generales, pero insuficientemente elaboradas.

Uso y difusión de resultados

Los resultados de las pruebas 2000-2001 no han tenido mayor difusión, excepto por la distribución de informes a autoridades regionales y escuelas de la muestra. La utilización de los datos al interior del Ministerio de Educación también ha sido insuficiente. Se han realizado algunas actividades de capacitación sobre la base de los resultados, pero la motivación mayor para el operativo fue la de realizar investigaciones institucionales sobre niveles de logro y alcances de la implementación del nuevo currículo.

Sistemas subnacionales de evaluación

No se han desarrollado sistemas subnacionales de evaluación.

Pruebas internacionales de aprendizaje

Panamá no ha participado en este tipo de pruebas.

Avances observados

Vale destacar la movilización civil organizada a favor de la institución de un sistema idóneo y eficiente de evaluación (Mesas de Diálogo Intersectorial y Diálogo para

la Transformación Integral del Sistema educativo Nacional, 2002).

- Dificultades observadas** La dificultad más compleja radica en la escasa voluntad política por apoyar, institucional y económicamente, el desarrollo de un sistema idóneo de evaluación. Las dependencias técnicas del Ministerio vinculadas a la evaluación han funcionado hasta el momento en forma aislada, se observa discontinuidad en los emprendimientos, así como un fuerte vaciamiento de recursos técnico-profesionales. Según las últimas fuentes de información consultadas, se está intentando revertir esta situación para lograr mayores niveles de comunicación y trabajo colaborativo entre los diferentes equipos profesionales del Ministerio.
- Proyección a futuro** Algunas organizaciones civiles, así como la Comisión de Educación de la Legislatura Nacional, han realizado pronunciamientos y propuestas importantes en torno a la evaluación educativa nacional y a la participación en pruebas internacionales de aprendizaje. Se espera que, además de la creación de la nueva Dirección Nacional de Evaluación en el Ministerio de Educación, el gobierno ofrezca el apoyo institucional y destine los recursos necesarios para un adecuado funcionamiento. El nuevo Sistema Nacional de Evaluación Educativa comprenderá un Sistema Nacional de Pruebas de Logros y un Sistema de Evaluación del los Aprendizajes y Régimen de Calificación y Promoción para los dos primeros niveles de enseñanza.
-

PARAGUAY

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
PARAGUAY (SNEPE: Sistema Nacional de Evaluación del Proceso Educativo)	1996	6	L, M	MN	G, U, P		OREALC 1997
	1997	3, 9	L, M	MN	G, U, P		
	1998	6, 12	L, M, CN	MN	G, U, P		
	1999	3,9	L, M, CS	MN	G, U, P		
	2000	6, 12	L, M, CS	MN	G, U, P		
	2001	3, 6	L, M	MN	G, U, P		
	2001	3, 6		CN (en 1150 escuelas del Proyecto «Escuela Viva”)			
	2001	3 ^{er} año Formación Docente		CN			

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional	El SNEPE es un programa administrado desde el Departamento de Orientación y Evaluación Educativa del Ministerio de Educación, y es financiado principalmente a través del Programa de Mejoramiento de la Calidad de la Educación Primaria en convenio con el Banco Interamericano de Desarrollo.
Currículo y estándares	Las tablas de especificaciones y las pruebas son validadas con la participación de los equipos ministeriales de desarrollo curricular y están referidas al currículo nacional vigente. No existen actualmente propuestas de estándares, debido en parte a la percepción de que el establecimiento de metas demasiado altas generaría altos grados de frustración en la mayoría de las escuelas.
Tipos de pruebas	En las aplicaciones muestrales al nivel nacional se utilizan pruebas referidas a normas, mientras que las pruebas aplicadas censalmente en el grupo de escuelas participantes en el programa “Escuela Viva” (1.500 centros educativos) están referidas a criterios. Los instrumentos incluyen ítems de respuestas de opción múltiple y de redacción de texto en el área de lenguaje. Las pruebas se acompañan con cuestionarios a alumnos para recabar información sobre sus actitudes hacia las asignaturas evaluadas.
Tipos de reporte	Los datos se reportan en forma agregada a nivel nacional, y desagregada por departamento, por sector (público o privado), ámbito geográfico (rural o urbano), tamaño de escuela, sexo y por turno de escolaridad (mañana, tarde o noche). Los valores se expresan en medias aritméticas y porcentajes promedios de respuestas correctas, ambos desagregados por competencias evaluadas dentro de cada área curricular. Los informes devueltos a las escuelas participantes de la muestra permiten realizar comparaciones con escuelas

de sectores y zonas similares, y con el conjunto de escuelas al nivel nacional.

Uso y difusión de resultados

Los resultados se entregan a autoridades superiores del sector y a las escuelas (directores y profesores) mediante dos tipos diferentes de informes: uno para las escuelas participantes de la muestra y otro para todas las escuelas del país. Los resultados se entregan durante talleres regionales donde participa la comunidad educativa. También se han difundido los resultados a través de medios de prensa, publicaciones y talleres en diferentes departamentos geográficos para un total de 405 centros educativos. Los resultados de las pruebas censales dentro del programa Escuela Viva se consideran como uno de los indicadores principales de impacto programático. Para tal fin se aplicarán pruebas comparables anualmente, de manera tal de estimar el impacto del programa en la calidad y eficiencia de las escuelas al cabo de cinco años. Las pruebas no tienen altas implicancias para los alumnos de educación primaria, pero se planea introducir mecanismos de acreditación para el egreso de educación secundaria mediante pruebas estandarizadas en el mediano plazo. Este plan, sin embargo, solo será posible una vez culminado el proceso de implantación gradual de la nueva estructura curricular para ese nivel. Actualmente, los mayores esfuerzos están puestos en la devolución de los resultados a todas las escuelas, como señal de responsabilidad y transparencia del sistema ante la comunidad educativa. Los resultados no han sido utilizados aun para el diseño de políticas y programas de mejoramiento educativo específicos, aunque esto constituye uno de los principales objetivos del sistema a largo plazo.

Sistemas subnacionales de evaluación

No se han desarrollado sistemas subnacionales de evaluación.

Pruebas internacionales de aprendizaje

Paraguay participó en la prueba internacional comparada del LLECE- 1997.

Avances observados

Destaca la continuidad de las autoridades políticas del sector en sus funciones, lo cual repercute positivamente en la estabilidad de los equipos técnicos de evaluación.

Asimismo, se percibe voluntad política para hacer públicos los resultados de las evaluaciones y así “transparentar” las acciones de gobierno y sus efectos en materia de política educativa. Se reportan también mayores grados de comunicación, respecto de estas mismas prácticas en el pasado, entre diferentes órganos de gestión sectorial, lo cual podría favorecer el uso de los datos de las evaluaciones para un diseño más informado de las políticas educativas.

Dificultades observadas Existen sectores docentes que piensan que los contenidos o competencias evaluadas no coinciden con lo que ellos realmente enseñan o con lo que sus estudiantes realmente están en condiciones de aprender. Falta un mayor análisis sobre qué aumenta y qué restringe el aprendizaje, de manera que trascienda el sentido común o las especulaciones. Los informes de resultados se limitan a numerosos gráficos con comentarios breves. Lo mismo sucede con las “variables complementarias” o factores “contextualizadores”, no se plantea cuál podría ser la relación entre los resultados y los logros, desaprovechándose la oportunidad de levantar preguntas adicionales donde se podría mirar otros datos y hacer análisis más detallados. Las autoridades no parecen estar explorando los procesos de formulación de estándares nacionales. Las actuales definiciones del contenido del dominio evaluado son tan vagas que difícilmente dan pautas a docentes sobre qué o cómo se tiene que mejorar.

Proyección a futuro El SNEPE planea reforzar la difusión de los resultados al interior del Ministerio de Educación para lograr mayor impacto en la toma de decisiones sobre programas de mejoramiento y de capacitación y formación docente.

PERÚ

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
PERÚ (UMC-CRECEER)	1996	4	L, M	MN			OREALC 1997; PISA Plus
	1998	4, 6	L, M, CS, CN	MN	G, U, P		
		4M, 5M]	L, M				
	2001	4, 6, 4M	L, M	MN	G, U, P		
2004	2, 6, 3M, 5M	L, M, C	MN	G, U, P			

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera, C=Ciudadanía.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

La evaluación externa de aprendizajes se realiza desde la UMC, Unidad de Medición de Calidad, órgano dependiente del Ministerio de Educación, aunque una proporción importante del diseño y aplicación de las pruebas

se realiza con la participación y asesoramiento de consultores externos al Ministerio.

- Currículo y estándares** Los marcos curriculares nacionales vigentes fueron aprobados en los años 1997 y 1998 para educación primaria, y en 2001 para secundaria (este último actualmente en revisión). La reforma curricular de los noventa marcó la transición hacia un currículo basado en competencias, y sobre ellas se construyen las matrices de referencia para las evaluaciones estandarizadas. El diseño de esas matrices, sin embargo, requirió de un trabajo adicional de especificación y operacionalización de contenidos, pues algunas áreas del currículo presentan definiciones demasiado amplias sobre los aprendizajes esperados. Hasta el momento no se han realizado propuestas de estándares de currículo al nivel nacional, aunque algunas organizaciones no gubernamentales han ensayado propuestas exitosas con algunos grupos de escuelas seleccionadas.
- Tipos de pruebas** Las primeras aplicaciones se realizaron con pruebas referidas a normas, pero actualmente se aplican instrumentos con referencia a criterios. Las pruebas son muestrales, incluyen ítems de opción múltiple y de respuestas abiertas, y han sido aplicadas en español, quechua y aimara. Junto con las pruebas se aplican cuestionarios de contexto para el estudio posterior de factores asociados al rendimiento. Se estudian variables escolares (Ej. insumos pedagógicos, características de escuelas y profesores, actitudes hacia las asignaturas y hacia las lenguas originarias, tareas en el hogar), y extraescolares (principalmente género, NES, lengua materna, localización urbana o rural, administración pública o privada, y región geográfica).
- Tipos de reporte** Los informes generales más recientes presentan los resultados en forma criterial, expresando en forma de porcentaje las probabilidades de que diferentes grupos poblacionales (Ej. escuelas rurales, escuelas públicas urbanas o quechuahablantes) logren diferentes niveles de desempeño en las áreas evaluadas y sus diferentes componentes o ejes temáticos. También se presentan gráficos comparativos sobre los niveles de rendimiento

de cada uno de los estratos definidos para la muestra. Las variables escolares y extraescolares analizadas se presentan asociadas al rendimiento global por área curricular y también a las diferentes subáreas de conocimiento evaluadas (Ej. geometría, estadística, comprensión inferencial de textos, etc.).

Uso y difusión de resultados

La difusión y el uso de los resultados de evaluación externa han sido, y aun son, aspectos particularmente débiles de la evaluación externa. La devolución de los resultados de las pruebas nacionales e internacionales se vio fuertemente restringida durante la gestión presidencial de los años noventa. Los datos recogidos en las aplicaciones de 1996 y 1998, así como los resultados de la prueba internacional de UNESCO fueron divulgados recién a partir del año 2000. Actualmente, la UMC lleva publicada una serie de 20 boletines de difusión de resultados de aprendizaje (sobre los resultados de las pruebas de 1998), de análisis de factores asociados al rendimiento, de análisis de ítems y errores de respuesta de las pruebas nacionales para fines pedagógicos, y de los resultados de la prueba internacional de UNESCO. También se han realizado talleres para la presentación y discusión de resultados con docentes y especialistas de órganos intermedios de gestión. Adicionalmente, investigadores independientes han realizado estudios del currículo implementado y de oportunidades de aprendizaje mediante el análisis de cuadernos de trabajo y de cuadernos de clase en muestras de alumnos de escuelas urbanas polidocentes. Otros estudios se han centrado en los factores asociados al rendimiento y en el análisis de la producción de textos de los estudiantes. La dificultad principal en este terreno radica, hasta el momento, en lograr una divulgación efectiva de los informes oficiales, y en lograr que sus contenidos tengan impacto en la responsabilización política por los resultados, en el uso de los datos para el mejoramiento pedagógico y para la toma de decisiones de política educativa.

Sistemas subnacionales de evaluación

No se han desarrollado sistemas subnacionales de evaluación.

Pruebas internacionales de aprendizaje

Perú participó en la prueba internacional comparada del LLECE (1997) y Pisa Plus (2001).

Avances observados	Destaca la continuidad de los operativos, a pesar de la inestabilidad de los equipos técnicos frente a los cambios frecuentes de administración sectorial. Resulta positiva también la publicación de los resultados luego de cada aplicación, así como los esfuerzos recientes por redactar informes más útiles para el mejoramiento de la gestión pedagógica en las escuelas.
Dificultades observadas	Los funcionarios del sector no hacen uso sistemático de los datos para mejorar las condiciones de implementación en las áreas más críticas del currículo y en las escuelas más desfavorecidas. También se percibe un uso insuficiente de los datos en el ámbito escolar, en parte debido al hecho de que la devolución de resultados no ha logrado universalizarse a todos los centros educativos. Parte del problema radica en que los equipos técnicos de evaluación no disponen del tiempo suficiente para diseñar estrategias más amplias de difusión.
Proyección a futuro	Una nueva Ley General de Educación, promulgada en 2003, ordena la creación del Instituto Peruano de Evaluación, Acreditación y Certificación Educativa. Este será un organismo público autónomo que probablemente incorporará la tecnología y los equipos técnicos que actualmente se desempeñan en la UMC en el Ministerio de Educación.

REPÚBLICA DOMINICANA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
REPÚBLICA DOMINICANA (Sistema de Pruebas Nacionales)	1991-1996	8, 4M, 3 ^{er} ciclo de adultos	L, M, CS, CN	CN	U		OREALC 1997
	1997-2003	8, 3 Ed. Adultos y Bachillerato	L, M, CS, CN	CN	G	CD, DC	

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

Actualmente, la unidad de medición de rendimiento académico es el Departamento de Pruebas Nacionales, órgano dependiente de la Dirección de Evaluación y Control de la Secretaría de Estado de Educación. Originalmente, el proyecto de Pruebas Nacionales fue administrado y liderado por una Comisión Ejecutiva integrada por el Secretario de Estado de Educación y

Cultura, los rectores de las principales universidades nacionales y el Secretario General de la Asociación Dominicana de Profesores. Entonces, el proyecto era financiado principalmente con fondos de préstamo internacional (BID y Banco Mundial). En la actualidad, la mayor parte de los programas de evaluación se financian con el presupuesto regular de la Secretaría.

Currículo y estándares

Las Pruebas Nacionales se construyen en función de los contenidos del currículo nacional, y los equipos de desarrollo curricular de la Secretaría participan en la definición de la matriz de referencia para las evaluaciones y en la validación de los ítemes. Sin embargo, suelen existir diferencias de percepción sobre la relevancia o pertinencia del universo de saberes curriculares que serán objeto de evaluación. Tales diferencias se explican por el grado insuficiente de especificación que el currículo plantea respecto a las expectativas de aprendizaje y desempeño para las diferentes áreas de conocimiento. Existe la percepción de que es necesario acordar un conjunto de estándares curriculares al nivel nacional, y en esa dirección trabaja actualmente la Dirección General de Currículo, específicamente en las áreas de lengua española y matemáticas.

Tipos de pruebas

Desde 1998 se aplican pruebas referidas a normas en 4º año del Bachillerato, y pruebas referidas a criterios en 8º grado y 3º ciclo de adultos. Para el Bachillerato, existen pruebas comprensivas (comunes a todas las modalidades) y diferidas (diferenciadas para cada tipo de orientación académica). Ambas versiones están contenidas en el mismo cuadernillo de pruebas. Las pruebas para el área de Lengua Española se elaboran a partir de competencias, mientras que las de Matemática, Ciencias Sociales y Ciencias Naturales se realizan a partir de contenidos u objetivos. Todas las pruebas se basan en un modelo de respuesta de opción múltiple. No se aplican instrumentos especiales para el levantamiento de información de contexto, pero igualmente se dispone de información relacionada con las características de los estudiantes (sexo, edad y calificaciones promedios obtenidas en la escuela en las asignaturas de pruebas nacionales), y con las características geográficas de las

escuelas (zona urbana, urbana marginal, urbana turística, rural, rural aislada y zona de pobreza). Con estos datos se realizan estadísticas descriptivas y correlaciones entre las variables mencionadas y el logro académico de los estudiantes.

Tipos de reporte

La tabla de especificaciones utilizada permite reportar los resultados en función de competencias logradas. Los informes institucionales (escolares) presentan los porcentajes de alumnos que lograron cada una de las competencias y los informes individuales (por alumno) indican el nivel de desempeño logrado para cada competencia. Más específicamente, esos informes presentan los siguientes datos: Informe institucional (datos desagregados al nivel de curso o sección y por área de conocimiento): porcentajes de estudiantes promovidos y aplazados; grado de correlación entre calificación escolar final y calificación en prueba estandarizada; comparación de resultados con centros educativos semejantes, y con todos los centros educativos en los ámbitos nacional, regional y distrital. Informe para alumnos: puntaje obtenido por dominios en cada área de conocimiento; análisis narrativo de logros y dificultades principales, por nivel de desempeño y en forma comparada con el promedio nacional y con el promedio de las escuelas del mismo nivel socioeconómico o zona geográfica; recomendaciones para mejorar el rendimiento académico. Los niveles de desempeño fueron establecidos de acuerdo a cuatro categorías (cuartiles) atendiendo a los porcentajes de estudiantes que dominan o no las preguntas para cada una de las competencias o contenidos evaluados. Para los fines de la discusión de estos resultados, se utilizan dos categorías de desempeño: “dominio bajo” y “dominio”. El criterio para alcanzar tal calificación por los grupos de estudiantes, es la respuesta correcta a 50% de las preguntas relativas a las competencias o contenidos (dominio bajo), o la respuesta correcta a 50% o más de las preguntas (dominio). Esta información se presenta en tablas cruzadas, por categorías de dominio y nivel de agregación geográfica (nacional, regional y distrital). Los resúmenes ejecutivos preparados por el Departamento de Pruebas Nacionales también ofrecen un análisis comentado de los resultados, así como un conjunto

de conclusiones y reflexiones puntuales sobre las implicancias de política de los resultados.

Uso y difusión de resultados

Las Pruebas Nacionales se utilizan, en primer lugar, como indicador parcial de rendimiento académico para determinar la promoción y graduación de alumnos en los niveles y cursos evaluados. La respuesta correcta a todos los ítems de la prueba equivale al 30% de aprobación del curso, mientras que el puntaje máximo en el último año del curso, combinado con el promedio en los últimos cuatro años de educación media, equivalen al 70% restante. En el caso de 8° grado del Nivel Básico y la Educación de Adultos la línea de corte para la aprobación del curso es de 65/100 puntos, mientras que en el caso de 4° año de Educación Media es de 70/100 puntos. Los alumnos que no alcanzan el puntaje mínimo de aprobación en alguna asignatura tienen la posibilidad de volver a presentarse al examen. Los estudiantes de 8° grado y de Educación de Adultos que no alcanzan el puntaje mínimo en las cuatro asignaturas deben repetir el curso. Las calificaciones finales se hacen llegar a los estudiantes por la vía documental (actas escolares de calificaciones), telefónica, o electrónica a través de una página web oficial. Los equipos de evaluación han convocado a otras unidades del sector (Dirección de Curriculum, institutos de capacitación docente, directores de escuelas, maestros y estudiantes, autoridades intermedias, etc.) para realizar lecturas críticas de los resultados y discutir sus implicancias para la redefinición de programas y líneas de acción para el mejoramiento de la calidad de los aprendizajes. Asimismo, se entregan resúmenes ejecutivos sobre los resultados a autoridades políticas y equipos técnicos centrales y locales. Sin embargo, se considera que los resultados han tenido poco impacto en la toma de decisiones de política educativa en general. La prensa local publica y comenta los resultados anualmente, aunque es difícil lograr que se haga un tratamiento sustantivo del tema más allá de la presentación de los promedios generales al nivel nacional.

Sistemas subnacionales de evaluación

No se han desarrollado sistemas subnacionales de evaluación.

- Pruebas internacionales de aprendizaje** República Dominicana participó en el estudio comparado del LLECE, 1997.
- Avances observados** La devolución de los resultados, tanto a las escuelas como a los estudiantes individualmente, se realiza de una forma que resulta útil curricularmente para mejorar los resultados de los aprendizajes. Los operativos de evaluación están debidamente normados por la Ley General de Educación y por un conjunto de Ordenanzas nacionales, lo cual garantiza continuidad en las aplicaciones y otorga mayor legitimidad formal a los procesos evaluativos.
- Dificultades observadas** Existen problemas de organización al interior de la Secretaría de Estado de Educación que impiden un flujo adecuado de información y el trabajo sinérgico entre las diferentes unidades de gestión. Suelen superponerse los esfuerzos de evaluación a cargo de diferentes dependencias y los resultados de las pruebas nacionales no impactan mayormente en la toma de decisiones para el diseño de políticas integrales o complementarias de mejoramiento de la calidad.
- Proyección a futuro** Se prevé que el programa de Pruebas Nacionales seguirá operando como hasta el momento. Desde el punto de vista institucional, el Consejo Nacional de Educación reglamentó recientemente que Sistema Nacional de Evaluación de la Calidad de la Educación comenzará a funcionar como un organismo descentralizado, pero aun no se han elaborado los lineamientos y las normativas específicas para concretar este cambio de status.
-

URUGUAY

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
URUGUAY (UMRE)	1996	6	L, M, AA	CN	G, U, P	CD, FA	PISA 2003
	1998	3	L, M, CN, CS	MN, autoaplicación universo escuelas	U, P	CD	
	1999	6	L, M, AA	MN, ME, autoaplicación universo escuelas	G, U, P	CD	
	1999	3M	L, M, CN, CS	CN	G, U, P	FA	
	2001	4					
	2001	preescolar, 1, 2	L, M, desarrollo cognitivo y afectivo	MN	G, U	CD	
	2002	6	L, M	MN, CN			

[2] Los grados de educación media se indican con la letra M.

[3] L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4] CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

La UMRE, Unidad de Medición de Resultados Educativos, fue creada en 1994 dentro del marco de un convenio de cooperación financiera entre la ANEP (Administración Nacional de la Educación Pública) y el Banco Mundial para el desarrollo del Proyecto de Mejoramiento de la Calidad de la Educación Primaria (MECAEP). En tal sentido, la UMRE no fue concebida como un sistema nacional de evaluación en sí mismo, sino como un proyecto de evaluación nacional de aprendizajes que sentaría las bases para el desarrollo posterior de un sistema de evaluación más integral y administrado desde la estructura funcional de la ANEP. Las acciones de evaluación llevadas a cabo por la UMRE se han realizado en colaboración con otras unidades administrativas de la ANEP, y si bien depende políticamente de la administración educativa central, goza de altos grados de autonomía técnica y administrativa. Desde 2001 la UMRE se institucionalizó en el marco de la entonces creada “Gerencia de Investigación y Evaluación” en la ANEP, con el nombre “Programa de Evaluación de Aprendizajes”. Tiene a su cargo las evaluaciones en educación primaria y media, las evaluaciones internacionales y la elaboración de las pruebas de ingreso a los institutos de formación docente. La unidad sigue gozando de autonomía técnica pero no administrativa.

Currículo y estándares

El currículo oficial nacional vigente en Uruguay data del año 1957 pero ha sufrido revisiones y actualizaciones en las últimas décadas. En los años setenta fue reformado según las corrientes conductistas de la época y los contenidos se presentaban con un alto grado de prescripción y detalle para su cobertura. En los años ochenta

ta, a partir de la restitución del orden democrático, se percibe la necesidad de remover del currículo los cambios introducidos durante el gobierno militar y se elabora una versión levemente modificada del documento de 1957. Esta versión propone un currículo más flexible que permite adecuaciones locales e institucionales pero, si bien se modificaron los postulados filosóficos y epistemológicos para la enseñanza, los contenidos disciplinares siguen siendo los mismos. La sucesión de cambios ha producido un referente curricular ambiguo que presenta contradicciones entre su fundamentación teórica y el tipo de contenidos académicos prescritos. Esa ambigüedad ha constituido una dificultad importante para los equipos de evaluación de la UMRE al momento de definir las tablas de especificaciones, pero se ha resuelto satisfactoriamente mediante la consulta sistemática a docentes y expertos de todo el país sobre la relevancia y adecuación de los tópicos y tipos de competencias que ameritan ser evaluados. Más aun, el diseño de las matrices de referencia y el modelo de evaluación utilizado tienen como uno de los objetivos centrales contribuir a una definición más pertinente y actualizada de las metas curriculares nacionales. En la práctica, los marcos conceptuales de las pruebas y los libros de textos gratuitos que se entregan a todas las escuelas con sus guías didácticas han ido constituyéndose en un referente alternativo al currículo nacional.

Tipos de pruebas

Se han realizado operativos censales y muestrales. Las pruebas incluyen ítemes de opción múltiple y también preguntas para el desarrollo de respuesta abierta. Junto a las pruebas se aplican cuestionarios de contexto socio-educativo a docentes, directores, alumnos y familias para el estudio de factores asociados al rendimiento. Se analizan variables escolares (Ej. infraestructura y equipamiento, dotación de recursos humanos, experiencia docente, gestión, concepciones pedagógicas), así como variables extraescolares (Ej. condiciones de la vivienda, composición de la familia, bienes materiales y culturales, niveles educativos y ocupacionales de los padres).

Tipos de reporte

Se presentan los resultados clasificados según los porcentajes de alumnos que alcanzan los diferentes niveles

de suficiencia predefinidos (altamente satisfactorio, satisfactorio, insatisfactorio, muy insatisfactorio), y se relacionan en tablas que comparan el rendimiento por departamento, según subsistemas (Montevideo, interior del país, colegios públicos y privados, colegios urbanos y rurales) y el contexto sociocultural de las escuelas. La UMRE ha puesto especial énfasis en la comparación de resultados entre escuelas de contextos socioeconómicos similares como una manera de evitar inferencias inadecuadas sobre la calidad educativa de escuelas que trabajan con alumnados de diferentes contextos socioculturales. (Por tanto, los informes que se distribuyen a las escuelas presentan sus resultados en comparación con escuelas del mismo contexto sociocultural del departamento, y del mismo contexto sociocultural del país. Los informes globales se difunden públicamente y presentan los resultados en forma agregada al nivel nacional, y en forma desagregada al nivel de departamentos y escuelas. La información desagregada para cada escuela es confidencial y no pretende establecer *rankings* entre ellas. Cada escuela recibe en sobre cerrado sus propios resultados, y los del conjunto de otras escuelas semejantes (Ej. todas las escuelas de contexto desfavorable). Los únicos agentes extraescolares que acceden a los resultados por escuela son los inspectores de zona, quienes podrían potencialmente difundir información comparada, pero mantienen el compromiso de confidencialidad y solo usan los datos para realizar recomendaciones pedagógicas. La presentación de datos sobre niveles de suficiencia en el desempeño académico se basa en una línea de corte (nivel *satisfactorio*) equivalente a la respuesta correcta por alumno al 60% del total de los ítems y en el porcentaje de alumnos que alcanza ese nivel de rendimiento. Los informes de 6° grado presentan también comparaciones interanuales (en lapsos de tres años) sobre el progreso de la población estudiantil en el logro de las competencias evaluadas. Tales datos se expresan en forma de puntos porcentuales de diferencia entre los años 1 y 3 de aplicación, en forma general para toda la población y todas las áreas evaluadas y desagregada solamente según el nivel socioeconómico de las escuelas.

Uso y difusión de resultados

Los destinatarios principales de los resultados de las evaluaciones son las escuelas. La devolución es universal, independientemente de la participación en la muestra, y es frecuentemente acompañada por presentaciones y talleres a cargo de los equipos técnicos de la UMRE. Tanto las presentaciones como los materiales de orientación publicados centran la atención en el análisis de los ítemes y las respuestas de los estudiantes, de manera de estimular la reflexión sobre los aspectos curriculares y pedagógicos que deben mejorarse para lograr mejores resultados. La experiencia hasta ahora indica que estas actividades son altamente valoradas por los docentes, lo cual se manifiesta en forma de solicitudes a la UMRE de más herramientas de evaluación y análisis para mejorar su práctica pedagógica. Además de los datos, la UMRE facilita también un conjunto importante de ítemes utilizados en las aplicaciones y orientaciones para su aplicación en forma autónoma por los docentes en sus cursos. Los docentes, entonces, disponen de los instrumentos de evaluación que les permite una aplicación independiente, y las informaciones necesarias para poder comparar los datos obtenidos con los de la aplicación nacional. Esta estrategia es particularmente útil cuando los operativos son muestrales y por lo tanto no todas las escuelas cuentan con una devolución individual de resultados. La responsabilización o “rendición de cuentas” no es un objetivo del sistema de evaluación. Por el contrario, los esfuerzos por legitimar el sistema ante la comunidad docente se centró en la confidencialidad y en el uso estrictamente pedagógico y curricular de los datos. En términos de responsabilización política sectorial, los datos no han generado debate público o académico, ni reacciones por parte de la sociedad civil, excepto por algunos grupos que demandan su utilización para impulsar mecanismos de responsabilización y descentralización. Los datos han sido utilizados para el desarrollo de programas de capacitación de docentes en servicio, coordinados frecuentemente en colaboración con los cuerpos de Inspectores de Primaria y con el Proyecto MECAEP de capacitación docente, aunque aún no se ha logrado incidir sobre los programas de formación magisterial inicial. Los resultados fueron utilizados recientemente para fundamentar la continuación y ampliación de un programa

de equidad para Escuelas de Tiempo Completo. Por el tipo de modelo de evaluación que se realiza en este país, los datos sirvieron para demostrar que este tipo de escuelas logran mejores resultados que otras en contextos socioeconómicos semejantes.

Sistemas subnacionales de evaluación	No se han desarrollado sistemas subnacionales de evaluación.
Pruebas internacionales de aprendizaje	Uruguay participa en el proyecto PISA y realizó la primera aplicación de instrumentos en 2003.
Avances observados	El sistema de evaluación ha logrado niveles importantes de legitimación entre los docentes, quienes valoran los datos recogidos como insumos de información válidos y útiles para el mejoramiento de las prácticas pedagógicas y de la gestión escolar. Destaca también la continuidad de los operativos en el tiempo, así como la estabilidad de los equipos técnicos en sus funciones. Vale remarcar los esfuerzos por construir matrices de referencia e ítemes de evaluación que, si bien alineados con el currículo nacional, trascienden la prescripción oficial e introducen nuevas metas que generan debate sobre la pertinencia y validez de los contenidos de aprendizaje en las escuelas.
Dificultades observadas	Los datos e informaciones difundidos por la UMRE han tenido escaso impacto en la opinión pública.
Proyección a futuro	Si las próximas administraciones de gobierno continúan apoyando el sistema nacional de evaluación, se espera adoptar la metodología de medición TRI, incorporar ítemes abiertos a las pruebas de 6° de primaria, mantener el ritmo trienal de aplicaciones, utilizar la prueba internacional PISA como referente para la evaluación y monitoreo del 3° de educación media y evaluar el final del Bachillerato.

VENEZUELA

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
VENEZUELA (SINEA-Sistema Nacional de Medición y Evaluación del Aprendizaje)	V	6	L, M	MN	G, U, P		OREALC 1997

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

El SINEA, Sistema Nacional de Medición y Evaluación del Aprendizaje, es un programa dependiente del Vice-ministerio de Asuntos Educativos del Ministerio de Educación, aunque no está incluido formalmente en el organigrama oficial del sector. Al momento de la aplicación muestral de 1998, el SINEA era administrado por la Ofi-

cina Sectorial de Planificación y Presupuesto. Originalmente se financió con fondos de préstamo del Banco Mundial pero actualmente se sostiene con el presupuesto ordinario del Ministerio de Educación en convenio con el CENAMEC, Centro Nacional para el Mejoramiento de la Enseñanza en Ciencia. La aplicación de 1998 se realizó con el apoyo técnico de consultores de la Universidad Católica de Venezuela, pero actualmente se cuenta con escasos recursos técnicos y humanos para el desarrollo de nuevos operativos.

- Currículo y estándares** Las matrices de referencia para el diseño de las pruebas fueron elaboradas sobre la base del currículo nacional vigente para educación primaria. Hasta el momento no existen propuestas para el desarrollo de estándares nacionales de currículo.
- Tipos de pruebas** La evaluación de 1998 fue muestral y las pruebas se basaron en un modelo de respuestas de opción múltiple. Se intentó evaluar también mediante preguntas abiertas pero los equipos de correctores no lograron acordar criterios uniformes para la corrección de respuestas. Las pruebas se construyen y validan siguiendo un modelo normativo, pero los resultados se reportan en forma criterial, utilizando una línea de corte equivalente al 70% de respuestas correctas.
- Tipos de reporte** Los resultados de 1998 se reportaron utilizando puntajes de corte sobre una escala gradual de desempeños. La escala definía un nivel de insuficiencia o *no logro* (0 a 40 puntos), *logro parcial* (40 a 70 puntos), y de suficiencia o *logro* (70 a 100 puntos). Este reporte criterial se presentó desagregado al nivel de entidades federales. También se reportaron los datos siguiendo un modelo normativo que expresaba el rendimiento relativo de diferentes regiones geográficas sobre la media aritmética. El documento consultado explica que en la muestra se definieron estratos para escuelas públicas y privadas, urbanas y rurales, y marginales y no marginales, aunque no se reportan resultados desagregados según ese diseño. En informes dirigidos a docentes se presentan los ítemes aplicados en la evaluación, así como un análisis somero de las dificultades presentadas por los estudiantes. Finalmente, se pre-

	<p>senta un conjunto de conclusiones sobre los resultados y sus implicancias pedagógicas generales.</p>
Uso y difusión de resultados	<p>Para la difusión de las pruebas de 1998 se produjeron informes diferenciados para el gobierno central, para cada uno de los estados y para los docentes. Estos últimos se enviaron a las direcciones de educación de cada estado, pero se desconoce qué alcance e impacto efectivo tuvieron al nivel de las escuelas. El Ministerio de Educación no utilizó los resultados para el diseño de políticas, ni tampoco lo hicieron las instancias intermedias de gestión, aunque sí fueron utilizados por el CENAMEC para el desarrollo de programas de capacitación docente. La prensa no hizo una cobertura formal de los resultados, aunque algunas notas periodísticas han citado los datos del SINEA entre otros indicadores sobre la baja calidad del sistema educativo público.</p>
Sistemas subnacionales de evaluación	<p>No se han desarrollado sistemas subnacionales de evaluación.</p>
Pruebas internacionales de aprendizaje	<p>Venezuela participó en la prueba del LLECE -1997.</p>
Avances observados	<p>Aunque ya discontinuado, el programa de evaluación contribuyó inicialmente al desarrollo de capacidades técnicas para la evaluación externa.</p>
Dificultades observadas	<p>Resalta principalmente el insuficiente grado de institucionalización logrado por el SINEA, cuyas actividades fueron discontinuadas al culminar el convenio financiero con el Banco Mundial. Actualmente se cuenta con recursos humanos y técnicos insuficientes, y las autoridades del Ministerio de Educación no han destinado nuevos recursos ni tiempo de gestión para el sostenimiento y desarrollo del programa.</p>

CASOS SUBNACIONALES

MINAS GERAIS, BRASIL

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
MINAS GERAIS (SIMAVE)	2000	4, 8, 3M	L, M	CR	G, U, P	CD, FA	
	2001	4, 8, 3M	CS, CN	CR	G, U, P	CD, FA	
	2002	4, 8, 3M	L, M	CR	G, U, P	CD, FA	
	2003	4, 8, 3M	L	CR	G, U, P	CD, FA	

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]
CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

El SIMAVE, Sistema Minero de Evaluación de la Educación Pública, es uno de los componentes centrales del PROEB, Programa de Evaluación de la Red Pública de Educación Básica, y está inserto en la estructura funcio-

nal de la Secretaría de Educación del Estado de Minas Gerais. Este sistema de evaluación fue creado dentro del marco de una fuerte redefinición de políticas educativas a partir de la nueva administración sectorial iniciada en el año 1999. Las inequidades en el acceso al conocimiento al interior del estado, así como los múltiples problemas de repitencia y promoción automática requerían de un sistema de evaluación diagnóstica de aprendizajes que permitiera focalizar apoyos y compensar esas inequidades. La evaluación se inserta dentro de un conjunto de programas de “Acompañamiento Pedagógico” y de fortalecimiento de la capacitación y la formación docente inicial. Una de las características más novedosas del sistema es su trabajo colaborativo formal con 29 instituciones de educación superior del estado. Ese trabajo involucra el diseño y aplicación de instrumentos, así como el procesamiento de datos y su devolución a las escuelas. Los operativos son coordinados por una de estas universidades en colaboración con las demás, en un proceso que pretende contribuir a mayores grados de descentralización de la gestión educativa en su conjunto, así como a mayores grados de participación de los actores locales. Motiva también esta asociación la necesidad de acercar mayores insumos de información a las instituciones formadoras de docentes, y de consolidar gradualmente una mayor cultura evaluativa dentro del sistema. El SIMAVE colabora con los operativos del SAEB (ver Brasil) y se sirve del banco nacional de ítems para la elaboración de sus propios instrumentos, pero no utiliza los datos nacionales para el diseño de políticas estatales.

Currículo y estándares

El referente curricular principal para el estado son los Parámetros Curriculares Nacionales (PCN). Sin embargo, la Secretaría de Educación estadual posee una propuesta curricular propia alineada con esos parámetros. A partir de estos dos referentes, las escuelas elaboran sus propios diseños curriculares, aunque este proceso está ocurriendo muy lentamente. Para la elaboración de las matrices de referencia del SIMAVE se utilizan los PCN, la propuesta curricular del estado, y la matriz de contenidos del SAEB. Se trata de priorizar los saberes más importantes para cada grado. Sin embargo, esa prioriza-

ción es realizada por los equipos técnicos de evaluación, con escasa participación de los equipos de desarrollo curricular. Estos últimos demandan, por tanto, mayores grados de apertura por parte del SIMAVE para lograr más pertinencia y legitimidad en las evaluaciones y sus contenidos.

Tipos de pruebas

Las pruebas aplicadas por el SIMAVE son censales y estrictamente normativas, pues están diseñadas básicamente para la comparación de rendimiento entre diferentes grupos de estudiantes. A pesar de eso, como se puede observar en el apartado siguiente, se intenta reportar los resultados en forma criterial desagregando resultados por tipos de competencias logradas en cada área curricular evaluada. Se aplican también cuestionarios para relevar las características socioeconómicas de las escuelas, pero no se realizan análisis estadísticos sobre factores asociados al rendimiento académico.

Tipos de reporte

Cada escuela recibe un informe individual. Los informes presentan gráficos separados para cada grado y área curricular evaluada, y cada uno de ellos muestra el porcentaje de alumnos que se ubica en cada uno de los tres niveles de desempeño (proficiencia) predeterminados (crítico, intermedio y suficiente). Los porcentajes se desagregan al nivel de las diferentes capacidades que conforman cada una de las competencias evaluadas en cada área curricular. Además, se ofrece un gráfico de barras comparando el desempeño promedio de la escuela respecto del municipio correspondiente, la superintendencia regional y el estado. Los informes se complementan con una revista explicativa para orientar a los profesores en la lectura e interpretación de los datos.

Uso y difusión de resultados

Los informes de resultados apuntan a dos audiencias principales: autoridades sectoriales y equipos escolares (directores y docentes). Las primeras reciben informes sobre los rendimientos globales de diferentes regiones, superintendencias y municipios para diseñar políticas y focalizar apoyos materiales y pedagógicos en las regiones y superintendencias con bajos niveles de desempeño. Las escuelas reciben los informes descritos en el apartado anterior. El SIMAVE se propuso lograr legiti-

mación mediante un modelo evaluativo diagnóstico que no implicara sanciones ni premios para los actores locales, sino que sirviera principalmente para orientar y mejorar los procesos pedagógicos y curriculares. Otro de los objetivos principales es el de acercar los datos de las evaluaciones a los estudiantes de educación, y futuros profesores del sistema, a través de la participación de las universidades en todo el proceso evaluativo (los estudiantes son frecuentemente contratados como aplicadores). Se espera lograr que las nuevas generaciones de docentes tengan una mayor participación colegiada en temas substantivos de la educación y no solamente en términos de reivindicación salarial. Los resultados están disponibles para el público en general y para la prensa. Sin embargo, deben realizarse grandes esfuerzos para evitar que la prensa haga un uso “amarillista” de los datos, especialmente mediante la presentación de *rankings* y comparaciones que pueden resultar agresivas. Los padres de familia no reciben resultados individualmente, excepto los representantes de este grupo en los gobiernos escolares de cada institución. Estos representantes tienen acceso a los informes de resultados, y de su voluntad depende que sean socializados entre los demás padres. Los datos han sido utilizados para el diseño de capacitaciones docentes, con especial énfasis en los problemas de lectura detectados, y sobre lo cual se han diseñado programas específicos de apoyo a las escuelas. Los datos sirven también para seleccionar a las escuelas que recibirán atención prioritaria mediante esos programas. Esas escuelas son también objeto de investigaciones cualitativas que buscan indagar sobre las razones que explican los bajos rendimientos.

Pruebas internacionales de aprendizaje No se ha participado en forma independiente de las muestras nacionales en este tipo de pruebas.

Avances observados Si bien la comunicación de los equipos de evaluación con otras dependencias de la Secretaría es escasa, los equipos de desarrollo curricular están comenzando a utilizar los datos de las evaluaciones para proponer nuevas líneas de acción desde su ámbito de trabajo. Hasta el momento, se han concentrado principalmente en los problemas de lectura detectados. La relación entre el

SIMAVE y las universidades parece ser provechosa y existe voluntad de ambas partes por continuar trabajando en forma asociada.

Dificultades observadas Según algunas críticas, para lograr que las pruebas tengan mayor impacto, debería involucrarse a los docentes más directamente en el proceso evaluativo y especialmente en la validación de los contenidos evaluados por el SIMAVE. A esta falta de participación se atribuye, en cierta medida, que los profesores no se sienten responsables por los resultados y que tampoco alientan a los alumnos a mejorar su rendimiento.

Proyección a futuro Dado que el sistema ha sido recientemente implementado, se espera lograr una mayor institucionalización y continuidad en los operativos. Asimismo, se espera tener mayor incidencia en los programas de capacitación y formación docente.

PARANÁ, BRASIL

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
PARANÁ (Programa de Evaluación del Sistema Educativo de Paraná)	1995	4	L, M	CR	G, U, P	CD, DC	
	1996	8, 2M	L, M, CN, CS	CR	G, U, P	CD, DC	
	1997	4, 8	L, M, CN	CR	G, U, P	CD, DC	
	1998	4, 8	L, M, CN	CR	G, U, P	CD, DC	
	2000	4, 8	L, M, CN	CR	G, U, P	CD, DC	
	2002	4, 8, 3M	L, M, CN	CR	G, U, P	CD, DC	

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

El Programa de Evaluación del Sistema Educativo de Paraná es parte de la estructura orgánica de la Secretaría de Educación del estado y se inscribe dentro de un pro-

grama más amplio denominado “Proyecto de Calidad de la Enseñanza”. Ese proyecto se inició en el año 1994 con apoyo financiero del Banco Mundial. Los requisitos de monitoreo impuestos por el Banco impulsaron el diseño de los primeros instrumentos de evaluación del sistema, entre los cuales cobró particular importancia, y continuidad, el de evaluación de los aprendizajes. El Programa dispone de escasos márgenes de autonomía, pues todas las decisiones técnicas y políticas son tomadas finalmente por las autoridades máximas de la Secretaría de Educación. Además de evaluar los procesos de implementación e impacto de programas de calidad específicos, tales como la Aceleración de Aprendizajes y la dotación de bibliotecas, material pedagógico y cursos de capacitación docente, el estado precisaba estudiar con mayor profundidad dos de las grandes problemáticas que se presumían fuertemente asociadas al rendimiento académico: repitencia y extraedad y deserción escolar. Por esa razón es que los resultados de aprendizaje, como se discute más abajo, se reportan diferenciadamente por franjas etarias al interior de cada grado evaluado. Adicionalmente, la Secretaría de Educación precisaba contar con indicadores de calidad que le permitieran monitorear los procesos de descentralización de las escuelas y dar apoyo focalizado a aquellas que más lo necesitaran. Existe una relación de intercambio técnico con el SAEB (ver Brasil), especialmente de consulta sobre matrices curriculares y metodologías de medición, pero el estado de Paraná no hace uso de los datos recogidos por la agencia de evaluación nacional.

Currículo y estándares Las matrices curriculares de las pruebas están referidas al “Currículo Básico de la Escuela Pública” del estado. Estos, a su vez, están alineados a los Parámetros Curriculares Nacionales (PCN). Los equipos de currículo de la Secretaría participaron en la elaboración de los PCN, y tienen participación en la elaboración de las matrices curriculares de evaluación del estado. Hasta mediados de la década anterior los docentes guiaban su práctica básicamente por los libros de texto comerciales e ignoraban los parámetros curriculares nacionales y estaduales. Por medio de las evaluaciones, y de las capacitaciones diseñadas sobre la base de los resultados, la

Secretaría de Educación apunta a revertir gradualmente esa tendencia. Según la opinión de los equipos de currículo, los contenidos de las matrices están debidamente validados, en términos de alineación con la propuesta curricular estadual, aunque se cuestiona el grado de exhaustividad que la matriz permite en el sondeo de las capacidades desarrolladas por los alumnos.

Tipos de pruebas

Las pruebas aplicadas son normativas, y son diseñadas mediante procedimientos derivados de la Teoría Clásica y de TRI. Las aplicaciones son censales (en todas las escuelas con más de 50 alumnos en los grados evaluados) y cada alumno responde 30 preguntas por área curricular evaluada. Se aplican cuestionarios a docentes y directores de centro para realizar análisis de factores asociados al rendimiento. Los principales factores analizados son: gestión y administración escolar, y nivel económico y social de alumnos y familias. Los operativos se realizan con amplia participación de equipos docentes locales y escolares, así como de padres de familia designados para monitorear los procesos de evaluación y actuar como difusores de información de los resultados posteriormente.

Tipos de reporte

Los datos que se devuelven a las escuelas consisten en frecuencias simples y medidas de tendencia central de aciertos en las respuestas, agregados al nivel estadual y desagregados al nivel de municipio, núcleo escolar, escuela y franjas etarias al interior de cada grupo de estudiantes evaluado. Los informes también presentan los resultados en forma de porcentajes de alumnos que alcanzaron cada uno de los cinco niveles de dificultad incluidos en la escala de desempeño propuesta en el modelo. Adicionalmente, los informes explicitan cuáles de las 30 preguntas respondidas por los alumnos en cada área corresponde a cada uno de esos niveles de desempeño. Con esos datos, y siguiendo orientaciones y formatos prescritos por la Secretaría, las escuelas redactan sus informes institucionales de rendimiento estudiantil, identifican problemáticas y proponen vías de acción pedagógica posibles. Los demás informes generales producidos por el Sistema de Evaluación presentan los resultados en forma desagregada por municipio, y de acuerdo

a otros estratos definidos tales como escuelas públicas y privadas, escuelas estatales y municipales y horario escolar diurno y nocturno. Se presenta también una síntesis de los contenidos curriculares más críticos (bajo desempeño) y un análisis pedagógico de algunos de los ítemes de las pruebas para orientar a los docentes en su planificación y en la resolución de los problemas de aprendizaje detectados. El mismo tipo de análisis se difunde para los ítemes de respuesta abierta (producción de textos), de manera de ofrecer indicadores más claros sobre el tipo de desempeños que se espera de los alumnos en esa área. Adicionalmente, la Secretaría de Educación devuelve los resultados de las evaluaciones en forma de “Boletines Escolares”, que incluyen informaciones específicas del rendimiento de los alumnos de cada centro educativo según la escala de desempeño utilizada, y en forma comparada con otras escuelas del municipio y del estado en su totalidad. Además de los datos de rendimiento, se presentan los resultados de las encuestas que el mismo operativo de evaluación realiza para recoger las opiniones de padres y alumnos sobre diferentes aspectos de la “vida escolar” de cada institución. Un segundo conjunto de informaciones complementarias en este boletín incluye datos institucionales sobre tasas de aprobación, reprobación y deserción, así como otros datos sobre el perfil de los docentes, todo lo cual puede ser comparado con las demás escuelas del municipio y del estado.

Uso y difusión de resultados

El uso de los datos más enfatizado por el Sistema es el que hacen las escuelas mediante los mecanismos descritos en el apartado anterior. Adicionalmente, la Secretaría de Educación utiliza los datos en todos los eventos de capacitación que se realizan en el estado. Resulta importante destacar que a todos esos eventos de capacitación, que usualmente se realizan en puntos estratégicos para reunir a docentes de varios municipios, asisten representantes de los padres de familia de cada escuela. Se apunta con esto a desarrollar una cultura evaluativa sobre la base de la demanda de información por parte de las familias. El uso de los datos para la toma de medidas de política, si bien era uno de los objetivos principales al momento de la creación del sistema, es relativamente

débil y se atribuye a la falta de tiempo por parte de los equipos de evaluación para realizar análisis más profundos de los datos recogidos. Las bases de datos están disponibles para el medio académico e investigadores externos, y pueden accederse mediante solicitud formal a la Secretaría de Educación. En general, la demanda por ese tipo de información es baja. También la prensa tiene acceso a los datos, aunque resulta difícil evitar el tratamiento “amarillista” de ellos. La repercusión en la opinión pública, según las autoridades del Sistema de Evaluación, es baja.

Pruebas internacionales de aprendizaje	No se ha participado en forma independiente de las muestras nacionales en este tipo de pruebas.
Avances observados	Se observan acciones muy definidas a favor de la utilización efectiva de los datos recogidos, y a favor de una apropiación real del sistema de evaluación por parte de los usuarios centrales y locales. Se aprecia también un grado relativamente alto de estabilidad de los cuadros técnicos, lo cual se atribuye a la estabilidad de las autoridades políticas del sector a lo largo de las últimas administraciones de gobierno. También vale destacar el funcionamiento orgánico entre los equipos de evaluación, capacitación docente y desarrollo curricular.
Dificultades observadas	Hasta la actualidad, la frecuencia anual de aplicaciones ha dejado poco tiempo disponible para que los equipos técnicos realicen análisis más profundos de los datos sobre rendimiento escolar.
Proyección a futuro	Se planea espaciar las aplicaciones, quizás a una frecuencia bianual, para dar más tiempo para el análisis y la difusión de datos.

SÃO PAULO, BRASIL

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
SÃO PAULO	1996	3, 7	(L, M) (L, M, CS, CN)	CR	G, U, P	CD, FA, DC, CG	
SARESP	1997	4, 8	(L, M) (L, M, CS, CN)	CR	G, U, P	CD, FA, DC, CG	
	1998	5, 1M	L, M, CS, CN	CR	G, U, P	CD, FA, DC, CG	
	2000	5, 7, 3M	L, M, CN	CR	G, U, P	CD, FA, DC, CG	
	2001	4, 8	L	CR	G, U, P	CD, FA, DC, CG	

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

El SARESP, Sistema de Evaluación del Rendimiento Escolar del Estado de São Paulo, es administrado en forma conjunta por la Secretaría de Educación del estado y la Fundación para el Desarrollo de la Educación (FDE), un órgano dependiente de aquella Secretaría pero que goza de altos grados de autonomía administrativa. La FDE recibe financiamiento de la Secretaría pero también se sustenta mediante la venta de servicios a otras agencias. Esta autonomía financiera, aunque sea parcial, le permite mayor agilidad y estabilidad en sus gestiones y provisión de servicios evaluativos al estado. Sin embargo, todas las decisiones sobre cobertura de las evaluaciones, así como sobre los usos de los datos recogidos, son tomadas principalmente por la Secretaría de Educación. La administración estadual de educación decidió la creación de un sistema de evaluación propio para contribuir a la resolución de dos problemáticas principales, las de equidad y autonomía escolar, que son los pilares de la gestión educativa actual. Los datos de las evaluaciones serían usados para la focalización de apoyos en las escuelas con desempeños más críticos (equidad), y el proceso evaluativo mismo sería realizado con la participación permanente de las escuelas, en proceso de descentralización, para garantizar el uso de los datos en su gestión pedagógica y administrativa (autonomía escolar). El sistema de evaluación también se propone evaluar longitudinalmente el impacto de las políticas educativas estaduais. El SARESP colabora con las aplicaciones nacionales del SAEB (ver Brasil) pero sus instrumentos y operativos son totalmente independientes. El soporte técnico para esas actividades (modelo psicométrico, procesamientos estadísticos y elaboración de informes) fue provisto, desde el primer momento, por la Fundación Carlos Chagas. Esta Fundación también está cooperando actualmente, mediante capacitaciones, a la transferencia de la coordinación técnica del SARESP a la Secretaría de Educación del estado.

Currículo y estándares

Los referentes curriculares principales para la elaboración de los instrumentos de evaluación son los Parámetros Curriculares Nacionales (PCN) y la propuesta curricular y pedagógica de la Secretaría de Educación del estado. Para la construcción de la matriz de referencia se

consultó ampliamente a la CENP, Coordinación de Estudios y Normas Pedagógicas, una de las dependencias de mayor protagonismo en la gestión educacional de la Secretaría.

Tipos de pruebas

Se aplican pruebas referidas a normas diseñadas sobre la base de un conjunto de habilidades definidas para cada área curricular evaluada. El diseño también incluye la definición de una escala de cinco niveles de logro por área, y bajo los cuales se clasifica el desempeño de cada alumno, así como de las diferentes subpoblaciones y de la población general en forma agregada. Las aplicaciones fueron originalmente muestrales, pero se decidió transitar hacia un modelo censal para estimular la responsabilización de las escuelas por sus resultados. Las pruebas se aplican en el último grado de cada ciclo escolar de cuatro años. Los cuestionarios aplicados a directores, supervisores y docentes permiten recabar información sobre las características de la escuela, la cual es utilizada posteriormente para el análisis de factores asociados al rendimiento. Algunos de esos factores analizados son: trabajo infantil; etnicidad, edad, NES, género y trayectoria escolar de los alumnos; características del cuadro docente; clima escolar; participación de la escuela en programas especiales. Adicionalmente, los padres delegados para la observación de las aplicaciones completan una planilla de preguntas sobre el desarrollo y transparencia del proceso de aplicación. Los datos son procesados por la escuela misma, de acuerdo a procedimientos pautados centralmente, y luego son elevados a la Secretaría de Educación. Existen mecanismos de control para monitorear la veracidad y consistencia general de los datos al nivel de las escuelas.

Tipos de reporte

Los informes elaborados por las escuelas presentan el porcentaje de aciertos (y el nivel de desempeño correspondiente a ese porcentaje) para el conjunto de alumnos evaluados. La escuela eleva los resultados institucionales a la Secretaría pero retiene los datos de rendimiento por alumno. Desde la FDE, se elaboran informes sobre el desempeño relativo y comparado de las diferentes subpoblaciones escolares, así como de los factores asociados al rendimiento para la población general. Los da-

tos se presentan desagradados por edad, tipos de escuelas, etnicidad y horario escolar (diurno y nocturno), entre otros. Para los años 1996 y 1997 se presentaron los datos en forma comparada (misma cohorte). La comparabilidad fue asegurada mediante la utilización de ítemes de anclaje para ambas aplicaciones.

Uso y difusión de resultados

Las pruebas aplicadas por el SARESP son de bajas y altas implicancias. El uso principal de los datos está orientado a la toma de decisiones de política pedagógica y curricular, en la forma de programas especiales de apoyo y dotación de insumos a las escuelas que presentan bajos resultados de aprendizaje (programas de refuerzo y recuperación de aprendizajes durante el año y en vacaciones). Los resultados son de altas implicancias para docentes y alumnos. Los primeros pueden recibir una bonificación salarial extraordinaria según los resultados globales de su escuela, mientras que los segundos ven comprometida su promoción al siguiente ciclo de escolaridad de acuerdo a su desempeño en las pruebas anuales. En el primer caso, los resultados del SARESP constituyen uno entre otros indicadores de desempeño (principalmente retención escolar). En el caso de los alumnos, un desempeño insuficiente en las pruebas del SARESP (menos de 50% de acierto) puede ser revisado por las autoridades si el rendimiento del estudiante ha sido muy bueno durante el ciclo y se presume que por alguna razón especial no pudo rendir al nivel de su potencial el día de la aplicación. En ese caso, el alumno tiene una segunda oportunidad para presentarse a la prueba. Al momento de esa publicación, el uso de las evaluaciones con fines de bonificación y promoción estaba siendo reconsiderado. Los resultados de las pruebas son devueltos a la sociedad civil mediante informes a la prensa, pero se desincentiva la presentación de resultados en forma de *ranking*. La Fundación Carlos Chagas ha realizado análisis independientes de los datos, con fines de investigación, dado que las bases de datos están disponibles para cualquier usuario externo al ámbito oficial. Entre las medidas de política educativa diseñadas a partir de los datos del SARESP destaca un programa de capacitación delegado por la Secretaría de Educación a tres universidades (UNESP, USP y UNICAMP). La ofer-

ta de las capacitaciones depende de la solicitud que realizan las escuelas mismas y que justifican a partir de resultados insatisfactorios en las pruebas estatales. Uno de los énfasis principales de las capacitaciones ha sido el de desarrollo de las habilidades de comprensión lectora. Los resultados también son utilizados para la evaluación de impacto de programas especiales de la Secretaría, especialmente del Programa de Aceleración de Aprendizajes. Las autoridades del CENP consideran que el progreso en el aprendizaje de saberes escolares, y observado a través del SARESP, se debe al impacto de un conjunto de medidas y programas impulsados por el estado. No se realizan seguimientos cualitativos sobre las causas que podrían explicar los resultados de las pruebas.

Pruebas internacionales de aprendizaje	No se ha participado en forma independiente de las muestras nacionales en este tipo de pruebas.
Avances observados	Destaca en este sistema la integración orgánica que se mantiene con otras dependencias de la gestión educativa, especialmente con los equipos de normas pedagógicas para el diseño de programas de apoyo y recuperación de aprendizajes. Resulta valioso también el alto grado de participación que tiene la comunidad educativa local a lo largo de todo el proceso evaluativo.
Proyección a futuro	Se espera seguir consolidando el sistema de evaluación y fortaleciendo las capacidades técnicas en evaluación de los equipos profesionales al interior de la Secretaría de Educación.

BOGOTÁ, COLOMBIA

País y agencia evaluadora	Calendario de aplicación	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
BOGOTÁ Subdirección de Evaluación y Análisis (Pruebas de competencias básicas)	A	1998	3, 5	L, M, CN, C	CR	G, U, P	CD, DC, CG, FA	
	B	1999	3, 5, 7, 9					
	A	1999	7, 9					
	A	2000	3, 5					
	B	2001	3, 5, 7, 9					
	A	2001	7, 9					
	A	2002	3, 5					
B	2003	3, 5, 7, 9						

[2]
Los grados de educación media se indican con la letra M.

[3]
L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera, C=Ciudadanía/Valores

[4]
CN=Censo Nacional, CR=Censo Regional o Distrital Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]
G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]
CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

El sistema de evaluación externa de aprendizajes en el Distrito Capital se coordina desde la Dirección de Evaluación y Acompañamiento, órgano dependiente de la Subsecretaría Académica de la Secretaría de Educación de Bogotá. Bajo esa Dirección trabajan conjuntamente la Subdirección de Evaluación y Análisis, y la Subdirección de Mejoramiento Educativo. El programa de Evaluación de Competencias Básicas se diseñó y comenzó a implementarse en el marco de importantes reestructuraciones de la administración del sector y los resultados de las evaluaciones se consideran insumos de información clave para la toma de decisiones tendientes al cumplimiento de los Planes Sectoriales de Gestión 1998-2001 y 2001-2004.

Currículo y estándares

Las escuelas del Distrito Capital orientan su trabajo en función de los marcos curriculares nacionales (ver Colombia). Sin embargo, dada la amplitud del universo de conocimientos prescritos en esos referentes, la Secretaría de Educación local optó por definir un conjunto de competencias básicas que clarificaran las expectativas de aprendizaje en cuatro áreas de conocimiento disciplinar. El trabajo de diseño de competencias fue realizado por especialistas académicos, varios de los cuales habían participado en la elaboración de los marcos curriculares nacionales en años anteriores. Estas competencias, cuyos contenidos disciplinares convergen con los lineamientos e indicadores de logro curricular nacionales, constituyen la matriz de referencia para el diseño de los instrumentos de evaluación estandarizada. Desde los puntos de vista técnico y político, las competencias básicas propuestas pueden ser consideradas como estándares de currículo, en tanto ofrecen una descripción operacional clara sobre los niveles de desempeño posible para cada ítem de contenido evaluado, a la vez que proporcionan una herramienta de comunicación efectiva entre todos los actores sectoriales sobre las expectativas de aprendizaje para todos los estudiantes del sistema.

Tipos de pruebas

Las pruebas de rendimiento son referidas a criterios, y se aplican censalmente en todos los grados de escolaridad evaluados. Incluyen ítemes de opción múltiple, así como preguntas para el desarrollo de respuestas abiertas. Las preguntas para el área de conocimiento sobre Comprensión

y Sensibilidad Ciudadana se enfocan en temas de desarrollo moral, representaciones sociales, actitudes democráticas, y conocimientos sobre el Estado colombiano.

Tipos de reporte

En los informes que distribuye la Secretaría a cada escuela, los resultados para cada asignatura se reportan desagregados por competencias, por niveles de desempeño alcanzados para cada una de ellas, y se expresa el valor promedio logrado por los alumnos del plantel en una escala de 0 a 306. En los informes generales de difusión pública, los porcentajes promedio de logro se presentan desagregados por asignatura, por tipo de plantel (oficial o privado), y por zonas o localidades geográficas al interior del Distrito. Los informes ofrecen también ejemplos y análisis de algunos ítemes evaluados, así como un conjunto de conclusiones sobre los resultados globales de cada aplicación.

Uso y difusión de resultados

Los resultados de las evaluaciones han sido utilizados con varios propósitos, principalmente para la capacitación docente y de directivos de escuelas, y para la focalización de apoyos a planteles. Los materiales y eventos de capacitación se centran en el desarrollo de habilidades para el análisis de los datos sobre aprendizajes, así como para el diseño de estrategias de mejoramiento pedagógico y curricular en función de los resultados. Se utiliza material impreso, así como videos y DVD de presentación de resultados, entrevistas a docentes y recomendaciones específicas. Tales recomendaciones se presentan como “rutas de mejoramiento” posibles, que pueden ser utilizadas por los planteles como modelo de diseño curricular y planeamiento pedagógico. En cuanto a la focalización de apoyos, destacan los proyectos implementados por la Secretaría de Educación que buscan rescatar de las instituciones con mejores resultados, ideas y herramientas de gestión pedagógica que puedan ser replicadas en otras escuelas con mayores necesidades de mejoramiento. Asimismo, las escuelas con bajos resultados tienen acceso a otros programas de apoyo, tales como expediciones culturales por la ciudad, o la nueva red de bibliotecas inaugurada en el último quinquenio. También se aprecia la utilización de las evaluaciones de aprendizajes, junto con otros indicadores, para la toma de decisiones sobre inversiones en infraestructura, equipamiento y reorganización institucional. En términos de responsabili-

zación política por los resultados de aprendizaje, el Plan Sectorial 2001-2004 explicita la meta de lograr un promedio mínimo de alcance de 180 puntos (en una escala de 0 a 306) y que todas las escuelas superen los 100 puntos de logro en la misma escala. Estas metas se han hecho públicas y corresponderá a las autoridades locales responder por los resultados al cabo del período mencionado. Se alienta a las escuelas a hacer públicos ante la comunidad sus avances hacia el logro de las competencias evaluadas, lo cual se materializa mediante carteleras impresas que los equipos docentes van completando conforme reciben los informes anuales de resultados. Esas carteleras se exhiben en lugares visibles de la escuela, tales como la recepción o pasillos de tránsito para alumnos y padres. Resulta importante destacar que las calificaciones escolares (internas) de los alumnos, que ellos y sus familias reciben en forma de boletines periódicos, se reportan de acuerdo al mismo conjunto de competencias básicas que evalúa la Secretaría de Educación en forma externa y estandarizada. Los resultados se utilizan también para monitorear, junto con el uso de otros indicadores, la calidad de las escuelas privadas que albergan matrícula oficial (estrategia utilizada por la Secretaría para garantizar el acceso a la escolaridad mientras se construyen nuevos establecimientos oficiales). Los datos son también utilizados como uno de los indicadores de desempeño docente dentro del nuevo esquema de regulación para la permanencia en el cargo.

- Pruebas internacionales de aprendizaje** Bogotá ha participado en pruebas internacionales como parte de la muestra nacional, pero no en forma independiente.
- Avances observados** Entre los avances más destacables, vale mencionar el desarrollo de criterios curriculares claros y operacionales para la evaluación de aprendizajes, los esfuerzos por socializar y legitimar el sistema de evaluación y las informaciones producidas, y la estabilidad de los equipos técnicos y de los operativos a lo largo del último quinquenio. Resalta también el uso efectivo que se da a los datos de las evaluaciones para el mejoramiento de los procesos de enseñanza-aprendizaje, así como para la toma de decisiones de política educativa en general.
-

AGUASCALIENTES, MÉXICO

País y agencia evaluadora	Años de aplicación	Grados/ años [2]	Áreas disciplinares [3]	Tipo de estudio [4]	Publicidad de resultados para conocimiento de [5]	Uso de resultados para [6]	Pruebas internacionales/ Años de participación
AGUAS- CALIENTES (Sistema Estatel de Evaluación)	2000-2003	6	L, M	CR	G, U, P	CD, DC, CG, FA	PISA TIMSS (parte de la muestra nacional)

[2]

Los grados de educación media se indican con la letra M.

[3]

L=Lenguaje, M=Matemáticas, I=Informática, CS=Ciencias Sociales, CN=Ciencias Naturales, ES=Educación para la Salud, EA=Educación Artística, AA=Aspectos Actitudinales, LE=Lengua Extranjera.

[4]

CN=Censo Nacional, CR=Censo Regional, MN=Muestra Nacional, MR=Muestra Regional, ME=Muestra Experimental.

[5]

G=Gobierno [se refiere a instancias de alta dirección del sistema (ministerios, secretarías de planeamiento, etc.)]. U=Usuarios [se refiere a la comunidad escolar, incluyendo a docentes, directivos, alumnos y padres de familia]. P=Público en general y medios de comunicación.

[6]

CD=Capacitación Docente (ofrecida por la alta dirección o desarrollada por docentes a partir de los resultados), CG=Capacitación para la Gestión, DC=Desarrollo Curricular, FA=Focalización de apoyos a estudiantes y centros educativos.

Marco institucional

Las evaluaciones educativas se administran desde el Sistema Estatal de Evaluación, una unidad técnica de la Dirección estatal de Formación, Actualización, Capacitación y Superación Profesional del Magisterio. Las líneas de acción en materia de evaluación responden a las prioridades de política educativa expresadas en el Plan Estatal de Desarrollo 1998-2004, y más específicamente del Programa de Desarrollo Educativo 1999-2004. El

programa de medición y evaluación de aprendizajes se inscribe dentro de un proceso de reestructuración del sistema de supervisión escolar, especialmente de redistribución de profesionales según demandas y necesidades de cada zona educativa, para hacerlo menos burocrático-administrativo y más eficaz en cuanto a sus funciones pedagógicas. Ese proceso de reestructuración fue política y administrativamente complejo, pero preparó el terreno para hacer posteriormente un uso efectivo de los datos recogidos por las evaluaciones. Un coordinador pedagógico general, que responde directamente al Secretario de Educación del estado, asegura la articulación de tareas entre los directores de diferentes unidades de gestión, y entre estos y los supervisores escolares locales. Todos los programas de mejoramiento en la actualidad están centrados en la calidad académica de las escuelas. Se parte de la convicción de que el nivel socioeconómico de los estudiantes no es el determinante principal de los logros educativos, y que las políticas pedagógicas adecuadas pueden marcar mayor diferencia en los aprendizajes que las condiciones contextuales en que operan las escuelas.

Currículo y estándares

Así como todos los demás estados de la Federación, Aguascalientes utiliza el currículo nacional como referente principal para la enseñanza. La matriz de referencia para el diseño de instrumentos de evaluación se basa en ese currículo, y los ítemes de las pruebas se toman del banco de ítemes que facilita la Dirección General de Evaluación de la Secretaría de Educación Pública nacional. No se realizan validaciones de los contenidos curriculares evaluados al nivel estatal, aunque sí se someten a validación local los ítemes seleccionados para las pruebas, así como el diseño de los instrumentos que se realiza mediante consulta a expertos locales. Las autoridades perciben que la validación de los contenidos no es un aspecto crítico del sistema de evaluación, puesto que el currículo nacional de referencia no es cuestionado por los docentes, sino su legitimación a través de la transparencia y la idoneidad técnica. Si bien no se han elaborado estándares al nivel estadual, los objetivos de evaluación seleccionados para las pruebas constituyen en sí mismos la medida de expectativa “oficial” respecto a

qué contenidos mínimos deben ser logrados por todos los alumnos de 6° grado del sistema.

Tipos de pruebas

Las pruebas de aprendizaje aplicadas por el estado son referidas a criterios y censales en el nivel de 6° grado de primaria. La evaluación se basa en un modelo de progreso o “ganancia educativa”, según la denominación oficial. Este modelo implica la aplicación de una prueba diagnóstica a principios del año académico y una prueba sumativa al final. La comparación de ambos resultados arroja un índice de progreso en el aprendizaje de los alumnos a lo largo de ese año. Curricularmente, el modelo es viable pues los contenidos en lenguaje y matemáticas de 5° y 6° grado son muy similares, diferenciándose solamente en la profundidad con que se tratan. Las pruebas incluyen ítemes de opción múltiple, y actualmente se están incorporando reactivos de **producción de textos escritos**. El objetivo es dar cuenta de la capacidad de intervención de la escuela en el desarrollo de habilidades para la composición escrita de textos. Se elaboró una escala de 20 aspectos a evaluar de cada texto, y la tarea de corrección recaerá en los supervisores escolares, directores y maestros, quienes están siendo debidamente capacitados en la corrección de textos según criterios uniformes. Los datos sobre aprendizaje se complementan con cuestionarios de contexto escolar y extraescolar aplicados a alumnos, docentes y directores. También se aplica los exámenes federales IDANIS, EXANI I y II, estándares nacionales de educación primaria y secundaria, y el proyecto de autoevaluación escolar (modelo escocés). Todos estos instrumentos, sin embargo, son diseñados centralmente desde la SEP y el estado de Aguascalientes solo opera como aplicador y consumidor de la información recabada. El programa de autoevaluación escolar, si bien está coordinado desde la SEP, involucra equipos de investigadores locales que deciden cuáles escuelas, usualmente las de más bajos resultados en las pruebas estatales, se beneficiarán más de la participación en ese programa.

Tipos de reporte

Los informes de las pruebas estatales se reportan por niveles de desempeño logrados en cada área. Los niveles definidos se corresponden con una partición en cuartos

de una escala de 1 a 100 y se clasifican en *crítico e insuficiente* (no dominio), y en *aceptable y deseable* (dominio). Los resultados de “ganancia educativa” a fin de año se reportan en una escala de -4 a 4, dependiendo de los avances o retrocesos que muestra la comparación entre pruebas inicial y final. Por ejemplo, si un alumno comenzó el año en un nivel *insuficiente* y culminó el año en un nivel *deseable*, la ganancia, según la fórmula utilizada, se calcula en 3 puntos sobre 4. Si el resultado fue inverso, es decir que comenzó en un nivel *deseable* y decreció a un nivel *insuficiente*, la “ganancia” se expresa en forma negativa, es decir como -3 puntos en la escala. Estos datos se presentan agregados al nivel estatal, y desagregados a los niveles de zona educativa, escuela, grupo (curso) y alumno en forma individual.

Uso y difusión de resultados

Los resultados se difunden a autoridades máximas de la gobernación y de la Secretaría de Educación, a coordinadores y supervisores de zonas educativas, a directores, profesores, alumnos y familias. Las escuelas reciben un informe sobre los niveles de desempeño y de ganancia educativa, desagregados por curso y por alumnos individualmente. Los niveles de desempeño en ambas áreas evaluadas se desagregan, para cada grupo y estudiante evaluado, en 50 ítems de conocimiento. Es decir, el docente puede conocer exactamente cuáles objetivos curriculares específicos logró cada alumno y en qué grado de profundidad fue logrado. Los supervisores reciben la información desagregada por escuela y por grupo de alumnos. También se presentan los resultados de los cuestionarios de contexto, en forma de porcentajes de respuesta a las opciones seleccionadas por cada grupo (por ejemplo, sobre educación de padres o actitud hacia la materia), pero no se presentan correlaciones ni otros tipos de análisis inferenciales. La Secretaría de Educación, con la participación de los supervisores de zona, realiza monitoreos continuos sobre los niveles de ganancia o progreso educativo en las diferentes escuelas, y solicita a los directores que realicen propuestas pedagógicas para superar las dificultades encontradas. El Secretario de Educación en persona visita escuelas, sin previo aviso, y verifica si el director ha leído los informes y conoce las dificultades reveladas por las pruebas.

Los proyectos de mejoramiento que se pide a las escuelas requieren el uso de las estadísticas disponibles y el análisis de los factores escolares y extraescolares que muestran incidencia sobre el rendimiento. Se enfatiza el análisis de los factores intraescolares, pues es sobre ellos que las escuelas pueden incidir directamente mediante el diseño de nuevas estrategias pedagógicas y curriculares. En las propuestas de mejoramiento diseñadas por las escuelas debe expresarse el tipo de progreso cuantitativo esperado para el año siguiente, típicamente en forma de crecimiento en los porcentajes de aciertos logrados por cada curso evaluado. Hasta el momento, se reporta que la mayor parte de las escuelas muestra progresos de año a año, que un 15% de ellas permanece estable, y que un 5% puede bajar de rendimiento. Se observa que las escuelas que mayor “ganancia” muestran son las que empezaron en los niveles más bajos de desempeño. Los datos se utilizan también para el diseño de capacitaciones a docentes, directores y coordinadores y supervisores de zona. Para los docentes, las capacitaciones se centran en aspectos curriculares y de metodología y didáctica de la enseñanza (desarrollo de habilidades del pensamiento, tácticas de interacción verbal, técnicas y modelos de planificación de contenidos, procesos de pensamiento relacionados con esos contenidos, y estrategias para articular contenidos y habilidades de pensamiento). Las capacitaciones a supervisores están orientadas a fortalecer su función pedagógica por sobre la función puramente administrativa que se busca superar. Un componente importante de la difusión es el acceso que se brinda a los padres sobre los resultados anuales de la escuela. Según opiniones de autoridades centrales y zonales, esta estrategia ha incentivado un mayor grado de participación de las familias en la vida escolar. La prensa local ha hecho una divulgación típicamente “amarillista” de los resultados, aunque gradualmente se está logrando que se haga un tratamiento más sustantivo, educacionalmente hablando, de los datos publicados.

Pruebas internacionales de aprendizaje

Aguascalientes ha participado en las evaluaciones comparadas de PISA y TIMSS como parte de la muestra nacional mexicana.

- Avances observados** Destaca entre los avances observados la organicidad del sistema de evaluación, en tanto promueve el flujo de información y el trabajo conjunto entre escuelas, supervisores y diferentes unidades de gestión de la Secretaría de Educación. También se observa un grado importante de institucionalización externa del sistema, cada vez que docentes y grupos sindicales apoyan la iniciativa y prestan su colaboración para la realización de los operativos. Se aprecia, además, un desarrollo importante de capacidades técnicas locales para la medición de aprendizajes y análisis de datos.
- Dificultades observadas** Los equipos técnicos locales expresan cierto grado de preocupación por la magnitud de las tareas de evaluación en relación a la exigua cantidad de recursos humanos disponibles. Algunos supervisores zonales también expresan preocupación por la actitud de algunos grupos de padres de familia, quienes utilizan los datos de las evaluaciones para enfrentarse con los equipos docentes en vez de buscar formas más eficaces de trabajo colaborativo.
- Proyección a futuro** La insuficiencia de presupuesto para ampliar la cobertura poblacional y curricular de las aplicaciones ha motivado la decisión de concentrar parte importante de los esfuerzos técnicos en lograr que las escuelas den el mayor uso posible de los datos disponibles para el 6° grado. El sistema aspira a utilizar los resultados de los operativos para evaluar a los docentes, aunque aún no se ha diseñado los mecanismos ni los esquemas de incentivos profesionales que acompañarían a esas evaluaciones. Ante la propuesta, los docentes de 6° grado cuestionan que las pruebas sean indicadores válidos sobre su desempeño profesional, y sostienen que los instrumentos deberían aplicarse también en los grados inferiores de escolaridad. Consideran también que no deben ser ellos los únicos actores del sistema que rindan cuentas sobre el rendimiento académico de los estudiantes. *A partir del año escolar 2003-2004 se iniciará la evaluación de conocimientos de manera censal en el 4° grado de educación primaria.*
-

REFERENCIAS BIBLIOGRÁFICAS

- FERRER, GUILLERMO (2004). Las reformas curriculares de Perú, Colombia, Chile y Argentina: ¿Quién responde por los resultados? GRADE. Lima, Perú.
- FERRER, GUILLERMO y PATRICIA ARREGUI (2003). Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: Criterios para guiar futuras aplicaciones. PREAL, Documento N° 26.
- FERRER, GUILLERMO (1999). Aspectos del curriculum prescrito en América Latina: Revisión de tendencias contemporáneas en curriculum, indicadores de logro, estándares y otros instrumentos. Mimeo. <http://www.grade.org.pe/gtee-preal/docs/Ferrer.pdf>
- RAVELA, PEDRO (Ed.) (2001). Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina? PREAL, Documento N° 20.
- RAVELA, PEDRO (2001). ¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina? PREAL.
- ROJAS, CARLOS y JUAN MANUEL ESQUIVEL (1998). Los sistemas de medición del logro académico en Latinoamérica. Banco Mundial, LCS-HD Paper Series.
- WOLFF, LAURENCE (1998). Evaluación educacional en América Latina: Progresos actuales y desafíos futuros. PREAL.

ENTREVISTAS A AUTORIDADES SECTORIALES, CUADROS TÉCNICOS Y CONSULTORES (2000-2004)

Casos Nacionales

ARGENTINA	Margarita Poggi Jorge Fasce Hilda Lanza Silvia Montoya Juan Cruz Perusia Lucrecia Tulic
BOLIVIA	Susana Barrera Carlos Gutiérrez

	Martha Méndez Esther Balboa Nicole Nucinkis
BRASIL	Alejandra Shulmeyer Equipo técnico INEP
CHILE	Cristián Cox Lorena Meckes Silvia Elgueta Patricia Aninat Beatrice Ávalos Leonor Cariola Erika Himmel
COLOMBIA	Carlos Pardo Héctor Fernández
COSTA RICA	Juan Manuel Esquivel Sandra Arauz
CUBA	Héctor Valdés
REPÚBLICA DOMINICANA	Leo Valeirón
ECUADOR	Martha Grijalva Gonzalo Barreno
EL SALVADOR	Graciela de Salgado Hilda Álvarez
MÉXICO	Felipe Martínez Rizo Jesús Mejía Víctor Velásquez
PANAMÁ	Laura G. de Vergara Ruth Nativí
PARAGUAY	Martha Lafuente
URUGUAY	Pedro Ravela Beatriz Picaroni Magela Figarola

VENEZUELA José Mirás
Julia Montoya (CNU)

Casos Subnacionales

MINAS GERAIS, BRASIL Cleuza Pereira
María Inés Barrozo

PARANÁ, BRASIL María Luiza Marques Dias
Marlene C. de Araujo

SÃO PAULO, BRASIL Cristina Amoroso
María Julia Filgueiras
Yara Espósito

BOGOTÁ, COLOMBIA Eliana Iannini
José Medina
Álvaro Leuro

AGUASCALIENTES, MÉXICO Javier Anaya
Josefina Mercado
Flabiano Jiménez