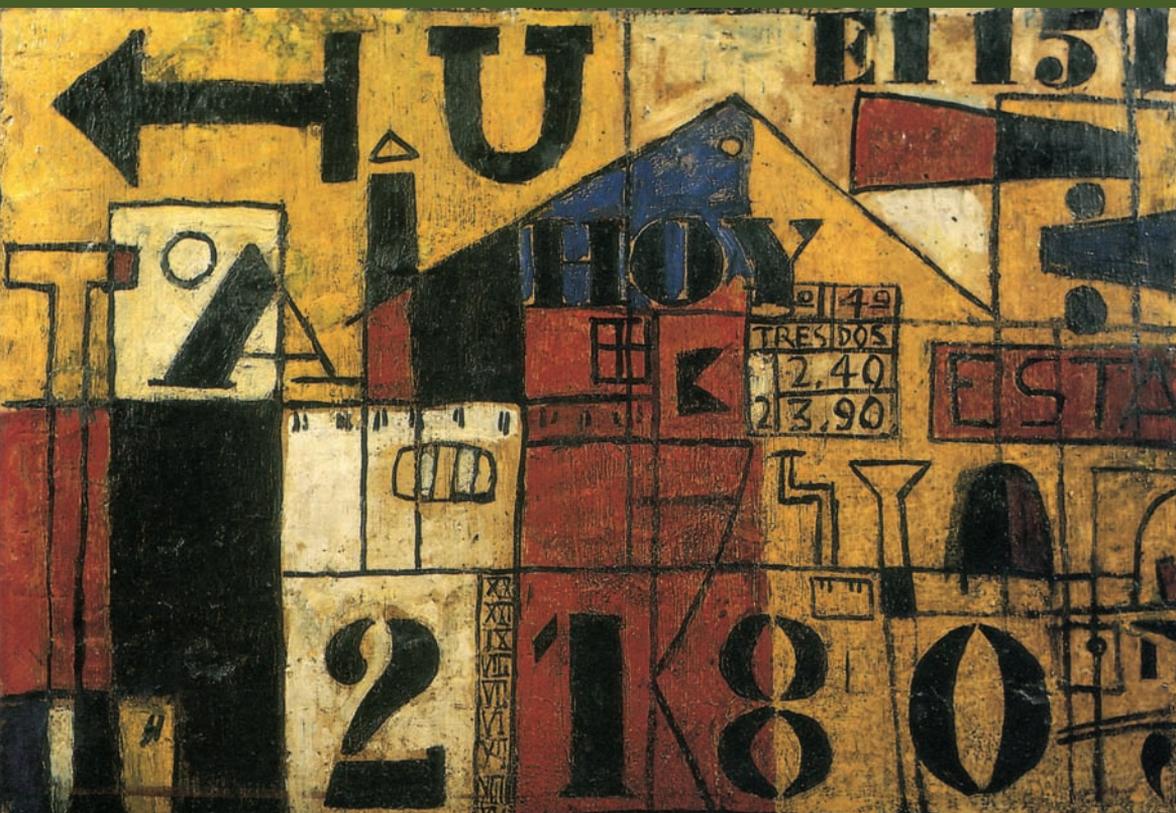


SOBRE ESTÁNDARES Y EVALUACIONES EN AMÉRICA LATINA



**GRUPO DE TRABAJO
SOBRE ESTÁNDARES
Y EVALUACIÓN**



SOBRE ESTÁNDARES Y EVALUACIONES EN AMÉRICA LATINA

Patricia Arregui, *Editora*

Juan Manuel Esquivel

Guillermo Ferrer

Pedro Ravela

Gilbert Valverde

Richard Wolfe

Laurence Wolff

**GRUPO DE TRABAJO
SOBRE ESTÁNDARES
Y EVALUACIÓN**



Sobre estándares y evaluaciones en América Latina

© Preal

Editora:

Patricia Arregui

Diseño portada:

Paulina Montalva

Imagen portada:

"Mural", Gonzalo Fonseca, 0000

Primera edición:

1.000 ejemplares

Diciembre 2006

I.S.B.N.:978-956-8589-05-9

Registro de Propiedad Intelectual:

Inscripción N° 160.234

Diseño e impresión:

Editorial San Marino

E-mail: sanmarino@adsl.tie.cl

Reconocimientos

La publicación de este libro, como las actividades del PREAL, son posibles gracias al apoyo de la United States Agency for International Development (USAID), el Banco Interamericano de Desarrollo (BID), el Banco Mundial, la International Association for the Evaluation of Educational Achievement (IEA), The Tinker Foundation, GE Foundation, entre otras.

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

Índice

Presentación	7
Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos LAURENCE WOLFF	13
Los próximos pasos: ¿Cómo avanzar en la evaluación de aprendizajes en América Latina PEDRO RAVELA, RICHARD WOLFE, GILBERT VALVERDE y JUAN MANUEL ESQUIVEL.	53
Aspectos del currículo prescrito en América Latina: Revisión de tendencias contemporáneas en currículo, indicadores de logro estándares y otros instrumentos GUILLERMO FERRER, con la colaboración de JUAN MANUEL ESQUIVEL y GILBERT VALVERDE	123
¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina? PEDRO RAVELA	181
Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: Criterio para guiar futuras aplicaciones GUILLERMO FERRER y PATRICIA ARREGUI	249
Los resultados de las evaluaciones educativas en la prensa PEDRO RAVELA	295



Presentación

Esta antología reúne algunos de los principales estudios sobre políticas y prácticas de determinación de expectativas y evaluación de logros de aprendizaje en América Latina realizados por el Grupo de Trabajo sobre Estándares y Evaluación del PREAL en el período 2000-2004.

El Grupo de Trabajo –conocido como GTEE/PREAL– es una red regional de especialistas y educadores que empezó a constituirse a fines de 1998. Su fin principal es promover la clarificación de las expectativas de aprendizaje escolar y la evaluación de logros educativos en los países de la región. Su propósito último, es contribuir al fortalecimiento de los procesos de mejoramiento de la calidad, pertinencia, equidad y eficiencia de los sistemas educativos latinoamericanos.

El GTEE/PREAL considera que expectativas claras –denominadas frecuentemente estándares de aprendizaje– y la evaluación son instrumentos clave para impulsar las reformas educativas, siempre y cuando lleguen efectivamente a incidir en lo que sucede en las aulas estimulando el desempeño, orientando los esfuerzos de mejora programática, curricular y pedagógica y responsabilizando a todos los actores que deben participar en esos esfuerzos. Por ello, sus objetivos generales son contribuir a ampliar, consolidar, legitimar y utilizar el espacio que hacia fines del decenio pasado ya había comenzado a adquirir la medición de resultados de aprendizaje en la definición de políticas educativas en los países de la región y promover la discusión y generación de consensos básicos sobre el tema de estándares de contenido y desempeño –así como de estándares de oportunidades de aprender– que, por el contrario, estaban escasamente presentes en la agenda educativa latinoamericana.

Actualmente, están afiliados a esta red más de 1.200 personas, a las cuales se les hace llegar información periódicamente y a quienes se con-

voca a diversas actividades de producción y difusión de ideas y experiencias, ya sea presenciales o virtuales. En el curso de dichas actividades, coordinadas desde el Centro de Análisis (GRADE) en Perú, se ha ido conformando un grupo de colaboradores asiduos que desde sus diversos países de residencia— que incluyen a algunos fuera de la región —se han constituido en un núcleo que también participa en la evaluación y planificación del programa de trabajo.

Los miembros de ese núcleo son los autores de la mayor parte de los trabajos que contiene este libro, que ahora publica el Programa de Promoción de la Reforma Educativa en América Latina y el Caribe —PREAL— proyecto conjunto del Diálogo Interamericano con sede en Washington y la Corporación de Investigaciones para el Desarrollo (CINDE), con sede en Santiago de Chile. Este tiene como principales objetivos promover un diálogo regional informado sobre política educacional, fortalecer el tema de la reforma educativa como una prioridad en la agenda política de los países de la región, crear espacios para la búsqueda de consensos y difundir experiencias exitosas en materia educativa. La ejecución de sus programas está a cargo de centros de investigación y políticas públicas en diversos países de la región que, como contrapartes colaboradoras del PREAL, realizan estudios, organizan debates y promueven diálogos sobre temas de política educacional y reforma educativa.

El libro que ahora presentamos incluye en primer lugar un artículo que antecede a la conformación del Grupo y que fuera publicado por el PREAL en julio de 1998, como uno de los primeros textos (No. 11) de su serie “Documentos”. Se trata de **Las Evaluaciones Educativas en América Latina: Avance Actual y Futuros Desafíos**, de Laurence Wolff, entonces consultor del Banco Interamericano de Desarrollo y ahora frecuente colaborador del PREAL. El texto realiza un rápido repaso de los avances de las evaluaciones internacionales y de la escasa participación de países latinoamericanos hasta ese momento, así como de experiencias en algunos estados norteamericanos, países europeos y países en desarrollo, para luego describir los avances de Chile, México, Costa Rica, Colombia, Brasil y Argentina, así como el esfuerzo pionero del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa impulsado por UNESCO/OREALC y finalizar con un resumen de la experiencia de esos países, las lecciones de ella desprendidas y los desafíos pendientes de confrontación.

Un segundo documento es el editado por Pedro Ravela y escrito colaborativamente por Richard Wolfe, Gilbert Valverde y Juan Manuel

Esquivel en un taller realizado en Lima en agosto de 1999, titulado **Los próximos pasos: ¿cómo avanzar en la evaluación de aprendizajes en América Latina?**, que fue publicado en versión resumida por PREAL como Documento No. 20. Se trata de un texto que pretendía aportar a la reflexión regional sobre la relación entre las finalidades de política educativa que los sistemas de evaluación pueden proponerse y sus implicaciones técnicas, de manera de enriquecer el abanico de alternativas a considerar en los diversos países latinoamericanos cuando cada uno de ellos se abocara a trazar o retrazar el rumbo por el cual seguirían impulsando el desarrollo de sus sistemas de evaluación de aprendizajes.

El documento se concentra en seis temas que, en opinión de sus autores, representan los principales retos que confrontarán: el problema de los fines y el diseño global del sistema nacional de evaluación que es tratado en la introducción del texto, asuntos referidos al diseño de las mediciones de calidad, que es tratado tanto en el primer capítulo del texto como en el quinto, el problema de la validez de los instrumentos y de la interpretación y uso de los resultados de las mediciones, las alternativas de enfoques de medición (pruebas referidas a “normas” y pruebas referidas a “criterios” y/o estándares, y, por último, el análisis de los factores asociados a los resultados escolares, concluyendo con una serie de recomendaciones sobre los “próximos pasos” requeridos para consolidar los sistemas de evaluación en la región.

Aspectos del currículum prescrito en América Latina: revisión de tendencias contemporáneas en currículum, indicadores de logro, estándares y otros instrumentos, publicado electrónicamente en el año 2000 y escrito por J. Guillermo Ferrer –entonces recién graduado de la Universidad del Estado de Nueva York en Albany como magíster en Administración en Política Educativa y en Desarrollo Curricular y Tecnología Educativa, luego investigador visitante de GRADE y actualmente consultor basado en Buenos Aires, Argentina–, el tercer documento incluido en esta antología, es otro producto temprano del GTE/PREAL.

El texto, que sólo ha estado disponible en versión electrónica hasta el momento, consiste en una revisión y diagnóstico de las políticas curriculares en América Latina durante el decenio anterior y del estado en ese momento del diseño de documentos curriculares. Se analizan los procesos de desarrollo y reforma curricular seguidos por los países de la región, así como algunos factores críticos para su implementación. Se resumen las referencias más importantes que se hacen a la evaluación y medición del desempeño y rendimiento académico en los documentos

curriculares vigentes, y se describe el estado actual del debate sobre la elaboración de estándares educativos en Latinoamérica.

El cuarto estudio, también realizado por Pedro Ravela, es **¿Cómo presentan sus resultados los sistemas de evaluación educativa en América Latina?** Fue publicado electrónicamente en agosto del 2001 y, en la versión resumida aquí reproducida, apareció como el No. 22 de la serie Documentos del PREAL en Febrero del 2002. Basado en el análisis de aproximadamente 70 informes nacionales producidos por unidades de evaluación de logros de aprendizaje de 10 países de la región, el estudio analiza el tipo de información que producen y difunden los sistemas de evaluación así como la manera en que la presentan a diferentes audiencias. Identifica los propósitos explícitos de las evaluaciones y de la publicación de los resultados y analiza si la información presentada parece o no servir efectivamente para esos propósitos, para luego formular propuestas respecto de cómo mejorar los modos de reportar los resultados.

Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: criterio para guiar futuras aplicaciones, es el quinto texto incluido en la antología. Resume los resultados del estudio realizado por J. Guillermo Ferrer y Patricia Arregui en el 2001, con apoyo del Consorcio Peruano de Investigación Económica y Social y el IDRC de Canadá, y fue publicado como Documento No. 26 del PREAL en abril del 2003. Tras una revisión sucinta del contexto de las reformas educativas de los ochenta y noventa en el cual se enmarca el surgimiento de las evaluaciones de logros académicos en la mayor parte de los países latinoamericanos, así como sus primeras participaciones en pruebas internacionales, se describe algunos elementos del debate internacional vigente sobre los beneficios y problemas que puede suscitar dicha participación. Luego, para los casos específicos de las pruebas del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa –LLECE– de UNESCO/OREALC, el Tercer Estudio de Ciencias y Matemáticas –TIMSS– y los estudios de Educación Cívica de la Asociación Internacional para la Evaluación del Logro Académico, y PISA de la Organización para la Cooperación y el Desarrollo Económico- OCDE, se revisan las experiencias de 7 países latinoamericanos y la de los Estados Unidos, en términos de cómo y por qué tomaron la iniciativa de participar, cómo se organizaron los procesos de desarrollo, aplicación y análisis de las pruebas, la interpretación y difusión de sus resultados y, por último, las decisiones que se tomaron (o no) en base a sus resultados y el impacto político que tuvieron. En base a esas experiencias, se delinear

recomendaciones para futuras oportunidades de participación en dichas pruebas que permitirían hacerlas más productivas.

Por último, **¿cómo aparecen los resultados de las evaluaciones educativas en la prensa?** es un texto producido en el 2003 por Pedro Ravela, quien contó con los aportes de Tacyana Arce, Julio Valeirón, Claudio Sánchez, Marta Grijalva, Sarah González de Lora, Renán Rápalo, Gilbert Valverde y Patricia Arregui, reunidos en un taller realizado en Santo Domingo para discutir las principales ideas surgidas del análisis de una rica colección de recortes periodísticos reunida por la coordinación del GTEE/PREAL.

El documento focaliza su atención en cómo los resultados son recogidos y entregados a la opinión pública a través de los medios de prensa, su principal informante, con el objetivo tanto de aportar a los periodistas elementos que les ayuden a informar mejor sobre resultados de las evaluaciones educativas, como de aportar sugerencias sobre cómo los Ministerios de Educación y otras agencias evaluadores podrían trabajar mejor con los medios de prensa.

Es deseo de los editores de este volumen que los textos aquí reunidos no sólo den testimonio de las contribuciones del Grupo de Trabajo de Estándares y Evaluación del PREAL al conocimiento y debate del tema en la región, sino que sirvan de estímulo a mayores y más profundas reflexiones sobre las acciones que sería necesario emprender en América Latina para que la evaluación pueda realmente contribuir a aquello que es su razón de ser: el mejoramiento de la calidad de la educación en nuestros países.

PATRICIA ARREGUI
Coordinadora de GTEE/PREAL

JEFFREY M. PURYEAR
MARCELA GAJARDO
Codirectores PREAL

Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos^(*)

LAURENCE WOLFF

I. INTRODUCCIÓN

La educación es fundamental para el desarrollo económico, social y cultural, como asimismo para la estabilidad política, la identidad nacional y la cohesión social. Además, las empresas de hoy, que utilizan tecnología de vanguardia, no pueden avanzar sin personas capaces de incorporar habilidades analíticas, creativas y de cooperación al lugar de trabajo. La disponibilidad de trabajadores con estas características también puede tener un gran impacto sobre la capacidad de un país para atraer inversión extranjera. Los logros económicos de los países del Este de Asia, solo por dar un ejemplo, pueden atribuirse en gran medida a la calidad y nivel superiores de la educación en la región, la que actualmente se vanagloria de exhibir cuatro de los cinco más altos récords mundiales en matemáticas de octavo grado.

En los últimos años, las naciones del mundo han convenido en la importancia de medir el rendimiento educacional. Estiman que la evaluación de

^(*) Este documento fue originalmente publicado por PREAL en la serie PREAL Documents N° 11, julio de 1998.

La traducción al español fue realizada por Maritza Blajtrach Roldán.

los actuales niveles de logro y la identificación de los obstáculos al progreso les permitirá mejorar el tipo, la profundidad y el alcance de la educación que ofrecen. Las evaluaciones educacionales pueden definirse como medidas del grado en el cual se han logrado los objetivos curriculares, ya sean establecidos por las autoridades gubernamentales o por expertos nacionales e internacionales. Las evaluaciones nacionales miden el avance de las instituciones en todo el país. Difieren significativamente de los exámenes finales o de ingreso destinados a seleccionar a los alumnos para el siguiente nivel de educación. Las evaluaciones internacionales comparan los logros en el aprendizaje entre diversos países. Es obvio que el mero hecho de medir el aprendizaje de los alumnos no aumentará su rendimiento al igual que el hecho de pesar el grano no aumenta la producción agrícola. Sin embargo, es una condición necesaria para establecer metas cuantitativas, evaluando las ventajas comparativas de las estrategias de asignación de recursos y tanto como combinaciones de insumos, la asignación de recursos y los esfuerzos con miras al logro de las metas establecidas. Para cerciorarse de que las evaluaciones educacionales no solo mejoren el aprendizaje de los alumnos, sino que también sean eficaces en función de los costos, se requiere de compromiso, financiamiento adecuado, conocimientos técnicos, experiencia administrativa y habilidad política.

Las evaluaciones educacionales son solo un medio para supervisar el avance en el logro de los objetivos. También es importante evaluar los resultados cuantitativos (por ejemplo, el número de alumnos matriculados, de alumnos que han completado un nivel determinado o de alumnos que han sido promovidos), la suficiencia de los insumos, tales como los textos escolares, los profesores, los coeficientes profesor-alumno y la capacitación docente; las interacciones y la pedagogía en la sala de clases y el desempeño en el mercado laboral (por ejemplo, cuántos egresados obtienen trabajo, con qué sueldo, etc.).

América Latina no es ajena a las evaluaciones educacionales. A fines de los años setenta, el Programa de Estudios Conjuntos para la Integración Económica Latinoamericana (ECIEL) completó un estudio comparativo del aprendizaje en cinco países, utilizando los instrumentos desarrollados por la Asociación Internacional para la Evaluación del Rendimiento de los Alumnos (IEA). Diversos países de América Latina han participado en evaluaciones internacionales patrocinadas por la IEA como asimismo por el Servicio de Pruebas Educacionales (Educational Testing Service - ETS). Sin embargo, los resultados de estas evaluaciones han sido decepcionantes; los puntajes obtenidos por los países de América Latina han sido

significativamente más bajos que aquellos de los países de América del Norte, Europa y Asia. En el estudio de 1989 de la IEA (Cuadro 1), Venezuela obtuvo un puntaje más bajo que todos los demás países participantes, incluido Indonesia, Trinidad y Tobago obtuvo un puntaje algo mejor que Venezuela, pero aún muy por debajo de los países desarrollados. En un estudio de matemáticas de 1991 (Cuadro 2), implementado por el ETS, las ciudades de Fortaleza y São Paulo, Brasil, obtuvieron un puntaje más bajo que el de todos los países y ciudades participantes, con excepción de Mozambique. Como se observa más adelante en este informe, Colombia, el único país latinoamericano participante, también obtuvo un puntaje bajo en el más reciente estudio de matemáticas de la IEA.

En 1991, Costa Rica, México, Chile y Colombia contaban con sistemas de evaluación. Desde entonces, casi todos los países latinoamericanos han dado inicio a algún tipo de programa. A través de un aporte del BID, la UNESCO ha apoyado un programa regional de evaluación para los alumnos de tercer y cuarto grado en lectura y matemáticas. Además, se ha incrementado la profundidad de las evaluaciones internacionales como asimismo el número de países que participan en ellas. En este informe se analizan los avances en las evaluaciones educacionales, tanto a nivel internacional como en seis países latinoamericanos; también se comentan algunas posibles nuevas orientaciones en cuanto a políticas y programas de evaluación en la región. El análisis se basa en estudios internacionales y nacionales recientes (identificados en la bibliografía), incluidos diversos estudios preparados para una conferencia de PREAL en Río de Janeiro en diciembre de 1996.

CUADRO 1

Estudio de la habilidad de lectura de la IEA, 1989

Países de América Latina y el Caribe	Media	Desviación Standard
Trinidad y Tobago	451	79
Venezuela	383	74
Otros países		
Finlandia	569	70
Estados Unidos	547	74
Hong Kong	517	71
Singapur	515	72
España	504	78
Alemania Occidental	503	84
Indonesia	394	59

Fuente: W.B. Elley, *How in the World do Students Read?* (Newark, Delaware: International Reading Association, 1992).

CUADRO 2**Evaluación internacional del progreso educacional prueba de matemáticas y ciencias, 1991**

Prueba de matemáticas		
País	Media	Desviación Estándar
Corea	73	0,6
Taiwán	73	0,7
España	55	0,8
Estados Unidos	55	1,0
Portugal	48	0,8
São Paulo, Brasil	37	0,8
Fortaleza, Brasil	32	0,6
Maputo y Beira, Mozambique	28	0,3
Prueba de ciencias		
País	Media	Desviación Estándar
Corea	78	0,5
Taiwán	76	0,4
España	68	0,6
Estados Unidos	67	1,0
Portugal	63	0,8
São Paulo, Brasil	53	0,6
Fortaleza, Brasil	46	0,6

Fuente: Educational Testing Service, The International Assessment of Educational Progress: Learning Mathematics (Princeton, N.J., ETS, 1992).

II. RECIENTES AVANCES EN LAS EVALUACIONES EDUCACIONALES INTERNACIONALES Y NACIONALES

Programas internacionales

El acontecimiento más importante relacionado con las evaluaciones internacionales en los últimos años ha sido el Tercer Estudio Internacional de Matemáticas y Ciencias (Third International Mathematics and Science Survey - TIMSS) de la IEA. La IEA es un consorcio de evaluación que opera en todo el mundo a través de su oficina central en Amsterdam, Holanda. Es conocida por sus estudios internacionales de programas de

matemáticas, ciencias, lectura, alfabetización y ciencias sociales. Los programas de la IEA son financiados por los países participantes.

El TIMSS tiene como objetivo medir, comparar y explicar el aprendizaje en ciencias y matemáticas en 41 países. Actualmente se dispone de los resultados de los exámenes de matemáticas y ciencias para niños de los grados cuarto, séptimo, octavo, y duodécimo. Además, recientemente, el TIMSS llevó a cabo un innovador análisis sobre la “oportunidad de aprender” de los niños, que clasifica en categorías y compara los currículos, los textos escolares y la pedagogía, en la sala de clases entre los 41 países participantes. Dichos análisis demuestran que los exámenes de rendimiento no deben simplemente concentrarse en la identificación de los países que obtienen los puntajes más altos o más bajos, sino que pueden ser utilizados como una herramienta para medir el progreso educacional de un país, redefinir sus objetivos curriculares y modificar las prácticas en la sala de clases.

En 1991, once países latinoamericanos (Costa Rica, Perú, Argentina, República Dominicana, Colombia, Guatemala, Venezuela, Chile, Ecuador, Brasil y México) asistieron a una reunión regional preliminar del TIMSS, pero solo Colombia y México participaron hasta el final. Además, únicamente se informaron los puntajes de Colombia en la publicación del TIMSS. En el último minuto, México decidió no permitir la divulgación de sus puntajes. Argentina y República Dominicana participaron en el análisis de los currículos. La razón de la baja participación de los países de América Latina (y de todos los países en desarrollo) parece ser una combinación de falta de recursos técnicos y financieros adecuados como, tal vez, la preocupación de algunas autoridades ante la posibilidad de obtener los peores puntajes. En el Cuadro 3 se presentan los puntajes obtenidos en las pruebas de rendimiento de matemáticas de octavo grado del TIMSS. Entre los 41 países que informaron sus puntajes, Colombia obtuvo el puntaje más bajo de todos los países, con excepción de Sudáfrica. Los datos muestran que solo el 5% de los estudiantes colombianos obtuvieron un puntaje ubicado entre el 50% superior de los estudiantes del mundo. La amplia mayoría de los participantes fueron países desarrollados con un ingreso per cápita mucho más alto que el de Colombia y la mayoría de ellos gastaba significativamente más en su sistema educacional. A su vez, países con un ingreso per cápita similar al de Colombia —específicamente Bulgaria, Lituania, Filipinas, Rumania, Latvia, Irán, Eslovaquia, Rusia y Tailandia— obtuvieron puntajes más altos que dicho país.

CUADRO 3**Rendimiento en matemáticas en el Tercer Estudio Internacional de Matemáticas y Ciencias****Porcentaje de alumnos que lograron los niveles de referencia internacionales, 8° grado**

País	Rendimiento en matemáticas, 8° grado	Nivel del 10% superior	Nivel de la mitad superior	Diferencias de rendimiento entre el 7° y el 8° grado
Singapur	643	45	94	42
Corea	607	34	82	30
Japón	605	32	83	34
Hong Kong	588	27	80	24
Bélgica (Fl)	565	17	73	8
República Checa	564	18	70	40
República Eslovaca	547	12	64	39
Suiza	545	11	65	40
Austria	539	11	61	30
Francia	538	7	63	46
Hungría	537	11	60	35
Eslovenia	541	11	61	43
Federación Rusa	535	10	60	35
Holanda	541	10	63	25
Bulgaria	540	6	58	26
Canadá	527	7	58	33
Irlanda	527	9	57	28
Bélgica (Fr)	526	6	58	19
Austria	530	11	57	32
Tailandia	522	7	54	28
Israel	522	6	56	-
Suecia	519	5	53	41
Alemania	509	6	49	25
Nueva Zelanda	508	6	48	36
Inglaterra	506	7	48	30
Noruega	503	4	46	43
Estados Unidos	500	5	45	24
Dinamarca	502	4	47	37
Escocia	498	5	44	36
Latvia (LSS)	493	3	40	32
España	487	2	36	39
Islandia	487	1	37	27
Grecia	484	3	37	44
Rumania	482	3	36	27
Lituania	477	1	34	49
Chipre	474	2	34	28
Portugal	454	0	19	31
Irán,				
República Islámica	428	0	9	27
Kuwait	392	0	3	-
Colombia	385	0	4	16
Sudáfrica	384	0	2	7

Fuente: Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS) (Chestnut Hill, Mass. Boston College-TIMSS International Study Center, 1996).

La metodología del TIMSS permite que quienes desarrollan los currículos y quienes diseñan las políticas determinen si existe correspondencia entre los currículos, los textos escolares y la enseñanza y aprendizaje en la sala de clases. De este modo, los detalles de los resultados obtenidos por Colombia en una diversidad de temas de matemáticas, como se muestra en el Cuadro 4, permiten formarse una buena idea de las fortalezas y debilidades del currículo en el área de las matemáticas. Así, Colombia puede utilizar sus resultados en cuanto a rendimiento para comparar la calidad de sus objetivos educacionales, currículos y textos escolares con aquellos de otras naciones y, de este modo, identificar las áreas que deben ser mejoradas.

A pesar de que el principal programa del TIMSS está próximo a ser concluido, ya se encuentra en curso una serie de actividades de seguimiento. Los investigadores han comenzado a establecer “puntos de referencia” para los currículos y pruebas actuales de los países sobre la base del examen del TIMSS y sus posteriores análisis. Además, ya se ha llegado al acuerdo de repetir el TIMSS en 1999.

CUADRO 4
¿Cómo varía el rendimiento en Colombia
de acuerdo con el currículo que se pretende aplicar
y el currículo realmente implementado?

Contenido	Escala internacional
Perímetro, área y volumen	1
Patrones, relaciones y funciones	41
Geometría: transformaciones	41
Números reales, sus propiedades y subcategorías*	12
Números racionales y sus propiedades*	12
Fracciones y decimales	40
Ecuaciones y fórmulas*	40

* Enfatizados en los textos escolares colombianos y por los profesores.

Fuente: Variations in Mathematics Performance in Colombia in Accordance with the Intended and Implemented Curriculum, presentación de Gilbert Valverde, “Las políticas educacionales nacionales y la participación en estudios internacionales”, reunión de PREAL, Río de Janeiro, diciembre de 1996.

La experiencia de Estados Unidos

Estados Unidos se encuentra a la vanguardia en el desarrollo de una amplia variedad de evaluaciones educacionales destinadas a medir el desempeño institucional, establecer niveles de competencia mínimos para el egreso de la escuela y establecer puntos de referencia para medir el rendimiento educacional en comparación con los estándares internacionales. El Recuadro 1, que presenta información acerca de cuatro programas de evaluación que se aplican en el Condado de Montgomery, Maryland, muestra la extensión, la complejidad y, tal vez, los excesos de los programas de evaluación desarrollados en Estados Unidos. El primer programa consiste en una prueba desarrollada por el condado, que entrega a los padres información acerca de la medida en que sus hijos que cursan los grados tercero a octavo dominan los principales elementos del currículo a nivel del grado correspondiente. Sirve como una herramienta de diagnóstico que permite identificar las debilidades y fortalezas de un alumno e incluye preguntas de selección múltiple, preguntas de interpretación abierta y preguntas basadas en el rendimiento. El segundo programa, el Programa de Evaluación del Rendimiento Escolar de Maryland, evalúa la medida en que cada escuela cumple con los estándares de rendimiento establecidos por el estado para los grados tercero, quinto y octavo. Utiliza principalmente preguntas de selección múltiple o solo pide respuestas breves. El tercer programa de evaluación, las Pruebas Funcionales de Maryland, consiste en un “examen de competencia mínima” en lectura, matemáticas, redacción y estudios cívicos, que debe ser aprobado por los alumnos para recibir un certificado válido de educación secundaria. El cuarto programa de evaluación consiste en las Pruebas Globales de Habilidades Básicas, recientemente lanzado y que se aplicará a una pequeña muestra de alumnos que cursan los grados segundo, cuarto y sexto. Servirá como un indicador de los niveles de logro de las habilidades básicas en relación con las normas nacionales e internacionales.

En virtud de su historia y de sus leyes, y en contraste con la mayoría de los países, Estados Unidos no cuenta con un currículo obligatorio a nivel nacional. Y dado que la mayoría de los estados solo entregan pautas generales, cada distrito escolar tiene considerable libertad para establecer el currículo. De acuerdo con un reciente informe de 1997 del Centro Nacional de Investigaciones de Estados Unidos (U.S. National Research Center), los currículos de los estados generalmente no están definidos adecuadamente y no se los aplica a nivel local. Como resulta-

RECUADRO 1**LAS EVALUACIONES EDUCACIONALES EN ESTADOS UNIDOS****Programa de evaluaciones del Condado de Montgomery, Maryland**

El sistema de escuelas públicas del Condado de Maryland, una zona suburbana cercana a Washington, D.C., utiliza una amplia batería de pruebas para medir el rendimiento de sus estudiantes. La siguiente descripción del programa de evaluaciones del condado ha sido extraída de un volante informativo enviado a los padres de los niños matriculados en las escuelas del mismo.

EVALUACIONES DEL CONDADO DE MONTGOMERY

Pruebas de Criterio (CRT). Las Pruebas de Criterio (CRT), desarrolladas por las Escuelas Públicas del Condado de Montgomery, son aplicadas cada primavera a los alumnos que cursan el tercer y octavo grado con el fin de medir la calidad del aprendizaje y el grado en que están aplicando las habilidades enseñadas en los currículos de matemáticas y lectura/idioma/arte.

Además del formato tradicional basado en la selección múltiple para evaluar los conocimientos de los contenidos concretos, las pruebas de criterio incluyen también preguntas de interpretación abierta y preguntas de desarrollo. Estas preguntas reflejan el compromiso, tanto a nivel local como del estado, con respecto a instrumentos más modernos no basados en la selección múltiple que evalúan habilidades más ricas y complejas, tales como la habilidad de comunicarse a través de la escritura y de resolver problemas de la vida diaria.

En el otoño, los padres reciben un informe individualizado que muestra el rendimiento de sus niños en las preguntas de selección múltiple de las pruebas de criterio y ofrece una comparación de dicho rendimiento con los puntajes a nivel de la escuela y del condado. El informe incluye los puntajes de años anteriores para ayudar a estimar el progreso del alumno y de la escuela, identificar las fortalezas y debilidades, y evaluar el efecto de cualquier medida tomada con el fin de mejorar el rendimiento durante el año.

Los planes actuales contemplan el uso de una escala única que combina el rendimiento de

los alumnos en las preguntas de selección múltiple e interpretación abierta para informar a las escuelas y a los padres lo que los alumnos saben y lo que son capaces de hacer.

LAS EVALUACIONES**DEL ESTADO DE MARYLAND****El Programa de Evaluación del Rendimiento Escolar de Maryland (MSPAP).**

Estas pruebas utilizan preguntas de respuesta breve y actividades complementarias para determinar si las escuelas han cumplido con los estándares de rendimiento del estado en las áreas de lectura, redacción, uso del lenguaje, matemáticas, ciencias y ciencias sociales. Los alumnos de tercero, quinto y octavo grado rinden estas pruebas cada primavera. En la actualidad, no se informan los puntajes individuales. Los resultados de las escuelas se publican en el mes de enero.

Las Pruebas Funcionales de Maryland.

Estas pruebas evalúan las habilidades de competencia mínimas en lectura, matemáticas, redacción y estudios cívicos. La aprobación de estas pruebas es un requisito para recibir el diploma de educación secundaria de Maryland. Los alumnos dan una prueba de lectura y matemáticas en el séptimo grado, de redacción en el octavo grado y de estudios cívicos en el décimo grado. Los alumnos tienen varias oportunidades de aprobar estas pruebas durante el año. Se informan los puntajes individuales obtenidos por los alumnos en cada prueba.

Las Pruebas Globales de Habilidades Básicas (CTBS).

Maryland utiliza esta prueba como un indicador de los niveles de logro de las habilidades básicas en lectura, idioma y matemáticas en relación con las normas nacionales. A partir de la primavera de 1997, Maryland aplicará esta prueba en los grados segundo, cuarto y sexto cada año. Las escuelas del Condado de Montgomery administrarán esta prueba a una muestra de 250 alumnos por grado. Los puntajes de cada alumno serán informados a los padres en el otoño.

Fuente: "Celebrating Knowledge through Assessment", Escuelas Públicas del Condado de Montgomery (mimeo), 1996.

do, lo que se enseña realmente en la sala de clases varía enormemente de una escuela a otra y de un distrito a otro, y los alumnos y los profesores a menudo tienen una noción exagerada del nivel de logro que obtienen. Debido a estos problemas, las autoridades han hecho un gran esfuerzo por establecer estándares más sólidos en cuanto a aprendizaje y logros a nivel de los estados, como asimismo estándares nacionales posiblemente voluntarios. Por otra parte, varios estados –Maryland entre ellos– están intentando basar sus currículos y pruebas en los estándares internacionales, especialmente en relación con el TIMSS. Cabe señalar que las dos principales asociaciones gremiales de profesores, especialmente la Federación Norteamericana de Profesores (American Federation of Teachers - AFT), han apoyado permanentemente los esfuerzos a nivel nacional y estatal tendientes al mejoramiento de la evaluación y la calidad de la educación.

La experiencia francesa

Desde mediados de los años ochenta, en Francia se ha desarrollado un sofisticado sistema de evaluación de los alumnos. Si bien Francia cuenta con un currículo obligatorio a nivel nacional, las escuelas elementales y secundarias tienen un alto grado de libertad en el uso de sus fondos y sus técnicas pedagógicas. La responsabilidad por la supervisión y evaluación de la educación está en manos de los más de 200 técnicos de la Direction de l'Evaluation et de la Prospective (DEP), una división del Ministerio de Educación de Francia.

Las principales autoridades de gobierno están firmemente comprometidas con el establecimiento de un sistema de medición de los resultados educacionales basado en estándares claros y en una difusión completa y abierta de los resultados, asegurando de este modo la independencia de la DEP.

Según el sistema francés, los estándares de logro esperados se establecen de manera centralizada y se toman pruebas anuales a todos los alumnos de grados específicos de la educación primaria y secundaria del nivel inferior. El examen de bachillerato (egreso) administrado a los alumnos del nivel superior de la educación secundaria también se utiliza con propósitos de evaluación. La DEP proporciona retroalimentación detallada y fácil de comprender a todas las partes involucradas: la opinión pública, los padres, los alumnos, los profesores y administradores y el ministro de Educación. El organismo ha desarrollado también una

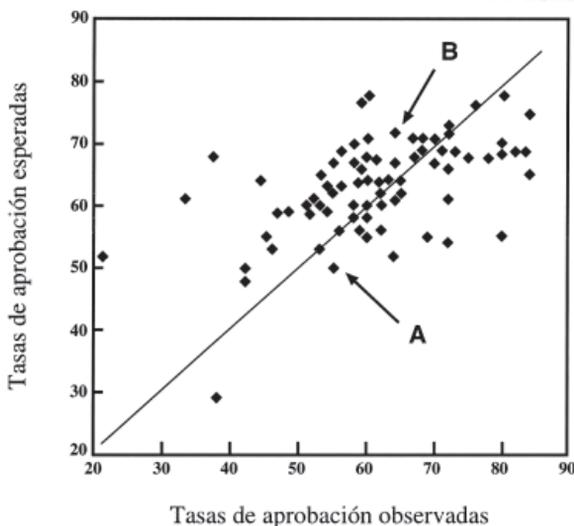
amplia variedad de materiales de enseñanza y aprendizaje opcionales orientados al rendimiento de los alumnos y a las fortalezas y debilidades de los profesores. Al igual que diversos organismos en Estados Unidos, la DEP ha intentado medir el rendimiento escolar en comparación con los estándares internacionales.

Como ocurre en todo el mundo, las escuelas de las comunidades más pobres de Francia obtienen puntajes más bajos que las escuelas que se encuentran en las áreas de mayor riqueza. La DEP ha intentado identificar las escuelas que se encuentran en áreas más pobres y están exhibiendo un buen rendimiento con el fin de definir de mejor manera el “valor agregado” de la escolaridad. En el Gráfico 1 se presenta un método

GRAFICO 1

Medición del valor agregado en la educación: un ejemplo de Francia

Las escuelas “eficaces” producen resultados mejores que lo esperado de una determinada población de alumnos



Fuente: Adaptado de “L’état de l’école” n° 5, octubre de 1995. Direction de l’Evaluation et de la Prospective (DEP), Ministère de l’Education Nationale, de l’Enseignement Supérieur, de la Recherche, et de l’Insertion Professionnelle (MEN), pág. 51. Los datos utilizados para preparar esta ilustración no corresponden necesariamente a los resultados reales de las pruebas.

estadístico de medición del desempeño de las escuelas en relación con el nivel socioeconómico de sus alumnos. Las escuelas que muestran un rendimiento mejor de lo esperado en base a su población escolar se definen como escuelas “eficaces”. Estudiando a dichas escuelas, las autoridades debieran ser capaces de determinar por qué son eficaces y luego establecer condiciones similares en las escuelas que presentan un rendimiento más bajo. Este tipo de análisis ya se está llevando a cabo en algunas partes de Estados Unidos (véase, por ejemplo, una descripción del “valor agregado” en el sistema escolar de Dallas, según lo descrito en Álvarez y Ruiz-Casares, 1997).

En este gráfico se presenta un método utilizado por el gobierno francés para medir el desempeño de las escuelas en relación con el nivel socioeconómico de sus alumnos. Cada punto representa una escuela y el diagrama muestra las tasas de aprobación reales (observadas) y esperadas en el examen de bachillerato. Todos los liceos que están ubicados bajo la línea ofrecen a sus alumnos mayores probabilidades de aprobar el examen comparado con lo que se podría esperar en función de la edad y el origen social de los alumnos.

La escuela A ha obtenido valor agregado en términos de éxito en el examen, debido a que su tasa observada es más alta que su tasa esperada (está bajo la línea). Este no es el caso de la escuela B, que no ha alcanzado su nivel de éxito esperado, a pesar de que su tasa de aprobación observada es realmente más alta que la de la escuela A.

Los países en desarrollo

Los países en desarrollo también están haciendo esfuerzos por establecer sistemas de evaluación educacional a nivel nacional. Aparte de América Latina, se han descrito y analizado sistemas de evaluación en Tailandia, Egipto, Corea del Sur y Jordania. A pesar de ser un país pequeño y relativamente pobre, Jordania ha desarrollado un programa de evaluación sistemático y técnicamente competente. Su éxito en esta área puede explicarse por las siguientes razones: (a) la participación directa de las más altas autoridades (por ejemplo, el príncipe heredero) y su compromiso de largo plazo con el mejoramiento de la calidad de la educación y la información abierta de los resultados; (b) el establecimiento de un organismo independiente y bien financiado, aparte del Ministerio de Educación, responsable de las evaluaciones; y (c) un fuerte liderazgo técnico, unido a asistencia del exterior. En 1991, Jordania llevó a cabo varias series de pruebas basa-

das en las evaluaciones internacionales y desarrolladas por el Servicio de Pruebas Educativas (la prueba “IAEP”) y actualmente está proporcionando asistencia técnica en esta área a los países vecinos. Una serie de evaluaciones aplicadas recientemente identificó significativos mejoramientos en el aprendizaje de los alumnos como resultado de los esfuerzos de reforma de los currículos escolares y la capacitación de los profesores que se vienen realizando desde hace una década en Jordania.

III. LA EXPERIENCIA LATINOAMERICANA

Con anterioridad a 1991, Chile y México eran los únicos países de la región que contaban con bastante experiencia en el área de las evaluaciones. El programa de evaluación de Chile estaba siendo aplicado desde 1980 y a pesar de que México contaba con un programa de evaluación desde hacía varios años, las autoridades hicieron poco o nada por difundir los resultados. Costa Rica llevó a cabo evaluaciones nacionales entre 1986 y 1990. Colombia contaba con un sistema de pruebas nacionales de larga data utilizado principalmente para la selección de los postulantes a la educación superior.

En los años siguientes, otros países latinoamericanos comenzaron también a avanzar en esta dirección. Hacia 1996, su experiencia se había incrementado considerablemente y en la actualidad prácticamente todos los países de América Latina han iniciado algún tipo de programa de evaluación. En este informe se extraen conclusiones a partir de los progresos hechos en Chile, Costa Rica, México, Argentina, Brasil y Colombia. También se describe el programa regional UNESCO/OREALC.

Chile

Como ya se ha mencionado, Chile ha estado involucrado desde hace largo tiempo en el desarrollo de evaluaciones educacionales. Su programa fue concebido en 1978, cuando el Ministerio de Educación solicitó a la Pontificia Universidad Católica que diseñara e implementara un sistema de información para la educación. En 1988, con la transferencia de las escuelas públicas a las municipalidades, el programa recibió el nombre de Sistema de Medición de la Calidad de la Educación (SIMCE).

La función del SIMCE es ayudar al Ministerio de Educación y a las autoridades regionales y provinciales a supervisar el sistema de educación,

a evaluar a las escuelas y a apoyar la capacitación docente en el servicio. El programa evalúa a los niños que cursan los grados cuarto y octavo en las áreas de español y matemáticas y al 10% de ellos en ciencias naturales, historia y geografía. También evalúa el desarrollo personal y las actitudes de los alumnos, las actitudes y antecedentes de los profesores y los padres, y la eficacia de las escuelas. Las evaluaciones de ambos grados se aplican en años alternados. A partir de 1991, el Ministerio de Educación se hizo cargo de la administración del programa.

A contar de 1988, el programa se ha vuelto más eficaz y eficiente, luego de la introducción de mejoramientos en la capacidad técnica, el sistema de procesamiento computarizado y la administración. Actualmente, los puntajes se entregan a las escuelas en forma mucho más rápida y se han simplificado los informes para asegurar la fácil comprensión de los resultados. Estos esfuerzos parecen estar teniendo impacto en la planificación pedagógica en muchas escuelas. Quienes diseñan los currículos y los materiales pedagógicos, por ejemplo, están dando énfasis al dominio de los objetivos que parecen ocasionar problemas a los alumnos. Algunos padres están utilizando los resultados del SIMCE para seleccionar para sus hijos las escuelas que muestran un mejor rendimiento. El costo del programa es de aproximadamente 5 dólares por alumno, lo que es comparable con los estándares internacionales.

Los puntajes en el SIMCE también han constituido la base de un programa pedagógico dirigido inicialmente a 900 y posteriormente a 1.200 de las escuelas chilenas que exhiben el peor rendimiento. Las escuelas que participan en este programa han recibido materiales pedagógicos, bibliotecas, libros, infraestructura y capacitación docente en servicio. Dependiendo de su condición inicial, se da preferencia a algunas escuelas en el otorgamiento de subvenciones para programas de mejoramiento locales. Por otra parte, las escuelas que han mejorado sus puntajes de un año al otro han recibido recompensas financieras.

Las pruebas del SIMCE revelan lo siguiente: aquellas escuelas que tienen niños que provienen de familias pobres y con un bajo nivel de educación o de áreas rurales obtienen los peores puntajes en las pruebas del SIMCE; las escuelas municipales públicas y las escuelas rurales obtienen puntajes más bajos que las escuelas privadas, especialmente aquellas instituciones con una larga tradición; las escuelas privadas muestran un rendimiento mejor que las escuelas públicas, aún después de controlar el factor del nivel socioeconómico de los padres y, finalmente, los puntajes de las 900 escuelas muestran un mejoramiento significativo durante

los últimos años. Entre los actuales problemas se encuentra el hecho de que si bien muchas escuelas están haciendo uso de los resultados del SIMCE con el fin de mejorar las condiciones locales, aún se espera que sean las autoridades centrales las que inicien las acciones correctivas. Algunas escuelas han informado un aumento del número de alumnos provenientes de hogares en situación desmedrada, en un aparente esfuerzo por mostrar que sus logros relativos (valor agregado) han mejorado. Las mediciones del área afectiva no han tenido éxito y quizás debieran ser descartadas. Tal vez también sería apropiado reducir la proporción de pruebas al universo y recurrir más al muestreo. Finalmente, debido a problemas técnicos, la comparabilidad de los resultados de un año a otro es inadecuada.

En general, Chile cuenta en la actualidad con el sistema de evaluación más amplio y mejor administrado de América Latina, y el SIMCE ha servido como una poderosa herramienta para la implementación de un programa de reforma que promueve la descentralización, la responsabilidad por los resultados y el aumento del aprendizaje y la escolaridad. Las autoridades de gobierno buscan mejorar aún más el sistema. En particular, actualmente están considerando agregar pruebas de desarrollo en las evaluaciones (hoy en día todas las pruebas son del tipo selección múltiple). Dado que aún no se comprenden cabalmente las causas del bajo rendimiento, se encuentra en marcha un esfuerzo de investigación más sofisticado. Finalmente, las autoridades chilenas han decidido participar en el estudio TIMSS de seguimiento programado para 1999 como una manera de establecer puntos de referencia para la comparación del rendimiento de sus alumnos con el resto del mundo.

Costa Rica

Costa Rica ha estado involucrada en las evaluaciones educacionales desde 1986. Su actividad en esta área ha pasado por tres etapas. Durante la primera etapa –entre 1986 y 1990– los objetivos planteados fueron medir el grado de aprendizaje de los conceptos básicos por parte de los niños y jóvenes, estimular a los padres y profesores a utilizar el tiempo de enseñanza en forma más eficaz, promover una discusión nacional acerca de la calidad de la educación, destacar el hecho de que todos los costarricenses son responsables del mejoramiento de la calidad de la educación y demostrar la importancia del restablecimiento de los exámenes de certificación nacional al final de la educación secundaria. El programa fue

formulado en su mayor parte por el Instituto de Investigación para el Mejoramiento de la Educación Costarricense (IIMEC), una institución autónoma de la Universidad de Costa Rica. Inicialmente, el programa introdujo pruebas de español y matemáticas para todos los alumnos de los grados tercero, sexto, noveno, undécimo y duodécimo. En 1987 y 1988, las pruebas se ampliaron cubriendo ciencias físicas, ciencias sociales, inglés y francés. En 1988, se estableció un examen de egreso de la escuela secundaria. Este primer programa generó una controversia pública, debido al bajo nivel de rendimiento de los alumnos y se volvió a las pruebas globales objetivas.

Durante el período 1988 a 1993, la concentración se dirigió principalmente al examen de egreso de la escuela secundaria. Las pruebas, preparadas por el IIMEC y aplicadas por el Ministerio de Educación representaban el 60% de la nota final y los puntajes se asignaron a nivel local. Cuando un estudio de 1991 indicó que el 30% de las pruebas habían sido evaluadas de manera incorrecta, casi en todos los casos en favor de los alumnos, las autoridades decidieron eliminar la mayoría de las preguntas de desarrollo y asignar el puntaje de casi todas las pruebas de manera centralizada utilizando lectores ópticos. Además, los exámenes de sexto grado, preparados por las autoridades regionales, representaban el 50% de la nota final, pero no se tomaron las medidas suficientes como para asegurar la confiabilidad y la validez de los resultados informados por las autoridades regionales.

A contar de 1993, bajo un nuevo gobierno y Ministro de Educación, Costa Rica se comprometió firmemente con las evaluaciones y recibió ayuda en esta área a través de un proyecto financiado por el Banco Mundial. Se volvieron a establecer las pruebas de diagnóstico para los grados tercero, sexto y noveno en cuatro materias, se efectuaron otros tipos de evaluaciones y se fortaleció el IIMEC mediante nuevo equipo y personal. El nuevo sistema consiste en la aplicación de una diversidad de pruebas y evaluaciones, la mayoría de las cuales son preparadas por el IIMEC.

Entre las evaluaciones formativas se incluyen las siguientes:

- 1) pruebas de rendimiento diagnósticas para tercer grado en las materias básicas. Estas fueron preevaluadas en 1995 y aplicadas a una muestra amplia en 1996.
- 2) evaluaciones diagnósticas iniciales, aplicadas al 10% de los niños que ingresan al primer grado para obtener información acerca de su condición física, cognitiva y socioemocional. Los resultados serán utilizados como una manera de establecer pautas para la entrega de expe-

- riencias de aprendizaje apropiadas en la educación preescolar y en primer grado.
- 3) pruebas de habilidades de resolución de problemas, aplicadas en 1996 a una muestra de alumnos del noveno grado. El objetivo de estas pruebas era medir las capacidades cognitivas en esta área como asimismo las variables socioeconómicas y académicas que influyen en el rendimiento.
 - 4) evaluaciones de la capacidad física, para medir la aptitud física de una muestra nacional de niños que cursaban los grados tercero, sexto y noveno.

También se han desarrollado diversas pruebas sumativas. Se aplica un examen de egreso preparado a nivel nacional a todos los alumnos que terminan el noveno grado en todas las materias básicas. Este examen consiste en preguntas de selección múltiple cuyo puntaje se asigna por medio de lectores ópticos, junto con preguntas de desarrollo que son evaluadas por profesores especialmente capacitados. Los resultados representan el 25% de la nota del segundo semestre. A contar de 1988, en Costa Rica también se ha aplicado un examen de egreso de la educación secundaria, preparado y administrado a nivel regional. Representa el 60% de la calificación final y consta tanto de preguntas objetivas como de desarrollo. Además, las oficinas regionales del Ministerio de Educación preparan y administran exámenes de egreso de sexto grado, los que, sin embargo, difieren en calidad, validez y confiabilidad.

En la etapa de prueba inicial (1988), los niños mostraron un rendimiento muy por debajo de las expectativas del currículo nacional. Las escuelas privadas urbanas obtuvieron los mejores puntajes en estos exámenes. Si bien el IIMEC distribuyó información resumida a cada escuela, las instituciones no hicieron ningún intento explícito de utilizar los resultados. Sin embargo, una encuesta hecha a los profesores en 1989 mostró que el 70% de estos tenían conocimientos acerca del material y aproximadamente el 35% de ellos utilizaban el material de una manera u otra en la sala de clases.

En la segunda etapa, la información acerca de los resultados de los exámenes de egreso de la educación secundaria fue entregada a algunos alumnos como también a algunas escuelas secundarias. El uso de dicha información quedaba a completa discreción de la escuela.

Se espera que la tercera etapa refleje un enfoque más sistemático en lo que respecta a la divulgación, utilización y retroalimentación. Se prepararán informes para los diferentes niveles y se solicitará a las unidades

educacionales regionales que incorporen los resultados de las evaluaciones en sus planes educacionales. También se hará el intento de medir el grado en el cual ha mejorado la educación. A pesar de estos avances, los padres no reciben ninguna otra información aparte de los puntajes de los alumnos, no se emiten comunicados de prensa especiales sobre educación, los resultados no se utilizan para preparar materiales específicos destinados a corregir las debilidades en el aprendizaje y las instituciones de capacitación no han recibido ninguna pauta ni información específicas acerca de las pruebas. Sin embargo, se entregarán informes a las escuelas, especialmente a nivel de primer grado y educación preescolar, junto con sugerencias para enriquecer el currículo.

En general, en Costa Rica los esfuerzos de evaluación se han visto dificultados por un enfoque “intermitente”. Si bien el actual ministro de Educación apoya decididamente el programa, algunos educadores y líderes políticos no lo hacen, y no está claro si esta situación continuará bajo la nueva administración, que asumió en mayo de 1998. Dado su reducido tamaño, es posible que Costa Rica haya emprendido un programa demasiado ambicioso. Quizá sería más productivo reducir el número de pruebas y concentrar los esfuerzos en la divulgación y utilización de los resultados. El examen secundario de alta exigencia de Costa Rica, que representa el 25% de la nota final como un medio de aumentar el rendimiento en este nivel puede ser de interés para otros países latinoamericanos que deseen aumentar el aprendizaje.

Colombia

En 1986, el gobierno de Colombia estableció al interior del Ministerio de Educación un departamento responsable de la evaluación de las instituciones y programas educacionales. Luego, en 1990, puso en marcha un sistema nacional de evaluación con el propósito de construir un modelo de evaluación operacional que permitiera tomar decisiones tendientes al mejoramiento de la calidad. Además, se esperaba que el sistema evaluara los conocimientos de los alumnos y determinara de qué manera estaban siendo motivados por los profesores, las escuelas y los materiales pedagógicos, generara un mayor sentido de la importancia de la escolaridad, comunicara este sentido a toda la sociedad, proporcionara el marco teórico para los conceptos tales como la calidad y la evaluación del rendimiento educacional y apoyara las investigaciones que permitieran al país lograr estos objetivos. El programa fue llevado a cabo por un equipo

interinstitucional, conformado por profesionales del Ministerio de Educación, la Universidad Pedagógica Nacional, el Centro de Estudios Sociales y el Instituto SER.

A partir de 1968, el Instituto Colombiano para el Mejoramiento de la Educación Superior (ICFES) ha sido responsable de la formulación de un examen nacional de ingreso a las instituciones de educación superior de la nación. En 1980, estos exámenes pasaron a ser obligatorios y, de ahí en adelante, fueron aplicados por el Servicio Nacional de Pruebas del ICFES. El programa, que se centra en las pruebas y el ingreso de los alumnos a las instituciones de educación superior más que a la evaluación del sistema educacional en su conjunto o sus subsistemas, ha ayudado a crear una competencia técnica en el área de la administración de pruebas.

A diferencia de lo que se hace en Chile y Costa Rica, en Colombia se da énfasis al muestreo y a la investigación como una manera de identificar las causas del bajo rendimiento. Por lo tanto, no se entrega información a cada padre, alumno o escuela relativo al rendimiento observado. Colombia está considerando actualmente ampliar la muestra considerada con el fin de entregar retroalimentación a un mayor número de municipalidades y escuelas.

El sistema colombiano comenzó con una prueba a los alumnos de tercer y quinto grado en las áreas de matemáticas y español. Posteriormente, se desarrollaron pruebas en las áreas de ciencias naturales y ciencias sociales para los grados séptimo y noveno. Estas pruebas median explícitamente habilidades de razonamiento de orden superior tales como el uso de algoritmos y las habilidades de resolución de problemas en matemáticas o la habilidad de extraer el significado a partir de material escrito en español.

Cuando se miden los logros cognitivos de los alumnos de tercer y quinto grado en comparación con el currículo nacional, los resultados son decepcionantes. Pocos alumnos son capaces de utilizar las operaciones básicas requeridas para resolver problemas concretos y son pocos los que comprenden cabalmente lo que leen, ya sea en un sentido crítico o reflexivo. Los mejores puntajes se obtienen en las escuelas urbanas y privadas y en algunas regiones. Incluso un estudio que controlaba el factor socioeconómico, sugirió que el hecho de asistir a una escuela privada otorga una ventaja en términos de logros. El rendimiento fue más alto en aquellos alumnos que habían asistido a la educación preescolar, no habían repetido, no tenían muchas ausencias, tenían libros en sus hogares y cuyos padres

tenían un mayor nivel de educación. Si se contaba con textos escolares y se los utilizaba, el rendimiento era más alto. Los alumnos que tenían profesores mejor capacitados, más textos escolares, docentes de sexo femenino y asistían a escuelas completas también exhibían un buen rendimiento. Las pruebas confirmaron también que los alumnos de la “Escuela Nueva”, un innovador programa para las pequeñas escuelas rurales, mostraban un rendimiento significativamente mejor que los alumnos de las escuelas rurales que no participaban en el programa.

Las conclusiones de estas evaluaciones han influido en los objetivos y el contenido de la nueva ley de educación, aprobada en 1994. Sin embargo, la información obtenida a partir de las actuales evaluaciones basadas en una muestra presta poca utilidad a las municipalidades y escuelas específicas, que se han vuelto cada vez más responsables de sus escuelas bajo el nuevo sistema de educación descentralizada. Si bien las actuales autoridades estiman que sería demasiado caro ampliar la evaluación a todos los alumnos, Colombia pretende trabajar con una muestra más amplia y espera establecer una base de datos y preguntas que pueda ser utilizada directamente por las escuelas y las autoridades educacionales. Como se mencionara anteriormente, Colombia es el único país latinoamericano que participó completamente en el programa del TIMSS. Si bien los resultados fueron decepcionantes, cabe aplaudir la voluntad de Colombia de aceptar el escrutinio externo. Actualmente, Colombia está intentando utilizar los resultados del TIMSS para reforzar el sistema de evaluación nacional y para reformar el currículo.

El programa de evaluación colombiano ha incluido un fuerte elemento de investigación que ha influido en las políticas globales, específicamente en la redacción de la Ley de Educación de 1994, y ha confirmado el éxito de la “Escuela Nueva”. Sin embargo, los resultados aún no han sido utilizados sistemáticamente para mejorar el rendimiento de las escuelas mismas ni para reformar el currículo. En particular, el currículo de matemáticas y ciencias requiere una reforma. Al mismo tiempo, la asociación entre las instituciones públicas y privadas ha funcionado bien.

Brasil

Brasil comenzó tardíamente el desarrollo de herramientas de evaluación a nivel nacional. En 1990 se estableció el Sistema Nacional de Evaluación de la Educación Básica (SAEB) y solo se comenzó a tomar mues-

tras en 1993 y 1995. Uno de los objetivos del SAEB ha sido estimular a los estados y municipalidades a iniciar sus propias evaluaciones. En particular, los estados de Paraná y Minas Gerais han iniciado programas de evaluación recientemente.

En 1995, el SAEB anunció una serie de innovaciones. Su estudio incluiría tanto la educación secundaria como las instituciones privadas. También adoptó métodos de medición más sofisticados, introdujo instrumentos que entregarían información acerca de los antecedentes de los alumnos y redujo el tiempo de demora de la publicación de los resultados. El estudio de 1995 se centró en los grados cuarto y octavo de la educación básica y en el tercer grado de la educación secundaria.

Los resultados del estudio, publicados en 1995, indicaron que se evaluaron 90.499 alumnos de cuarto y octavo grado de la educación básica y segundo y tercer grado de la educación secundaria. Esta muestra de alumnos fue tomada de 2.289 escuelas públicas y 511 escuelas privadas. Las preguntas del estudio se basaron en el nivel de aprendizaje esperado por los profesores y especialistas en educación. En el área de lectura se evaluó el rendimiento de los alumnos en comprensión, extensión y examen crítico del significado. En el área de las matemáticas, el estudio se centró en tres categorías: comprensión de los conceptos, comprensión y aplicación de los procedimientos, y resolución de problemas. En el análisis estadístico, la principal tarea fue determinar los niveles de rendimiento esperados y reales de los niños en diversos grados.

En la prueba de 1995, los niños de todo el país obtuvieron un puntaje significativamente inferior a los niveles que los profesores y especialistas esperaban de ellos. En matemáticas, solo el 21% de los estudiantes que cursaban el cuarto grado obtuvieron un puntaje sobre el nivel esperado, solo el 15% de los alumnos de octavo grado obtuvieron un puntaje igual o superior al nivel esperado y solo el 4% de los estudiantes secundarios obtuvieron un porcentaje superior al nivel esperado. En idioma, los puntajes fueron aún más bajos. Solo el 22% de los alumnos de cuarto grado, el 14% de los alumnos de octavo grado y el 1% de los alumnos de educación secundaria obtuvieron puntajes superiores al nivel esperado. Los puntajes más altos se observaron en el sur, el sudeste y el centro-oeste del país. En el norte y el noreste se observaron los puntajes más bajos. Los alumnos de las ciudades principales obtuvieron mejores puntajes que aquellos del interior. Los niños que tenían padres con mayor nivel de educación y aquellos que asistían a escuelas privadas también obtuvieron puntajes más altos. Los estudiantes que asistían a las escuelas

nocturnas obtuvieron puntajes inferiores. Los alumnos de mayor edad obtuvieron puntajes más bajos que los alumnos más jóvenes. Los estudiantes blancos y orientales obtuvieron puntajes más altos que aquellos de raza mixta o negros. Y los alumnos con profesores mejor capacitados obtuvieron puntajes más altos que aquellos con profesores con menor capacitación. Estos resultados tienen importantes implicaciones en el área de las políticas educacionales, pero las autoridades recién están comenzando a tomarlas en consideración. Aún más, todavía no han incorporado técnicas más sofisticadas en sus análisis, tales como el control del factor socioeconómico en los alumnos que ingresan.

Recientemente, en el estado de São Paulo se inició un amplio programa de evaluación. Se evalúa a todos los alumnos de grados seleccionados de la educación básica y del nivel inferior de la educación secundaria en las áreas de matemáticas e idiomas, y se envían los resultados a los padres, profesores y escuelas. Se entrega un análisis de las preguntas en las cuales los alumnos han obtenido el peor puntaje. Se incluyen “preguntas ancla” para asegurar la compatibilidad de los resultados de un año al otro. El programa de São Paulo se aplica principalmente bajo contrato con organismos de pruebas privados sin fines de lucro. También parece ser que el programa de evaluaciones de Minas Gerais está avanzando en la utilización de sus resultados, ya que se están informando los puntajes individuales a las escuelas.

De manera bastante independiente de lo anterior, en virtud de una ley aprobada en 1995, el Ministerio de Educación desarrolló un sistema para evaluar las instituciones de educación superior. El propósito de este sistema es informar a los estudiantes y a la sociedad en general acerca de la calidad de las instituciones de educación superior. Las escuelas que obtengan un bajo puntaje tendrán que diseñar programas para fortalecer la calidad, los que serían apoyados por el gobierno federal.

En resumen, el programa de evaluación brasileño todavía se encuentra en sus primeras etapas. Las autoridades brasileñas están recién comenzando a considerar de qué manera podrían utilizarse las evaluaciones para mejorar las políticas educacionales y el currículo. Existe la necesidad de llevar a cabo análisis más rigurosos para identificar los factores que afectan el rendimiento. Dado que la educación brasileña es altamente descentralizada y no se rige por un currículo nacional, el gobierno está considerando establecer algunos estándares nacionales voluntarios como asimismo una prueba que permita medir si estos han sido alcanzados. El programa de evaluación de la educación superior, el primero de este tipo en la región, constituye una importante innovación.

México

En 1970, México estableció una oficina en la Unidad de Planificación Educacional de la Secretaría de Educación Pública, que finalmente se transformó en la Subdirección de Evaluación y Acreditación, con el propósito de estudiar las características y la calidad del sistema educacional del país. Los profesionales involucrados llevaron a cabo estudios de aptitud en niños de sexto grado de la educación básica y establecieron un examen de ingreso a la educación secundaria.

Durante el período 1976 a 1982 la Subdirección investigó el aprendizaje en una muestra representativa de alumnos que cursaban el cuarto y quinto grado. Los resultados de esta evaluación aparecieron en publicaciones científicas y especializadas, no siendo divulgadas de ninguna otra manera, y las autoridades les prestaron poca atención. En efecto, la información con respecto a la evaluación se transformó en un “secreto de Estado” solo conocido por un reducido número de profesionales de la Secretaría. Este enfoque dificultó el desarrollo técnico como asimismo la utilización para el diseño de políticas.

Durante el período 1983 a 1988, en México se desarrolló un examen para egresados de las escuelas de capacitación de profesores. Posteriormente, en 1989, se tomó la decisión de aplicar el concepto de evaluación de manera más amplia con el fin de mejorar la enseñanza y el aprendizaje y de publicar los resultados. En 1992 el gobierno federal y las Asociaciones Gremiales Nacionales de Profesores acordaron un programa tendiente a modernizar la educación básica a través de su descentralización a nivel de los estados, dejando en manos del gobierno federal la responsabilidad de medir y evaluar el aprendizaje y asegurar la calidad de la educación básica y la capacitación de los profesores. Con este propósito, la Secretaría de Educación Pública se comprometió a apoyar las evaluaciones de los profesores, las evaluaciones en la sala de clases y aquellas realizadas a nivel nacional.

En 1994, después de cinco años de evaluaciones de la calidad de la educación en México, la Secretaría emitió un informe acerca de los conocimientos y las habilidades de 480.000 profesores como asimismo el rendimiento de 2,8 millones de niños de educación básica y secundaria. Sus principales conclusiones fueron que los niños que habían asistido a la educación preescolar obtenían puntajes más altos que aquellos que no lo habían hecho; los niños que habían repetido el sexto grado como asimismo aquellos que estaban trabajando exhibían un menor rendimien-

to que aquellos que nunca habían repetido y aquellos que no trabajaban; los niños que asistían a escuelas urbanas o privadas obtenían mucho mejores resultados que aquellos que asistían a escuelas rurales y públicas; los puntajes más bajos se encontraban entre los niños que asistían a escuelas indígenas y comunitarias, que contaban con instalaciones inadecuadas y profesores con menor nivel de capacitación, y los niños que obtenían los mejores puntajes asistían a escuelas urbanas y tenían padres que exhibían un mayor nivel de educación. Mientras que los niños de primer y segundo grado obtenían puntajes cercanos a lo que los investigadores y diseñadores de currículos esperaban, sus puntajes bajaban sucesivamente cada año en términos del porcentaje de respuestas correctas. Las autoridades mexicanas informaron también que era imposible medir sistemáticamente el desempeño de los profesores en la sala de clases, debido a que las poblaciones de alumnos eran extremadamente diversas y aún existían dificultades técnicas.

Como se mencionara anteriormente, México participó cabalmente en el TIMSS, pero en el último momento decidió no publicar los resultados. Esta decisión no debiera sorprender si se considera la renuencia general de México a divulgar los resultados de los exámenes en el pasado, tratándose en ocasiones como secretos de Estado. Si bien las actitudes han cambiado desde entonces y actualmente se divulgan los resultados de las evaluaciones correspondientes a la escuela básica y secundaria, esto aún se hace con alguna ambivalencia. La característica más distintiva del sistema de evaluaciones de México es que ha evaluado sistemáticamente los conocimientos y la capacidad de los profesores.

Argentina

Al igual que Brasil, Argentina comenzó a realizar evaluaciones en forma tardía, pero ha avanzado con rapidez en el establecimiento y la utilización de las evaluaciones para mejorar la calidad educacional. De acuerdo con una nueva ley de educación federal que descentralizó la educación, el Ministerio de Educación estableció en 1993 un sistema de evaluación nacional con oficinas en la Secretaría de Programación y Evaluación Educativa. Los objetivos del sistema de evaluación eran promover la descentralización, proporcionar información clave acerca de la situación de la educación, supervisar el avance en el logro de los objetivos de la reforma, identificar las desigualdades e insuficiencias como asimismo las áreas en las que se requieren programas compensatorios para las

poblaciones en situación desventajosa y estimular la participación de sectores más amplios en la toma de decisiones en el área de la educación. A partir de 1993, en Argentina se han administrado pruebas a todos los niños que cursan el último año de la educación primaria y secundaria tanto en idioma como en matemáticas. En 1995, las evaluaciones se ampliaron cubriendo el tercer grado de la educación primaria y el segundo grado de la educación secundaria. También se comenzó a administrar pruebas en las áreas de las ciencias sociales y las ciencias naturales. Se aplicaron cuestionarios complementarios a los profesores, directores, alumnos y padres. Las pruebas se basaron en un análisis detallado de los logros curriculares esperados y un cuidadoso programa de pruebas piloto. Se confirmó la confiabilidad y la validez de estas pruebas a través de análisis estadísticos y revisiones por parte de expertos de la medida en la cual las preguntas correspondían a los objetivos curriculares. Todo el programa es administrado por el Ministerio de Educación. Los resultados de las evaluaciones se incorporan en un sistema de información educacional más amplio, que se utiliza para controlar y supervisar el sistema educativo en todos los niveles.

Entre la información relativa a las evaluaciones se incluyen detalles acerca del rendimiento escolar y los perfiles de los alumnos, profesores y modelos administrativos asociados al rendimiento. Esta información es dada a conocer a las autoridades nacionales, estatales, locales y escolares. Sin embargo, los puntajes individuales de cada escuela no se informan a las escuelas mismas. Diversos informes han abordado también el tema de los contenidos de las materias e incluyen recomendaciones tendientes a introducir mejoramientos en la pedagogía, tanto a nivel central como a nivel de cada escuela. Los resultados de las pruebas también han sido utilizados para desarrollar manuales orientados al mejoramiento del aprendizaje y la capacitación en el servicio, y estas pautas han sido entregadas, junto con asistencia técnica, a los 24 estados del país.

Si bien las preguntas de las pruebas fueron elaboradas en base a los niveles de respuesta mínimos esperados, en promedio los alumnos respondieron solo el 50% de las preguntas correctamente. Los factores escolares representan el 40% de la varianza asociada a los puntajes en matemáticas y el 28% de la varianza en los puntajes en idioma. Los niños de nivel socioeconómico más alto obtuvieron los mejores puntajes. Sin embargo, las diferencias entre las escuelas fueron mayores que las diferencias al interior de las escuelas, lo que sugiere que el sistema educacional argentino está altamente segmentado. Los niños pobres que

asistían a clases heterogéneas obtuvieron puntajes más altos que los niños pobres de clases homogéneas. En general, los resultados muestran que se requiere introducir mejoramientos significativos en las escuelas.

El programa de evaluación de Argentina, si bien se inició solo en 1993, está bien concebido y tiene una estrecha relación con su estrategia de descentralización y mejoramiento de la calidad de la educación en general. Al igual que Chile, Argentina está realizando todos los esfuerzos posibles por utilizar los resultados de las evaluaciones. Ya se han distribuido ampliamente informes nacionales y regionales, muchos de los cuales abordan los problemas más significativos y entregan materiales pedagógicos destinados a fortalecer la enseñanza.

Programa de Evaluación Regional UNESCO/OREALC

A través de un subsidio del BID, la Oficina Regional de la UNESCO (Oficina Regional para América Latina y el Caribe, OREALC) se encuentra implementando un programa de evaluación regional bajo la denominación de “Laboratorio Latinoamericano de Medición de la Calidad de la Educación”. El programa, iniciado en 1995, busca fomentar la coordinación y fortalecer la capacidad de los organismos de evaluación nacionales en la región. En 1997 se administraron pruebas de matemáticas y lectura, acompañadas de detallados cuestionarios de antecedentes, a una muestra de alumnos de tercer y cuarto grado de 15 países de América Latina. Las pruebas fueron desarrolladas por un comité en el que se encontraban representados todos los países participantes y se basan en un análisis informal de los objetivos curriculares en el área de las matemáticas y la lectura en la región. Los resultados se dieron a conocer aproximadamente a mediados de 1998.

En 1993, la UNESCO implementó una prueba similar, en calidad de piloto, para alumnos de cuarto grado de siete países latinoamericanos. La mayoría de estos alumnos obtuvieron un puntaje muy por debajo de los niveles mínimos esperados por los expertos regionales en el área de los currículos y las pruebas. En promedio, los niños pudieron responder correctamente alrededor de la mitad de las preguntas. Venezuela y Costa Rica obtuvieron los mejores puntajes, en tanto que República Dominicana, Bolivia y Ecuador obtuvieron los puntajes más bajos. En el Cuadro 5 se entrega un resumen de los resultados.

La UNESCO tiene proyectado encargar un estudio que compare el grado de dificultad y el contenido curricular de la prueba de matemáticas

CUADRO 5**Puntajes obtenidos en los exámenes regionales de cuarto grado en matemáticas e idiomas, 1994**

Porcentaje de alumnos que respondieron las tres cuartas partes (o más) de las preguntas correctamente

País	Matemáticas	Idiomas
Venezuela	15,5	25,2
Costa Rica	10,4	23,8
Chile	5,0	23,2
Argentina	1,3	18,5
República Dominicana	1,4	5,0
Bolivia	2,1	4,9
Ecuador	0,4	7,7
Región	5,5	15,8

Fuente: UNESCO/OREALC, "Medición de la Calidad de la Educación: Resultados" (Santiago, Chile, 1994).

de la UNESCO con la prueba de matemáticas del TIMSS. Esto entregará evidencia directa del nivel en que se encuentran los países latinoamericanos en comparación con los países de fuera de la región. También se espera que, con el apoyo de la UNESCO, los expertos del TIMSS ayuden a países seleccionados de América Latina a implementar la metodología del TIMSS para medir la "oportunidad de aprender", como por ejemplo, la revisión y comparación del currículo oficial, el currículo aplicado, los textos escolares y la pedagogía utilizada en la sala de clases.

El programa de la UNESCO constituye un paso importante en el logro de una coordinación regional. Para asegurar el fortalecimiento a largo plazo de las capacidades regionales, los esfuerzos futuros requerirán el apoyo a los centros regionales de excelencia, el establecimiento de relaciones con las instituciones públicas y privadas, y el mantenimiento de vínculos más fuertes que en el pasado con la IEA y los principales centros mundiales de investigación y desarrollo.

IV. LECCIONES APRENDIDAS Y FUTUROS DESAFÍOS

RESUMEN DE LA EXPERIENCIA EN SEIS PAÍSES

Chile y Argentina cuentan con los sistemas de evaluación más amplios y mejor administrados de América Latina. Chile ha demostrado también el mayor compromiso de largo plazo con el desarrollo de evaluaciones. Si bien Chile y Argentina han hecho los mayores avances en la utilización de las evaluaciones en el área del diseño de las políticas, la reforma curricular y el mejoramiento de las escuelas, aún quedan muchos aspectos que mejorar, incluso en estos países. En particular, todavía no se han establecido objetivos de aprendizaje claros a nivel nacional y solo recién se están iniciando esfuerzos sistemáticos por “compatibilizar” el currículo, los textos escolares y la pedagogía utilizada en la sala de clases. Chile y Argentina también han hecho avances en la integración de las evaluaciones en un sistema amplio de supervisión y evaluación.

Entre los principales problemas que requieren atención se encuentran la renuencia de México a divulgar los resultados de las evaluaciones; el enfoque “intermitente” de Costa Rica con respecto a las evaluaciones y el excesivo número de pruebas aplicadas; y el lento progreso de Brasil en la utilización de las evaluaciones para la reforma de los currículos y las políticas. Con excepción de Costa Rica, los seis países analizados en este trabajo han utilizado preguntas de selección múltiple más que preguntas de rendimiento o de interpretación abierta.

Otro punto que debe tomarse en cuenta es que los seis países informaron resultados similares en los exámenes. En particular, los alumnos que cursaban los últimos años de la educación básica y aquellos que asistían a la educación secundaria obtuvieron puntajes muy por debajo de las expectativas de los educadores e investigadores profesionales. Los estudiantes con los mejores puntajes provenían de escuelas urbanas y privadas y tenían padres con un mayor nivel de educación. Pocos países de la región han emprendido análisis multivariados detallados de estos resultados con el fin de identificar sus causas. Si bien diversos estudios han mostrado una correlación entre los profesores mejor capacitados y un mejor rendimiento de los alumnos, la mayoría de estos estudios no controlaban el factor del nivel socioeconómico de los alumnos o la ubicación de la escuela lo suficientemente como para confirmar la importancia de la educación de los profesores. Además, aún existe poca infor-

mación con respecto a si los niveles de rendimiento de los alumnos han cambiado a través del tiempo.

Entre las innovaciones importantes en la región se incluye el intento de Brasil de evaluar las instituciones de educación superior; las evaluaciones de la capacidad de aprendizaje efectuadas en Costa Rica entre los niños que entran a la educación primaria y sus exámenes para los alumnos que terminan la educación secundaria; las evaluaciones aplicadas en México para medir los conocimientos y habilidades de los profesores; el énfasis dado en Colombia a la investigación; el uso que se hace en Chile de las evaluaciones para orientar la asignación de recursos y los esfuerzos de Argentina por utilizar las evaluaciones para la reforma curricular. El programa UNESCO/OREALC también constituye una importante innovación que tendrá mayor impacto a medida que se vaya aumentando la coordinación con las instituciones como la IEA.

En general, entre las principales lecciones aprendidas y desafíos para el futuro se incluyen los siguientes: (a) la importancia de un consenso nacional y un compromiso de largo plazo; (b) la importancia de centrarse explícitamente en el uso de las evaluaciones como herramientas para mejorar el aprendizaje; (c) la necesidad de generar capacidades y competencia técnica; y (d) la necesidad de obtener beneficios de los adelantos e innovaciones provenientes tanto de América Latina como de otras regiones, especialmente a través de programas de pruebas internacionales como los de la IEA.

Generación de consenso y compromiso

Generación de un consenso nacional. Tal vez la lección más importante a la fecha es que la necesidad de llevar a cabo evaluaciones ya no debiera ser un tema de debate en los países de América Latina. Por el contrario, estos países necesitan establecer objetivos educacionales, determinar si los niños, las instituciones y los sistemas escolares están cumpliendo sus objetivos y, luego, establecer programas tendientes a asegurar que dichos objetivos se cumplan finalmente. Las evaluaciones no servirán para mejorar la calidad de la educación a menos que todos concuerden en la importancia de mejorar la calidad y en la necesidad de informar los resultados de las evaluaciones de una manera amplia, oportuna y fácilmente comprensible a todos los participantes en el proceso (es decir, de manera “transparente”). Este consenso debe representar una coalición entre los profesores, los padres, los administradores y

los líderes empresariales y políticos. Debe persistir en el tiempo y contar con el respaldo de las más altas autoridades de gobierno; sin embargo, sus decisiones no se deben ver influidas por los compromisos políticos. Los programas de evaluación deben ser llevados adelante con decisión y no se debe permitir que sean suspendidos y luego reiniciados en forma intermitente (como en el caso de Costa Rica) ni que sus resultados sean ocultados a la opinión pública (como en el caso de México).

Integración de los profesores en el proceso de evaluación. Es preciso incluir a los profesores en el proceso de evaluación desde el comienzo si se desea que las evaluaciones tengan impacto en la educación. En América Latina, las asociaciones gremiales de profesores tienden a ser observadores pasivos o incluso a oponerse a las evaluaciones. Es preciso convencerlas de que el hecho de adoptar un enfoque más profesional con respecto a su trabajo, perfeccionar los procesos de la sala de clases y participar cabalmente en todos los programas de evaluación va en su propio interés. Las asociaciones gremiales no deben temer que las evaluaciones sean utilizadas para recompensar o castigar a los profesores, puesto que el alumnado varía tanto de un año al otro que la calidad de la enseñanza siempre será difícil de medir con precisión, como lo demuestra la experiencia en Francia, Estados Unidos y México. Si las evaluaciones se visualizan como una herramienta para evaluar a las escuelas y a los sistemas escolares locales en lugar de los profesores, tanto los padres como los profesores y los administradores estarán mejor capacitados para trabajar conjuntamente como equipo con el fin de mejorar los puntajes en el rendimiento a nivel de las escuelas.

Un enfoque sistémico. Un compromiso con la transparencia también requiere un enfoque sistémico para abordar la supervisión y la evaluación. Al respecto, debe entenderse que las evaluaciones solo miden el rendimiento de los niños y es poco lo que nos dicen acerca de la magnitud y las causas de la deserción. Otras herramientas para medir el rendimiento de los alumnos y del sistema son las siguientes: estudios de la deserción y la repitencia, mediciones de los insumos escolares y estimaciones de los recursos mínimos que cada escuela debiera tener, observaciones sistemáticas de los procesos escolares y estudios del desempeño de los egresados de las escuelas en el mercado laboral.

Usos de las evaluaciones

El mayor desafío para el futuro en América Latina será asegurar que las evaluaciones se utilicen precisamente como herramientas para mejorar la calidad de la educación. Las evaluaciones pueden influir en muchos elementos del sistema educacional, incluidas las políticas educacionales nacionales, los programas de reforma educativa, los currículos educacionales, las decisiones de los padres, estudiantes y profesores, las políticas educacionales locales y regionales y la pedagogía y los programas a nivel de las escuelas. Tres áreas de particular importancia en América Latina en las cuales pueden utilizarse las evaluaciones son las siguientes: el refuerzo de la “compatibilización”, la orientación de los recursos a las escuelas más necesitadas y el estímulo a la responsabilidad y la recompensa de los logros a nivel de las escuelas.

Refuerzo de la “compatibilización”. Por sobre todo, las evaluaciones pueden ayudar a los países de América Latina a lograr una mayor compatibilización del currículo que se pretende aplicar, el currículo que realmente se aplica, los textos escolares, los conocimientos de los profesores, la pedagogía en la sala de clases y el aprendizaje. Los ajustes resultantes permitirán establecer objetivos más altos pero factibles, mejorar la capacitación docente y la pedagogía, y asegurar que el currículo revisado sea aplicado en la práctica. Esto requerirá un gran trabajo de análisis, difusión y capacitación. Con una buena retroalimentación y recursos, las evaluaciones podrán ser utilizadas directamente como una herramienta para mejorar la enseñanza y el aprendizaje. Argentina ha hecho los mayores avances en este proceso a la fecha.

Orientación de los recursos. Dada la gran disparidad en los logros educacionales en América Latina, las evaluaciones deben ser utilizadas para canalizar los recursos adicionales hacia las escuelas más necesitadas de la región, particularmente aquellas que se encuentran en las áreas rurales y en los barrios más pobres de las zonas urbanas. Dicha actividad puede llevarse a cabo siguiendo el modelo del “Programa de las 900 escuelas” aplicado en Chile, que está explícitamente dirigido al mejoramiento del rendimiento de las escuelas que han obtenido los peores puntajes.

Estímulo a un mejor rendimiento en las escuelas. Las evaluaciones deben proporcionar información útil a las escuelas, padres y profesores. Esto requerirá, entre otros elementos, determinar el “valor agregado” de

la escolaridad, identificando aquellas escuelas que obtienen un puntaje mayor que el esperado en atención a los antecedentes socioeconómicos de los alumnos, como se hiciera en Francia, o midiendo el grado de mejoría en los puntajes de las escuelas en períodos sucesivos. Estos enfoques conducen a un reconocimiento y recompensa de las escuelas “eficaces” y motivan a las demás escuelas a reproducir las características de dichos establecimientos. Este enfoque elimina también el sesgo en contra de las escuelas que se encuentran en barrios pobres.

Hasta la fecha, Chile es el país que más ha avanzado en el reconocimiento y estímulo a dichas escuelas.

Creación de capacidades y competencia técnica

Fortalecimiento de las instituciones independientes. En la mayoría de los países de América Latina sería muy difícil para el gobierno central implementar un programa de evaluación directamente en el largo plazo, como se hizo en Francia. Entre los obstáculos a un sistema de ese tipo se encuentran los posibles conflictos de intereses, los sueldos insuficientes y los procedimientos de contratación y administración inflexibles. Una estrategia más apropiada para la región sería fomentar la creación de organismos de pruebas competentes independientes del gobierno, manteniendo un organismo gubernamental pequeño, especializado y altamente eficiente en la función de supervisar dichas actividades. Colombia parece haber desarrollado una sólida asociación de este tipo entre el sector público y el sector privado.

Garantía de conocimientos especializados adecuados e investigación de apoyo. Si se desea que las evaluaciones sean eficientes como herramientas para medir lo que se pretende que midan, se requiere un alto nivel de conocimientos especializados. De lo contrario, es muy fácil que presenten deficiencias: puede que no discriminen adecuadamente entre los aprendientes; la comparabilidad de un año al otro puede verse afectada; las escuelas y regiones pueden engañar; el marco de la muestra puede ser inexacto, con lo cual toda la evaluación puede resultar inútil; técnicas analíticas más sofisticadas pueden mostrar que las supuestas relaciones “causales” son falsas. Muchos de estos problemas ya han sido detectados en las evaluaciones hechas por los países. Por lo tanto, es esencial capacitar y remunerar adecuadamente a expertos en evaluación de currículos, metodología de estudios muestrales y técnicas analíticas. Como se ha hecho en Jordania, es importante consultar a las autoridades

más conocidas y respetadas del mundo en el área de las evaluaciones, ya que no existe una metodología “regional” de pruebas.

Los gobiernos y fundaciones también necesitan fomentar centros de posgrado de excelencia en el área de la docencia y la investigación en la región. En especial, existe la necesidad de apoyar las investigaciones independientes basadas en los datos de las evaluaciones, que constituyen una rica fuente de información sobre todos los aspectos de la educación relacionados con los alumnos, los profesores y las escuelas. Para obtener los mejores resultados, estas investigaciones deben basarse en las técnicas analíticas más modernas. Otro aspecto igualmente importante, es que los investigadores como asimismo quienes diseñan las políticas deben estar de acuerdo en los objetivos de las investigaciones y estar dispuestos a divulgar los resultado, aun si estos son ambiguos o negativos.

Nuevos enfoques

Participación en los programas de evaluación internacionales. El TIMSS ofrece a los países de América Latina una enorme oportunidad para el establecimiento de estándares, la compatibilización de los currículos y los textos escolares, el perfeccionamiento del profesorado y los avances en el aprendizaje. Toda la región puede obtener beneficios de los conocimientos especializados y la metodología del TIMSS.

Actualmente, la IEA tiene planes de duplicar el TIMSS en 1999. Chile y Brasil ya han expresado su intención de participar. Los países latinoamericanos podrían también tomar parte en el nuevo programa de educación cívica de la IEA, que será particularmente importante por los aspectos relacionados con el fortalecimiento de la sociedad civil en la región. Actualmente, Chile y Colombia están comprometidos con el estudio relativo a la educación cívica.

Como alternativa, los investigadores del TIMSS podrían llevar a cabo un análisis de los currículos en los textos y las pautas curriculares y compararlos con los currículos aplicados en otras regiones del mundo. También podrían hacer el intento de determinar si los exámenes nacionales y otros son compatibles con el currículo nacional que se pretende aplicar. La metodología del TIMSS ya ha sido utilizada para analizar el contenido curricular, los textos escolares y la pedagogía en Argentina, Colombia, México y República Dominicana. Es posible solicitar asistencia técnica en cualquiera de estas actividades tanto a los profesionales internacionales como a los investigadores regionales del TIMSS.

Método basado en el universo *versus* método basado en una muestra. Antes de emprender una evaluación del universo, los países deben estar conscientes de que esta es una tarea de alto costo y que requiere una administración minuciosa por parte de profesionales competentes. También se necesita un sistema de retroalimentación detallado y apoyo a las escuelas. En Chile, el costo de la aplicación de pruebas al universo es de aproximadamente 5 dólares por alumno, lo que representa aproximadamente el 2% de los costos unitarios totales. Incluso esta reducida cantidad puede ser difícil de justificar en los países con graves restricciones fiscales y que carecen de fondos para material pedagógico.

En comparación, los estudios muestrales cuestan menos, pero pueden proporcionar información suficiente para el establecimiento de políticas nacionales y para la identificación de los problemas de compatibilización entre el currículo, los textos escolares y la enseñanza. Sin embargo, estos estudios requieren precisas técnicas de muestreo y una administración rigurosa. Siempre que sea posible, los países de América Latina también deben utilizar los exámenes de selección existentes para efectos de evaluación.

Pruebas de rendimiento. En los últimos años, Estados Unidos ha ido incorporado cada vez más las preguntas de interpretación abierta en sus pruebas de rendimiento, en tanto que Europa se ha concentrado en mejorar la confiabilidad y validez de los instrumentos de medición. Por el contrario, la mayoría de los países de América Latina todavía favorecen las evaluaciones del tipo “selección múltiple”. Los países de la región debieran por lo menos explorar los enfoques más nuevos con respecto a las pruebas. Sin embargo, no podrán hacerlo si aquellas personas que califican las pruebas no cuentan con una cabal capacitación en los procedimientos pertinentes.

Exámenes de competencia mínima. Puede que los países de América Latina deseen considerar la aplicación de “exámenes de competencia mínima” para el nivel secundario. Estos ya han sido desarrollados en algunas partes de Estados Unidos y en Costa Rica, como una manera de estimular el aprendizaje. Este tipo de exámenes serían especialmente apropiados para los países descentralizados de gran tamaño, como Brasil.

Opciones para los países más pequeños y más pobres. Los países más pequeños y más pobres simplemente no cuentan con los conoci-

mientos técnicos especializados ni los recursos necesarios para implementar un sistema de evaluación en forma sostenida. Como lo sugiere la experiencia de Costa Rica, los conocimientos especializados locales pueden ser insuficientes. Por lo tanto, los países más pequeños deberían emprender solo un mínimo de pruebas (por ejemplo, solo matemáticas e idioma) en unos pocos grados. Otra opción es operar en consorcios (compuestos, por ejemplo, por los países centroamericanos), hacer uso de los programas internacionales y aprovechar la asistencia técnica regional (por ejemplo, de Chile y Argentina).

Cooperación regional. Para fortalecer el proceso de coordinación regional será necesario desarrollar vínculos más estrechos con la IEA y con los principales centros de investigación y desarrollo del mundo en el área de las pruebas y evaluaciones, junto con establecer una amplia participación, que incluya las instituciones no gubernamentales, en los foros y comités regionales. También son necesarios los esfuerzos a nivel regional para: analizar los resultados de las evaluaciones de la UNESCO/OREALC de manera de identificar los factores asociados a un alto rendimiento, comprender las interacciones entre los objetivos curriculares y el aprendizaje a través de la medición de las “oportunidades para el aprendizaje” y, lo que es más importante, identificar y apoyar los centros regionales de excelencia dedicados a las pruebas, las mediciones y las investigaciones educacionales.

BIBLIOGRAFÍA

- Aldana, Eduardo. 1996. “Medición del logro educativo: el caso colombiano”. Seminario de PREAL sobre Evaluación y Estándares en la Educación Latinoamericana, Río de Janeiro.
- Álvarez, Benjamín, y Mónica Ruiz-Casares, eds. 1997. *Assessment and Education Reform: Policy Options*. 1997. Washington D.C.: Academy for Educational Development.
- Capper, Joanne. 1996. “Testing to Learn, Learning to Test”. Newark, Del.: International Reading Association.
- Esquivel, Juan Manuel. 1996. “Medición de logros del aprendizaje y el empleo de los resultados en Costa Rica”. Seminario de PREAL sobre Evaluación y Estándares en la Educación Latinoamericana, Río de Janeiro.

- Horn, Robin, Laurence Wolff y Eduardo Vélez. 1991. *Developing Educational Assessment Systems in Latin America: A Review of Issues and Recent Experience*. Washington D.C. Banco Mundial.
- Ministerio de Educación y Cultura, Brasil. 1996. “Seminario Internacional de Avaliação da Educação”, octubre de 1995, Río de Janeiro.
- De Moura Castro, Claudio, y otros. 1984. “Determinantes de la Educación en América Latina: Acceso, Desempeño y Equidad”. ECIEL (Programa de Estudios Conjuntos para la Integración Económica Latinoamericana). Río de Janeiro: Fundação Getulio Vargas.
- Murphy, Paul, Vincent Greaney, Marlaine E. Lockheed y Carlos Rojas, eds. 1996. *National Assessments - Testing the System*. Washington, D.C. Banco Mundial, Economic Development Institute.
- National Research Council. 1995. *National Science Education Standards*. Washington, D.C. National Academy Press.
- Rodríguez Schuller, Cristina. 1996. “Sistema de Medición de la Calidad de la Educación: características y uso de los resultados de la evaluación para mejorar la calidad (Chile)”. Seminario de PREAL sobre Evaluación y Estándares en la Educación Latinoamericana, Río de Janeiro.
- Schmelkes, Silvia. 1996. “Propuesta para el diseño y desarrollo de un sistema nacional de evaluación integral de la educación básica en México”. Seminario de PREAL sobre Evaluación y Estándares en la Educación Latinoamericana, Río de Janeiro.
- Schmidt, William, Curtis McKnight y Senta A. Raizen. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht: Kluwer Academic.
- Schmidt, William A. y otros. 1997. *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science*. Dordrecht: Kluwer Academic.
- U.S. National Research Center for the Third International Mathematics and Science Study. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht: Kluwer Academic.
- Valverde, Gilbert A. 1996. “Las políticas educacionales nacionales y la participación en estudios internacionales”. Seminario de PREAL sobre Evaluación y Estándares en la Educación Latinoamericana, Río de Janeiro.
- Vianna, Heraldo Marelim. 1996. “Programas de avaliação de sistema de educação estadual do estado de São Paulo”, Seminario de PREAL sobre Evaluación y Estándares en la Educación Latinoamericana, Río de Janeiro.

ANEXO: UNA MIRADA A LOS SISTEMAS DE EVALUACIÓN DE SEIS PAÍSES

ARGENTINA	BRASIL
Fecha de la primera evaluación: 1993.	Fecha de la primera evaluación: 1990.
Fecha del programa de evaluación más reciente: 1995.	Fecha del programa de evaluación más reciente: 1995.
Grados evaluados (programa más reciente): Grados 2, 3, 5 y 7 en educación primaria.	Grados evaluados (programa más reciente): Grados 4 y 8 de la educación primaria y grados 2 y 3 de la educación secundaria (además, Brasil ha iniciado un programa de evaluación de todos los estudiantes universitarios).
Materias evaluadas: Matemáticas y lectura y, en el 7º grado, también ciencias naturales y sociales.	Materias evaluadas: Matemáticas y lectura.
Muestra o universo: Universo de estudiantes.	Muestra o universo: Muestra de 90.499 alumnos de 2.289 escuelas públicas y 511 escuelas privadas.
Resultados importantes: Las preguntas se seleccionaron en base a los niveles mínimos esperados, pero los escolares respondieron solo el 50% de las preguntas correctamente; los factores escolares representaron el 40% de la varianza asociada a los puntajes en matemáticas.	Resultados importantes: Los escolares generalmente obtuvieron puntajes muy por debajo de los niveles de aprendizaje esperados definidos por los profesores, investigadores y especialistas; los puntajes más altos se observaron en el sur del país, en las ciudades capitales, en las escuelas urbanas y privadas y en las escuelas diurnas de jornada completa; los estudiantes más jóvenes, los estudiantes blancos y orientales y los estudiantes con profesores con mayor capacitación obtuvieron los mejores resultados.
Tipos de investigaciones realizadas: Correlaciones simples y tabulaciones cruzadas.	Tipos de investigaciones realizadas: Correlaciones simples y tabulaciones cruzadas.
Información disponible acerca de los costos: —	Información disponible acerca de los costos: —
Elementos técnicos de interés: Selección múltiple e interpretación abierta.	Elementos técnicos de interés: Selección múltiple.
Estructura institucional: Administrado por el Ministerio de Educación, pero dado en contrato a diversas instituciones privadas.	Estructura institucional: Administrado por el Ministerio de Educación, pero dado en contrato a diversas instituciones privadas.
Retroalimentación: Programa amplio que entrega información acerca del rendimiento escolar, los perfiles de los alumnos, los profesores y los modelos administrativos asociados al rendimiento; identifica el contenido con mayor dificultad e incluye recomendaciones para mejoramientos en la pedagogía.	Retroalimentación: Los informes fueron puestos a disposición de la prensa, pero no fueron utilizados por las escuelas mismas.
Impacto sobre las políticas: Manuales metodológicos para mejorar el aprendizaje como asimismo capacitación en el servicio basada en los resultados de las pruebas distribuidos a todas las escuelas; asistencia técnica entregada a 24 estados.	Impacto sobre las políticas: A la fecha no se ha tomado ninguna decisión específica sobre la base del programa de evaluación.
Aspectos políticos y otros: —	Aspectos políticos y otros: No se han considerado en forma adecuada los usos de los resultados de las evaluaciones en el área de las políticas, tales como la concentración en las escuelas con peor rendimiento.
Relación con programas de evaluación internacionales: Participó en el análisis de currículos del TIMSS y emprendió estudios demasiado tarde como para ser incluidos en el informe; ha entregado asistencia técnica a Bolivia y Paraguay.	Relación con programas de evaluación internacionales: Participó en el estudio de la IEA en matemáticas y lectura en 1993 y tiene planes de participar en el TIMSS-R.

(*) Sin perjuicio de la existencia de programas de evaluación más recientes, hay que tener en consideración que el documento fue elaborado en 1998.

CHILE

Fecha de la primera evaluación:

1981.

Fecha del programa de evaluación más reciente:

1991-96.

Grados evaluados (programa más reciente):

Grados 4 y 8.

Materias evaluadas:

Lectura y aritmética en los grados 4 y 8; muestra del 10% en ciencias naturales, historia y geografía.

Muestra o universo:

Principalmente universal, excepto lo observado anteriormente.

Resultados importantes:

Los resultados continúan mostrando un bajo rendimiento en las escuelas públicas municipalizadas y en las escuelas rurales. Alta correlación entre el nivel socioeconómico y el rendimiento; se han producido significativos mejoramientos en los puntajes en las escuelas participantes en el programa de "900 escuelas" que se concentra en las escuelas con peor rendimiento.

Tipos de investigaciones realizadas:

Correlaciones simples y tabulaciones cruzadas; trabajo analítico más sofisticado actualmente en curso.

Información disponible acerca de los costos:

US\$ 5 por estudiante evaluado.

Elementos técnicos de interés:

Selección múltiple hasta ahora; se considera la introducción de más preguntas de rendimiento, interpretación abierta y ensayo; se está revisando el método estadístico para asegurar una mayor confiabilidad de un año al siguiente.

Estructura institucional:

La capacidad de pruebas ha sido creada por el Ministerio de Educación más que por un organismo autónomo.

Retroalimentación:

Los informes están siendo constantemente simplificados y ajustados a receptores específicos, especialmente a nivel de las escuelas; los directores de las escuelas, los profesores y los padres utilizan los resultados de las pruebas; algunos padres utilizan los resultados de las pruebas para seleccionar las escuelas; los diseñadores del currículo nacional están utilizando los informes para revisar el currículo.

Impacto sobre las políticas:

Los diseñadores del currículo están dando énfasis al dominio de las materias en las áreas de debili-

dad; se ha continuado con los programas pedagógicos dirigidos a las escuelas de peor rendimiento (programa de "900 escuelas").

Aspectos políticos y otros:

El programa está tan institucionalizado que se están volviendo más importantes aspectos de "segundo orden"; por ejemplo, los profesores podrían estar "enseñando con miras a la prueba" de manera exagerada y las escuelas podrían no estar considerando el nivel de pobreza de los estudiantes.

Relación con programas de evaluación internacionales:

Ninguna directamente hasta el momento, pero existen planes de participar en el TIMSS-R.

COLOMBIA

Fecha de la primera evaluación:

1990.

Fecha del programa de evaluación más reciente:

1994.

Grados evaluados (programa más reciente):

Inicialmente los grados 3 y 5 y posteriormente los grados 7 y 9.

Materias evaluadas:

Inicialmente matemáticas y lectura; posteriormente, también ciencias naturales y sociales.

Muestra o universo:

Solo muestra.

Resultados importantes:

Los niveles cognitivos de los alumnos de los grados 3 y 5 son bajos comparados con las expectativas, especialmente en el área de las habilidades críticas y reflexivas. Los mejores puntajes se obtienen en las escuelas urbanas y privadas, entre quienes han asistido a la enseñanza preescolar, no han repetido, tenían libros en sus casas y tenían padres con niveles de educación relativamente altos; los alumnos de clases con profesores con mayor nivel de educación, disponibilidad de textos escolares, docentes de sexo femenino, de la escuela nueva y de escuelas completas obtuvieron mejores resultados.

Tipos de investigaciones realizadas:

Correlaciones, tabulaciones cruzadas y algún análisis multivariado.

Información disponible acerca de los costos:

—

Elementos técnicos de interés:

Selección múltiple.

Estructura institucional:

Empresa conjunta entre el Ministerio de Educación, el Instituto Ser, la Universidad Nacional de Profesores y el Centro de Estudios Sociales.

Retroalimentación:

Los informes y resúmenes de las investigaciones están disponibles pero no son útiles para la mayoría de las escuelas y municipalidades debido a la base de la muestra.

Impacto sobre las políticas:

La Ley de Educación de 1994 incorpora los resultados de las evaluaciones tales como la concentración en la educación preescolar, la educación de los profesores y los textos escolares.

Aspectos políticos y otros:

Se considera que la expansión al universo sería demasiado cara y superaría las capacidades institucionales, pero se ampliará la muestra de manera de entregar más información a las escuelas y regiones evaluadas.

Relación con programas de evaluación internacionales:

Pleno participante en el TIMSS con información de los resultados; tiene planes de participar en el estudio de ciencias sociales de la IEA.

COSTA RICA

Fecha de la primera evaluación:

1987.

Fecha del programa de evaluación más reciente:

1993-97.

Grados evaluados (programa más reciente):

Grados 1, 3, 6, 9, 12.

Materias evaluadas:

Grado 1, capacidad de aprendizaje; grados 3, 9 y 12, todas las materias; también muestras sobre las capacidades cognitivas y físicas en diversos grados.

Muestra o universo:

Principalmente universo, con excepción de las capacidades físicas y cognitivas y la capacidad de aprendizaje.

Resultados importantes:

Aún no se encuentran disponibles para las pruebas actuales; los resultados anteriores mostraban que los alumnos tenían un rendi-

miento muy inferior al esperado en comparación a los objetivos del currículo nacional; los alumnos de las escuelas privadas, urbanas y académicas obtuvieron los mejores resultados. Los alumnos obtuvieron puntajes progresivamente peores en comparación con los objetivos del currículo a medida que aumentaba su tiempo de permanencia en la escuela.

Tipos de investigaciones realizadas:

Tabulaciones cruzadas simples.

Información disponible acerca de los costos:

US\$ 1,50-US\$ 7,50 por alumno evaluado en 1991.

Elementos técnicos de interés:

Principalmente selección múltiple.

Estructura institucional:

Se está probando actualmente una empresa de cooperación entre el Ministerio de Educación y el IIMEC en la Universidad de Costa Rica.

Retroalimentación:

Los anteriores programas de evaluación tuvieron poca difusión; en el nuevo programa, los informes serán preparados para múltiples receptores.

Impacto sobre las políticas:

Reintroducción en 1989 de las pruebas nacionales o regionales que representan un alto porcentaje de las notas finales de los alumnos, utilizadas como justificación para el aumento de los fondos para educación.

Aspectos políticos y otros:

Enfoque intermitente con respecto a las evaluaciones, dependiendo de la actitud de los sucesivos ministros de educación; el actual ministro está a favor de las evaluaciones.

Relación con programas de evaluación internacionales:

Ninguna directamente.

MÉXICO

Fecha de la primera evaluación:

1976.

Fecha del programa de evaluación más reciente:

1994.

Grados evaluados (programa más reciente):

Todos los grados de la educación básica y secundaria, como también los profesores.

Materias evaluadas:

Matemáticas y lectura en las escuelas; se evaluó a los profesores en las materias y los conocimientos pedagógicos.

<p>Muestra o universo: Universo en 1994, anteriormente muestras.</p> <p>Resultados importantes: A medida que los alumnos avanzan en la escuela, obtienen puntajes progresivamente más bajos que lo esperado; los puntajes más altos fueron obtenidos por los alumnos que habían asistido a la enseñanza preescolar, no estaban repitiendo el año, asistían a escuelas urbanas o privadas y tenían padres mejor educados en áreas urbanas; los niños de las escuelas indígenas y comunitarias con instalaciones escolares inadecuadas y profesores con menor capacitación obtuvieron los peores puntajes.</p> <p>Tipos de investigaciones realizadas: Correlaciones simples y tabulaciones cruzadas.</p> <p>Información disponible acerca de los costos: US\$ 35 por alumno en las muestras tomadas en 1990.</p> <p>Elementos técnicos de interés: Selección múltiple.</p>	<p>Estructura institucional: Administrado por el Ministerio de Educación (Secretaría de Educación Pública).</p> <p>Retroalimentación: Los informes se entregan a la opinión pública pero no son fáciles de comprender; hasta 1990 los resultados de las evaluaciones tuvieron carácter confidencial.</p> <p>Impacto sobre las políticas: Hasta hace poco ningún impacto sobre las políticas; las evaluaciones a los profesores están relacionadas con el convenio con la asociación gremial nacional para mejorar la calidad de la educación básica.</p> <p>Aspectos políticos y otros: Difusión y utilización inadecuadas a nivel de las escuelas.</p> <p>Relación con programas de evaluación internacionales: Pleno participante en el TIMSS, pero en el último momento solicitó que no se informaran los resultados.</p>
--	---

Los próximos pasos: ¿Cómo avanzar en la evaluación de aprendizajes en América Latina?*

PEDRO RAVELA
RICHARD WOLFE
GILBERT VALVERDE
JUAN MANUEL ESQUIVEL

INTRODUCCIÓN

Durante la década de los '90, especialmente a partir de 1992, al menos 20 países de América Latina y el Caribe pusieron en funcionamiento algún tipo de sistema nacional de evaluación de aprendizajes. Este esfuerzo se apoya en algunas premisas generales que son ampliamente compartidas por académicos y responsables de la implementación de políticas educativas, tales como:

La educación es por naturaleza propia una actividad “opaca” en cuanto a sus resultados, comparada con otras actividades humanas donde es más sencillo para la sociedad “ver” los resultados de lo que se hace. El hecho de que los niños estén o no aprendiendo lo que se espera no puede ser percibido directamente por la sociedad y por las familias.

Los juicios que las familias pueden hacerse acerca de la calidad de la escuela a la que asisten sus hijos, normalmente se basan en aspectos como el orden existente, el trato que reciben los niños o la proposición de tareas para realizar en el hogar, pero difícilmente tienen una visión clara de los conocimientos y competencias que sus hijos están adquiriendo.

(*) Este documento fue originalmente publicado, en versión resumida, por PREAL en agosto de 1999 en la serie PREAL Documentos N° 20, agosto de 2001.

Los resultados suelen ser “opacos” también para el propio maestro que, si bien puede tener una visión cabal acerca de lo que sus alumnos son capaces de hacer, por lo general no cuenta con una referencia externa acerca de los conocimientos y competencias que logran adquirir los niños en otras escuelas del país o de la región.

Para las autoridades y otros tomadores de decisiones en materia de política educativa, ya no son suficientes los indicadores tradicionales sobre matrícula, cobertura, repetición y deserción. En un contexto en el que el desafío principal para la política educativa es cómo garantizar el acceso equitativo a los conocimientos y competencias fundamentales para el desempeño social, un sistema de evaluación que produzca información relevante sobre este aspecto adquiere hoy importancia estratégica para la gestión educativa.

En este sentido, existe un consenso amplio respecto a la necesidad de contar con mecanismos que permitan producir información sobre lo que efectivamente se enseña y se aprende en las escuelas, de modo de dotar de mayor transparencia a los sistemas educativos y hacerlos más responsables ante la sociedad. Se asume y espera que esto contribuirá a mejorar la calidad de la educación.

Variedad de experiencias

Las experiencias impulsadas en este ámbito en la Región han tenido distintos propósitos y enfoques.

En una buena parte de los casos, la creación de sistemas nacionales de evaluación de aprendizajes ha sido impulsada por los organismos internacionales de crédito, como parte de sus convenios de préstamo con los países. Sin embargo, las características concretas que en cada país asume el sistema de evaluación parecen depender más de las capacidades técnicas y las decisiones políticas locales que de directivas específicas de dichos organismos.

En algunos países, las pruebas se realizan a nivel censal en ciertos grados, en tanto en otros se trabaja con muestras de escuelas y/o grupos determinados.

Entre quienes trabajan a nivel censal, unos han optado por publicar los resultados en la prensa, atribuyendo al sistema de evaluación la función principal de entregar información a las familias con el fin de que exista un control del usuario sobre la gestión escolar. Otros han optado por devolver la información a cada escuela con carácter confidencial,

apostando al uso de la información como instrumento de aprendizaje profesional para los educadores. Algunos países han desarrollado experiencias de utilización de la información de resultados a nivel grupal, como elemento de evaluación de la labor del maestro y como parte del sistema de incentivos económicos.

Los países que trabajan sobre la base de muestras suelen presentar información de resultados generales a nivel nacional, con algunos niveles de desagregación por área geográfica y tipo de escuela. En algunos casos se realizan importantes esfuerzos por acompañar la devolución de resultados con materiales de orientación didáctica para los docentes, explicitando las áreas de menor logro y los problemas de aprendizaje y de enseñanza que pueden estar involucrados. En otros, después de varios operativos de evaluación y varios años de trabajo, nunca se han dado a conocer los resultados en forma pública.

Prácticamente todos los países evalúan logros en lenguaje y matemática, y existe una importante variedad de situaciones en cuanto a la evaluación de otras áreas del aprendizaje (ciencias naturales, ciencias sociales, autoestima, etc.), así como en cuanto a los grados y niveles evaluados y la periodicidad de las evaluaciones.

La UNESCO, por su parte, ha impulsado el Laboratorio Latinoamericano de Medición de la Calidad de la Educación, que ha permitido desarrollar una primera experiencia de evaluación internacional en varios países de la Región.

Solo como ejemplo de esta diversidad de enfoques, se puede mencionar que Chile ha optado por un esquema de evaluaciones de carácter censal con publicación de resultados en los medios de prensa, con el objetivo principal de informar a los usuarios acerca de la calidad del servicio que brinda cada establecimiento; mientras que México ha optado por un sistema masivo pero voluntario para los maestros, con el objetivo principal de usar la información como indicador de la calidad del trabajo del maestro y de allí derivar incentivos de carácter económico. En ambos casos las pruebas que se aplican son absolutamente confidenciales. Uruguay, por su parte, ha adoptado un enfoque centrado en el uso de la información como instrumento de aprendizaje al interior del sistema educativo, priorizando la devolución de resultados a cada establecimiento en forma confidencial y haciendo completamente públicas las pruebas luego de los operativos. Argentina ha desarrollado evaluaciones anuales de carácter muestral, con un fuerte énfasis en la producción de “cuadernos” de recomendaciones metodológicas para los docentes.

El panorama sucintamente presentado permite afirmar que **se ha dado un primer paso de enorme trascendencia para la Región: los sistemas se han instalado en los países, se ha generado una cierta capacidad para la implementación de operativos nacionales de evaluación a gran escala, y la sociedad y los cuerpos docentes comienzan a valorar y comprender la necesidad de este tipo de evaluaciones.**

Revisión de la experiencia y búsqueda de nuevos caminos

No obstante lo anterior, al cabo de una primera etapa de implantación de su sistema nacional de evaluación, muchos países se encuentran ingresando en una etapa de evaluación de su propia experiencia, y de consideración y estudio de nuevas alternativas para el desarrollo y rediseño de sus sistemas de evaluación hacia el futuro.

Ello obedece a la constatación de tres grandes tipos de insuficiencias en lo realizado hasta el momento:

- Insuficiente aprovechamiento de la información producida por los sistemas de evaluación, lo que tiene como consecuencia el deficiente impacto de las evaluaciones en el conjunto del sistema educativo.
- Insuficiente calidad y capacidad de evaluación de aprendizajes complejos en las pruebas que están siendo aplicadas.
- Debilidades técnicas en los procesos de desarrollo y validación de los distintos instrumentos de medición.

De hecho, la mayoría de los sistemas de evaluación de la Región fueron contruidos sobre un marco limitado de saberes respecto al diseño de pruebas, así como respecto al universo de opciones técnicas posibles, en un contexto caracterizado por la inexperiencia que había a principios de los 90 tanto en los Ministerios de Educación como en otras instituciones o centros especializados. A ello es preciso agregar el hecho de que normalmente en las comunidades académicas del mundo educativo en la Región existía un fuerte rechazo, de carácter más ideológico que técnico, a cualquier intento de medición en el área educativa.

En muchos países de la Región ha sido insuficiente la reflexión acerca de los fines específicos que se espera cumplan los sistemas de evaluación dentro de un país, así como de las definiciones técnicas más adecuadas para cada fin. Como se puede observar en el Recuadro 1, los sistemas de evaluación pueden perseguir distintos fines y es imposible que un mismo diseño sirva para satisfacerlos todos.

RECUADRO 1

Finalidades de los sistemas de evaluación de aprendizajes

Un sistema nacional de evaluación de aprendizajes puede proponerse alguna o varias de las siguientes finalidades:

1. ***Evaluar la productividad de los maestros para establecer un sistema de incentivos.*** En este caso se parte del supuesto de que un modo de lograr que los maestros enseñen mejor y los niños aprendan más, es estableciendo incentivos monetarios o de otro tipo para los maestros cuyos alumnos exhiben mejores niveles de logro. El papel principal del sistema de evaluación es producir información sobre los aprendizajes logrados en cada grupo escolar.
2. ***Brindar a los padres de familia información que les permita evaluar la calidad de las escuelas.*** En este caso, se espera que el sistema de evaluación proporcione una medida de la calidad de la enseñanza que ofrece cada una de las escuelas, de modo que los padres estén en mejores condiciones para controlar su labor y para elegir la escuela que consideren mejor para sus hijos. Se supone que esto obligará a las escuelas a esforzarse por mejorar su trabajo.
3. ***Devolver información a las escuelas y maestros para que estos examinen los resultados de su trabajo.*** En este caso el sistema de evaluación debe producir información útil y detallada sobre lo que los alumnos están aprendiendo, para enriquecer la discusión técnica de los docentes y la búsqueda de nuevos caminos para mejorar la práctica pedagógica.
4. ***Establecer la acreditación de los alumnos que finalizan un determinado nivel de enseñanza.*** En este caso, la evaluación debe permitir decidir si un estudiante individualmente considerado, ha logrado los conocimientos y competencias indispensables para completar cierto nivel de enseñanza y obtener el certificado correspondiente.
5. ***Seleccionar u ordenar a los estudiantes.*** El sistema de evaluación puede tener como objetivo ya no constatar si los alumnos dominan ciertos conocimientos y competencias, sino simplemente ordenar o jerarquizar a un conjunto dado de alumnos de acuerdo a sus niveles de dominio, por ejemplo, con vistas a un proceso de selección para el ingreso a distintas modalidades de educación superior o de formación para el trabajo.
6. ***Informar a la opinión pública y generar una cultura de la evaluación.*** Aquí el propósito principal de la evaluación es producir información adecuada para rendir cuentas periódicamente ante la opinión pública de un país acerca de la marcha del sistema educativo, en términos de los niveles de aprendizaje que alcanzan los estudiantes en diferentes áreas disciplinarias y niveles del sistema, y su evolución a lo largo del tiempo. Esto ayuda a generar una cultura de evaluación del sistema educativo en la sociedad en general.
7. ***Contribuir a establecer estándares de calidad para el sistema educativo.*** El sistema de evaluación puede tener como propósito explícito o implícito dar una señal a las escuelas y maestros acerca de qué conocimientos y competencias se espera que los alumnos dominen al finalizar un grado o nivel de la enseñanza o, en caso de que los mismos estén explícitamente definidos bajo la forma de estándares o indicadores de logro, evaluar el grado en que los mismos se alcanzan, a modo de mecanismo de control de la calidad del sistema educativo.
8. ***Construir un “mapa de situación” del sistema educativo con el fin de identificar áreas prioritarias de intervención y tipos de intervenciones necesarias.*** La evaluación nacional de aprendizajes puede servir para detectar las regiones, distritos o establecimientos en que las dificultades para lograr los aprendizajes esperados son mayores, con el fin de facilitar el diseño de estrategias de intervención focalizadas y apropiadas. Del mismo modo, puede servir para identificar regiones, distritos o establecimientos con resultados especialmente buenos, con el fin de conocer y difundir sus modos de trabajar.

9. Evaluar el impacto de políticas, innovaciones o programas específicos. En el marco de los procesos de reforma y cambio educativo en curso en todo el mundo, los Ministerios de Educación desean contar con información sobre los resultados de un nuevo currículo que ha sido implementado en un conjunto de escuelas, un plan de capacitación de maestros o una inversión en nuevos materiales didácticos.

10. Realizar estudios de tipo costo-beneficio. Una expectativa que muchas veces existe es que los sistemas de evaluación proporcionen información útil para evaluar los costos y beneficios en términos de inversión económica y resultados educativos de distinto tipo de intervenciones. Se busca por este camino apoyar los procesos de toma de decisiones, con el fin de que los recursos disponibles sean utilizados de manera efectiva y eficiente.

11. Contribuir a la generación de conocimiento. Los sistemas de evaluación de aprendizaje generan importantes bases de información útiles para investigaciones que contribuyan a la acumulación de conocimiento sobre el funcionamiento de los sistemas educativos, las prácticas de enseñanza, el impacto de las variables sociales sobre el aprendizaje de los niños y los tipos de intervenciones más efectivos para mejorar los aprendizajes.

Muchos países han trabajado a partir de un propósito general de informar sobre los resultados del sistema educativo para contribuir a su mejoramiento, pero sin diseñar una estrategia más específica. Por otra parte, es bastante común que las autoridades ministeriales comiencen a demandar, sobre la marcha, que las evaluaciones sirvan para nuevos propósitos o que aporten información para fines para los que no fueron diseñadas.

Asimismo, normalmente no se cuenta con un plan de trabajo detallado de largo plazo respecto al desarrollo del sistema de evaluación y sus objetivos, que permita diseñar las estrategias adecuadas a los diferentes tipos de fines y ordenar las decisiones técnicas sobre la conformación de las bases de datos, la conformación de los bancos de ítems y la comparabilidad de las evaluaciones, entre otras cosas.

En el presente, sin embargo, **parecen estar dadas las condiciones para realizar un “salto cualitativo” en materia de evaluación, una vez que se ha transitado por las primeras experiencias, que se ha superado la preocupación inicial por las enormes exigencias de la implementación de los operativos de medición a gran escala, y que existen sistemas funcionando que permiten debatir el tema ya no en abstracto, sino a partir de experiencias en marcha.**

Dar un “salto cualitativo” exige desarrollar un proceso de estudio y análisis en dos grandes planos, íntimamente relacionados entre sí:

- La discusión sobre las **opciones de política** en materia de evaluación nacional de aprendizajes: qué impactos específicos se espera que tengan los sistemas de evaluación en el sistema educativo, más allá de la definición genérica de la responsabilidad ante la sociedad y la mejora de la calidad.
- La discusión sobre las **opciones técnicas**: qué abanico de tipos de pruebas, instrumentos complementarios, procesamientos y análisis de información existen y cuáles son los más adecuados para los fines propuestos.

Propósitos de este documento

El presente documento pretende aportar a la reflexión sobre la relación entre las finalidades de política educativa que los sistemas de evaluación pueden proponerse y sus implicaciones técnicas. A través del mismo, se espera enriquecer el repertorio de alternativas disponibles a la hora de reflexionar sobre los rumbos a seguir por los sistemas de evaluación de aprendizajes de la Región en el futuro.

El documento fue elaborado en el marco de un Taller de Trabajo convocado por GRADE, el cual se desarrolló en Lima (Perú) entre el 17 y el 20 de agosto de 1999. Se convocó a un conjunto de especialistas en el tema con conocimiento directo de distintos sistemas nacionales de evaluación de la Región, así como de los dilemas, opciones y revisiones a las que muchos de los países se han visto enfrentados.

Cuatro grandes temas o problemas fueron seleccionados en el Taller para organizar el documento, a cada uno de los cuales corresponde un capítulo del mismo:

1. El problema del diseño global del sistema nacional de evaluación.

Se intenta ofrecer una visión sistemática acerca de la relación entre los diversos fines del sistema de evaluación y las decisiones técnicas relacionadas con la cobertura de las mediciones, tanto en términos poblacionales (tamaño de las muestras y/o censo) como en términos de contenidos (nivel de detalle de los conocimientos y competencias a evaluar dentro de un área o disciplina). Junto con abordar este tema en el primer capítulo, se analiza en el capítulo quinto la relación entre los fines del sistema de evaluación y los modos de reportar los resultados, y se presenta un conjunto de alternativas técnicas para esos efectos.

2. **El problema de la validez de los instrumentos de medición.** Un tema hasta ahora insuficientemente atendido por los sistemas nacionales de evaluación de la Región, que está directamente vinculado con su valor como insumo para la toma de decisiones.
3. **El problema de los paradigmas de construcción de pruebas de medición de aprendizajes o logros.** Se examinan las características, exigencias, posibilidades y limitaciones que caracterizan a los paradigmas de pruebas referidas a “normas” y pruebas referidas a “criterios”, y se reflexiona sobre el problema de la validez en cada uno de los paradigmas.
4. **El problema de los factores “asociados” a los resultados escolares.** La mayor parte de los sistemas de evaluación de la Región, junto con la aplicación de las pruebas, recogen un importante volumen de información de carácter contextual, tanto sobre las características socio-culturales de los alumnos como sobre las propias escuelas y maestros. Sin embargo, este tipo de información es escasamente difundida y poco utilizada en el análisis de los resultados de aprendizaje.

El documento fue elaborado a través de aproximaciones sucesivas de discusión colectiva, redacción individual por parte de los participantes, lectura y discusión de lo producido, y un nuevo proceso de redacción individual. Se realizaron tres ciclos de este tipo durante el Taller y luego se continuó con ajustes y correcciones durante un mes.

Si bien todos los contenidos del documento fueron discutidos en forma colectiva, la redacción de cada uno de los capítulos estuvo a cargo de una persona que se indica junto al título de cada capítulo. Participaron también Patricia Arregui y Santiago Cueto en las instancias de discusión colectiva del trabajo y en la revisión de los sucesivos borradores.

Es el deseo de todo el grupo de trabajo que el documento sea útil para enriquecer la discusión sobre los sistemas de evaluación de aprendizajes en la Región y para ampliar la mirada hacia el futuro en un área estratégicamente central para el mejoramiento de los sistemas educativos.

CAPÍTULO I

EL DILEMA DE LA “GRANULARIDAD” EN EL DISEÑO DEL SISTEMA DE EVALUACIÓN: COBERTURA CURRICULAR VS. COBERTURA POBLACIONAL

Richard Wolfe^(*)

En las evaluaciones nacionales, ¿es preferible trabajar con muestras o hacerlo a nivel censal?

¿Es preferible emplear una única prueba o diferentes formas con distintos ítems?

¿Con qué grado de desagregación es posible y deseable reportar los resultados?

¿Con qué grado de profundidad es posible y deseable medir los conocimientos y competencias adquiridos por los alumnos?

¿Es adecuado el modo en que los diseños de los sistemas de evaluación toman en cuenta todos estos aspectos?

INTRODUCCIÓN

Los diseños de los sistemas nacionales de evaluación educacional en América Latina son muy variados y sus características dependen de las filosofías, estructuras, costumbres burocráticas e historias de la educación específicas de cada país, de las etapas de reforma educativa en que se encuentren y de los estados de desarrollo de la investigación y planificación educativa.

Pero al mismo tiempo, cuando se examinan en detalle los diferentes sistemas de evaluación, se observan algunas características comunes derivadas del hecho de tener objetivos fundamentales y requerimientos técnicos similares.

^(*) **Richard Wolfe.** Especialista en medición, evaluación y estadística educacional, graduado de la Universidad de Chicago. Actualmente es Profesor Asociado del Departamento de Currículo, Enseñanza y Aprendizaje del Ontario Institute for Studies in Education de la Universidad de Toronto, Canadá. Ha asesorado en las áreas de su especialidad a varias instituciones gubernamentales e internacionales en las Américas.

El propósito de este capítulo es examinar un asunto crítico en el diseño de los sistemas nacionales de evaluación: las cuestiones sobre la denominada “granularidad”, es decir, la cantidad de detalle con que el sistema recoge y luego reporta los datos. Por ejemplo, puede haber enormes diferencias en el costo y en el modo de utilización entre sistemas de evaluación que solo proporcionan resultados nacionales y aquellos que suministran resultados de todos los estudiantes o escuelas individualmente. De igual manera, puede haber grandes diferencias entre las evaluaciones que dan información general sobre temas amplios –tales como los logros en matemáticas o lenguaje– y aquellas que brindan información detallada sobre lo que los estudiantes pueden y no pueden hacer en esas asignaturas.

Para analizar el tema de la granularidad es necesario abordar sus dos dimensiones: quién es evaluado y qué es evaluado.

¿Quién es evaluado?

Si bien los estudiantes –y a menudo los padres, los profesores, los directores escolares y otros– son las fuentes primarias de datos en un sistema de evaluación educacional, no suelen ser la principal unidad para la cual se calculan los resultados y se hacen los reportes, salvo en los casos en que se trata de exámenes de certificación o graduación. La granularidad de los reportes, es decir, la unidad más pequeña respecto a la cual se brinda información sobre sus resultados, suele establecerse en niveles superiores de la estructura educativa.

En los sistemas de evaluación de América Latina, comúnmente se encuentran los siguientes niveles de análisis y de reporte:

- Poblaciones nacionales (o internacionales), tales como la población de escolares matriculados en tercer grado de educación primaria.
- Los principales estratos definidos educacional, política y socialmente, tales como estudiantes en escuelas públicas, en las escuelas rurales, o estudiantes en programas bilingües.
- Principales divisiones regionales, tales como regiones geográficas, provincias, o estados.
- Jurisdicciones menores, tales como ciudades o municipalidades.
- Escuelas.
- Salones de clase (o profesores).
- Estudiantes.

Además de la selección entre estos niveles de reporte, en todos los niveles se diferencia según el grado escolar.

La elección del nivel o los niveles de reporte debería depender del objetivo y de los usos del sistema de evaluación. También debería determinar la metodología general para llevar a cabo la evaluación. Por ejemplo, si se van a devolver los resultados a cada estudiante o si cada profesor o escuela va a ser calificado individualmente, es obvio que se necesita una aplicación censal. Por otro lado, si solo son necesarios los resultados nacionales o los principales resultados subnacionales para rastrear la productividad general y el cambio en el sistema educacional, se puede usar una encuesta por muestreo, lo cual es mucho más económico.

En efecto, **existe una relación sumamente importante entre la granularidad del reporte y su costo. Cuanto más detallada información se requiera, más costoso es suministrarla.** Es un hecho básico del muestreo estadístico que el tamaño requerido de una muestra para un nivel dado de precisión es principalmente una función del tamaño de la muestra y no, como intuitivamente muchos pensarían, del tamaño de la población. Por ejemplo, si es necesario obtener información igualmente precisa para cada provincia de un país, entonces los requerimientos de tamaño de la muestra serán igualmente altos para las provincias con pocos estudiantes como para las provincias con muchos estudiantes. El tamaño de la muestra agregada será muy grande en comparación con lo que sería necesario si el único requerimiento fueran estadísticas nacionales precisas. Por otra parte, un muestreo proporcional o una muestra simple al azar rinde buenos resultados generales y resultados razonables para amplias subpoblaciones (e.g., grandes provincias), pero obtener una alta precisión para pequeñas subpoblaciones requiere un sobremuestreo costoso.

¿Qué es lo que se evalúa?

¿Con qué tipo de detalle se calculan y presentan los resultados del potencial dominio de contenido? Este es otro aspecto de la granularidad en el diseño de los sistemas de evaluación, que también tiene conexiones importantes con los objetivos de un sistema de evaluación e implicancias para la metodología.

En los diferentes tipos de sistemas de evaluación, nos encontramos con estos niveles de reporte de la información:

- Resultados globales, incluyendo matrícula, participación en la evaluación, sin una verdadera evaluación de contenidos.
- Éxito o fracaso en general, culminación del plan de estudios, graduación, certificación, tal vez basados en evaluaciones de diferentes asignaturas y otra información.
- Puntajes en asignaturas, tales como el logro general en matemáticas o en lenguaje.
- Puntajes en áreas de asignaturas, tales como solución de algoritmos, álgebra, o geometría en matemáticas y comprensión de lectura, expresión escrita o convenciones gramaticales en lenguaje.
- Logro de niveles particulares de desempeño en diferentes estándares en un área o asignatura, tales como la competencia para aplicar métodos geométricos en la solución de problemas de distancia, o interpretaciones a nivel de principiante en lecturas literarias.
- Estadísticas de respuestas para ítems específicos, tales como el porcentaje correcto en un ítem de opción múltiple o el porcentaje de calificaciones que se sitúan en cada nivel de una tarea de desempeño.
- Registro detallado de las respuestas a una prueba, incluyendo patrones de distribución de las respuestas a los ítems, transcripciones de desempeños o resultados cognitivos en laboratorio.

La granularidad del contenido de una evaluación determina fuertemente nuestra capacidad de interpretar y comprender la calidad del logro educacional y de tomar medidas para mejorarla. Por ejemplo, los resultados generales de logro respecto a un plan de estudios pueden ayudar a localizar áreas de éxito general relativamente alto o bajo (por Ej., mejores escuelas o tipos de escuelas). Pero un conocimiento más detallado de la substancia y el contenido de esos logros permite evaluar la importancia y las consecuencias de tales diferencias, y legitimará, a su vez, medidas tales como la selección o el establecimiento de incentivos.

La granularidad del contenido también determina nuestra capacidad para usar información de la evaluación para diseñar ajustes al currículo y a la enseñanza. Con información detallada y en profundidad acerca de los contenidos, se puede llegar a comprender cuáles aspectos de un currículum son aprendidos exitosamente y se pueden hacer recomendaciones específicas sobre la secuencia curricular y las prácticas de enseñanza. Ello implica un desplazamiento de preguntar *cuánto* saben los estudiantes a preguntar *qué* saben y *qué* son capaces de hacer.

Las evaluaciones de grano más fino son generalmente más costosas, porque el número de ítems requerido para cubrir en detalle un área de contenido es alto. Primero, habrá un número relativamente grande de subcontenidos dentro de un área de contenido. Por ejemplo, en matemáticas tenemos áreas generales como aritmética, geometría o álgebra, y áreas más específicas tales como fracciones, adición de fracciones, formas geométricas, congruencia, series y secuencias. Segundo, cada subcontenido requiere un número suficiente de ítems, tal vez cinco o diez, para suministrar una muestra adecuada de los posibles desempeños o niveles de desempeño. Asimismo, en la evaluación será necesario contar con una muestra adecuada de respuestas de los escolares a cada uno de los ítems.

El requerimiento de un número mínimo de ítems tiene un fundamento sustantivo y otro estadístico. En cuanto a lo sustantivo, se necesita ver un número suficiente de ejemplos de lo que es difícil y de lo que es fácil para comprender los tipos de conocimiento y destrezas que poseen los estudiantes. Estadísticamente, se necesita obtener una buena medida del desempeño promedio de los estudiantes, de la variación entre los ítems y de la interacción entre los ítems y los estudiantes (algunos estudiantes hacen algunas cosas bien, otros estudiantes hacen otras cosas bien).

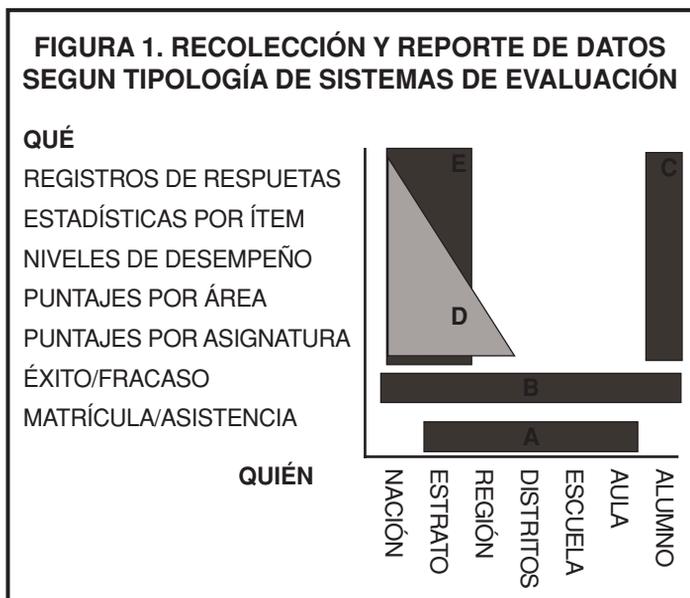
Por otro lado, una muestra relativamente pequeña de contenidos y de ítems puede ser suficiente si el propósito del reporte es suministrar unos promedios simples que resumen la situación en un dominio de contenidos, aunque ello no constituiría un diagnóstico del currículo.

Otra cuestión es que la evaluación de estándares educacionales importantes requerirá a menudo determinar si los estudiantes pueden llevar a cabo tareas complejas e integradas. Por ejemplo, en matemáticas no basta con interesarse solo si los estudiantes pueden sumar o restar, sino más bien si pueden usar la aritmética en contextos novedosos y realistas para resolver problemas. Desde la perspectiva de la evaluación, este aspecto es simultáneamente de “grano grueso”, porque se refiere a un estándar general de la educación y atraviesa diferentes contenidos y áreas; y de “grano fino”, porque requiere la definición de tareas de desempeño particulares y la recopilación, calificación y análisis de registros de desempeño.

TIPOLOGÍA DE SISTEMAS DE EVALUACIÓN

A partir de estas dos dimensiones de la granularidad –quién es evaluado y qué es evaluado– se puede definir una tipología de sistemas de evaluación cuya granularidad se grafica en la Figura 1.

- A. Estadísticas educacionales.** Representan una forma simple de sistema de evaluación, común a todos los países. La mayoría de las estadísticas educacionales contemplan cantidades de estudiantes (tal vez diferenciados por grado, sexo, inicación de *repetidor* o *promovido*); información respecto adónde se ubican los estudiantes (en escuelas, distritos y niveles más altos del sistema educativo). Estas estadísticas suelen ser recopiladas mediante una metodología de censo, puesto que todos los estudiantes y escuelas deben ser contabilizados y se requiere una desagregación muy fina de la ubicación. Sin embargo, no hay ningún contenido que diferenciar.
- B. Programas de evaluación y certificación.** Desde la perspectiva de la granularidad, un programa de este tipo está muy cercano a un sistema de estadística educativa, pero refinado en ambas dimensiones (el qué y a quién). El reporte de los datos se extiende hasta el nivel del individuo. Hay, además, una medición apropiada de contenidos, otorgada por una prueba de graduación o salida. Una mayor desagregación de los contenidos puede ser deseable, pero no es esencial. Puede también realizarse un resumen estadístico a distintos niveles de la jerarquía educacional por requerimiento administrativo. Los programas de evaluación y certificación son realizados de manera censal en las poblaciones afectadas. A menos que haya un propósito secundario de diagnóstico del éxito o fracaso estudiantil, o criterios múltiples para la certificación, hay poco interés en la diferenciación del contenido.
- C. Programas de evaluaciones diagnósticas.** Tienen el propósito específico de diagnosticar algunas características educacionales y psicológicas de los individuos (por ejemplo, dislexia). Suelen buscar identificar a los estudiantes con dificultades en el aprendizaje, para asignarlos a programas especiales o remediales. Los datos que se reportan son estrictamente individuales. Existe cierta diferenciación de contenido relacionada con los tipos de dificultades de aprendizaje que están siendo diagnosticados. Estos programas generalmente no son considerados evaluaciones nacionales de educación, salvo tal vez en pruebas de despistaje a gran escala y suelen ser implementados por personal local especializado.



D. Evaluación nacional por muestreo. Es un tipo de evaluación realizada en diversos países mediante relevamientos muestrales de ítems y de individuos. Lo usual es dividir un conjunto muy grande de ítems en múltiples formas de prueba para su administración, y se aplica cada forma a muestras paralelas de estudiantes, a través de algún sistema de rotación en la aplicación. En la Figura 1, la recolección y reporte de datos de este tipo de evaluación se muestra como un triángulo, porque suele ser posible una mayor diferenciación de los contenidos en los niveles más altos de agregación. Esto es simplemente una consecuencia de la precisión de las muestras. Como el número de estudiantes que responden un ítem particular es relativamente pequeño, no se tendrá una precisión adecuada para reportar estadísticas acerca de las respuestas dadas a ese ítem por diferentes poblaciones. En cambio, si se agregan los puntajes correspondientes a varios ítems, es posible trabajar conjuntamente los resultados de los estudiantes para proporcionar información más detallada sobre unidades más finas. Se podría incluso calcular puntajes individuales razonables en base a todo el conjunto de contenidos.

E. Estudio de investigación curricular mediante muestreo. Busca establecer los tipos de aprendizaje y enseñanza que se dan en una asignatura y estudiar las relaciones entre enseñanza y aprendizaje, así como los efectos de los contextos educacionales y sociales. En la Figura 1, el área de la recopilación y reporte de datos está representada como un rectángulo en el cual todo tipo de diferenciación de contenidos es importante, pero no hay gran interés por informar sobre logros que no sean a nivel nacional o de los estratos más altos del sistema. Esto no quiere decir que las variables referidas a los estudiantes y las escuelas no sean importantes, sino que estas unidades (estudiantes, profesores, escuelas) son anónimas. Puede ser muy importante, por ejemplo, estudiar el impacto de las características de la escuela, las prácticas del profesorado y los antecedentes de los estudiantes en los resultados educacionales, pero no interesa informar individualmente sobre las escuelas, clases o estudiantes.

El dilema de la granularidad

El propósito de detallar el tema de la granularidad y de suministrar una tipología de los sistemas de evaluación, ha sido exponer un dilema crucial que debe ser confrontado en el diseño de un sistema de evaluación educacional.

El dilema es que las dos dimensiones de la granularidad están en conflicto. Ello significa que, para un costo y esfuerzo fijos, un incremento en la granularidad de un tipo debe correr de la mano con una disminución de la granularidad del otro tipo. Por ejemplo, si se necesita obtener medidas detalladas sobre cada municipalidad del país, probablemente habrá que usar una prueba muy breve y simple, y la cantidad de detalle en los contenidos será mínima. Si, por otro lado, se quiere tener una gran profundidad en la medición de un dominio de contenidos, habrá que usar rotaciones de ítems y múltiples formas de prueba, que harán que el detalle de información quede muy disperso en los niveles más finos de reporte, en especial para los estudiantes individuales y posiblemente para las aulas, escuelas y niveles intermedios.

Un desafío central para el diseño de una evaluación es desarrollar metodologías que permitan combinar diferentes propósitos de la mejor manera. Por ejemplo, para el TIMSS se desarrolló un sistema muy intrincado y cuidadosamente diseñado de formas de prueba. Cada forma contiene una muestra estratificada de ítems que provienen de todo el campo

de contenidos, además de algunos ítems constantes que sirven para la calibración. Muchas de las formas contienen tareas de desempeño que insumen la mayor parte del tiempo de administración, mientras que otras formas incluyen casi exclusivamente ítems de opción múltiple o de respuesta corta. La administración de las formas fue llevada a cabo con una rotación y balance cuidadosos dentro de cada aula y escuela de la muestra. Si bien el tamaño de la muestra total de TIMSS no era grande (alrededor de 200 aulas y 7.000 estudiantes por grado en cada país), hay una enorme cantidad de información disponible para un análisis y examen detallados de los aprendizajes en matemáticas y ciencias y sus relaciones con los factores asociados.

A manera de contraste, se puede decir que hay varios sistemas nacionales de evaluación en América Latina que trabajan con muestras de mucho mayor tamaño o que son llevados a cabo como operaciones censales, que tienen mucho menor detalle de los contenidos y que, simultáneamente, no reportan información más allá de los estratos de agregación más altos.

En este sentido, parece recomendable que se otorgue mayor atención al análisis cuidadoso de estos aspectos de la granularidad en el diseño de las evaluaciones de rendimiento escolar en la Región.

CAPÍTULO II

LA INTERPRETACIÓN JUSTIFICADA Y EL USO APROPIADO DE LOS RESULTADOS DE LAS MEDICIONES

Gilbert Valverde^(*)

- ¿Qué significan los resultados que obtienen los estudiantes en nuestras pruebas nacionales de rendimiento?
- ¿Estamos realmente sacando conclusiones apropiadas, significativas y útiles a partir de los resultados de las evaluaciones?
- ¿En qué medida podemos justificar la manera en que interpretamos estos resultados?
- ¿Se usan los resultados de las evaluaciones de manera apropiada en la toma de decisiones?

INTRODUCCIÓN

Cuando los sistemas de evaluación conducen sus actividades, su interés es descubrir, describir e interpretar facetas del sistema educativo. Un propósito que comparten todos los sistemas de evaluación en América Latina es el de comprender qué capacidades académicas adquieren los niños y las niñas como resultado de su asistencia y participación en las escuelas y colegios del país. En el lenguaje curricular y evaluativo, a esas capacidades adquiridas como resultado de la escolarización comúnmente se las denomina *logro*.

Es posible que la estrategia óptima para comprender cómo se da el logro sea registrar el tipo de éxito que el estudiante experimenta al enfrentar cada una de las situaciones de aprendizaje en los que participa al año. Esto es lo que los docentes comúnmente intentan hacer como parte de su labor de evaluación en el aula.

^(*) **Gilbert A. Valverde**. Ph.D., Department of Educational Administration and Policy Studies, School of Education, University St Albany, State University of New York, Education 313 A, Albany, New York 12222. Fono: (518) 4425089, Fax: (518) 4425084 E-mail: valverde@uamall.albany.edu.

Pero también hay diversos actores que esperan obtener otros tipos de información de las evaluaciones, tales como:

- Las autoridades políticas y la sociedad civil tienen interés por obtener información acerca del sistema educativo. Por ejemplo, cómo es la calidad de la educación en comparación con los estudiantes de otros países o con grupos de estudiantes en generaciones anteriores en su propio país, o en relación con los propósitos académicos que el sistema mismo se ha fijado para sí mismo.
- Hay quienes están preocupados por los aspectos de equidad y necesitan descubrir si el sistema educativo favorece en forma desigual a distintos grupos económicos, culturales o lingüísticos.
- Otros desean información útil para juzgar la eficacia de distintos tipos de inversiones o intervenciones que se proponen hacer en el ámbito nacional en la educación.

Resulta casi evidente que la estrategia “óptima” mencionada anteriormente no se ajustaría a sus requerimientos, ya que sería imposible realizar un seguimiento a todos los estudiantes de un país o a un número representativo de ellos.

Hasta la fecha, la estrategia que se sigue en todos los sistemas de evaluación en América Latina es la de plantear una situación relativamente novedosa a los estudiantes, que dura uno o dos períodos lectivos, donde estos deben demostrar que han adquirido un número significativo de las capacidades esperadas. En todos los países de la Región, el tipo de situación que plantea el sistema de evaluación es una *prueba escrita* con preguntas que, a criterio de los autores de la prueba, exigen que los estudiantes utilicen lo que aprenden en la escuela para contestarlas correctamente. Las representaciones que más típicamente arrojan las pruebas en América Latina, son números llamados *promedios* o *notas*, cuyo significado debe ser bien entendido por las personas encargadas de interpretar estos números.

Ahora bien, ¿cuán fieles son estas representaciones?

Interpretar correctamente y usar apropiadamente la información que dan las pruebas, significa que hay que preocuparse por entender el tipo de representación del logro que estas permiten. Esto significa preocuparse por lo que, en medición, se llama *validez*.

La validez no es una propiedad intrínseca de las pruebas o encuestas, sino una propiedad de las interpretaciones y los usos que se propone dar a los datos que se obtienen de ellas. Es así que actualmente se define la

validez, como *el grado en que la evidencia empírica y la teoría dan sustento a las interpretaciones de los resultados de una medición*. Asimismo, la validez se refiere al *ámbito del legítimo uso de esas interpretaciones* y también *al grado en que el uso de la prueba no produce un impacto negativo no deseado sobre el sistema educativo*. En otras palabras, la validez se refiere a la calidad de las conclusiones que se toman a partir de las mediciones y a las consecuencias que las mediciones generan en los procesos que se proponen medir¹.

La validación de pruebas en América Latina. Ejemplos.

Algunas situaciones que se dan en América Latina sirven para ejemplificar algunos tipos de preocupación por la validez de las evaluaciones que se realizan en la actualidad:

Caso 1.

¿Habilidades de resolución de problemas o simple memorización?

El Ministerio de Educación se encuentra implementando un nuevo currículo nacional de matemáticas, cuyo enfoque principal es que los estudiantes

¹ Durante mucho tiempo, la concepción de validez más vigente y extendida, y que dominó el mundo académico y de las prácticas de medición evaluativa tanto en América Latina como en gran parte del mundo, fue la propuesta en 1949 por L.J. Cronbach en su libro *Essentials of Psychological Testing* (New York: Harper and Row), cuya versión quizás más conocida fue la ofrecida por A. Anastasi en su *Psychological Testing*, publicado en 1954. Desde entonces, la evolución de la teoría y métodos de las mediciones psicológicas y educacionales ha llevado a una nueva conceptualización y a la estandarización de la misma entre los profesionales de esas disciplinas. Así, en la tercera edición del texto *Educational Measurement* de R.L. Linn, publicada también por MacMillan en 1989, apareció la propuesta de Samuel Messick. Revisiones de esa propuesta llevaron a la acepción de validez actualmente establecida y que está documentada en los *Standards for Psychological and Educational Measurement*, publicados conjuntamente por la Asociación Americana de Investigación Educativa, la Asociación Psicológica Americana y el Consejo Nacional de Medición Educativa de los Estados Unidos en 1999. Es esta concepción, que se refiere a las acciones, decisiones e inferencias que se toman a partir de las mediciones –es decir, a cómo se usan– la que se ha utilizado en este capítulo.

Fuentes bibliográficas adicionales recomendables para este tema son Campbell, L. J. y Fiske, D. W. (1959) "Convergent and Discriminant Validity in the Multitrait-Multimethod Matrix" en *Psychological Bulletin*, 56: 81-105; Cronbach, L.J. (1989) "Construct Validation after Thirty Years" en R. L. Linn (Ed.) *Intelligence: Measurement Theory and Public Policy* Urbana: University of Illinois Press; Messick, S. (1989b) "Meaning and Values in Test Validation: The science and ethics of Assessment" en *Educational Researcher*, 18(2), 5 - 11; Messick, S. (1994) "The Interplay of Evidence and Consequences in the Validation of Performance Assessments" en *Educational Researcher*, 23 (2), 13-23; Moss, R.A. (1995) "Themes and Variations in Validity Theory" en *Educational Measurement: Issues and Practice*, 14 (2). 5-12.

aprendan a resolver problemas novedosos de la vida real utilizando elementos de razonamiento matemático. Sin embargo, para medir el logro, se administra una prueba escrita cuya mayoría de preguntas o reactivos exigen a los estudiantes que recuerden términos y principios matemáticos, o solo requieren que ellos apliquen procedimientos rutinarios para resolver problemas o ejercicios muy parecidos a los que aparecen en sus libros de texto.

En este caso, el Ministerio de Educación claramente no cuenta con un instrumento apropiado para descubrir si los estudiantes han logrado dominar las capacidades que persigue el nuevo currículo nacional. Sería injustificado concluir que los estudiantes que obtienen un alto promedio en esta prueba poseen la capacidad de resolver problemas novedosos de la vida real, porque las preguntas no exigen que recurran a este tipo de habilidades.

Caso 2.

¿Medición de conocimientos o de habilidad lectora?

Se escribe una prueba para descubrir si los estudiantes de 7 años de edad están adquiriendo conocimientos acerca de ciencias naturales. En las aulas se enseñan estos contenidos sin texto escolar, usando elementos del entorno natural de la escuela. La prueba contiene muchas preguntas cuya comprensión exigiría que los niños y las niñas posean gran habilidad para comprender textos escritos y un vocabulario altamente desarrollado.

En una prueba de esta naturaleza, el significado de los promedios es sumamente difícil de descubrir. ¿Acaso un bajo promedio indica la no adquisición de los conocimientos que se pretendía medir, o más bien mide la habilidad lectora de los niños? En el caso de niños pequeños, ¿en qué medida son las supuestas pruebas de ciencias (o de matemáticas, ciencias sociales, etc.) en realidad pruebas de lectura?

Caso 3.

¿Un problema de eficacia educativa o de acceso a recursos?

Se administra una prueba de logros a todos los estudiantes de octavo grado en un país. El Ministerio de Educación utiliza los resultados obtenidos por los estudiantes en cada escuela para calcular el promedio de logro para cada establecimiento. Comparando los promedios de los establecimientos según estos sean privados o públicos se descubre que los promedios de las escuelas privadas son más altos que los de las públicas. Se concluye que las primeras son más eficaces que las segundas, aun cuando ocurre que estas no cuentan con textos que aborden uno de los temas más importantes de la prueba.

Aquí es muy problemática la interpretación que se propone para los resultados, ya que un recurso esencial para el aprendizaje de un área de contenido o competencia específico (libro de texto que cubra temas medidos en la prueba) no se encuentra repartido equitativamente en los establecimientos. ¿Se justifica la interpretación de un bajo promedio como indicador de falta de eficacia del establecimiento? ¿No habría que interpretarlo como indicador de una falta de equidad en la distribución de los recursos?

Caso 4.

¿Dominio general de conocimientos o la suerte de dominar el área de contenidos que más se midió?

En un país se utiliza una prueba a final de la educación secundaria o media para avalar un diploma que se otorga al egreso de ese nivel. Se interpreta que pasar esta prueba indica que un estudiante ha logrado dominar todos los objetivos del currículo propuestos para cada año en ese nivel. En la prueba se miden algunos aspectos del currículo con una variedad de preguntas, otros con muy pocas. Se otorga el diploma correspondiente a todos los estudiantes que aprueban.

Preocupa en este caso si la conclusión de que un estudiante domina los objetivos del nivel se puede defender si no se mide con igual rigor los distintos componentes del currículo.

Caso 5.

Pertinencia de la evaluación educativa y el peligro de las comparaciones generales.

En un país no existe un currículo nacional, sino que cada provincia tiene el suyo propio. La Secretaría de Educación administra una prueba en todas las provincias. Para garantizar que la prueba es justa para todas, se decide poner solo preguntas sobre aquellos temas que se enseñan en todas ellas, lo que significa que se evalúa un subconjunto de las cosas que en cada provincia se pretende enseñar: Comparando los promedios de cada provincia, se encuentra que en algunas se obtienen resultados muy superiores que en las demás. Se concluye que es mayor la eficacia de los establecimientos en aquellas provincias. Sin embargo, ocurre que en las provincias de alto rendimiento, se pretende enseñar muy pocos temas que no están en la prueba nacional. En las provincias de más bajo rendimiento, los temas que se evalúan en la prueba nacional representan solo una pequeña parte de los temas que se proponen

enseñar, y no se les dedica mucho tiempo lectivo ni espacio en los libros de texto.

¿Es pertinente hacer una comparación entre los resultados de las provincias cuando en algunas de ellas se está enseñando una mayor proporción de los temas evaluados que en otras? ¿Acaso los promedios diferentes obtenidos de esta manera indican diferencias en eficacia educativa? ¿No será más bien que estos distintos promedios indican diferencias en la pertinencia de la prueba para cada una de las provincias?

Caso 6.

Decisiones de inversión a partir de limitadas evidencias.

Se diseña una prueba de lenguaje que entre sus preguntas contiene una sola en la cual los estudiantes escriben un texto propio. Al revisar este texto, se califican aspectos de ortografía, gramática y otras características de la escritura. El Ministerio de Educación desea distribuir material de apoyo pedagógico para docentes de lenguaje, pero para usar mejor su presupuesto, pretende descubrir los aspectos más débiles de los logros de los estudiantes, para lo cual se fija en los resultados de la prueba. Observa que la mayor parte de los estudiantes tuvieron mal rendimiento en la pregunta donde se pedía que escribieran su propio texto. En consecuencia, se escriben módulos de apoyo pedagógico y se proporciona capacitación a los docentes para ayudarlos a enseñar mejor gramática y expresión escrita.

¿Acaso la falta de éxito en contestar una sola pregunta es suficiente para concluir que los estudiantes no dominan esas capacidades? Si el Ministerio cuenta con recursos limitados para esfuerzos de refuerzo pedagógico y trata de utilizar los resultados de la evaluación para sacar provecho máximo de su inversión en ella, ¿ha utilizado en forma apropiada los resultados de la evaluación? Por otro lado, si los docentes adquieren, mediante los módulos y capacitaciones, la convicción de que deben dedicar mucho más esfuerzo a enseñar gramática y expresión escrita, ¿ha sido apropiada la información para ocasionar ese cambio en las prioridades de los docentes?

Opciones para la validación de mediciones en educación

Las situaciones anteriores ejemplifican los problemas que existen en torno a las interpretaciones justificadas y al uso apropiado de la información que arrojan las mediciones. A menudo se distorsionan los significa-

dos reales, lo que afecta su validez y, en consecuencia, su pertinencia como insumo para la toma de decisiones. Dado que este es un riesgo ineludible en la medición, **es importante explicitar de antemano los tipos de uso para los cuales los resultados podrán ser empleados legítimamente, así como los fines para los cuales los resultados NO podrán utilizarse de manera justificada.**

El proceso de acumulación de evidencias que dan sustento a las interpretaciones que se proponen para una medición se denomina *validación*. Existe una gran cantidad de opciones en cuanto al tipo de evidencia que se puede acumular y reportar. Cada tipo de evidencia ilumina o da apoyo a distintas facetas de la validez, pero no representa un tipo distinto de validez. La validez es un concepto unitario que obliga a los diseñadores y usuarios a evaluar de manera integral toda la evidencia disponible sobre cuán bien están justificadas las interpretaciones de los datos y las maneras de utilizar la información recogida durante la aplicación de la medición.

En el caso de las pruebas de logro, sean estas referidas a normas o a criterios, se pretende derivar conclusiones que van más allá de las preguntas que componen las pruebas. Es decir, en ambos casos se reconoce que las preguntas que contiene la prueba representan solamente una pequeña muestra de todas las preguntas posibles que se podrían formular para conocer si los y las estudiantes poseen ciertas capacidades. De los análisis de cualquiera de los dos tipos de pruebas mencionadas se concluye que si los estudiantes contestan con éxito 80 por ciento de las preguntas formuladas en la prueba, serían también capaces de contestar con éxito 80 por ciento de todas las preguntas posibles que se podrían formular para medir esa capacidad.

Una forma obvia de proceder para sustentar esta conclusión es mediante una definición clara de lo que se quiere medir. Una vez que se cuenta con esa definición, es posible comparar cada pregunta que se propone para la prueba y juzgar su concordancia con la definición. Si las preguntas de la prueba se han escrito de acuerdo a una definición precisa de lo que se pretende medir, las inferencias que se realicen con respecto al desempeño de los y las estudiantes en esas preguntas serán más válidas que en el caso contrario. Desde este punto de vista, la validación es un proceso inherente al procedimiento que se sigue para diseñar pruebas referidas a criterios (ver el capítulo al respecto en este mismo volumen), puesto que la definición del dominio (en términos de campo de conocimientos o habilidades) y el esfuerzo por asegurar la concordancia de las

preguntas con el dominio definido son dos de sus preocupaciones centrales. Cuando se desarrolla y aplica este tipo de pruebas, la documentación de las definiciones de los dominios, los juicios acerca de la concordancia de las preguntas con los dominios y los pasos seguidos para asegurar que los dominios representen con justicia el currículo o los estándares, sirven a dos propósitos: guían el desarrollo de la prueba y documentan la evidencia de la validación de la medición propuesta.

Frecuentemente se propone también que las pruebas sean interpretadas con relación a un criterio externo. Esto es típico, por ejemplo, de las pruebas de admisión a la educación superior. En esos casos se establece (con mayor o menor grado de fundamento) que un promedio determinado *predice* una exitosa carrera universitaria. En algunos países, se pretende establecer que un diploma de educación secundaria –avalado por una prueba de bachillerato– certifica que el diplomado posee ciertas capacidades básicas como posible empleado, de modo tal que se supone que el éxito en la prueba predice una exitosa carrera como trabajador.

Aun en los casos en que no existe un criterio externo propuesto explícitamente para la prueba, la utilización de referentes externos puede reforzar la validación de las pruebas. Por ejemplo, cuando se compara dos formas de medir la misma competencia y ambas formas arrojan resultados semejantes, esto puede dar evidencia para la validación.

Para la validación de los resultados que generan las pruebas, es de suma importancia que los servicios nacionales de evaluación educativa publiquen informes técnicos que contesten las siguientes preguntas con claridad:

- **¿Acerca de cuáles capacidades o destrezas se *derivarán conclusiones*?** En esos informes se debe incluir una definición explícita de las capacidades que interesan, así como de aquellas que se pretende evitar que debiliten la validez de la medición de las primeras. Por ejemplo, debe explicarse cómo se ha procurado que la habilidad para leer no obstaculice la oportunidad que tienen niños de corta edad de demostrar lo que saben en una prueba de ciencias naturales.
- **¿Cómo se aseguró concordancia entre las *preguntas y las capacidades o destrezas que se propuso medir*?** Es necesario documentar los procedimientos del caso y describir en detalle el resultado de su uso. Por ejemplo: ¿cómo se utilizaron las definiciones a la hora de escribir preguntas o cómo procedieron los jueces para asegurar la concordancia entre las preguntas y los dominios a medir? ¿De qué manera se recogieron y analizaron sus juicios?

- ***¿Qué tipos de preguntas permiten comprobar que se dominan las capacidades?*** Por ejemplo, si se quiere comprobar si los estudiantes pueden resolver problemas novedosos de la vida real en matemáticas o producir textos propios legibles, coherentes y persuasivos, ¿se puede usar preguntas en las cuales los estudiantes escogen la opción correcta entre cuatro o cinco posibilidades? ¿Acaso la habilidad de reconocer la respuesta correcta entre distintas opciones es idéntica a la generación de una respuesta propia? ¿O se necesitan más bien preguntas que les pidan demostrar los pasos que siguen para resolver problemas o para escribir textos? ¿Por qué? Algunas destrezas o capacidades, para ser medidas, requieren del uso de más de un tipo de preguntas, en cuyo caso habrá que documentar cuáles tipos, cuántos de cada tipo y justificar el peso que se asignará a cada tipo a la hora de calcular promedios, entre otras cosas.
- ***¿Cómo se evidencia que lo que predice la prueba ocurre en realidad?*** Cuando el propósito de una prueba es predecir el éxito académico o el éxito en la vida laboral, se debe acumular y reportar evidencias acerca de la relación entre puntajes o promedios obtenidos por los estudiantes en las pruebas con lo que ocurre de hecho durante su carrera académica o laboral.
- ***¿En qué medida son compatibles los resultados obtenidos con un instrumento y los obtenidos con otro?*** A menudo existen distintos instrumentos que pretenden medir cosas semejantes. Por ejemplo, pueden existir provincias que desean medir el logro de sus estudiantes con el propósito de reportarlo a cada familia. Si existiera simultáneamente una prueba nacional que se usa con el fin de evaluar logros promedio en el ámbito nacional en las mismas áreas, se puede comparar los resultados de los mismos estudiantes en las dos pruebas para acumular evidencia acerca de la convergencia de los resultados. De la misma forma, las pruebas internacionales pueden servir para propósitos técnicos de validación de las mediciones nacionales. Otra estrategia de validación es contrastar los resultados de una prueba con una observación directa a estudiantes o con el análisis de sus tareas y proyectos realizados en clase.
- ***¿Cómo se aseguró que las posibilidades que tienen los estudiantes de demostrarlo que saben no está mediada por factores ajenos a su control?*** Es importante describir cómo se asegura que todos los estudiantes estén en igualdad de condiciones para demostrar lo que saben. Por ejemplo, tener evidencia de que las preguntas son interpretadas

de la misma forma en distintas partes del país o entre distintos grupos lingüísticos, culturales y socioeconómicos. Por otro lado, si se pretende utilizar los resultados de las pruebas para evaluar programas de estudio, opciones pedagógicas o currículo, es importante describir cómo se hará para discriminar entre las ocasiones en que los estudiantes no pueden contestar preguntas que versan sobre cosas que les fueron enseñadas en clase, de aquellas ocasiones en que no pueden contestarlas porque no les fueron enseñadas. Esto es importante, puesto que existen serios problemas éticos cuando se responsabiliza a los estudiantes por contenidos que no han tenido la oportunidad de aprender, o cuando se responsabiliza a los docentes por el logro de sus estudiantes, no habiéndoseles proporcionado materiales o capacitación para enseñar esos contenidos.

- *¿Cómo se aseguró una relación óptima entre los contenidos que se pretende enseñar en el grado evaluado y los contenidos evaluados?* Es importante documentar la relación entre el currículo o los estándares y el contenido de las pruebas. ¿Cómo se aseguró congruencia entre ambos? ¿Hubo participación o consulta de las unidades responsables de elaborar el currículo o planes de estudio durante el proceso de construcción de la prueba? ¿Cómo se procedió?

Estas son solamente algunas de las evidencias de validez que los sistemas de medición en América Latina deben considerar en sus estrategias de validación y que en la actualidad raramente se reportan.

Consideraciones finales

Como se estableció anteriormente en la definición formal, la validez es cuestión de *grado*. **No existen mediciones perfectamente válidas o que reproduzcan fielmente todas aquellas facetas de la realidad educativa que pretenden medir. Lo que existen son mediciones más o menos válidas, dependiendo de las conclusiones que se pretende tomar a partir de ellas o del uso que se pretende hacer de la información que arrojan.**

Es importante recordar que las responsabilidades con respecto a la validación de las mediciones corresponden tanto a los diseñadores de las mismas como a sus usuarios. Quienes diseñan mediciones tienen la responsabilidad de reportar con claridad para qué sirven y para qué no

sirven. Deben reportar toda la información pertinente para que los usuarios tengan elementos de juicio para evaluar su validez. Por su parte, los usuarios tienen la responsabilidad de usar los resultados de acuerdo a los criterios de validez que tienen o, si proponen un uso nuevo para las mediciones, les corresponde validarlas para ese nuevo uso.

Debe señalarse también que en América Latina se pretende a menudo que una misma evaluación sirva para más de un propósito. Frecuentemente se espera que una misma prueba permita, por ejemplo, distinguir entre estudiantes que logran o no logran los objetivos académicos de un nivel y que, al mismo tiempo, sirva para juzgar la eficacia de distintas escuelas y la eficacia de diversos programas en las cuales participan dichas escuelas. La validación es específica de acuerdo al uso, es decir, validar un propósito de una prueba no equivale a validarla para otro.

También es cierto que la validez es específica a las poblaciones: una prueba validada para un país o una provincia determinada, no puede ser considerada como validada para otras poblaciones.

Adicionalmente, hay que tener en cuenta que el tiempo cambia las características de los fenómenos y que, por lo tanto, la validación es una tarea continua y una forma de asegurar que nuevos factores que puedan aparecer con el tiempo no atenúen la validez de las mediciones.

Finalmente, dado que su objetivo es asegurar la congruencia de la medición con la realidad educacional que se supone se está midiendo, la validación constituye una actividad *científica*. También se trata de una *actividad técnica de desarrollo*, porque la tarea de acumular evidencia de validez a menudo trae como consecuencia el rediseño o el afinamiento de los instrumentos o de sus sustentos teóricos.

Es necesario reconocer que en América Latina puede no ser posible diseñar evaluaciones específicas para cada propósito para el cual se necesita contar con información para tomar decisiones. Esto genera un dilema importante que deben confrontar los países. Pongamos un ejemplo: si no existiera actualmente una prueba que se haya validado específicamente para ser usada para distinguir entre la eficacia de centros educativos que utilizan un programa de estudios y la de centros que utilizan otro, y es necesario decidir cuál de los programas debe ser difundido y promovido por el Ministerio, ¿significa acaso que no debemos utilizar las pruebas existentes para ese propósito? No hay respuesta simple. Para decidir sobre este asunto, será necesario determinar en qué medida es mejor la decisión que tomaríamos utilizando los resultados de la prueba, en comparación con la decisión que tomaríamos sin usarla. Si el posible

mayor valor de una decisión tomada sobre la base de la prueba se juzga suficiente, sería sin duda un insumo que se debe usar. Pero es necesario tener presente que esto no significa que la hemos validado para este propósito. El valor de los resultados de las pruebas como insumos para la toma de decisiones solo puede optimizarse cuando se asume la responsabilidad de validarlos para ese propósito. Tomar una decisión basada en una inferencia inválida equivale a tomar una decisión sin fundamento².

² En este capítulo se ha abordado solamente el tema de la validez. Otra cuestión técnica asociada a la validez es el tema de la consistencia de las mediciones, denominada confiabilidad. Esta se refiere a tres cosas interrelacionadas:

(1) A la noción de la estabilidad de la medición. En este sentido, nos preguntamos si las pruebas o encuestas arrojan resultados similares siempre que se aplican a sujetos similares en condiciones similares. (2) A su nivel de precisión, esto es, la relación de los resultados de la medición con la "realidad" que mide. (3) A la confiabilidad, es decir, a la cantidad de error (llamada varianza sistemática) que contiene la medición. Si una prueba no mide con confiabilidad lo que se propone medir, no es válida. Es importante señalar, sin embargo, que la confiabilidad es una condición necesaria, mas no suficiente, para la validez. Sin confiabilidad no hay validez, pero la confiabilidad no es garantía de validez.

CAPÍTULO III

EL DISEÑO DE LAS PRUEBAS PARA MEDIR LOGRO ACADÉMICO: ¿REFERENCIA A NORMAS O A CRITERIOS?

Juan Manuel Esquivel^(*)

¿Qué opciones existen para medir, a escala nacional, los conocimientos que los estudiantes adquieren en las escuelas?

¿Se hacen las mediciones en Latinoamérica con el propósito de comparar los logros de grupos de estudiantes con otros grupos de estudiantes?

¿Se realizan, en cambio, para medir si estos han logrado los aprendizajes que el sistema educativo pretende que ellos logren?

¿Cuáles son las diferencias y similitudes conceptuales y metodológicas entre esas dos formas de realizar la medición?

¿Cuál es el papel de la evaluación del desempeño y evaluación auténtica en los sistemas de medición de la región?

INTRODUCCIÓN

A los funcionarios de ministerios de Educación y entidades encargadas de los sistemas de medición de logro de los países de América Latina se les presenta la siguiente disyuntiva al desarrollar las evaluaciones educacionales:

- elaborar pruebas que permitan comparar el logro de grupos de estudiantes con otros grupos; o
- elaborar pruebas que permitan descubrir qué aspectos, conocimientos u objetivos específicos logran los estudiantes.

Responder a este dilema implica desarrollar pruebas en base a paradigmas con fundamentaciones teóricas diferentes y, en ciertos aspectos,

^(*) **Juan Manuel Esquivel.** Profesor de Educación Media en Química. B.Sc. en Química M. Sc. Medición e Investigación Educativa. Ph.D. en currículo. Catedrático de la Facultad de Educación de la Universidad de Costa Rica. Investigador retirado del IMEc-ucR. Consultor en Evaluación y Medición. Actualmente Director de dos proyectos educativos en la coordinación Educativa y cultural centroamericana. E-mail: sgcecc@racsa.co.cr

contradictorios. En el primer caso, se puede trabajar dentro del paradigma de medición referida a *normas*, mientras que en el segundo se debe recurrir al paradigma de medición referida a *criterios*.

En la mayoría de los países de la Región se ha recurrido a las pruebas referidas a normas, en las que se privilegia la función de ordenamiento o “discriminación” entre grupos o individuos. Este enfoque está fuertemente marcado por su función principal, que históricamente ha sido la de seleccionar individuos para el ingreso al ejército o a las universidades. En esos casos no importaba tanto si el individuo dominaba o no ciertos campos del conocimiento, sino distinguir a los individuos más aptos de los menos aptos. El enfoque de pruebas referidas a criterios, en cambio, se propone comprobar principalmente si los individuos dominan un cierto campo de contenidos y/o destrezas.

Respecto a la disyuntiva planteada, en este capítulo se analiza un ejemplo de lo que típicamente se encuentra en la Región en cuanto a desarrollo y validación de pruebas de conocimiento dentro del paradigma referido a normas. Luego, se revisan algunas de las diferencias entre los paradigmas señalados y se da un ejemplo de desarrollo de una prueba de acuerdo con el paradigma referido a criterios. Finalmente, se hace una referencia al impulso que, en algunos países desarrollados, se ha comenzado a dar al empleo de las denominadas *pruebas de desempeño* y a la *evaluación auténtica*.

Las pruebas referidas a normas

Como se mencionó, **la mayoría de los países latinoamericanos ha desarrollado pruebas para sus sistemas de medición del logro dentro del modelo psicométrico de las pruebas referidas a normas.**

La información producida por estas pruebas generalmente se ofrece en forma de promedios basados en el número de preguntas correctas obtenidas por los estudiantes o como una escala derivada de esta información básica, por ejemplo, el porcentaje de respuestas correctas o la nota en términos de la escala de calificación empleada en cada país. Estos promedios, aunque tienen una utilidad innegable para realizar comparaciones entre los diferentes niveles de desagregación de las variables de interés en las muestras (por ejemplo: urbano-rural, público-privado), tienen escaso sentido pedagógico, pues no entregan información real sobre el logro de conocimientos, destrezas o habilidades específicas de

parte de los estudiantes. ¿Qué información de utilidad le comunican los promedios a un maestro de aula que le permita mejorar su trabajo con los niños? ¿Qué utilidad tienen para un curriculista en el Ministerio de Educación?

Entre las razones por las cuales se ha recurrido al enfoque referido a normas están:

- La abundancia de experiencia e información internacionalmente disponible sobre los procedimientos que tradicionalmente se han seguido en la elaboración y validación de pruebas de este tipo.
- La limitada formación y capacitación académica en el área de la medición a la cual han podido acceder los funcionarios ministeriales encargados de desarrollar las pruebas, se ha dado sobre los principios de la teoría de pruebas referidas a normas.
- La disponibilidad de paquetes estadísticos de cómputo que permiten realizar análisis de ítems tradicionales o novedosos y otros análisis técnicos.

En el Recuadro 2, se presenta un caso que se podría encontrar típicamente en los sistemas de medición del logro en la Región, sobre el cual se harán los siguientes comentarios y observaciones puntuales referidas a las diferentes etapas de su desarrollo.

RECUADRO 2**La medición del logro en América Latina:
Un ejemplo típico de prueba bajo el paradigma referido a normas.****Etapa 1.**

Durante la preparación de un proyecto de préstamo o de donación con un organismo internacional, se detectó la necesidad de tener un sistema de medición del logro académico de los estudiantes. Una vez financiado el proyecto, se inició la preparación de pruebas para tercero y sexto grado de primaria en matemáticas y lenguaje, utilizando como base el currículo prescrito. Como primer paso del proceso, se definió una tabla de especificaciones, en la cual los contenidos del currículo se dividieron por áreas y estas, a su vez, se subdividieron en contenidos más específicos. La tabla se balanceó de acuerdo con la complejidad cognitiva (nivel taxonómico) que se pretende tengan los ítems y con el número de ítems con los que se quería medir esos conocimientos.

Etapa 2.

Se procede a la elaboración de los ítems para llenar las expectativas establecidas en la tabla de especificaciones. Para ello, se ha acostumbrado realizar grandes operativos, como talleres con maestros en diferentes lugares del país, a quienes se les pide redactar ítems de preguntas de selección múltiples para diferentes contenidos y en distintos niveles taxonómicos. Así se recogen enormes cantidades de ítems, lo que se justifica bajo el argumento de que le da validez curricular a la prueba. En otros lugares, la elaboración de ítems está a cargo de un número reducido de personas que tiene una vasta experiencia y/o capacitación en estas tareas.

Etapa 3.

Se someten los ítems a una revisión, tarea cuya complejidad y sistematización varía desde una revisión relativamente informal para detectar defectos gruesos en la estructura hasta revisiones con hojas de calificación preparadas con ese propósito expreso y, en unos pocos casos, revisiones de relación ítem-contenido llevadas a cabo por jueces independientes.

Etapa 4.

Se prepara una prueba piloto de ítems. Se selecciona una muestra de escuelas o aulas de acuerdo con un plan de muestreo previamente establecido. En la mayoría de los casos el muestreo es intencional, procurando que incluya escuelas representativas de los diversos estratos de interés del sistema de medición y que cada ítem sea respondido por entre 150 y 300 estudiantes. Los ítems disponibles se agrupan en diversos formularios que se aplican simultáneamente. El propósito fundamental de la aplicación piloto es poder hacer análisis estadísticos para conocer índices que caracterizarán a los ítems. Generalmente estos análisis se ejecutan con paquetes estadísticos comerciales. Con los parámetros obtenidos (dificultad, correlación ítem-puntaje total de la prueba –discriminación–, frecuencia de respuesta según opción, etc.), se realiza una selección de los ítems que constituirán las pruebas definitivas. Esto generalmente se rige por los principios de la teoría de pruebas referidas a normas, que establecen como preguntas ideales aquellas que tienen una dificultad cercana al 50% y una discriminación por encima de 0.40 o correlaciones ítem-puntaje total positivas y significativamente mayores que cero.

Etapa 5.

Una vez aplicada la prueba, se reportan resultados en términos de los promedios obtenidos en cada una de las áreas en que se dividió la tabla de especificaciones. Estos se reportan como puntajes de logro y se interpretan en términos de dominio de cada una de esas áreas, cuando se supera cierto puntaje de corte. Además, se reportan resultados de promedios totales de las pruebas, que resultan muy convenientes para establecer comparaciones entre los diversos niveles de desagregación de las variables de la muestra (grupos de alumnos o escuelas).

Etapa 1.

Preparación de las especificaciones de los contenidos a medir:

Cabe destacar tres hechos relevantes respecto a este tema:

- La decisión que generalmente se hace de tomar el currículo prescrito como base de las pruebas. Al hacerlo, se asume que este es conocido y comprendido por todos los maestros, que ellos han recibido la capacitación adecuada para ejecutarlo, que tienen acceso a los mismos materiales de enseñanza y, por lo tanto, los niños han tenido alguna oportunidad para aprenderlo. Dado que asumir estas condiciones no es siempre justificable, al menos se debería planear la ejecución de un estudio paralelo a las pruebas para conocer en qué medida esas condiciones realmente están dadas en cada caso.
- La decisión, que normalmente se toma, de basar el diseño de la prueba en una tabla de especificaciones dividida por áreas y estas, a su vez, en contenidos más específicos y de emplear alguna taxonomía para catalogar la complejidad cognitiva con que se quiere medir los contenidos. La complejidad cognitiva de los contenidos, en general, incluye desde los conocimientos memorísticos hasta el empleo del razonamiento lógico. Se presentan diferentes maneras de denominar estos niveles de complejidad cognitiva de acuerdo con la taxonomía que se emplee, aunque la más popular es la taxonomía de Bloom y sus colaboradores. Los niveles taxonómicos señalan la complejidad cognitiva que se pretende que tengan los ítems con que se medirán los contenidos determinados en la tabla de especificaciones. Debe tenerse en mente que la tabla de especificaciones es un instrumento que se emplea con el propósito de tener alguna seguridad de que la prueba será una muestra representativa de los contenidos considerados para ser medidos con la prueba y los niveles taxonómicos con que se quiere medir esos contenidos. Dentro del paradigma de pruebas referidas a normas, la evidencia de que la prueba es una muestra representativa de la totalidad de los contenidos considerados para ser medidos en una prueba es una información fundamental para establecer la validez de la interpretación de los resultados. La intención de dividir el contenido por áreas, en la tabla de especificaciones, se hace evidente cuando se leen los informes de resultados, pues en ellos se reportan los promedios de estas áreas como puntajes de logro. Esto

significa que se hace una interpretación propia de la medición referida a criterios (ver más adelante) con una prueba que se ha diseñado como referida a normas.

- La enorme debilidad que tiene cualquier sistema taxonómico como medio de balancear una tabla de especificaciones. Esta debilidad deriva del hecho que el nivel taxonómico que se le atribuye a un ítem determinado está definido por la experiencia de enseñanza que tiene la persona que lo juzga. Así, un mismo ítem recibirá una variada gama de niveles taxonómicos cuando se somete a juicio de varias personas.

Etapa 2.

Elaboración de ítems

- La escritura de preguntas de selección u opción múltiple de buena calidad es un trabajo especializado que requiere para su ejecución de personal con experiencia y capacitación en esta labor. El típico constructor de ítems debe tener dos características: (i) un excelente dominio de la materia sobre la que escribirá ítems, y (ii) una amplia y probada experiencia en labores de escritura y revisión de ítems. Las preguntas de opción múltiple, si están adecuadamente elaboradas, permiten medir habilidades complejas y, por lo tanto, la crítica de que con ellas solamente se mide memoria y habilidades simples es bastante infundada, o es válida únicamente para pruebas mal diseñadas.
- Se debe reconocer que los operativos para que los maestros escriban ítems de las pruebas, como ocurre con frecuencia, pueden tener beneficios políticos para la aceptación del sistema de medición y para capacitar maestros. Sin embargo, hay que tener claro que estos operativos no necesariamente contribuyen a la calidad de la prueba. El argumento de que así se le da validez curricular a la prueba, no tiene sentido, dado que a los maestros se les pide redactar ítems sobre contenidos y niveles taxonómicos que han sido decididos en el nivel central. Estos contenidos en esos niveles taxonómicos pueden nunca haber sido enseñados por los maestros que han construido ítems para medirlos. Por otra parte, se ha comprobado que la inmensa mayoría de los ítems escritos por maestros en estos operativos, son desechados en la primera revisión.

Etapa 3.

Revisión de ítems que se incluirán en las pruebas

- En general, en la Región el desarrollo y validación de las pruebas empleadas han sido limitados, pues no se ha puesto la debida atención a los procesos de comprobación de la calidad de la estructura de los ítems y a establecer claramente la relación de cada ítem con el contenido que pretende medir. Esto es particularmente serio para efectos de establecer la evidencia de validez necesaria para la interpretación y uso del resultado de la medición. La revisión de los ítems deberla dividirse en dos aspectos: (i) una revisión estructural sistemática, hecha por jueces especialistas que dominen la disciplina para la que se escribieron los ítems y tengan amplia experiencia en la escritura de ítems de selección múltiple; y (ii) una comprobación de la relación entre el ítem y el contenido que se supone que este mide, también a través de jueces que tengan una experiencia reciente en la enseñanza en el nivel en el que se aplicará la prueba y dominio de los conocimientos de la disciplina. Ellos trabajarán independientemente, juzgando si cada ítem mide o no el contenido que se supone que mide. Una mayoría calificada de los jueces (alrededor de un 75%) tendrá que mostrar acuerdo en la relación de cada ítem con un contenido.

Etapa 4.

Aplicación de prueba piloto

En esta etapa se observan dos debilidades básicas:

- Casi nunca se establece como objetivo de la aplicación de la prueba piloto la obtención de retroalimentación sobre la prueba por parte de los estudiantes y de los docentes. La recolección de información cualitativa acerca del contenido que cubren las preguntas de la prueba y sobre la claridad y comprensión de los ítems sería de vital importancia para el esfuerzo por acumular evidencia de validez. Esto podría lograrse mediante una discusión con los estudiantes y una conversación con el maestro.
- La selección de los ítems que formarán las pruebas definitivas sobre la base de las preguntas que muestran una dificultad cercana al 50% y una discriminación de 0.4, como se suele hacer, tiene algunas

consecuencias, pues debido a ello comúnmente los resultados reportados como promedios de los puntajes totales de las pruebas están alrededor del 50%. En otras palabras, sería absolutamente imposible obtener resultados que no estuvieran en el rango de alrededor de la mitad del puntaje posible. Por otra parte, este procedimiento implica descartar los ítems que resultan muy difíciles o muy fáciles, aunque los mismos sean buenos desde el punto de vista pedagógico y midan competencias relevantes, lo que implica perder la posibilidad de recoger información valiosa sobre las capacidades y conocimientos de los estudiantes.

Etapa 5.

Reporte de los resultados una vez aplicadas las pruebas

- Dados los procedimientos seguidos en su desarrollo, las interpretaciones de logro que suelen darse a los resultados por área carecen de sustento teórico y empírico. Esto significa que se está realizando una interpretación referida a criterios para una prueba referida a normas.

Resumiendo algunos aspectos que se derivan de lo analizado, cabe destacar lo siguiente:

1. Las pruebas referidas a normas tienen un espacio en los sistemas de medición del logro si su desarrollo es congruente con el objetivo que se pretende alcanzar al aplicar las pruebas, como sería la comparación del rendimiento general de los estudiantes de acuerdo a variables tales como sexo, rural-urbano, sostenimiento de las escuelas, regiones geográficas, etc.
2. Existen usos apropiados para las pruebas referidas a normas y para las pruebas referidas a criterios, dependiendo del grado de la “granularidad” (ver capítulo 1 de este documento) de lo que se mide y a quién se mide.
3. De acuerdo con el análisis aquí realizado, existiría aún mucho espacio para mejorar el desarrollo y validación de las pruebas referidas a normas que actualmente se emplean en la Región.

Las pruebas referidas a criterios

Unos pocos sistemas de medición del logro en la Región han desarrollado pruebas referidas a criterios y las han sometido a un proceso de validación. Esta alternativa de medición tiene la gran ventaja de que permite obtener información con mucho significado pedagógico, pues evalúa los conocimientos, destrezas y habilidades específicas que un grupo de estudiantes logra dominar.

RECUADRO 3

Ejemplo de una prueba referida a criterios desarrollada en América Latina

El Ministerio de Educación define como objetivo de las pruebas de final de la Educación Primaria brindar información específica sobre el logro de los objetivos fundamentales del currículo por parte de los estudiantes. Para ello, el Departamento de Pruebas Nacionales (DPN) del Ministerio decide elaborar y someter a proceso de validación pruebas referidas a criterios en las cuatro asignaturas básicas.

La primera pregunta que plantea el DPN a las autoridades políticas es: ¿Se quiere medir el currículo prescrito o el currículo enseñado? La respuesta de dichas autoridades es el currículo prescrito o sea el “deber ser” curricular. Las siguientes preguntas clave que se hace el DPN son: ¿cuáles son los objetivos fundamentales del currículo? y ¿quiénes son las personas más indicadas para realizar la selección de los objetivos fundamentales? Para responder la primera pregunta se realizó una lectura e interpretación de los programas de estudio (fundamentados en el humanismo-constructivista) de cada una de las asignaturas. Esto dio origen a listados de entre 50 y 70 objetivos de aprendizaje por asignatura. Esta interpretación hecha por pares de especialistas en cada asignatura, miembros del equipo de la DPN, se sometió al juicio de grupos de 10 especialistas en cada asignatura del Departamento de Currículo del Ministerio. A estos jueces se les solicitó, en primer lugar, manifestar su acuerdo o desacuerdo con la interpretación hecha de los programas de estudio y participar en una discusión para, mediante el consenso, llegar a una interpretación única de los programas de estudio. En segundo lugar, se les solicitó que trabajando independientemente señalaran los 30 objetivos que consideran más importantes de lograr por un estudiante que termina la Educación Primaria. Luego, se les solicitó realizar la priorización de 20 objetivos de los 30 seleccionados. Mediante un procedimiento estadístico, se definieron los 20 objetivos que tuvieron las más altas prioridades promedio, los cuales constituyeron los objetivos fundamentales de cada una de las asignaturas, y para cada uno de estos objetivos se desarrollaron especificaciones de contenido.

A continuación, el DPN procedió a seleccionar la técnica mediante la cual se definirían las especificaciones de contenido. En este caso se escogió la técnica de los objetivos amplificadas. Los pares de técnicos por asignatura fueron capacitados en el empleo de la técnica en un taller de una semana. Durante ese período y las dos semanas siguientes, estos técnicos desarrollaron las veinte especificaciones de su asignatura. El paso siguiente consistió en la validación de los objetivos amplificadas escritos. Para ello, se solicitó a cinco especialistas en la asignatura que escribieran un ítem para cada objetivo amplificado, de acuerdo con las condiciones establecidas en ellos. Si los ítems producidos por los cinco especialistas para cada uno de los objetivos amplificadas resultaban muy similares, se comprobaría que los objetivos amplificados cumplen adecuadamente la doble función de limitar el contenido y establecer y comunicar las reglas de estructuración de los ítems.

La fase siguiente fue la escritura de los ítems. Se decidió que se escribirían 12 ítems para cada objetivo amplificado. La escritura de los ítems estuvo a cargo de los pares de especialistas por

asignatura y de dos personas más seleccionadas entre el grupo de escritores de ítems, ya capacitados, que tiene el DPN. Estas dos personas adicionales por asignatura fueron profesores de educación primaria que en ocasiones anteriores habían probado su habilidad para escribir ítems de selección múltiple.

A continuación se llevaron a cabo dos procesos de vital importancia:

1. La revisión estructural de los ítems. Para ello se contrataron dos personas en cada asignatura. Estos profesionales tenían la doble característica de poseer una gran experiencia en la preparación y revisión de ítems de selección múltiple, y un reconocido dominio de la materia. Para que sirviera de base al trabajo de estos pares de jueces, se preparó una hoja de cotejo que resumía las principales características de la estructura de los ítems que ellos debieron examinar. Como producto de esta revisión se modificaron algunos ítems y se desecharon unos pocos.
2. El establecimiento del índice de la congruencia de cada uno de los ítems con su objetivo amplificado. Para esto se contrataron 10 jueces por asignatura, quienes tenían un probado dominio de la materia y alguna experiencia de enseñanza en la Educación Primaria. Se prepararon formularios en los cuales los jueces vertieron su juicio independiente y se acondicionó un local para facilitar su trabajo y el control que sobre ese trabajo tuvieron que realizar los especialistas del equipo del DPN. Mediante la lectura óptica de los formularios empleados por los jueces, se capturó la información producida en esta fase de juicio. Con un programa de cómputo apropiado se calculó el índice de congruencia para cada ítem. Los ítems con índices de congruencia menores a 0,75 fueron desechados.

En cada asignatura, con los ítems que presentaban índices de congruencia iguales o mayores a 0,75 se constituyeron folletos o cuadernillos de prueba con 40 ítems cada uno; cuatro ítems por cada objetivo amplificado. Se constituyeron seis cuadernillos diferentes, puesto que se tenían 20 objetivos amplificados y hasta 12 ítems por objetivo. En los casos en que se hubiesen desechado ítems por los procesos antes descritos, y no se tuvieran los 12 ítems por objetivo, se repetían ítems en algunos de los diferentes cuadernillos. Se seleccionó una muestra no aleatoria de escuelas de diferentes características geográficas, sociales, de tamaño y de financiamiento (públicas-privadas-subvencionadas). La meta era que un mínimo de 200 estudiantes respondieran cada formulario de cada asignatura. Además, se planificó la recolección de información cualitativa sobre los ítems, mediante las discusiones que se tuvieron con los estudiantes acerca de la claridad y comprensión de los ítems. A los maestros de los grupos de estudiantes a los que se les aplicaron los cuadernillos de prueba, también se les solicitó y se registró su opinión acerca de los ítems. Esta información cualitativa sirvió para detectar deficiencias de lenguaje en los ítems y se utilizó para modificarlos. Con las respuestas a los ítems y empleando los programas de cómputo apropiados, se calcularon la dificultad y la discriminación de Brenann para cada ítem.

Para cada objetivo amplificado se hizo un banco de ítems con aquellos ítems que presentaban los índices de congruencia de más alto valor y que, además, tuvieran un valor de discriminación cercano a cero y una dificultad mayor al 50% (el índice de dificultad señala el porcentaje de alumnos que contesta correctamente, lo que significa que si es mayor al 50% es un ítem relativamente fácil). De este banco de ítems se seleccionaron aleatoriamente cuatro ítems para medir cada objetivo amplificado. De acuerdo con la definición de prueba en la medición referida a criterios, estos cuatro ítems constituían una prueba. En cada asignatura, se constituyeron dos folletos o cuadernillos, y en cada uno de ellos se reunieron 40 ítems pertenecientes a 10 objetivos amplificados. De esta manera se cubrieron los 20 objetivos amplificados para los que se desarrollaron ítems. Los folletos o cuadernillos se aplicaron a muestras de estudiantes seleccionados aleatoriamente o estratificadas por región educativa, tamaño de escuela y zona geográfica.

Una vez analizada la información, se escribieron varios tipos de informes. A las escuelas y autoridades técnicas regionales y centrales se les hizo llegar uno en que se ofrecía el porcentaje de estudiantes que había dominado cada uno de los objetivos fundamentales, de acuerdo con los diferentes niveles de las variables en que se estratificó la muestra.

Con el ejemplo del Recuadro 3, se ilustra el tipo de procedimientos que emplea este enfoque.

Diferencias entre pruebas referidas a normas y pruebas referidas a criterios

Es común escuchar y leer la errónea aseveración de que lo único que distingue a las pruebas referidas a criterios de las de normas es la interpretación de los resultados, interpretación que puede ser relativa (o sea con respecto a la media aritmética y la variabilidad) o referida al logro. Esta confusión se deriva de una concepción errada de las características técnicas de las pruebas referidas a criterios y de un desconocimiento de la teoría que sustenta este paradigma.

Existen, de hecho, importantes diferencias conceptuales y metodológicas entre ambos tipos de pruebas, como se resumen en el Recuadro 4.

RECUADRO 4

Diferencias conceptuales y metodológicas de las pruebas referidas a normas y las referidas a criterios

DIFERENCIAS CONCEPTUALES	Pruebas referidas a normas	Pruebas referidas a criterios
1. Paradigma de base	Se basa en el paradigma psicométrico, cuya premisa es que los resultados de la medición de cualquier característica humana en una población se comportarán de acuerdo con la curva normal. Como consecuencia, privilegia maximizar la variabilidad y así asegurarse que los resultados de la aplicación de las pruebas se comportan normalmente.	Se basa en el paradigma educativo, cuyo principio es que la educación persigue que todos los niños aprendan; por consiguiente, se espera una distribución de resultados sesgada hacia los valores más altos de la escala de puntajes. La variabilidad no es una característica que importe, por lo cual no preocupa su valor.
2. Tipos de información que busca	Privilegia la comparación entre estudiantes o entre grupos de estudiantes	Privilegia la comparación de los logros de los estudiantes con respecto a las metas de aprendizaje o a las competencias que el sistema educativo persigue que estos alcancen.

DIFERENCIAS METODOLOGICAS	Pruebas referidas a normas	Pruebas referidas a criterios
<p>1. La definición de lo que se va a medir</p>	<p>Suele ser una definición general y vaga. Generalmente consiste en listados de conocimientos a manera de temarios o en listados de objetivos más o menos definidos.</p>	<p>La definición tiene que ser clara y específica, detallando el “dominio del conocimiento” que abarca el contenido por medir y las reglas básicas de estructuración de los ítems con que se medirá ese dominio. A estas definiciones se las conoce como “especificaciones de contenido”.</p>
<p>2. La definición de prueba</p>	<p>La prueba es el conjunto de ítems que forman una muestra representativa de todos los conocimientos, destrezas y habilidades que se quiere medir. El criterio de representatividad puede ser muy variado.</p>	<p>La prueba es el conjunto de “n” ítems, aleatoriamente seleccionados de una población infinita de ítems, que se emplea para medir únicamente una especificación de contenido. En otras palabras, se mide un solo conocimiento, habilidad o destreza.</p>
<p>3. La evidencia de validez</p>	<p>Para esta evidencia será muy importante: (a) la certeza de que los ítems tienen las características adecuadas en su construcción; (b) el proceso de juicio por el cual se recoge información acerca de la relación del ítem con el contenido que pretende medir; y (c) la evidencia que se necesita obtener sobre el ajuste entre los ítems seleccionados para constituir la prueba definitiva y las especificaciones contenidas en la tabla correspondiente.</p>	<p>Aquí es importante: (a) la certeza de que los ítems estén contruidos de acuerdo con las características propias de las preguntas de selección múltiple; y (b) un valor aceptable en el índice de congruencia entre cada ítem y su especificación de contenido.</p>
<p>4. Los procedimientos de cálculo de la confiabilidad y de análisis de ítems.</p>	<p>a) Los modelos empleados para estos efectos se fundamentan en la maximización de la variabilidad de dos variables que están correlacionadas, lo que permite asegurarse que la magnitud de la correlación entre esas dos variables será más alta. La correlación es la técnica fundamental empleada en el cálculo de la confiabilidad. b) Los parámetros resultantes del análisis de ítems, principalmente la dificultad y la discrimi-</p>	<p>a) Se da más importancia a la consistencia entre los resultados obtenidos por un grupo de alumnos a los cuales se les aplica la misma prueba o una prueba paralela en dos oportunidades distintas. El índice de confiabilidad resulta de la proporción de estudiantes cuyas respuestas en ambas oportunidades demuestren que sí (o no) dominan una competencia o especificación de contenidos. b) La selección de ítems que constituirán la prueba definitiva</p>

DIFERENCIAS METODOLÓGICAS	Pruebas referidas a normas	Pruebas referidas a criterios
	<p>nación o el índice de correlación ítem-puntaje total de la prueba, se usan como indicadores fundamentales para la selección de los ítems posibles de incluir en la prueba definitiva.</p>	<p>se realiza considerando en primer lugar el índice de congruencia entre el ítem y la especificación de contenido y, luego, considerando la discriminación y la dificultad.</p>
<p>5. La interpretación de los resultados de la medición.</p>	<p>La interpretación es relativa. El puntaje tiene significado al ser comparado con la media aritmética y la desviación estándar (o con las normas, si la prueba ha sido estandarizada o normalizada).</p>	<p>La interpretación es absoluta. El resultado se interpreta en términos del logro o no logro de la especificación del contenido medido o sea, en términos del dominio del conocimiento, habilidad o destreza medida.</p>
<p>6. Otras diferencias.</p>	<p>a) No necesita establecer un puntaje de corte, entendido esto como el puntaje que define si un estudiante domina o no la especificación de contenido, como ocurre en las pruebas referida a criterios.</p> <p>b) El número de preguntas que constituyen una prueba estará determinado, entre otros factores, por el objetivo de la prueba, por el nivel de escolaridad de los niños que tomarán la prueba, la asignatura que se mide, la tabla de especificaciones y el tiempo del que se dispone para su aplicación.</p>	<p>a) Necesita establecer un puntaje de corte.</p> <p>b) El número de preguntas dependerá fundamentalmente del tipo de decisión que se va a tomar con la prueba (formativa o sumativa) y con respecto a quién se va a tomar esa decisión (un individuo o una muestra de individuos). En general se determina que cuando las decisiones son formativas y para muestras de individuos, el número de ítems varía entre 3 y 5, mientras que decisiones sumativas e individuales requieren entre 8 y 10 ítems.</p>

Nuevas perspectivas en la medición del logro: la evaluación del desempeño y la evaluación auténtica.

Una nueva alternativa en materia de evaluación, que ha cobrado fuerza durante la última década en los países desarrollados, es la llamada “evaluación del desempeño”, un enfoque de medición según el cual los estudiantes deben producir sus respuestas o ejecutar tareas, en lugar de simplemente seleccionar la respuesta correcta entre varias alternativas.

El desempeño de los estudiantes se juzga con criterios preestablecidos, basados en el discernimiento humano. Los medios que emplea la

evaluación del desempeño han sido en el pasado usados por los educadores en las aulas, y requieren un tiempo sustancial de parte del estudiante:

- Preguntas de respuesta o final abierto.
- Producción de un ensayo.
- Resolución de problemas.
- Producción de materiales o discursos para exhibición pública.
- Elaboración de artefactos y documentos.
- Producción de un portafolios o muestras de trabajos realizados a lo largo de un período.

Se dice que la evaluación del desempeño es *auténtica* cuando las tareas que el estudiante ejecuta tienen como contexto situaciones propias del mundo real o recrean un contexto del mundo real. **Lo novedoso de este enfoque radica en el énfasis en la medición de conocimientos y habilidades complejas y de alto nivel de pensamiento, preferiblemente en un contexto de mundo real en el que se emplean esos conocimientos y habilidades.**

Es necesario, sin embargo, tomar en cuenta algunas características de este enfoque cuando se aplica a muestras masivas de estudiantes, y que pueden ser consideradas como una limitación:

- ***El tiempo que requiere un estudiante para completar una tarea específica*** (algunas veces todo un curso, como es el caso de los portafolios).
- ***La limitación técnica de depender del juicio humano para juzgar la calidad de la ejecución de la tarea.*** Este factor se ha señalado como una debilidad fundamental del enfoque, pues es muy difícil cumplir con principios básicos que permitan considerar confiable la calificación. Esta realidad, a su vez, influye en la calidad de la evidencia empírica que permite interpretar los resultados de estas mediciones de la forma que se pretenden interpretar.
- ***Bajo poder de generalización.*** En la mayoría de los casos que se observan en la práctica de la evaluación del desempeño, se mide un conocimiento o habilidad compleja con una sola tarea, dado lo extenso y complejo de la misma y el tiempo que consume su ejecución. Esto hace evidente una limitación de esta opción, dado el poco poder de generalización del resultado. La decisión del logro que se hace sobre la base de una respuesta a una única tarea tiene poca validez.
- ***Costo.*** Esto es una seria limitante para el uso de este tipo de mediciones en la Región, pues no solo se requerirá emplear más tiempo en el

proceso de medición, con el costo que esto implica, sino que se requerirá más materiales para realizarla. El costo sube aún más al agregar el pago del personal que califica las tareas, su capacitación y el necesario control para que se mantengan los niveles mínimos de la confiabilidad de la calificación.

Sin embargo, resulta indudable que es necesario medir el conocimiento y las habilidades complejas, y la idea de aplicar algunos de los medios de la evaluación del desempeño en las pruebas de los sistemas de medición del logro la Región resulta atractiva. Una posibilidad es hacerlo en pequeñas submuestras de estudiantes para, de esta manera, mantener los costos bajos y poder tener información sobre el logro de conocimientos y habilidades complejas, las cuales no se pueden medir con las pruebas tradicionales con preguntas de selección múltiple.

Conclusiones

Con lo expuesto en este capítulo se ha pretendido dar respuesta a las preguntas con las que se abrió el mismo. Se puede resumir lo escrito de la siguiente manera.

1. En la región latinoamericana todos los países tienen alguna forma de sistema de medición de logro. La mayoría de las pruebas que se emplean se desarrollan bajo los principios de la medición referida a normas. En muchos casos se ha descuidado la recolección de evidencia empírica que sustente la interpretación válida de los resultados. Con la medición basada en normas no es posible tener información específica y válida sobre el logro de conocimientos, habilidades y destrezas. Solo es apropiada cuando el objetivo de la medición es realizar comparaciones acerca del rendimiento académico general entre diversos estratos de la población sometida a medición.
2. Unos pocos países han ensayado la elaboración de pruebas referidas a criterios. En estos casos, se ha tenido una mayor preocupación por sustentar la validez de la interpretación de los resultados. La medición basada en criterios permite llegar a conclusiones sobre el logro específico de ciertos conocimientos, habilidades y destrezas, lo cual entrega información valiosa para evaluar el cumplimiento de los objetivos curriculares.

3. Existen diferencias conceptuales profundas entre la medición referida a criterios y la medición referida a normas. Estas diferencias conceptuales, a su vez, dan lugar a diferencias en los procedimientos metodológicos del desarrollo y validación de las pruebas.
4. La *evaluación del desempeño* y la *evaluación auténtica* ofrecen medios para medir el logro de conocimientos y habilidades complejas. En muchos países de Latinoamérica se están llevando a cabo reformas curriculares profundas que dan mayor importancia al logro de habilidades complejas. Por ello, resultaría conveniente que los sistemas de medición incluyan alguno de los medios de la evaluación del desempeño en los instrumentos de medición que empleen. Esto daría validez curricular a los resultados de la medición y permitiría un mejor y mayor alineamiento entre la medición del logro y las reformas curriculares. Sin embargo, se debe tener en cuenta que, por sus características, la evaluación del desempeño presenta limitaciones de otra índole en cuanto a la confiabilidad y validez de sus resultados. Asimismo, el costo de su empleo es un factor que deberá considerarse cuando se planifica su uso en los sistemas de medición del logro en la Región.

CAPÍTULO IV

LA INFORMACIÓN SOBRE FACTORES SOCIALES E INSTITUCIONALES ASOCIADOS A LOS RESULTADOS

Pedro Ravela^(*)

- ¿Es necesario incluir cuestionarios de familia y encuestas a maestros en las mediciones de aprendizaje o alcanza con la aplicación de pruebas?
- ¿Para qué puede resultar útil la información sobre los contextos sociales e institucionales?
- ¿Es adecuadamente aprovechada la información que en el presente muchos países recogen junto con la aplicación de pruebas?
- ¿Es posible mejorar la calidad de los instrumentos de recolección de este tipo de información?

INTRODUCCIÓN

En la mayoría de los países latinoamericanos se aplica, junto con las pruebas de logro, cuestionarios dirigidos a recoger información de una enorme gama de variables relacionadas con las características de las familias y hogares en que viven los alumnos, así como acerca de las características de las escuelas a las que asisten y los maestros que los atienden. Sin embargo, generalmente esta información está siendo muy poco aprovechada y no forma parte de los reportes nacionales.

La mayor parte de los países suele limitarse a informar los resultados bajo la forma de porcentajes de respuestas correctas para las pruebas en su conjunto o para partes de ellas, por lo general desagregados por jurisdicción política/geográfica (provincia, región, estado, departamento) y tipo de escuela (urbano/rural, público/privada).

^(*) **Pedro Ravela.** Profesor de Educación Media en Filosofía, Magíster en Ciencias Sociales con Especialización en Educación; Gerente de Investigación y Evaluación en la Administración Nacional de Educación Pública (ANEP) de la República Oriental del Uruguay. Profesor de "Evaluación de la Gestión Curricular" en la Maestría en Educación de la Universidad Católica del Uruguay. E-mail: pravela@adinet.com.uy.

En casos excepcionales, se ofrece información sobre las variables relativas a las familias, es decir, externas a los sistemas educativos, más que acerca de las variables escolares sobre las cuales los Ministerios pueden tomar decisiones.

Adicionalmente, pocos países han desarrollado trabajos sistemáticos de investigación acerca de los factores escolares asociados a las diferencias de resultados de aprendizaje entre las escuelas.

A continuación se analizarán los problemas relacionados con la subutilización de la información “de contexto”, así como con la falta de investigación sobre la relación entre esta información y los resultados educativos.

Problemas relacionados con la falta de contextualización sociocultural de la información sobre resultados de las pruebas

La ausencia de caracterización sociocultural de las poblaciones a las que “enseñan” los distintos sectores del sistema educativo impide extraer conclusiones válidas acerca de la eficacia de dicha enseñanza. Normalmente aparecen como menos eficaces los sectores del sistema educativo que atienden a la población con mayores carencias, al tiempo que aparecen como mejores los sistemas educativos de las provincias o regiones cuya población está más alfabetizada y vive en mejores condiciones. Del mismo modo, normalmente se reportan mejores resultados en la educación privada en relación a la educación pública, pero no se analiza el tipo de selección social que uno y otro sector hacen del alumnado que atienden.

Al respecto, cabe tener presente las diferencias que se presentan por diversas categorías.

- **Educación pública y educación privada.** Graficaremos esto con un ejemplo: En un país, la comparación de resultados entre el conjunto de las escuelas públicas y el conjunto de las privadas muestra diferencias de 25 puntos porcentuales en la proporción de alumnos que logra un nivel satisfactorio en la prueba de matemática. Sin embargo, el análisis de los datos socioculturales indica que diferencias similares existen entre ambos tipos de escuelas, en variables tales como los niveles de educación alcanzados por los padres y madres de los niños, la existencia de libros en los hogares, el nivel general de equipamiento de los mismos y las condiciones de las viviendas. Cuando las diferencias de logro entre es-

cuelas se analizan controlando las variables socioculturales, las diferencias en la proporción de alumnos que logran un nivel satisfactorio en el conjunto de la prueba se reducen a unos 5 puntos porcentuales y, en algunos sectores sociales, son favorables a las escuelas públicas.

- **Diferencias territoriales.** Algo similar ocurre con la presentación de los datos en función de agregaciones político/geográficas. Cuando se presentan los resultados desagregados por provincia, estado, región o departamento, sin información adicional, la conclusión inmediata para el lector no especializado es que las escuelas y los maestros deben estar trabajando mejor en aquellas regiones en que los resultados son más “altos”. Sin embargo, normalmente estas serán las regiones con mayores tasas de alfabetización y con mejores indicadores de desarrollo en general.
- **Lo urbano y lo rural.** Los reportes nacionales suelen entregar la información desagregada en función del carácter urbano o rural de la escuela. Sin embargo, es preciso señalar que esta opción desconoce la enorme disparidad y heterogeneidad que normalmente existe al interior del mundo urbano. En dicha categoría quedan incluidas las escuelas pertenecientes a pequeños poblados del interior –probablemente muy similares a las rurales–, las escuelas ubicadas en zonas marginales de la periferia de las grandes ciudades y las de los barrios acomodados y altamente educados de esas mismas ciudades. Es discutible pues, la relevancia de comparar al conjunto de escuelas urbanas en relación a las rurales. Del mismo modo, es discutible en muchos países tratar a las escuelas rurales como una categoría homogénea, ignorando las diferencias culturales y lingüísticas que en algunos casos existen en su interior.

Recomendaciones

A partir de lo mencionado, parece necesario reflexionar **sobre cómo establecer formas relevantes de caracterización sociocultural de los niveles de desagregación de la información, de modo de poder hacer comparaciones entre establecimientos, departamentos o provincias que atienden poblaciones con algún grado de similitud.** En este sentido, es recomendable:

- Utilizar la información social recogida en los operativos de evaluación, u otra información sociocultural disponible a partir de los Cen-

sos o Encuestas Nacionales de Hogares, para caracterizar a los niveles de desagregación elegidos. Ello permitiría, junto con la comparación global, ofrecer comparaciones y generar “competencia” al interior de ciertos segmentos del sistema educativo que atiende a sectores de la población en cierto modo similares. En los casos en que la información se entrega desagregada a nivel de escuela, esto es aun más importante.

- Adoptar una metodología de “valor agregado” cuando se desea emitir un juicio sobre la calidad de una escuela o de una jurisdicción del sistema educativo. Se denomina enfoque de “valor agregado” a aquellas evaluaciones en las que se intenta medir la calidad de una escuela o jurisdicción no solo en función de sus “resultados absolutos”, sino principalmente en función de sus “resultados ajustados” por el tipo de alumnado que la escuela atiende: lo que logra por encima o por debajo de lo anticipable de acuerdo a la población con la que trabaja.
- Mejorar el diseño de instrumentos que recojan información sociocultural de base relevante, dado que en muchos países los cuestionarios están dirigidos principalmente a recoger opiniones de las familias sobre el sistema educativo y la escuela, pero no relevan datos “duros” que permitan caracterizar a esas familias.

Problemas relacionados con la información sobre las características de las escuelas y la enseñanza

Por lo general, todos los sistemas nacionales aplican en sus operativos encuestas a maestros y directores en las que se recoge información sobre materiales didácticos empleados, clima escolar y años de experiencia del maestro, entre otros. Sin embargo, en los Informes Nacionales se reporta muy poco sobre estos aspectos y su relación con los resultados de aprendizaje.

La ausencia de difusión de información respecto a las variables estrictamente escolares que están asociadas con los resultados de las pruebas, implica desaprovechar información sumamente valiosa para la adopción de decisiones de intervención y mejoramiento. Los aspectos estrictamente escolares y pedagógicos son los únicos susceptibles de ser modificados, en el corto plazo, desde la política educativa y desde las decisiones que cotidianamente maestros y directivos toman al interior de las escuelas. Su ausencia en los reportes nacionales puede contribuir a generar la imagen de que el sistema educativo nada puede

hacer ante la fatalidad de las diferencias sociales. Sin embargo, una vez que los niños ingresan a la escuela, lo que allí ocurre cuenta en términos de aprendizaje. De hecho, se ha mostrado que al interior de una misma categoría social existen diferencias en los niveles de logro de las escuelas, las que son atribuibles a lo que estas hacen o dejan de hacer.

Dos dificultades principales parecen plantearse en relación a la información sobre factores escolares:

- ***La sobreabundancia de información que al respecto se recoge***, por lo general sin un plan de análisis y difusión previos, que luego hace sumamente difícil decidir cómo organizar esa información y cómo vincularla con los datos sobre aprendizaje.
- ***La asistematicidad a través del cual se construyen los instrumentos de recolección de la información de contexto***. El diseño de este tipo de instrumentos se realiza en parte en base a la intuición, en parte para satisfacer requerimientos de información de distintas unidades de los Ministerios, en parte en base a la acumulación de conocimiento respecto a factores de efectividad y en parte en base a los modelos utilizados por otros países, pero sin que se desarrolle un proceso de pilotaje y análisis de lo que los instrumentos pueden o no rendir.

Recomendaciones

- Es necesario avanzar en el desarrollo de una metodología para el diseño de los instrumentos de relevamiento de “factores escolares” que incluya:
 - un marco conceptual explícito respecto al papel de los factores escolares, que sistematice y organice la investigación existente sobre escuelas y prácticas de enseñanza eficaces;
 - la identificación más precisa de las variables escolares que es *relevante y posible* medir en el marco de un operativo nacional de evaluación, teniendo en cuenta, aquellos aspectos sobre los cuales es posible la toma de decisiones tanto desde la política educativa como desde el interior de los establecimientos, así como las limitaciones propias de los cuestionarios autoadministrados;
 - mejores modos de formular las preguntas, así como el desarrollo *ex ante* de escalas dirigidas a medir aspectos específicos tales como el clima institucional, el empleo del tiempo en el aula, el currículo implementado, los enfoques didácticos, los tipos de actividades realiza-

- das por los niños, la utilización de los materiales y textos además de su mera existencia, entre otros; y
- el pilotaje y validación previa de los instrumentos.
 - Se requiere avanzar, en particular, en la recolección de información acerca de lo que realmente se enseña en las escuelas. Muchas veces los niveles de logro insatisfactorios en ciertas áreas no reflejan una enseñanza “no efectiva” sino, sencillamente, ausencia de enseñanza de ciertos temas y dominios.
 - Es preciso publicar la información existente en cada país sobre factores escolares, incluso simplemente bajo la forma de tablas descriptivas de la distribución de las diferentes variables. Ello contribuiría al conocimiento sobre lo que está ocurriendo al interior de los sistemas educativos, permitiría comenzar a realizar comparaciones entre sistemas educativos y, hacia el futuro, podría constituirse en una forma de preparar el terreno para la construcción regional de indicadores educacionales comparables. Asimismo, permitiría ir acumulando conocimiento para afinar el tipo de preguntas que es útil formular, aligerar los cuestionarios o ir pasando en evaluaciones sucesivas a indagar nuevos aspectos.
 - Se necesita diseñar mejores formas de reportar los resultados de las pruebas, junto con la información sobre contextos sociales y factores escolares. Esto implica preguntarse: ¿Cómo ofrecer a los diferentes destinatarios –autoridades, otras unidades ministeriales, maestros, opinión pública– información que permita una lectura más compleja de los datos, sin abrumar a los eventuales lectores? ¿Cómo diversificar los tipos de informes que se producen? ¿Es posible avanzar hacia ciertos “formatos tipo” más sofisticados que los existentes hasta el momento (porcentaje de respuestas correctas por jurisdicción político/geográfica y por tipo de escuela)?

Problemas relacionados con la falta de investigación sistemática acerca de factores escolares asociados con el aprendizaje

Además de introducir mejoras en las formas de reportar los resultados de las evaluaciones nacionales, **parece necesario mejorar el aprovechamiento de la información generada por los sistemas nacionales de medición con fines de investigación sobre el modo en que los diversos**

factores inciden sobre los aprendizajes. Al respecto hay que tener presentes algunos problemas centrales que se enfrentan en este terreno:

- ***Las investigaciones suelen utilizar evaluaciones que realizan los países sobre niveles de logro al final de ciertos grados o niveles de enseñanza, pero no necesariamente aprendizaje en un cierto período de tiempo.*** El nivel de logro de un estudiante al final de cierto grado escolar depende de múltiples factores ajenos a lo que ocurrió durante ese año en su aula. Tiene relación, por ejemplo, con la historia escolar anterior de los integrantes del grupo y con la acumulación de conocimiento con la que llegaron. En rigor, una evaluación de aprendizaje cuyo objetivo es investigar acerca de los factores que explican esos aprendizajes, exige contar con mediciones de conocimientos y competencias al inicio y al final del año escolar. Solo así se podrán establecer relaciones entre lo que ocurrió en la escuela ese año y el avance de los alumnos en términos de aprendizaje. Probablemente, la realización de operativos de evaluación al inicio y al final de un mismo año esté fuera del alcance de las posibilidades logísticas y económicas de los países de la Región. Sin embargo existen caminos intermedios a explorar, como realizar la medición inicial en muestras más pequeñas o, en países que evalúan grados sucesivos (por ejemplo, 50, y 60 grados de primaria), considerar como medida de aprendizaje de los alumnos del grado superior a la diferencia de logro con relación a los alumnos del grado inferior.
- ***La investigación sobre factores asociados implica la utilización de técnicas estadísticas sofisticadas de carácter multivariado y plantea severas exigencias en cuanto a la conformación y calidad de las bases de datos.*** Normalmente será necesario contar con información completa sobre todas las variables incluidas en el modelo para todos los alumnos, lo que no siempre es posible cuando los relevamientos se efectúan a través de cuestionarios autoadministrados y se trabaja con muestras muy grandes. Una de las principales limitaciones de los modelos estadísticos multivariados es que la posibilidad de que una variable “ingrese” al modelo depende del grado en que la misma varía en la realidad. Aquellas variables con menor variabilidad difícilmente ingresan, lo que no implica que no sean relevantes en la producción de los resultados.
- ***A través de la investigación de corte estadístico, aun la más sofisticada, se pueden construir ciertos tipos de conocimiento y de información, pero no es suficiente para la comprensión de los fenóme-***

nos. Es necesario desarrollar también estrategias de investigación de carácter cualitativo o “estudios de casos”, que permitan una mirada distinta sobre aspectos que los instrumentos usualmente utilizados no pueden captar. El hecho de contar con una medición de logros educativos y contextos sociales e institucionales, constituye un formidable “mapa” sobre el cual efectuar la selección de casos relevantes para un estudio en profundidad. Asimismo, la acumulación de conocimiento en el área parece requerir también de investigaciones de corte “cuasi experimental”, que permitan medir y controlar un conjunto de variables, como las relativas a las prácticas de enseñanza. Por otra parte, es necesario preguntarse acerca de qué tipo de decisiones de política educativa es posible tomar a partir de los resultados de un análisis estadístico. Difícilmente podrá o deberá establecerse una relación directa entre los resultados de un trabajo de investigación y la toma de decisiones de política educativa. Para decirlo en forma caricaturizada, normalmente un ministro no está esperando los resultados del análisis multivariado para decidir si compra libros o dicta una resolución para que los maestros dediquen más tiempo a enseñar quebrados. Es necesario un proceso de acumulación de conocimiento previo a la toma de decisiones, más allá de que esta está regida además por otro tipo de consideraciones y restricciones.

Recomendaciones

- Es necesario propiciar el establecimiento de asociaciones y convenios de colaboración entre las unidades de medición y centros de investigación especializados, de modo que estas últimas permitan potenciar el aprovechamiento de las bases de datos existentes mediante la realización de trabajos que las unidades no logran llevar adelante y, simultáneamente, colaborar en el mejoramiento de los instrumentos de medición y de la calidad de las bases de datos. Ello requiere, en primer término, voluntad política de parte de los Estados para facilitar el acceso a las bases de datos y, en segundo término, apoyar el desarrollo de las capacidades de investigación en estos temas. Esto incluye apoyar la capacitación de recursos humanos y la acumulación de conocimiento y experiencia en materia de investigación educativa no solo al interior de los Ministerios de Educación sino también en las universidades y centros no estatales.

CAPÍTULO V

ALTERNATIVAS TÉCNICAS EN RELACIÓN A LAS ESCALAS DE REPORTE DE LOS RESULTADOS DE LAS PRUEBAS DE RENDIMIENTO

Richard Wolfe

¿Cuál es el mejor modo de informar acerca de los resultados de una evaluación nacional: puntajes promedio para un conjunto de ítems, porcentaje de respuestas correctas a ítems individuales, porcentajes de alumnos que alcanzan cierto puntaje en una prueba, otros?

¿Cuándo es adecuado emplear unos u otros?

¿Cuál es el significado real de las cifras a través de las que se reportan los resultados?

¿Bajo qué condiciones es técnicamente válido realizar comparaciones entre las cifras obtenidas a partir de mediciones efectuadas en distintas áreas de contenidos o en distintos momentos en el tiempo?

INTRODUCCIÓN

El propósito de este capítulo es examinar las alternativas técnicas existentes con respecto a las escalas de reporte de los resultados de las evaluaciones nacionales.

Por “escala de reporte” se entiende, en primera instancia, a qué tipos de números se usan para presentar resultados (números o frecuencias simples, porcentajes, percentiles, puntajes en escalas, etc.), pero involucra también las maneras en que se registran, procesan, transforman, agregan y presentan los datos cualitativos y cuantitativos sobre las respuestas dadas por los alumnos. Las escalas de reporte están asociadas a:

- **La granularidad** (ver primer capítulo). Por ejemplo, si las estadísticas sobre los ítems de las pruebas que miden un contenido específico se informan por separado o se suman para conformar una medida a nivel de un área de conocimientos u otro tipo de escalas.
- **La representación cualitativa y numérica de los resultados**, tales como la presentación de ítems que realmente fueron aplicados en las

pruebas, la provisión de modelos ejemplares del trabajo de los alumnos, porcentajes simples para categorías de respuesta o cualquiera de las diversas maneras posibles de resumir, escalar y mostrar las distribuciones del logro escolar.

- **La consistencia y comparabilidad** de distintos contenidos y a lo largo del tiempo.

Históricamente, los reportes sobre logros educacionales se han derivado de las estadísticas educativas, con refinamientos sucesivos que progresaron desde las matrículas hacia las tasas de egreso y, luego, hacia algún tipo de “porcentaje de logro”.

Cuando se piensa más detenidamente en los reportes de logros, surge la pregunta respecto a qué significan esos porcentajes. Existe una desafortunada confusión en los modos de encarar las discusiones públicas y profesionales sobre los resultados de logro. Por ejemplo, es posible ver informes de evaluación en los cuales un logro de 50% es considerado bajo (o alto), cuando en realidad ello es simplemente una consecuencia arbitraria de un proceso de desarrollo y selección de ítems mediante el cual se eligió aquellos que tenían aproximadamente 50% de dificultad (ver tercer capítulo).

Asimismo, es de crucial importancia determinar en qué medida los resultados reportados de las evaluaciones son consistentes y comparables. Por ejemplo, ¿bajo qué condiciones tiene sentido decir que los logros en matemáticas son mayores o menores que los logros en lenguaje? ¿Cómo podemos decir que el desempeño en matemáticas es mayor o mejor en sexto grado que en tercer grado? ¿Cómo podemos producir reportes que demuestren que el desempeño ha mejorado de un año al otro?

Las investigaciones educacionales y psicométricas han proporcionado varios tipos de soluciones a estos problemas:

- El análisis de los dominios de los contenidos y de las pruebas, y la determinación de la “generalizabilidad” de los puntajes y porcentajes.
- El uso de métodos estadísticos para igualar y calibrar pruebas diferentes.
- El desarrollo de teorías de respuesta al ítem (IRT), que intentan colocar a los ítems y a los estudiantes en dimensiones o escalas latentes. (Se trata de un caso especial del método estadístico).

Todos estos enfoques tienen ventajas sustanciales, pero al mismo tiempo presentan peligros sustanciales. Antes de abordar las técnicas, es necesario considerar las maneras en que se reportan los datos de las evaluaciones y cómo ello depende del nivel de agregación, o granularidad (ver recuadros 5, 6 y 7).

RECUADRO 5

Enfoques básicos de los reportes (1)

Reporte sobre ítems individuales

Cuando se quiere comprender en detalle qué tipos de cosas son capaces de hacer los estudiantes, se puede reportar:

1. Registros del desempeño real de los individuos, tales como su éxito o fracaso en ítems o tareas específicos de las pruebas, calificaciones de los productos de sus tareas de desempeño, o los mismos productos de la prueba, tales como un ensayo escrito.
2. Ejemplos de desempeños individuales para ilustrar niveles de respuesta típicos o extremos dentro del conjunto de respuestas de los estudiantes, tales como ejemplos de desempeño *novato*, *competente* y *experto*.
3. Estadísticas sobre logros de grupos o poblaciones de estudiantes, tales como estadísticas de ítems (porcentaje de respuestas correctas) o distribuciones de desempeño (porcentaje de individuos con desempeño satisfactorio) o parámetros de los rasgos latentes derivados de los modelos de IRT.

Reporte áreas de contenidos más amplias

Para generalizar los hallazgos en aspectos o maneras diversas de considerar el aprendizaje, así como para resumir los logros por áreas de contenido más amplias, es necesario realizar agregaciones de la información que se reporta. Algunas alternativas son:

1. Dar a conocer las estadísticas de ítems, incluidas las distribuciones de las respuestas correctas y de las incorrectas. En el caso de ítems complejos, cuyas respuestas han sido calificadas por jueces, se puede reportar las distribuciones de sus puntajes. La validez de contenido depende de la interpretabilidad del ítem como representativo de un aspecto importante del aprendizaje de la asignatura, o del conjunto de ítems como representativo de una muestra amplia de una subárea, así como de la interpretabilidad de la calificación misma.
2. Informar sobre promedios entre ítems dentro de pequeñas áreas de contenidos (tópicos), tales como la multiplicación de números enteros o la extracción de información objetiva a partir de la lectura de un texto. La precisión y la interpretabilidad dependen de la calidad y el tamaño de la muestra de ítems.
3. Llegar a medidas compuestas de todo un dominio de contenido o incluso abarcando varios dominios.

RECUADRO 6**Enfoques básicos de los reportes (2)****Alternativas de métricas (escalas numéricas) para los reportes de resultados de las pruebas**

1. Reportar porcentajes de estudiantes y de ítems. El puntaje de un estudiante sería su porcentaje de respuestas correctas. La dificultad de un ítem sería el porcentaje de estudiantes que lo respondió correctamente.
2. Con un sistema de análisis a partir de la Teoría de Respuesta al ítem (IRT), asociar las dificultades de los ítems, así como las habilidades de los estudiantes, con valores de la escala IRT. Dicha escala tiene en su primera aplicación un rango arbitrario, que se puede fijar, por ejemplo, con una media de 500 y una desviación estándar de 150 (o de 50 y 15, o de 100 y 16). Pero en aplicaciones posteriores de acuerdo con el modelo IRT, la métrica puede mantenerse en el siguiente sentido: a través de un proceso de calibración, puede cambiarse la muestra de ítems sin cambiar los valores de la escala para los estudiantes, o puede cambiarse la muestra de estudiantes sin cambiar los valores de la escala para los ítems.
3. Reportar datos normativos, tales como el percentil en que se ubica el puntaje de un individuo con respecto a la distribución de puntajes de un universo de referencia de estudiantes, o el porcentaje de estudiantes de un grupo particular que está en o por encima de un porcentaje adoptado como criterio de referencia (e.g. 50%) para el universo.
4. Establecer estándares y clasificar a los individuos en categorías tales como *novicio*, *competente*, *proficiente* o *experto*, etc., de acuerdo al nivel de competencia que demuestran en las pruebas. así como informar qué proporción de un grupo (o de diversos grupos) de individuos alcanzan cada uno de esos niveles o categorías.

Por supuesto, cualquier resultado puede ser diferenciado según variables estratificadoras de los estudiantes (sexo, edad) o de las escuelas (públicas, privadas, región, etc.).

RECUADRO 7**Enfoques básicos de los reportes (3)****Comparaciones**

Además de lo señalado en los recuadros 5 y 6, hay otras comparaciones importantes que pueden y suelen aparecer en los reportes de las pruebas nacionales o que son inferidas por quienes los leen:

1. Comparaciones a lo largo de las áreas de contenido. Por ejemplo, si el desempeño en matemáticas es más alto que el desempeño en lenguaje.
2. Comparaciones a través de los grados. Por ejemplo, ¿es el desempeño en matemáticas de los alumnos de sexto grado relativamente más alto que el de los alumnos de tercer grado?
3. Comparaciones a lo largo del tiempo. El desempeño medido este año en un tópico y en un grado particulares, ¿es más alto que el desempeño medido el año anterior para ese mismo tópico y grado?

Muchos investigadores y psicometristas dirán que el primer tipo de comparación, entre contenidos, es casi imposible de hacer, y también es difícil en el caso de comparaciones entre grados escolares. Sin embargo, la comparación de un mismo contenido a lo largo del tiempo sí es susceptible de análisis riguroso y solución técnica. Si bien a menudo se realiza incorrectamente, con datos erróneos o análisis equivocados, cuando puede hacerse bien, suministra exactamente el tipo de información que el sistema educativo y el público necesitan para la evaluación y la corrección del progreso educacional.

Generalizabilidad de los resultados reportados

La teoría de la generalizabilidad (Cronbach, Gleser, Nanda y Rajaratnam, 1972) constituye un modelo integral para analizar los puntajes alcanzados por los estudiantes en las pruebas y los reportes de los mismos. Es una expansión y extensión del modelo tradicional de confiabilidad, erigida sobre la noción de que los ítems de las pruebas y las condiciones de aplicación y calificación de las mismas constituyen muestras extraídas de universos más amplios de ítems y condiciones de aplicación y calificación. Así, por ejemplo, en una prueba de capacidad de expresión escrita, se puede pedir a los estudiantes que respondan por escrito a cada una de varias preguntas o reactivos. Estos buscarían provocar la redacción de textos narrativos, informativos o de otra naturaleza. Esto significa que serían muestras de textos de una muestra de tipos de redacción. Más aún, la calificación de los textos redactados estaría a cargo de una muestra de docentes. El objetivo es generalizar el desempeño de los estudiantes a los universos correspondientes de preguntas o reactivos para la redacción, tipos de textos y docentes-calificadores.

En la teoría de la generalizabilidad, la confiabilidad de los puntajes de las pruebas es considerada desde la perspectiva de la generalización a partir de muestras a las poblaciones relevantes. El análisis de generalizabilidad implica la determinación de la variabilidad de la muestra en diferentes componentes medidos por el desempeño en las pruebas. En particular, el objetivo es determinar qué partes de la variación de los puntajes de las pruebas se deben a características estables de los individuos y grupos que son independientes de los ítems particulares y de las condiciones de aplicación, y qué parte de la variación obedece a características de los ítems y de las circunstancias de la aplicación de las pruebas. Esto último es considerado como “error de medición”.

La aplicación de la teoría de la generalizabilidad al diseño de evaluaciones educacionales requiere la definición cuidadosa y formal de las poblaciones o universos de ítems y de las condiciones de medición. Las tablas de especificaciones convencionales de las pruebas son un buen punto de partida, pero es necesario añadirles una definición formal de las poblaciones o universos de ítems que corresponden a cada una de las celdas de la tabla, y de una especificación igualmente explícita acerca de las condiciones de medición, incluyendo los tipos de ítems, las maneras de administrar la prueba, los procedimientos de calificación o las condiciones de replicación (si es que la prueba en cuestión ha sido ya administrada en anterior oportunidad).

Las rúbricas de calificación y las replicaciones se tornan especialmente complejas en los casos de tareas de desempeño.

El método estadístico apropiado para el análisis de la generalizabilidad es el Análisis de Varianza. Con un conjunto detallado de información sobre las categorías de los ítems y de las condiciones de medición (lo que a veces se denomina un estudio “G”), se obtiene información rigurosa acerca de la magnitud de los diferentes componentes de la varianza, lo que permite calcular la precisión de las mediciones obtenidas y, especialmente, de las comparaciones que es posible hacer a lo largo del tiempo.

Comparabilidad de los resultados reportados

Por lo general, no es factible usar una misma prueba en dos momentos diferentes en el tiempo. Pero hay métodos que permiten poner dos pruebas paralelas o superpuestas en la misma escala y luego tratar sus mediciones como si fueran resultantes de la misma prueba.

Cuando las pruebas son paralelas, en el sentido de que tienen el mismo contenido y estructura, ítem por ítem, el alineamiento estadístico es considerado una “equiparación” (*equating*). Que el contenido sea el mismo se logra mejor asignando de manera aleatoria pares de ítems a dos formas de pruebas.

Cuando las formas de las pruebas no son estrictamente paralelas, sino que tienen una superposición sustancial —es decir, ítems en común— se aplican diferentes procedimientos estadísticos de regresión y el resultado se denomina “calibración”.

Estos procedimientos para obtener comparabilidad entre las mediciones a lo largo del tiempo requieren obviamente una planificación de largo plazo para el diseño y la administración de las pruebas. No se pueden desarrollar pruebas, de novo, cada año.

La tecnología para la comparabilidad mediante la equiparación y la calibración es estadística. Por lo tanto, está sujeta a error estadístico. Es importante calcular y dimensionar el error de equiparación o calibración y no realizar comparaciones que excedan la precisión alcanzada.

La elaboración de escalas de reporte

La Teoría de Respuesta al Ítem (IRT) está siendo propuesta como una suerte de panacea para los problemas de construcción y diseño de pruebas.

Se basa en un modelo sofisticado que sugiere cómo se encuentran situados teóricamente los estudiantes y los ítems de las pruebas en una escala numérica común y cómo las respuestas de los estudiantes a los ítems están estadísticamente determinadas por sus posiciones relativas en dicha escala. De ese modelo se ha derivado toda una tecnología que parece resolver varios problemas difíciles en la equiparación de formas de prueba, en la calibración de los resultados a lo largo del tiempo y en la construcción de mediciones paralelas por vías más “fuertes” que los métodos estadísticos clásicos. Un importante ejemplo de esa fortaleza es que es posible establecer una relación entre los estudiantes en un nivel de puntaje particular y los ítems que ellos dominan. Esto se realiza de una manera referida a criterios, sin tener que tomar en cuenta poblaciones normativas de estudiantes o de ítems.

El hecho de que esta referencia funcione o que los problemas de comparabilidad sean realmente resueltos, depende de la adecuación de los modelos con respecto a un sistema de pruebas específico, es decir, con respecto a poblaciones particulares de ítems y de estudiantes. Un aspecto cuestionable del modelo, por lo menos en muchas aplicaciones educacionales, es que presupone que fundamentalmente existe una sola dimensión en la variación en la dificultad de los ítems y en la capacidad de los estudiantes.

El uso de los métodos IRT en el análisis estadístico y la interpretación de los resultados de las evaluaciones educacionales requiere programaciones y análisis estadísticos sofisticados, aunque esto es también cierto para los métodos convencionales. Es muy importante recordar que los sistemas de análisis, sean el clásico o el IRT, son solo instrumentos estadísticos para resolver algunos problemas técnicos de análisis de ítems, de equiparación y calibración de pruebas y de evaluación de la precisión estadística de los resultados. Los problemas fundamentales tanto conceptuales como pedagógicos son, por un lado, la definición de los universos de ítems y, por otro, la construcción y selección de una muestra de esos universos.

El dilema respecto a las escalas de reporte

El requerimiento clave para el desarrollo de escalas de reporte apropiadas para las evaluaciones educacionales es que la interpretación que se haga a partir de los resultados sea auténtica. Esto significa:

- que a las cifras reportadas se les atribuirá significados que estén justificados por las características de las poblaciones de contenidos y de estudiantes, y por el sistema empleado para la medición;
- que la precisión de los reportes será tomada en cuenta correctamente; y
- que las comparaciones al interior de y entre partes de la evaluación serán hechas solo en la medida en que estén justificadas tanto desde el punto de vista sustantivo como del estadístico.

El dilema en el diseño de escalas de reporte en educación es que las cuestiones técnicas son complejas, mientras que los conocimientos e intuiciones sobre las mismas que tienen los usuarios están muy arraigados, aunque errados (por ejemplo, piensan que un puntaje de 50% siempre significa un fracaso o un logro mínimo). Esto hace muy difícil encontrar los medios apropiados para la comunicación de los resultados.

CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

Pedro Ravela

El propósito de este capítulo es resumir las principales reflexiones formuladas a lo largo del documento y proponer algunas líneas de acción que los organismos regionales interesados en apoyar el desarrollo de los sistemas nacionales de evaluación de aprendizajes y PREAL, a través de su Grupo de Trabajo sobre Estándares y Evaluación, deberían impulsar en los próximos años.

Durante la década de los '90 se desarrolló en América Latina una primera fase de instalación de sistemas de evaluación de aprendizajes a nivel nacional. Tuvo lugar, además, una primera experiencia de evaluación a nivel regional conducida por UNESCO/ OREALC. El desarrollo de estas experiencias constituye una clara manifestación de la preocupación de los gobiernos por producir información sobre los aprendizajes que se logran al interior de los sistemas educativos. En un contexto internacional en que el conocimiento y las capacidades de los individuos serán cada vez más importantes para el desarrollo y competitividad de los países, es previsible que en los próximos años este esfuerzo se mantenga. Asimismo, en la medida en que crece la conciencia respecto a la necesidad de incrementar los recursos destinados al sector educación, será imprescindible contar con información que permita evaluar el impacto de la inversión adicional de recursos y monitorear en forma permanente y adecuada los avances o retrocesos en los resultados del sistema educativo.

El panorama resultante de esta primera década muestra una importante diversidad de enfoques y experiencias en cuanto al tipo de conocimientos y competencias que son evaluados, a la periodicidad de las evaluaciones, a los grados y áreas curriculares que abordan, al tipo de variables contextuales sobre las que se recoge información, y a los análisis y formatos de devolución de información, entre otros. Detrás de esta heterogeneidad de experiencias y enfoques existe, en muchos casos, un esfuerzo de reflexión y construcción de un modelo a nivel nacional. En otros, se han adoptado modelos en forma menos reflexiva. En todos los casos, se aprecia la fragilidad propia de los primeros pasos dados en un terreno nuevo y desconocido.

La mayoría de estos sistemas de evaluación se encuentra aún en una fase de institucionalización, ya que su continuidad en el tiempo está fuertemente atada a los cambios políticos y/o a los recursos aportados por organismos internacionales de crédito. Por otra parte, cabe destacar que los sistemas de evaluación compiten por recursos con otras actividades y necesidades igualmente importantes. Por lo tanto, su sostenibilidad depende de que se aprovechen y maximicen los beneficios que en cierto modo “prometen” a la política educativa.

En este contexto, el presente documento busca aportar a la reflexión sobre los próximos pasos a dar para fortalecer los sistemas de evaluación de aprendizajes en América Latina, con la convicción de que los mismos pueden constituirse en una herramienta de política educativa eficaz para promover un mejoramiento de los aprendizajes a los que acceden los niños de todos los estratos sociales. Y también con la convicción de que, para que esto último ocurra, es necesario ingresar en una nueva etapa de revisión y consolidación de los sistemas nacionales de evaluación, para lo cual se torna imprescindible abordar en profundidad un conjunto de debates sobre opciones técnicas y políticas en esta materia.

Se han priorizado cuatro grandes temas centrales

En primer término, se analiza un conjunto de alternativas técnicas relacionadas con el diseño global del sistema de evaluación y con los objetivos que se espera que el mismo cumpla. Por un lado, se plantea la existencia de una relación inversa entre cobertura curricular y cobertura poblacional. Cuanto más detalladamente se desee conocer qué aprenden los alumnos en cierto nivel del sistema educativo, menos factible será contar con información desagregada a nivel de distritos y establecimientos educativos. Por el contrario, si el propósito es generar información con estos últimos niveles de desagregación, solo será posible obtener mediciones más globales y menos detalladas de lo que los alumnos aprenden. En última instancia, la opción depende del rol que se espera que el sistema de evaluación desempeñe en la política educativa. En el quinto capítulo se analiza la existencia de múltiples alternativas técnicas para reportar los resultados, que también dependen del enfoque dado al sistema de evaluación, y que debieran ser discutidas en profundidad para mejorar la calidad y pertinencia de la información que están reportando los sistemas de evaluación en América Latina.

En segundo lugar, se alerta sobre la necesidad de analizar más cuidadosamente la información que producen los sistemas de evaluación, en el sentido de validar las conclusiones e interpretaciones que de dicha información se realiza. Ello implica, por un lado, una labor de “formación permanente” de los usuarios –autoridades, opinión pública, medios de comunicación, maestros– respecto a los usos válidos de los distintos tipos de información que aportan estas evaluaciones y respecto al tipo de interpretaciones y conclusiones que NO es posible extraer válidamente de ella. Para esto, el primer paso es que las propias unidades de evaluación mejoren el modo en que reportan los resultados e incluyan reportes técnicos completos y comprensibles acerca de las limitaciones y potencialidades de la información que producen, y acerca de los procedimientos de producción de la información.

En tercer lugar, se plantea la necesidad de profundizar la discusión respecto al enfoque de diseño de pruebas más adecuado. En la mayoría de los países de la Región, el diseño de pruebas ha estado fuertemente marcado por los principios y procedimientos propios de la elaboración de pruebas referidas a normas, en las que se privilegia la función de ordenamiento o “discriminación” entre grupos o individuos. Este enfoque está fuertemente marcado por su función principal, que históricamente ha sido la de seleccionar individuos para el ingreso al ejército o a las universidades. En esos casos no importaba tanto si el individuo dominaba o no ciertos campos del conocimiento, sino distinguir a los individuos más aptos de los menos aptos. El enfoque de pruebas referidas a criterios, en cambio, se propone como objetivo central comprobar si los individuos dominan un cierto campo de contenidos y/o destrezas, y se busca hacerlo del modo más exhaustivo posible. Ello implica que, en el diseño de las pruebas, no necesariamente deben ser descartados aquellos ítems que resultan fáciles o difíciles, dado que, aun cuando no sirvan para discriminar entre malos y buenos estudiantes, pueden aportar información relevante acerca del grado en que los conocimientos y competencias definidos como fundamentales están siendo logrados por los estudiantes en el sistema educativo. Este enfoque parece ser el más adecuado para sistemas de evaluación de aprendizajes que buscan producir información relevante para la mejora del currículo y la enseñanza. Para ello es imprescindible enriquecer y mejorar los procedimientos de diseño de pruebas e ítems. Asimismo, se requiere avanzar en el diseño y corrección estandarizada de pruebas de desempeño.

En cuarto lugar, se señala que es necesario mejorar sustancialmente no solo la calidad de los instrumentos de medición de aprendizajes, sino también los instrumentos de medición de aspectos relevantes del contexto social y escolar en que ocurren los aprendizajes. La mayoría de los países recoge información sobre variables sociales y escolares, pero en muchos casos la calidad de la misma no es suficiente y, menos aún, el aprovechamiento que de ella se hace para el análisis de los resultados de aprendizaje y la investigación. Es necesario, por un lado, mejorar la medición de variables de tipo sociofamiliar, con el fin de contextualizar socialmente el análisis y reporte de los resultados y evitar la falacia de atribuir a los establecimientos educativos el mérito o la culpa por resultados que en realidad obedecen a la selección social del alumnado. Por otro lado, es preciso mejorar la medición de variables institucionales y pedagógicas, para desarrollar investigaciones que permitan comprender mejor la compleja trama de factores que intervienen en el logro de los aprendizajes y, de este modo, enriquecer el horizonte conceptual y la base empírica de la toma de decisiones en materia de política educativa.

Las reflexiones y análisis anteriores permiten afirmar que, si bien al cabo de esta primera década de instalación se han dado pasos muy importantes, lo que se está haciendo no es suficiente. Es necesario inventar cosas nuevas. En este sentido, una posible agenda de trabajo para los pasos a dar en los próximos años, estaría centrada en tres ejes principales:

1. *Analizar el papel de los sistemas de evaluación en la política educativa*, es decir, la estrategia a través de la cual se espera que un sistema de evaluación nacional de aprendizajes tenga algún impacto en la mejora de los aprendizajes que se logran en el sistema educativo.

Se requiere propiciar instancias de discusión sobre el rol que se espera del sistema de evaluación en el marco de la política educativa. Si bien existe un consenso genérico en cuanto a que el simple acto de evaluar –por el mero hecho de dar “visibilidad” a los resultados– puede tener un efecto positivo sobre el sistema educativo, es necesario delinear a nivel nacional una estrategia más específica al respecto. En algunos casos se ha buscado que el impacto se produzca a través del control de los padres sobre la calidad de las escuelas; en otros, por la vía de utilizar las pruebas para certificar la aprobación de cierto nivel de enseñanza; mientras que en otros se ha utilizado la información principalmente para promover la actualización docente y el aprendizaje profesional al interior de las escuelas.

En este documento se ha planteado una amplia gama de alternativas de política en materia de evaluación, destacando que las diversas opciones técnicas dependen de decisiones de política relativas a la finalidad de las evaluaciones. En la mayoría de los países de la Región ha existido la premisa básica general de que “evaluar ayuda a mejorar”, pero ha faltado una reflexión en profundidad sobre las alternativas de política y estrategia en materia de evaluación de aprendizajes y sobre la necesidad de articular diferentes finalidades en un diseño coherente, técnicamente adecuado y pensado para el largo plazo. Esta carencia está relacionada, asimismo, con las urgencias que impone la puesta en marcha de un operativo nacional de evaluación, la necesidad de cumplir con plazos y compromisos asumidos con organismos internacionales.

La reflexión que se propone debería involucrar preguntas tales como:

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

En relación con este eje, en el futuro inmediato PREAL y otras entidades regionales deberían facilitar la realización de eventos de debate y presentación de experiencias nacionales en materia de diseño del sistema nacional de evaluación y su papel en la política educativa. Se podría desarrollar un “observatorio” internacional sobre el tema, propi-

ciendo la construcción de estudios de casos de países de la Región y del mundo desarrollado en los que se describa: diversas modalidades de diseño de los sistemas nacionales de evaluación; las condicionantes históricas, sociales y políticas que constituyeron el contexto de su desarrollo; las características técnicas del diseño; y los impactos, costos, beneficios y efectos perversos constatables en cada caso. Esta información podría luego difundirse bajo la forma de una serie de publicaciones o en el marco de seminarios de discusión. Un desafío importante en este ámbito será el de involucrar en estos debates a otros actores de la política educativa y no solo a los técnicos directamente relacionados con las evaluaciones.

2. Mejorar la calidad técnica de los diversos aspectos constitutivos de los sistemas de evaluación, en especial el diseño de los instrumentos de recolección de información y los modos de procesar y reportar los resultados.

A lo largo del documento se insistió respecto a que las decisiones técnicas que son aptas para ciertos fines no lo son para otros. Por tanto, es necesario garantizar la congruencia entre las opciones de política y las decisiones técnicas que definen el diseño del sistema de evaluación. Asimismo, se señaló que es imprescindible mejorar el diseño de los instrumentos de medición y garantizar una interpretación apropiada –válida– de los resultados que se obtienen. Todo ello exige intensificar los esfuerzos de capacitación de cuadros técnicos y la acumulación de conocimiento y experiencia en una materia que aún es nueva en la Región y sobre la que existe escasa “masa crítica”.

En este terreno PREAL y otras organizaciones de carácter regional podrían facilitar el contacto de los profesionales de la Región vinculados al área de la evaluación, con especialistas de la comunidad internacional. Esto se podría hacer a través de seminarios en los que participarían países que estén dispuestos a someter sus instrumentos y procedimientos de evaluación al escrutinio de otros, para analizar detalladamente las fortalezas y debilidades de los instrumentos y procedimientos empleados. Ello podría servir también para desarrollar procedimientos comunes a varios países para medir ciertos aspectos relevantes, tanto en el terreno de los aprendizajes como en el de las variables sociales y escolares.

También sería de gran utilidad impulsar la formulación de un conjunto de estándares técnicos que deberían cumplir las pruebas, los procedimientos de implementación de los operativos de evaluación, los procesos

de conformación y procesamiento de las bases de datos, y los reportes de resultados.

Finalmente, cabe destacar que toda iniciativa dirigida a propiciar la formación sistemática de cuadros en materia de evaluación, tanto a nivel de agentes estatales como de organizaciones no gubernamentales, será bienvenida.

3. *Discutir las estrategias de uso y difusión de los resultados de las evaluaciones*, que si bien está estrechamente relacionado con los dos puntos anteriores, merece una consideración específica.

Todavía es muy poco lo que se sabe acerca del uso e impacto que la información producida por los sistemas de evaluación de aprendizaje tiene en sus potenciales usuarios. Aún no se ha construido evidencia empírica acerca de:

- el modo en que los resultados son analizados y utilizados en las escuelas;
- el grado en que las familias y la opinión pública reciben y comprenden la información;
- el modo en que la misma es empleada como insumo en la toma de decisiones de política educativa por parte de los Ministerios de Educación; y
- el grado en que las bases de datos son aprovechadas por académicos y centros de investigación para producir conocimiento.

En este terreno, PREAL y otros organismos regionales deberían propiciar el desarrollo de trabajos de investigación que permitan recoger evidencia empírica acerca de los efectos que los distintos tipos de reportes de resultados de las evaluaciones nacionales tienen en diversos públicos.

En segundo término, sería sumamente útil realizar algún tipo de evento que permita “escuchar a los destinatarios”:

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

Ello permitiría analizar la demanda potencial de información, aprender acerca de los modos pertinentes de informar a los diversos tipos de usuarios potenciales y desarrollar diferentes tipos de formatos de informe adecuados a cada uno de ellos. Asimismo, dado que la mayoría de los potenciales destinatarios probablemente no tenga una noción cabal de lo que espera de un sistema de evaluación de aprendizajes, un evento como el descrito aportaría pistas para desarrollar estrategias de “formación de la demanda”. Es decir, ayudarla a pensar qué tipo de acciones desarrollar para ir conformando en nuestros países una cultura con relación a la evaluación de aprendizajes que incluya aspectos tales como la conciencia acerca de la necesidad de contar con esta información, los alcances y limitaciones de la misma, el tipo de interpretaciones válidas, los modos de utilizarla y los tipos de información que es posible demandar al sistema de evaluación para fines específicos.

Aspectos del currículo prescrito en América Latina: Revisión de tendencias contemporáneas en currículo, indicadores de logro, estándares y otros instrumentos

J. GUILLERMO FERRER^(*)

GILBERT A. VALVERDE^(**)

JUAN MANUEL ESQUIVEL ALFARO^(***)

INTRODUCCIÓN

En las aulas de América Latina tiene lugar una variedad de actividades mediante las cuales se crean oportunidades para que los estudiantes obtengan nuevos conocimientos, desarrollen nuevas capacidades y adquieran actitudes y disposiciones acerca de las diferentes áreas disciplinarias. Docentes y estudiantes encarnan sus papeles en el aula, respectivamente, como proveedores y beneficiarios de esas oportunidades. Sin embargo, no es en el aula solamente donde se decide qué oportunidades curriculares han de brindarse. La historia del currículo implementado en el aula comienza fuera de los escenarios de aprendizaje, comienza cuando aquellos encargados de moldear y configurar el currículo escogen metas, objetivos y visiones específicas para guiar lo que ocurre en el sistema educativo. A esta serie de decisiones acerca de los objetivos y metas del sistema educativo llamamos *currículo prescrito*¹ (Valverde, 1997). Estas

^(*) Este documento fue originalmente publicado en octubre de 1999 en el marco de las actividades desarrolladas por el grupo de trabajo sobre estándares y evaluación.

¹ Entendemos por currículo prescrito el conjunto de contenidos que un sistema educativo aspira a transmitir y que generalmente se declara en documentos oficiales de circulación pública.

decisiones se toman en una variedad de niveles del sistema educativo de acuerdo a la naturaleza de la decisión y al país o subsistema educativo del que se trate. Algunas se toman en ciertos países a nivel del Estado nacional, otras se toman a nivel departamental o provincial, y en muchas de nuestras naciones se toman decisiones importantes acerca de objetivos en el propio recinto escolar. En todas las naciones de América Latina, sin embargo, los ministerios o secretarías nacionales, departamentales y/o provinciales toman decisiones de alto impacto sobre los demás componentes del sistema. A veces, esta toma de decisiones es una responsabilidad política que cuenta con un marco constitucional o legal que determina que los otros componentes del sistema tienen, a su vez, responsabilidades definidas en cuanto al cumplimiento de las metas propuestas. Otras veces las metas tienen un valor persuasivo pero carecen de un marco jurídico o constitucional que las avale; su intención es la de guiar a los actores, proporcionando un paradigma educativo que actúa como modelo para la elaboración de los objetivos de los demás niveles del sistema.

En su informe “El futuro está en juego”, la Comisión Internacional sobre Educación, Equidad y Competitividad Económica (PREAL, 1998) establece cuatro recomendaciones fundamentales para el logro de una transformación real en la calidad de la educación en América Latina. La primera recomendación de la Comisión declara la necesidad de “Establecer estándares para el sistema de educación y medir el avance en su cumplimiento”. Específicamente, PREAL sugiere que:

Los gobiernos deben establecer estándares educacionales claros, introducir pruebas a nivel nacional y utilizar los resultados para corregir los programas y reasignar los recursos. Las naciones de América Latina y el Caribe deben hacer que sus estudiantes participen en pruebas a nivel internacional para poder comparar la calidad de sus escuelas con la de otros países.

En un primer intento por describir el estado actual de la elaboración de estándares educativos en la región, el Grupo de Trabajo sobre Estándares y Evaluación del PREAL procura en este informe examinar algunos de los instrumentos centrales del currículo intencional: los currículos o programas nacionales, provinciales o departamentales, los estándares educativos y, en particular, los indicadores que se proponen en estos como señales del cumplimiento –o del progreso hacia el cumplimiento– de objetivos curriculares específicos. En este informe hemos utilizado la noción de estándar educativo de manera tal que las apreciaciones sobre

el estado del diseño curricular actual en la región pudieran efectuarse en relación a un parámetro diferente al que se ha utilizado hasta el presente en el análisis del currículo en América Latina. Tomamos como parámetro de calidad en la elaboración de contenidos curriculares una noción de estándar que definimos como una *delimitación* específica de *los conocimientos conceptuales y procedimentales a ser adquiridos por la totalidad del alumnado, en términos estrictamente académicos, política y socialmente validados, que constituyen un lenguaje común para todas las instituciones y actores, y que están sujetos a una evaluación sistemática.* Junto con los instrumentos mencionados, aspiramos a enriquecer el debate con la participación de diferentes actores del sistema que actualmente planifican o ejecutan las transformaciones curriculares en los países de Latinoamérica. Nos ocupa principalmente la tarea de brindar al público una primera aproximación al estado de la discusión y elaboración de estándares educativos en la región, y dejar así iniciado un proceso de intercambio conceptual entre los diferentes agentes que hoy tienen la posibilidad, y la responsabilidad, de efectuar transformaciones reales y sostenibles a largo plazo en los sistemas educativos de nuestra región.

METODOLOGÍA Y ORGANIZACIÓN

Para la elaboración de este informe se utilizaron dos tipos de fuentes: *documentos curriculares y entrevistas* a funcionarios y personal técnico de las oficinas de currículo y evaluación de los ministerios de educación de la región. El proyecto se llevó a cabo en Costa Rica, Perú y los Estados Unidos de América, con una duración total de dos meses hasta la fecha de elaboración del informe.

La Secretaría General de la Coordinación Educativa y Cultural Centroamericana (SG/CECC), con sede en San José de Costa Rica, y el Grupo de Análisis para el Desarrollo (GRADE), en la ciudad de Lima, tuvieron a su cargo la tarea de reunir la mayor cantidad posible de documentos curriculares y normativos vigentes de la región. Del total de documentos recibidos en estos dos centros, se analizaron aproximadamente 40 (cuarenta), entre programas de estudio, información sobre implementación curricular de circulación oficial, leyes y decretos nacionales de educación. El análisis de los documentos se realizó sobre la base de un extenso cuestionario que permitió registrar la información disponible sobre los siguientes aspectos de los currículos relevados: fundamen-

tos filosóficos, epistemológicos, políticos y económicos; metas educacionales; organización de los contenidos disciplinares e indicadores de logro; factores críticos de implementación; evaluación; iniciativa y antecedentes políticos de las reformas². Toda la información se refiere a la educación del primer nivel³ en cada país, en las áreas curriculares de matemática, lengua y ciencias naturales.

Los centros de investigación SGCECC y GRADE solicitaron entrevistas personales y telefónicas a representantes de las oficinas ministeriales de currículo y evaluación en 17 (diecisiete) países. Del total de solicitudes, se concretaron 14 (catorce) entrevistas⁴ de aproximadamente treinta minutos de duración cada una. Las entrevistas se basaron en un protocolo estructurado de 13 (trece) preguntas sobre diseño e implementación curricular, sobre aspectos políticos y técnicos de los procesos de reforma educativa en cada sistema educativo y sobre el estado actual del debate y elaboración de estándares educativos.

La redacción de este informe está basado en la información obtenida en los dos tipos de fuentes mencionados, y se organizó siguiendo la estructura y tópicos de los cuestionarios y protocolos utilizados durante el trabajo de campo. En ocasiones, se citan fuentes bibliográficas que inspiran algunas de las conclusiones teóricas e inferencias basadas en la información recogida. El informe está organizado en cinco secciones. La primera sección hace referencia a las fuentes de información sobre necesidades de transformación utilizadas por los diferentes países de la región durante el proceso previo a la reforma educativa. También se discuten algunos de los fundamentos filosóficos, políticos, económicos y epistemológicos que se plantean como marco teórico de la transformación curricular en los diferentes documentos analizados. La segunda sección consiste en un breve análisis descriptivo de algunos currículos⁵

² Una lista de los documentos revisados y copias de los protocolos utilizados se encuentran en el apéndice de este informe.

³ En este informe utilizamos la Clasificación Internacional Normalizada de la Educación (CINE, 1997), versión revisada y aprobada en la Conferencia General de la UNESCO - París, 1998. El nivel 1 de la CINE se conoce bajo distintos nombres en América Latina, aunque comúnmente se llama "enseñanza primaria", "ciclo básico" o "enseñanza general básica".

⁴ Estas entrevistas se condujeron bajo un acuerdo de anonimato con cada entrevistado. Los entrevistados fueron en todos los casos personal técnico de ministerios o secretarías de educación. En el apéndice de este informe, se detalla en cuáles países se condujeron entrevistas.

oficiales; esta sección pretende ejemplificar el estado actual del diseño curricular en la región, específicamente con relación a la definición de indicadores de logro académico. La tercera sección plantea los principales factores que afectan la implementación curricular en América Latina. La sección cuarta resume las referencias más importantes que se hacen a la evaluación y medición del desempeño y el rendimiento académico en los documentos curriculares vigentes. En la quinta y última sección, nos referiremos al estado actual del debate sobre la elaboración de estándares educativos en Latinoamérica, según se desprende de las entrevistas realizadas y de los documentos relevados para este informe.

1. JUSTIFICACIÓN Y FUNDAMENTOS DE LAS REFORMAS CURRICULARES

Los documentos curriculares relevados, así como las entrevistas realizadas para este estudio, dan cuenta de que las reformas educativas y curriculares de la última década en América Latina se inspiran en informaciones tanto formales como informales sobre las necesidades de la región. En sus documentos oficiales, trátase de leyes, reglamentos, o programas de estudio, son pocos los países que hacen referencia a estudios específicos sobre la situación de sus sistemas educativos o de las necesidades sociales que la educación propone atender. En general, las necesidades se expresan como grandes problemáticas de orden social, político y económico, y se le atribuye a la educación una función crítica en la reversión de las profundas crisis que afectan a muchos de los países de la región.

Entre las fuentes formales utilizadas para informar los cambios impulsados figuran: censos sobre población estudiantil (matriculación, repitencia y deserción); consultas a instituciones nacionales e internacionales (universidades, ONGs, Banco Mundial, UNESCO, etc.); resultados de pruebas nacionales sobre rendimiento; encuestas a la sociedad civil.

⁵ Distinguimos entre currículos, documentos curriculares y documentos normativos. Currículos son los programas de estudio que especifican los contenidos y objetivos de las diferentes áreas curriculares o disciplinas. Los documentos curriculares pueden ser currículos pero incluyen también a los documentos que proponen una justificación para la reforma, una fundamentación teórica, o indicaciones específicas para directivos, docentes y el público en general. Los documentos normativos son aquellos que tienen fuerza legal, como leyes nacionales y provinciales, decretos y resoluciones.

Por fuentes informales entendemos aquellas que representan las voces de diferentes sectores de la sociedad, que han contribuido con su opinión a lo largo de las últimas dos décadas, pero que no constituyen evidencia documentada.

Los documentos normativos y curriculares relevados indican que las mayores dificultades que atraviesan los sistemas educativos en América Latina están relacionados con:

- Altas tasas de repitencia, especialmente en los primeros grados y en áreas rurales, y atribuidas al fracaso escolar en su función básica de alfabetización
- Altas tasas de deserción
- Ineficiencia escolar
- Escasez de recursos pedagógicos y didácticos
- Centralización

Una lectura diagonal de las justificaciones a los diferentes currículos de la región revela que si bien existe cierto consenso sobre la importancia de mejorar la calidad de la educación en términos de los conocimientos a ser alcanzados al cabo de la experiencia escolar, el acento recae primordialmente sobre los procesos de socialización de los alumnos, en lo que podría caracterizarse como una corriente reconstruccionista en la definición de las nuevas metas de la educación. Según las justificaciones a la mayor parte de las transformaciones curriculares de la región, la educación escolar tiene la responsabilidad de contribuir a contrarrestar, revertir y prevenir los grandes problemas de la sociedad moderna, tales como la pobreza, la violencia, la destrucción del medio ambiente y la enfermedad, entre otros. Para la elaboración de este informe, sin embargo, hemos dirigido los esfuerzos hacia la identificación y análisis de las secciones del currículo que describen los indicadores de desempeño y rendimiento académico, y que representan los aspectos más formales de la educación escolar. Es en estos aspectos donde la escuela puede ser considerada un agente insustituible, y en consecuencia donde mayor atención debe centrarse en la actualidad para lograr una auténtica mejora de la calidad del servicio educativo público.

Si bien la educación siempre se ha considerado como un factor crítico para el desarrollo de las naciones, los documentos oficiales actuales reconocen el rol decisivo del *currículo* en este proceso de transformación. México, por ejemplo, en una breve reseña histórica a modo de introduc-

ción de los nuevos programas de estudio, señala que “en décadas anteriores los esfuerzos del sector estuvieron centrados en garantizar equidad y acceso”, mientras que “hoy las demandas de la sociedad moderna hacen que sea necesario que el Estado y la sociedad en su conjunto realicen un esfuerzo sostenido para elevar la calidad de la educación que reciben los niños”. También Chile expresa la necesidad de mejorar los contenidos de los programas de manera de obtener las “competencias necesarias para facilitar la inserción de los jóvenes en el mundo laboral, productivo y de servicios, lo que facilitará, a su vez, la mejor inserción del país en los mercados mundiales”. Algunos fundamentos para la transformación se enuncian en un marcado tono patriótico, de reafirmación de la identidad nacional, mientras que otros hacen referencia a la necesidad de ajustarse a los procesos de globalización que experimenta el mundo entero.

Este enfoque sobre la transformación curricular como parte esencial de las reformas educativas está avalado por las iniciativas políticas y acuerdos para el mejoramiento educativo en la región, que ponen énfasis en la calidad de los aprendizajes y en la necesidad de marcar los niveles mínimos de rendimiento académico. Los acuerdos de PROMEDLAC IV, V y VI (Gajardo, 1999), entre otros objetivos que apuntan al mejoramiento de la infraestructura educativa y al acceso equitativo a la educación, señalan la necesidad de:

- Promover la transformación curricular a través de propuestas basadas en la satisfacción de las necesidades básicas de individuos y sociedades y modificar las prácticas pedagógicas.
- Dar prioridad a la educación inicial y a la calidad de los aprendizajes en los primeros años de la educación básica.
- Fortalecer los aprendizajes de los códigos básicos de la modernidad: lectoescritura y matemáticas elementales.
- Dar prioridad a los aprendizajes y la formación integral y complementarlos con el desarrollo de una cultura científica y tecnológica fundamental desde la educación básica.
- Desarrollar mediciones comparativas sobre niveles de competencia e indicadores para la evaluación del desempeño de las escuelas y generar modalidades para mejorar el uso de la información.

Antes de analizar la forma en que los diferentes currículos reflejan este nuevo enfoque en la calidad de los contenidos, nos detendremos en algunas características del cambio curricular compartidas por la mayoría

de los países en Latinoamérica. Distinguimos, con ese propósito, cuatro grandes grupos de ideas para la descripción de tales características.

i. Los contextos políticos y económicos del cambio

Según los documentos relevados, los cambios curriculares se inscriben en contextos de grandes transformaciones políticas y del orden económico regional y mundial, y se aspira a que los nuevos contenidos de aprendizaje contribuyan a la consolidación de las democracias, al afianzamiento de la paz y la estabilidad interna, y al logro de un desarrollo económico sostenible basado en la productividad y en la competitividad internacional.

Es importante destacar que muchos de los procesos de reforma educativa y curricular de la región se producen a partir de las transformaciones políticas globales de cada país. En Guatemala, por ejemplo, los movimientos de reforma coinciden con los Acuerdos de Paz de 1996. En El Salvador, la transformación curricular se inscribe dentro de un programa general de desarrollo llamado Plan de Desarrollo Económico y Social 1994-1999. En México, la elaboración de los nuevos planes de estudio comienza en los primeros meses de 1989 como tarea previa a la elaboración del Plan Nacional de Desarrollo 1989-1994. Como resultado de la consulta realizada en esos momentos surge el Programa para la Modernización Educativa 1989-1994 que establece como prioridad la renovación de los contenidos y los métodos de enseñanza, el mejoramiento de la formación de maestros y la articulación de los niveles educativos de la educación básica. Colombia, por su parte, experimenta su reforma educativa dentro de un proceso de transformación global, con la consagración de la nueva Constitución Política de 1991. Estos ejemplos explican algunas de las razones por las que los documentos curriculares de América Latina plantean ambiciosos objetivos de transformación de los individuos y de la sociedad, a veces restándole espacio al tratamiento de los contenidos conceptuales del currículo.

Las entrevistas realizadas confirman que el ímpetu para las reformas educativas y curriculares de la región es de origen eminentemente político. Sin embargo, se realza en todos los casos que las transformaciones impulsadas por el Poder Ejecutivo y sus ministerios se han validado con el consenso de la sociedad y de los diferentes actores del sistema educativo en cada nación. La iniciativa política se concreta con las acciones legislativas que a lo largo de las dos últimas décadas han marcado un

camino de transformación gradual de los sistemas educativos. Los hitos más destacados del proceso son las leyes generales y orgánicas de educación derogadas en la última década, pero que tuvieron lugar merced a una serie de programas especiales y medidas legislativas o ministeriales previas que sentaron precedente para el cambio. Ejemplos de estas medidas son la Consulta Nacional y Acuerdos Institucionales de 1991 en República Dominicana, la formación de la Comisión Nacional de Modernización de la Educación de 1994 en Chile, o la Ley de Transferencia de 1992 en Argentina (Gajardo, 1999; Pinto Cueto, 1999). El consenso social, por su lado, se manifiesta principalmente por medio de las consultas populares realizadas en varios países como paso inicial hacia la reforma, y en la participación de los diferentes grupos invitados para la elaboración de los contenidos curriculares, aunque estos últimos no se hayan constituido en organismos permanentes de consulta.

ii. Necesidad de consenso y validación

Una de las características comunes a los procesos de reforma curricular en la región es que todos los países, en mayor o menor medida, expresan la necesidad de arribar a consensos sobre qué metas y contenidos son deseables para su población y su sistema educativo. Los documentos y entrevistas realizadas por el autor revelan que la elaboración de los programas de estudio vigentes estuvo a cargo de diferentes actores en diferentes países; en algunos casos solo se mencionan los agentes que intervinieron en la elaboración de los contenidos, en otros se da cuenta detallada de los mecanismos utilizados para arribar a acuerdos locales, regionales y nacionales. Las explicaciones más generales sobre el proceso de consensuación ofrecen solamente una lista de las agencias protagonistas del cambio curricular: universidades, especialistas de área, pedagogos, maestros, padres, representantes de organizaciones sociales, organismos no gubernamentales (ONGs), e instituciones religiosas. Otros países, como Argentina, Brasil y México, explican en forma más acabada los procedimientos que se utilizaron para la elaboración consensuada de metas y contenidos. Estos procedimientos involucran a una amplia variedad de actores e implican la formulación y validación de los contenidos en diferentes etapas. En algunos casos se mencionan fechas y eventos clave en el proceso de elaboración, mientras que en otros casos simplemente se mencionan los procedimientos, tales como reuniones, mesas de trabajo, y encuestas populares y a especialistas. Las descripciones más detalladas del proceso de validación, que corresponden a Brasil y a México,

incluyen información sobre el número de encuestas realizadas, la cantidad de personas que participaron en los talleres de discusión y redacción, y el número de propuestas finales recogidas. En todos los casos en que se habla de *validación* de los contenidos, se hace énfasis en la complejidad del proceso llevado a cabo para la elaboración de los nuevos currículos, y que incluye varias instancias de elaboración como bosquejos, borradores, correcciones, compatibilización, revalidación con grupo control, redacción, ejecución, ajustes y evaluación.

Estos procesos de validación pueden definirse como instancias de legitimación técnica y política de los contenidos curriculares a través de sucesivas revisiones del currículo por parte de diferentes actores del sistema. En algunos países se le dio más espacio a los docentes en este proceso de validación, mientras que en otros se privilegió la acción de académicos en las diferentes disciplinas y de representantes de diferentes sectores de la sociedad civil, tales como grupos profesionales y empresarios del sector privado. En las entrevistas realizadas por el autor se mencionan algunos de los aspectos del currículo que fueron sometidos a validación en la región. En un caso, la principal dificultad para arribar a un acuerdo sobre los contenidos curriculares fue la decisión sobre qué contenidos habrían de descartarse cuando no podían ser cubiertos por el programa debido a restricciones de tiempo. Se trataba de casos donde había consenso sobre la importancia y calidad de todos los contenidos pero existían diferencias en cuanto a cuáles de estos contenidos priorizar. En otro caso en particular, el debate para la validación giró en torno a la inclusión de contenidos actitudinales en el currículo nacional. Debido a la inevitable carga axiológica de la discusión, los puntos de vista tendían a ser múltiples y más subjetivos que en el debate sobre contenidos conceptuales y procedimentales y esto sometía al currículo a un proceso mucho más complejo de validación. Con respecto al carácter técnico o político de los procesos de validación, encontramos que la iniciativa para emprender estos procesos es, con frecuencia, de origen político. Sin embargo, los procedimientos que tuvieron lugar durante las transformaciones curriculares pueden considerarse eminentemente técnicos. Aun dentro del marco de paulatina democratización que vive la región, percibimos que el público no ha adoptado una actitud política en relación a los contenidos curriculares o a los resultados del aprendizaje. En la actualidad, la participación popular para la validación ocurre solamente cuando los diferentes actores son convocados formalmente por las administraciones educativas centrales para discutir aspectos técnicos de la

elaboración de los contenidos curriculares. En resumen, observamos que la validación del currículo es un proceso de *iniciativa* política, pero de ejecución técnica.

Una consecuencia importante de esta diferenciación es que la iniciativa política no se consolida en una *práctica* política sostenida, como ocurre en países de otras regiones. La validación curricular en América Latina ocurre con frecuencia durante la elaboración inicial de los currículos pero no se ha institucionalizado como proceso permanente. Países como el Reino Unido, Dinamarca o España poseen organismos ejecutivos y consultivos permanentes para la producción y validación de los contenidos curriculares (Cariola, 1994). En América Latina, los procesos de validación para la transformación se realizaron con comisiones y grupos de trabajos creados ad hoc durante la planificación de los nuevos programas, pero que no se constituyeron en cuerpos consultivos permanentes. Existen en la región diferentes formas de consejos locales y distritales que participan en el diseño curricular, pero solamente durante los procesos de adecuación y recontextualización, es decir, una vez que los contenidos mínimos y obligatorios han sido establecidos con fuerza legal. En la práctica, en muchos de nuestros países cambiar los contenidos mínimos implica modificar leyes y decretos nacionales, una tarea que requeriría más energía, conocimiento y poder político de los que probablemente disponen en la actualidad los consejos consultivos locales. Los sindicatos docentes y asociaciones profesionales no se han constituido en agentes permanentes de promoción para la elaboración y renovación de contenidos, lo cual indica que la validación curricular es uno de los aspectos de las políticas educativas de la región que aún permanecen relativamente centralizados. A pesar de esto último, es importante destacar que se ha trascendido desde la prescripción exacta y rígida de los contenidos para todas las escuelas, hacia una concepción más flexible del diseño curricular que propone «lineamientos» curriculares generales que favorezcan la adecuación curricular y la autonomía institucional. Esta autonomía, a su vez, conlleva ciertas dificultades para el establecimiento de estándares claros de desempeño y rendimiento; este punto se discutirá con más detenimiento en la sección tercera de este informe.

iii. Los valores como contenidos curriculares

También es común a todos los países de la región una declaración explícita de los valores que la educación debe fomentar en sus ciudadanos

para construir naciones sólidas y democráticas. Es evidente la preocupación por consolidar la democracia, reafirmar los acuerdos de paz en los países más afectados por los conflictos armados internos, y la recuperación de los valores humanos y sociales más altos como el amor, la honestidad, la justicia, el trabajo, la tolerancia, la libertad, la solidaridad, la verdad y la igualdad, entre otros. En muchos países de la región, los nuevos currículos reconocen por primera vez como valor la pluriculturalidad y el plurilingüismo. Estos currículos a menudo incluyen un discurso que impele a los actores del sistema educativo a rescatar y valorar los aportes de los grupos lingüísticos y culturales minoritarios. Los currículos de América Central hacen mayor hincapié que los de América del Sur sobre el eje axiológico del diseño en cada área curricular. Es importante destacar estos aspectos de las justificaciones de todos los currículos dado que, como veremos más adelante, inspiran e impregnan el diseño de los programas de estudio en todos los casos a través de la inclusión de los contenidos actitudinales.

iv. Fundamentos epistemológicos

La gran mayoría de los documentos analizados expresan la convicción de que los objetivos de transformación educativa van a ser alcanzados en tanto se efectúe un cambio radical en la concepción sobre la naturaleza de los procesos de aprendizaje y de los procesos pedagógicos hasta ahora vigentes. En casi todos los documentos analizados este cambio de enfoque supone una alineación con los principios de la educación constructivista. Algunos países, como Nicaragua y Panamá, declaran explícitamente su adherencia al modelo mencionado. Otros hacen una descripción exhaustiva del modelo pero sin darle un nombre determinado. Gran parte de los documentos curriculares de la región presentan extensas y elaboradas explicaciones de los principios epistemológicos y pedagógicos que deben guiar la ejecución del currículo de la transformación. Si bien se trata de lineamientos curriculares cuya función básica es la de brindar indicadores de logro en las diferentes áreas de conocimiento, en muchos casos podrían confundirse con textos de estudio para cursos de formación y capacitación docente.

Los principios epistemológicos expresados por la mayoría de los documentos curriculares oficiales en América Latina comparten las siguientes características:

- Los contenidos pretenden la formación integral de los alumnos, en sus aspectos cognitivos, afectivos, sociales y psicomotrices.
- Se desplaza a la enseñanza como centro del proceso educativo y se ubica en su lugar al alumno y al aprendizaje mismo.
- Se privilegia el pensamiento crítico y reflexivo, así como la creatividad y la imaginación.
- Se asume que el conocimiento se construye y reconstruye en un diálogo permanente entre los saberes del docente y los del alumno.
- El aprendizaje debe ser significativo: los alumnos traen al aula conocimientos previos sobre los cuales han de construirse los nuevos saberes.
- Saber significa saber ser, saber hacer y saber aprender.
- Se promueve el desarrollo de capacidades, competencias y destrezas, no solamente del conocimiento declarativo.
- Las capacidades y competencias implican el desarrollo de procesos mentales de orden superior, tales como el análisis, la resolución de problemas, la toma de decisiones, la inferencia, la comparación, etc.
- El desarrollo de estas competencias complejas se hace posible mediante actividades de aula complejas: investigación, aprendizaje colaborativo, trabajo en equipo, etc.
- Los temas y las experiencias de aprendizaje deben centrarse en los intereses de los alumnos, en su realidad cotidiana y en sus necesidades.
- Los procesos pedagógicos deben adaptarse a los diferentes ritmos y estilos de aprendizaje de los alumnos.
- Los contenidos no son fines en sí mismos, sino medios para el desarrollo de capacidades y destrezas.
- Los contenidos conceptuales no pueden ser aislados de los valores y actitudes que la educación escolar debe promover.

En el análisis de contenidos y de indicadores de logro a continuación volveremos sobre estos principios y sobre la forma en que están representados en los diferentes diseños curriculares⁶ en América Latina.

⁶ Por diseño curricular entendemos la organización, clasificación y secuenciación de los contenidos disciplinarios. Consideramos que este diseño debe constituir una herramienta de comunicación efectiva entre curriculistas y otros actores del sistema educativo. Desarrollo curricular, por su parte, se refiere a los procesos técnicos y políticos que permiten que el currículo sufra transformaciones de acuerdo a las necesidades de cada nación o región.

2. CONTENIDOS E INDICADORES DE LOGRO

En esta sección nos proponemos realizar un análisis descriptivo de la organización de los contenidos en los diferentes currículos de la región, y de la forma en que esta organización se traduce, o no, en indicadores de logro para una apropiada evaluación de los saberes adquiridos. Dada la variedad de modelos de currículo y debido a las diferencias en la terminología usada para describir los distintos componentes del currículo, optamos por ordenar la información de esta sección en una escala que abarca las principales estrategias de organización de contenidos utilizadas por cada país; identificamos en un extremo de esta escala las tendencias más firmes, consistentes, y comunes a todos los países en la organización y secuenciación de los contenidos curriculares; en el otro extremo describimos las estrategias de organización de contenidos que presentan mayores dificultades técnicas y variabilidad entre los currículos de los diferentes sistemas educativos.

Currículo por competencias

Una lectura rápida de las justificaciones que se ofrecen para cada currículo revela una marcada tendencia a definir los productos finales del aprendizaje como *competencias*. Este término marca un cambio de paradigma en la concepción del aprendizaje y del papel de la educación en la sociedad moderna. En todos los documentos encontramos una ferviente declaración de la necesidad de establecer metas educativas que apunten a la formación de individuos sanos e íntegros, que participen activa y constructivamente en su medio social y en el mundo. El logro de estas metas requiere de la definición de un perfil de alumno diferente, y observamos que la nueva organización de los contenidos pretende contribuir a la formación de este perfil. En teoría, un alumno o ser humano *competente* es el que combina efectivamente cuatro tipos de saberes: un saber conceptual (o simplemente “saber”), un saber procedimental (saber hacer), un saber actitudinal (saber ser), y finalmente un saber metacognitivo (saber aprender) (Pinto Cueto, 1999). La persona demuestra una competencia, entonces, cuando decide poner el conocimiento conceptual a su servicio y del medio en el que se desenvuelve, en forma efectiva y eficiente. Además, es capaz de reflexionar sobre ese proceso, y así reproducir y mejorar su desempeño en nuevas situaciones o contextos⁷.

⁷ Utilizaremos esta definición para referimos al concepto de competencia.

Sobre la base de este principio se han diseñado la mayor parte de los currículos oficiales de los últimos años en América Latina. Tal y como ha sido empleado hasta ahora, este nuevo modelo presenta ventajas y desventajas. Una de las ventajas es la “integralidad” de las nuevas metas educativas. Se asume que el educador puede realizar mejor su tarea cuando conoce el perfil general del estudiante y del futuro ciudadano que se espera al cabo de la experiencia escolar. Una noción consensuada de los valores y actitudes a promover permite organizar mejor las experiencias de aprendizaje, y también legitima los juicios valorativos que se realizan en el proceso de enseñanza-aprendizaje. Por otro lado, la noción de competencia es demasiado compleja para ser transmitida por medio de documentos masivos donde se hace necesaria la síntesis gráfica de la información curricular. Los cuadros sinópticos que se ofrecen a directivos y maestros son limitados en la cobertura que hacen de los contenidos mínimos y no llegan a transmitir, u operacionalizar, la complejidad de una educación basada en competencias. Esta limitación puede obstruir la implementación curricular y tener efectos negativos en los procesos de planificación y evaluación de los aprendizajes.

Los currículos que hacen explícitos los objetivos por competencias organizan los contenidos en saberes conceptuales, procedimentales y actitudinales, generalmente en dos, tres, o más columnas diferenciadas. En algunos modelos encontramos al principio de cada área curricular (asignatura) una justificación axiológica para la adquisición y uso de los saberes conceptuales y procedimentales. Se explica en tales casos que los valores y actitudes no se pueden desagregar por temas o conceptos, y por ende se plantean como válidos para toda la disciplina. En otros casos, los valores y actitudes a desarrollar aparecen como saberes diferenciados correspondientes a temas y conceptos específicos, y se los presenta como un indicador de logro más.

Los países que presentan los contenidos actitudinales en forma transversal y no como indicadores de logro evidencian un diseño curricular más claro y, en general, más completo en términos de saberes conceptuales y procedimentales. Por el contrario, aquellos modelos que han intentado incluir los valores y actitudes a promover a lo largo del plan de estudios en forma detallada tienden a presentar los objetivos conceptuales, procedimentales y actitudinales mezclados desordenadamente, dificultando así la lectura del currículo.

Tomemos dos ejemplos que ilustran este último argumento. Brasil, en su programa de estudios para el segundo ciclo de matemáticas, dife-

rencia claramente entre objetivos generales de la asignatura para el ciclo, y luego procede al listado de los contenidos conceptuales y procedimentales de los diferentes bloques temáticos del área. Si bien estos contenidos son muy generales, ya que no describen exactamente qué conceptos han de incluirse en los procesos mencionados, proponen al menos un modelo claro para el docente que planifica el curso anual.

Ejemplo 1: Brasil. Matemáticas, Ciclo II. Magnitudes y medidas. “Parámetros Curriculares Nacionais de 1^a a 4^a Séries” y “Parámetros Curriculares Nacionais de 5^a a 8^a Séries”.

Contenidos conceptuales y procedimentales:

- Comparación de magnitudes de igual naturaleza y elección de una unidad de medida de la misma especie que el objeto a ser medido.
- Identificación de magnitudes mensurables en el contexto diario: masa, capacidad, superficie, etc.
- Reconocimiento y utilización de unidades usuales de medida como metro, centímetro, kilómetro, gramo, miligramo, kilogramo, litro, mililitro, metro cuadrado, etc.
- Reconocimiento y utilización de medidas usuales de tiempo y de temperatura.
- Etc.

Contenidos Actitudinales:

- Confianza en sus posibilidades de proponer y resolver problemas.
- Perseverancia, esfuerzo y disciplina en la búsqueda de resultados.
- Seguridad en la defensa de sus argumentos y flexibilidad para modificarlos.
- Etc.

Ejemplo 2: Nicaragua. Español, 3^{er} grado. “Programas Español y Matemática: Tercer Grado” Serie: Enseñemos y aprendamos activamente. 1996.

Nicaragua, cuyo programa de estudios para el área de español de tercer grado presenta una importante carga valórica, dedica menos espacio a los contenidos conceptuales y procedimentales.

Objetivos generales:

1. Cultive amor y respeto a Dios, a sí mismo, a la Patria, a la naturaleza, a la familia, al maestro y a sus semejantes.
2. Practique hábitos de orden, aseo y alimentarios para conservar la salud y prevenir enfermedades.
3. Incremente hábitos y habilidades de lectura comprensiva e interpretativa para apreciarla y utilizarla como fuente de información, conocimiento y desarrollo de valores, sentido de superación, autodisciplina, perseverancia, entre otros.
4. Etc.

Este documento presenta como objetivo de la primera unidad temática un único **logro de aprendizaje:**

1. Demuestre a través de la evaluación diagnóstica:
 - 1.1 Las habilidades, destrezas y conocimientos desarrollados en la expresión oral.

Como único **contenido básico** especifica:

1. Diagnóstico sobre desarrollo de habilidades, destrezas y conocimientos y práctica de valores:
 - 1.1 Expresión oral.

Luego de sugerir actividades para el estudiante, donde no se especifican contenidos conceptuales, se proponen los **procedimientos de evaluación:**

Observar, evaluar y registrar:

- Las actitudes que muestren los niños en las dinámicas de integración.
- El orden, diligencia y práctica de cortesía evidenciado en las actividades.
- La expresión coherente y el uso de vocabulario adecuado.
- La escucha atenta y respetuosa de narraciones y descripciones.

Observamos cómo en gran parte del currículo domina el eje axiológico, a expensas de los contenidos estrictamente disciplinarios.

Currículo por bloques o ejes temáticos

Una de las estrategias de organización curricular de mayor consistencia en la región es la división de las áreas por bloques o ejes temáticos. Muchos países coinciden en el uso de los mismos términos para la clasificación de bloques de contenido, especialmente en las áreas de matemática y ciencias naturales. Esto implica que los contenidos conceptuales y los procesos indicados para cada año o grado escolar en diferentes países guardan una mayor semejanza en esas áreas que en aquellas cuyos contenidos no están clasificados en ejes o bloques temáticos. Sistemáticamente, la mayoría de los países de la región usan los siguientes términos para clasificar los bloques de las áreas de matemática, ciencias naturales y español.

Matemáticas:

1. Números
2. Operaciones
3. Lenguaje gráfico y algebraico
4. Geometría
5. Mediciones
6. Estadística y probabilidad

Ciencias Naturales:

1. La vida/los seres vivos y sus propiedades
2. El mundo físico/Materia y energía

3. El cuerpo humano/conservación de la salud
4. El ser humano y el medio ambiente

En la enseñanza del español hay cierto consenso en llamar al área Lenguaje y Comunicación y, aunque con menos sistematicidad que en las otras áreas mencionadas, se coincide en utilizar bloques temáticos con la siguiente denominación:

Lenguaje y Comunicación:

1. Lengua oral
2. Lengua escrita
3. Discurso literario
4. Reflexión sobre la lengua

Si tenemos en cuenta que una de las funciones básicas de los estándares es la de posibilitar un lenguaje técnico común para diferentes sistemas educativos (McREL, www.mcrel.org), estamos en condiciones de argumentar que esta clasificación relativamente homogénea de ejes temáticos en la región constituye un avance considerable en la definición de estándares de contenido, aunque no se planteen como tales.

Currículo por expectativas de logro

Las expectativas de logro difieren de los objetivos generales en que las primeras hacen referencia al perfil del alumno que se espera hacia el final de su experiencia escolar en las diferentes áreas curriculares; los objetivos generales, por otro lado, anticipan o introducen los conocimientos conceptuales y procedimentales de cada año o grado escolar. El énfasis de las expectativas de logro recae generalmente en las competencias complejas que los alumnos deben demostrar en términos conceptuales, procedimentales y actitudinales.

Ejemplo 3: Argentina, Ciencias Naturales, Bloque 3. Expectativas de logro al final de la Educación General Básica. “Contenidos Básicos Comunes para la Educación General Básica”. (www.mcye.gov.ar)

Los alumnos y las alumnas deberán: Reconocer y analizar transformaciones químicas, su presencia universal en los procesos de la naturaleza desde la atmósfera y la corteza terrestre hasta las manu-

facturas y los seres vivos, así como tener una idea de la naturaleza y estructura microscópica de la materia. Relacionar, si bien solo a nivel descriptivo, los niveles de energía atómicos con la producción de energía electromagnética (luz) y entender el almacenamiento de energía en uniones químicas que resultan de la interacción entre átomos. Adquirir una perspectiva de la complejidad y naturaleza transdisciplinaria de muchos problemas reales; formar actitudes y familiarizarse con los procedimientos –como los aplicados en el esfuerzo de informarse, al consultar a expertos– necesarios en la toma de decisiones responsables sobre temas tan complejos del mundo actual.

Ejemplo 4: Perú, Área Lógico-Matemática. Competencias terminales al concluir el tercer ciclo (Inicial y Primaria). “Nueva Estructura Curricular: Lineamientos generales de planes de estudio”. Ministerio de Educación, julio 1998.

Interpreta, resuelve y crea problemas aritméticos y de geometría básica, tomados de situaciones cotidianas, demostrando flexibilidad y perseverancia, así como habilidad en el cálculo, en el razonamiento y en el uso de unidades de medida.

Estas expectativas de logro no constituyen indicadores aptos para la evaluación del rendimiento pero sí se plantean como orientadores para la planificación de las experiencias de aprendizaje y para la evaluación de aula e institucional. Observamos en los ejemplos dados que cada país elige diferentes grados de definición para las expectativas de logro. Mientras que Perú las propone solamente como expectativas generales para la totalidad del área, Argentina especifica sus expectativas para cada bloque temático de la asignatura.

Currículo por objetivos específicos⁸ o indicadores de logros

Los ejemplos en las secciones anteriores proponen parámetros para la medición de los logros educacionales pero no hacen referencia a los

⁸ Los objetivos específicos o indicadores de logro se refieren a la descripción más detallada que hacen los diferentes currículos de los saberes y capacidades que los alumnos deben demostrar haber adquirido en cada área curricular.

saberes conceptuales y procedimentales específicos que los alumnos deben alcanzar en cada área curricular. A continuación exponemos una serie de ejemplos de indicadores de logros presentes en los currículos de la región. Hacemos especial hincapié en este aspecto del currículo porque consideramos que es en este nivel del diseño donde pueden establecerse los estándares mínimos de contenido y desempeño. Los ejemplos incluidos en esta sección ilustran el nivel más profundo de concreción del currículo en cada país, y las observaciones que se hacen en cada caso implican una comparación con el ideal de estándar al que podríamos aspirar. Debido a la gran variedad de modelos y de calidad en los diferentes diseños curriculares, optamos por organizar los ejemplos en una escala que ilustra las dificultades y alcances más prominentes en la elaboración de indicadores claros de desempeño y rendimiento. En el extremo inicial de esta escala ubicamos los diseños curriculares que, por diferentes razones, no alcanzan a definir claramente los contenidos y competencias a desarrollar. Entre estas razones encontramos:

- **Documentos de difícil lectura:** La presentación gráfica de los contenidos no permite una visión clara de los objetivos a corto y a largo plazo, o de la articulación posible entre áreas o grados de escolaridad. No se utilizan marcadores gráficos y textuales que permitan una fácil identificación de áreas, ejes temáticos, o tipos de objetivos.
- **Terminología confusa:** La elección de términos para definir a los indicadores se presta a confusión, especialmente de los contenidos que se presentan como procesos o como competencias. También se confunden los indicadores de logro y rendimiento con las actividades de aula propuestas.
- **Falta de enfoque:** No se identifican prioridades claras en el diseño. El eje axiológico predomina sobre el conceptual y procedimental, dificultando la lectura de los últimos. Los contenidos conceptuales no llegan a operacionalizar los objetivos fundamentales planteados para ese currículo, como el desarrollo de competencias complejas o la interdisciplinariedad de la educación.
- **Débil articulación vertical:** Los contenidos no se complejizan o profundizan en función del desarrollo cognitivo y la madurez que los estudiantes ganarían año a año.

A medida que nos acercamos al extremo final de la escala propuesta en este informe, ofrecemos ejemplos de currículos que presentan una

definición más acabada de los saberes conceptuales y procedimentales a ser alcanzados. Además de una mayor definición y una mayor cantidad de contenidos, observaremos un diseño curricular más claro, gráfica y verbalmente, con mayor potencial para facilitar la tarea de planificación y evaluación de los aprendizajes a los establecimientos. Entre los **términos** utilizados en América Latina para referirse a los objetivos específicos o indicadores de logros de cada área curricular encontramos los siguientes:

- objetivos
- contenidos básicos
- rendimientos básicos
- destrezas
- habilidades
- actividades
- competencias
- procesos
- productos específicos
- contenidos conceptuales
- contenidos procedimentales
- contenidos actitudinales
- logros de aprendizaje
- objetivos instrumentales
- conocimientos
- objetivos fundamentales
- contenidos mínimos
- criterios de evaluación
- indicadores de evaluación
- capacidades
- experiencias de aprendizaje

Aunque muchas de estas clasificaciones no pueden ser consideradas como reales indicadores de logro, todas constituyen el nivel más alto de especificación de los conocimientos conceptuales, procesuales y actitudinales a ser adquiridos por los alumnos de educación básica en diferentes países. Los siguientes ejemplos ilustran algunas de las dificultades presentadas en la elaboración de indicadores de logros en los currículos de la región, así como los avances en la definición de indicadores de desempeño y rendimiento en las diferentes áreas.

Ejemplo: País 01. Matemáticas, Educación Básica.

Los contenidos curriculares básicos en este país se presentan como un apoyo al educador para que organice su trabajo diario. Los contenidos en matemáticas, por ejemplo, se dividen en bloques regulares para todos los grados: estadística, geometría, números y operaciones medidas. Cada subárea expresa una expectativa de logro para la conclusión de la Educación Básica.

Los contenidos del área están divididos en **conocimientos** y **habilidades y destrezas**. Los conocimientos, sin embargo, están expresados como habilidades o procesos. La cantidad de conceptos incluidos para la educación básica en cada bloque es menor que en otros países de la región.

BLOQUE: Medidas

Conocimientos:

- Identificación de instrumentos de medidas convencionales y no convencionales: regla, metro, compás, reloj, balanza, recipientes de litro y botella.
- Construcción del concepto de longitud y uso del metro, decímetro, centímetro y kilómetro.
- Identificación del área de superficies por conteo de unidades cuadradas.
- Reconocimiento de medidas de capacidad mediante el uso del litro, medio litro, botella y media botella.
- Reconocimiento de medidas al utilizar libras, medias libras y onzas.
- Reconocimiento de intervalos de tiempo: día, semana, mes, año, horas y minutos.
- Identificación de [moneda nacional], sus denominaciones y su equivalencia fraccionaria.

Habilidades y Destrezas (para el mismo bloque)

- Utilización adecuada de instrumentos de medida.
- Estimación y cálculo de medidas en situaciones lúdicas.
- Capacidad para detectar la conservación de la longitud, área y volumen.
- Toma de decisiones al seleccionar patrones de medida convencionales.
- Lectura correcta de horas y minutos en el reloj.
- Manejo adecuado del dinero.
- Elaboración de estrategias personales para llevar a cabo mediciones y estimaciones.
- Explicación verbal y escrita del proceso seguido al efectuar mediciones.
- Creatividad al organizar y representar en tablas y gráficas el resultado de sus actividades.

La Unidad Uno de matemática de tercer grado se denomina “Ubíquemonos en el espacio”. Sus dos únicos **objetivos** son:

1. *Desarrollar capacidades de ubicación, localización y orientación, en el plano y el espacio.*
2. *Discriminar posiciones relativas en dos, tres o cuatro rectas.*

En la descripción de la unidad no se listan conceptos específicos a adquirir sino que se describen ciertos procesos tales como “identificar y clasificar los triángulos por comparación de sus ángulos”. El resto de la unidad describe detalladamente el tipo de **actividades** de iniciación y de desarrollo para trabajar los escasos conceptos que se presentan. La descripción se asemeja a las recomendaciones que podríamos encontrar en un libro de didáctica.

Observamos también un cierto grado de contradicción entre los principios de adecuación curricular que se proponen al principio del documento y la exhaustiva descripción de las actividades para cada unidad. Se corre el riesgo con esta descripción de actividades que los docentes hagan exactamente lo que se propone como ejemplo, en vez de ser utilizado solo como modelo. Si bien otros documentos curriculares de este país sostienen que el énfasis debe centrarse en la producción/construcción del conocimiento, y no en la enseñanza, el programa de estudios hace más hincapié en las actividades de clase que en los conocimientos a ser desarrollados. No se especifican niveles de complejidad o de ejecución dentro de cada unidad. Los criterios de evaluación se limitan a reformular los pasos en la actividad sugerida para la clase.

Ejemplo: País 02. Tercero y Sexto Grados. Procesos y objetivos instrumentales.

Las guías curriculares⁹ de este país describen a los contenidos como **procesos**. Cada proceso es definido como una competencia cognitiva o socioafectiva. Por ejemplo:

⁹ Los países que han optado por establecer metas y objetivos generales para cada disciplina pero que dejan en manos de cada institución la elaboración de indicadores de logro específicos usan el término “guía curricular” en vez de currículo.

Medir: Cuantificar una magnitud de acuerdo a un patrón establecido.

Compartir: Dar a otras personas algo nuestro y recibir de otras personas (ideas, materiales, etc.)

El siguiente nivel de especificación de contenidos se expresa como **objetivos instrumentales** que permiten trabajar los procesos:

Los objetivos instrumentales son conductas específicas desprendidas del listado de procesos y persiguen adecuar los mismos al nivel de desarrollo de los estudiantes y plantearlos de manera que sean más sencillos de trabajar para el maestro y para los alumnos. Los objetivos instrumentales son, pues, un desglose de las acciones necesarias para cada proceso, haciendo un énfasis especial en el desarrollo del razonamiento verbal y del razonamiento lógico-matemático.

A cada una de las áreas curriculares le corresponde un número de procesos a desarrollar que se extraen de la *lista general de procesos para el tercer grado*. Finalmente, se mencionan, aunque no hay documentos que expliquen exactamente en qué consisten, las Unidades de Aprendizaje Integrado. “Naturalmente, tratándose de una planificación flexible, pueden tomarse objetivos instrumentales de diversos bloques para una misma unidad de aprendizaje integrado pero, si no hay un motivo especial, conviene mantenerse dentro de un solo bloque”. En cierta medida, los objetivos instrumentales constituyen indicadores de logros que describen conductas y competencias evaluables. Sin embargo, algunos objetivos son muy específicos mientras que otros son extremadamente generales. Los relacionados al área de matemática, como es el caso en la mayoría de los países, tienden a ser más específicos:

- Establecer series numéricas de 2 en 2, de 3 en 3, de 4 en 4, etc.
- Establecer relaciones de equivalencia entre unidades de mil, centenas, decenas y unidades: entre centenas y decenas o unidades.

En otras áreas, como Ambiente Social y Natural, los conceptos son tan amplios que la cobertura de los mismos podría variar considerablemente entre diferentes aulas o instituciones:

- Identificar los elementos necesarios para la supervivencia de los seres vivos
- Realizar investigaciones

Esta guía curricular no presenta diferenciación de bloques temáticos en cada área. Se ofrece una tabla con los números de los procesos a incluir en cada área. Sin embargo, la lista donde están redactados los procesos no está organizada por áreas o bloques. Tomando una serie de contenidos tal como se hallan listados en el documento oficial, nos encontramos con los siguientes objetivos instrumentales para el tercer grado:

55. Identificar las principales variaciones entre los organismos
56. Distinguir entre recursos renovables y no renovables
57. Identificar problemas ambientales de su localidad
58. Movilizar coordinadamente su cuerpo, siguiendo rutas definidas con obstáculos
59. Atender diferentes indicaciones verbales, para llevar a cabo formaciones y agrupaciones
60. Crear diseños, para los trabajos que lo requieran
61. Aplicar diversas técnicas artísticas
62. Desarrollar su sensibilidad auditiva y afectiva, a través de la música de compositores nacionales
63. Medir longitudes con metro, decímetro y centímetro
64. Establecer equivalencias entre las unidades convencionales de longitud

La desagregación y secuenciación de estos objetivos para lograr una apropiada articulación horizontal y vertical de contenidos siguiendo esta lista sería una tarea ardua de llevar a cabo para el docente o institución a cargo del diseño curricular anual.

En la Guía Curricular de Sexto Grado, los contenidos se plantean por áreas. A cada área le corresponde una serie de **Objetivos Generales** y otra de **Objetivos Instrumentales**. La lectura de estos objetivos para el docente se hace difícil ya que los objetivos instrumentales no constituyen una acabada descripción de las competencias que los alumnos deben

adquirir para el logro de los objetivos generales. Son más bien actividades a realizar que no siempre explican los conocimientos que deben ser adquiridos, y que además no agregan información substancial a lo expresado en los objetivos generales. Ejemplo en el área de lenguaje:

Objetivo general:

- Interpreta críticamente comunicaciones que le sean entregadas en forma oral.

Objetivo instrumental:

- Interpreta críticamente informaciones escuchadas o leídas.

Queda a discreción del docente o institución la elección del medio o canal en que se van a efectuar las «comunicaciones, así como la definición de qué constituye una interpretación crítica, o quién es el emisor de las informaciones escuchadas o leídas. Cabe preguntarse también qué temas y materiales se utilizan durante todo un año escolar para alcanzar estos objetivos instrumentales.

Ejemplo: País 03. Español, Sexto Grado.

El Programa de Sexto Grado (1999) establece los **objetivos generales** de la asignatura para la educación primaria y luego, por separado, los **objetivos generales** para español en el sexto grado. Estos últimos no agregan información más detallada, solo reformulan los objetivos generales de la asignatura. El cuadro de contenidos dispone tres columnas diferenciadas que reproducimos a continuación:

OBJETIVOS	CONTENIDOS (Conceptual, Actitudinal, Procedimental) [sic]	EXPERIENCIAS DE APRENDIZAJE Y DE EVALUACIÓN
1) Desarrollar fluidez en el pensamiento y dicción clara y correcta, condiciones básicas en el emisor. 2) Reconocer los aspectos básicos de la comunicación.	<ul style="list-style-type: none"> - Estudio y dominio de los elementos de la comunicación: emisor, receptor, mensaje, canal, código. - Fluidez de pensamiento y dicción clara y correcta. - Fluidez de pensamiento y dicción por medio de: <ul style="list-style-type: none"> - charlas - debates - Oratoria 	<ul style="list-style-type: none"> - Seleccionarán el temario propuesto por el grupo. - Investigarán el contenido del mismo mediante el uso del diccionario. - Formarán grupos de trabajo con un vocero (emisor); - Asignar temas para exponer una mesa redonda. - Conversar sobre los diferentes tipos de canal. - Planearán la adquisición de una experiencia en una excursión, una entrevista. - (Etc.)

Observamos que ninguna de las tres columnas ofrece una indicación clara de los logros que se espera del alumnado. Los contenidos conceptuales propuestos no pueden ser utilizados para desarrollar “fluidez de pensamiento” o una “dicción clara y correcta”. Se trata simplemente de conocimiento declarativo sobre los elementos de la comunicación de acuerdo al modelo de un autor. Lo mismo se aplica a las charlas y debates propuestos para el segundo objetivo, que no pueden considerarse procedimientos sino situaciones comunicativas que debieron incluirse en la columna de Experiencias de Aprendizaje. Por último, las experiencias de evaluación consisten en una serie de actividades que no ofrecen indicadores claros sobre los conceptos ni las conductas que se requieren para el logro del objetivo propuesto en la primera columna.

Los objetivos generales de sexto grado en el área de matemáticas son más específicos que en la de español. Si bien se confunden tipos diferentes de objetivos (de conceptualización, aplicación y valoración), hay una definición más clara de las destrezas a promover. Los objetivos específicos están mejor definidos, pero no substancialmente. Se mezclan nuevamente en forma indiscriminada los contenidos conceptuales, actitudinales y procedimentales, y finalmente se agregan las experiencias de aprendizaje y evaluación que tampoco especifican indicadores claros de realización.

Ejemplo: País 04. Español. Tercer Grado.

Los rendimientos básicos que presenta el currículo de este país están definidos como “los productos del aprendizaje que el alumno debe alcanzar al finalizar cada grado de la escuela primaria”. Estos **rendimientos básicos** están clasificados en **producto general, producto específico e indicadores de evaluación**: Por ejemplo, en español de tercer grado:

Producto General:

- Demuestra destrezas para hablar Si...

Producto Específico:

- ...Expresa comprensiblemente los mensajes implícitos en ilustraciones y los relaciona con su experiencia.

Indicadores de Evaluación:

- Explicará oralmente y con precisión el mensaje que comunican dos ilustraciones. Comentará verbalmente el mensaje de una ilustración y lo asociará con sus experiencias (aseo personal, letrización, vacunación, basura...)

Desde este enfoque, el docente debe inducir un comportamiento determinado a través de actividades muy específicas. La mayor parte de los productos de aprendizaje se refieren a conocimiento de tipo declarativo y memorístico, especialmente en ciencias naturales y estudios sociales. Los alumnos y alumnas deberán narrar eventos y procesos, identificar datos, etc., pero no se requiere de ellos la aplicación de conocimientos en situaciones nuevas o la resolución de problemas. En ciencias naturales, también para el tercer grado, hallamos los siguientes objetivos:

Producto General:

- Comprende la importancia del agua Si...

Producto Específico:

- Explica el ciclo del agua en la naturaleza.

Indicadores de Evaluación:

- Presentada la lámina del texto, describirá correctamente el ciclo del agua en la naturaleza.
- Dibujado el ciclo del agua, resumirá correctamente en no más de cinco renglones cada uno de los cambios que sufre.

En matemática, casi todos los indicadores se refieren a la realización de operaciones con exactitud; no se exige la resolución de problemas en relación a la vida cotidiana, ni se observan esbozos de interdisciplinariedad. En español, se apela a ciertas habilidades cognitivas más sofisticadas, como la interpretación (no solamente comprensión literal) de textos literarios o la producción creativa de una narración. Los conceptos y procedimientos de cada asignatura no se agrupan en bloques temáticos, y por lo tanto van variando los enfoques de grado en grado.

Ejemplo: País 05. Lenguaje y Matemática.

Los nuevos programas de estudio en este país organizan los contenidos por ciclos y por competencias solamente. Las competencias incluyen conceptos, procedimientos y actitudes pero no están separadas en categorías, ni se aprecia un criterio definido para la secuenciación de estas competencias. Tampoco hay una división de los contenidos por bloques o ejes temáticos. Se van intercalando competencias orales, escritas y de valoración sin un orden aparente. Algunos objetivos proponen indicadores relativamente claros de rendimiento, como por ejemplo en el área de lenguas originarias como segunda lengua del segundo ciclo:

Competencia:

- Escucha comprensivamente y comunica en forma oral: instrucciones para realizar alguna actividad; expresiones de sentimientos y de sensaciones propias, de otras personas y de grupo; descripciones de procesos y funciones; expresiones metafóricas de uso frecuente; narraciones de mediana extensión referidas a diversos temas, con apoyo de otros códigos: gestos, ademanes, etc.

Otros objetivos son más generales, desatendiendo a los conceptos específicos que deberían adquirirse para desarrollar las competencias propuestas. Por ejemplo, en el bloque de lengua materna del primer ciclo:

Competencia:

- Produce textos escritos, atendiendo a la diagramación, de acuerdo al tipo de texto, la intención comunicativa, y usando el alfabeto y puntuación de su lengua, de acuerdo a sus necesidades comunicativas.

Se puede observar en este último caso que para un ciclo escolar de 3 años pudieron hacerse más explícitos ciertos indicadores como una clasificación de signos de puntuación a utilizar, los tipos de necesidades comunicativas propios de la edad, o el tipo de textos que se puede exigir de acuerdo a las diferentes necesidades comunicativas en cada año. En relación a la articulación vertical de los contenidos, observamos que en el tercer ciclo del área de competencia arriba mencionada no se proponen conceptos y procesos más complejos, como habría de esperarse para un nivel mayor de madurez cognitiva. La débil especificación de contenidos puede también observarse en los siguientes ejemplos:

Competencia:

- Produce textos escritos, respetando los aspectos formales de su lengua y los presenta en forma clara y atractiva para favorecer los diferentes espacios de comunicación.

No se indica en ningún punto del currículo qué aspectos formales de la lengua deben enfatizarse: ¿Sintaxis? ¿Morfología? ¿Ortografía? ¿Longitud del texto? ¿Noción de audiencia? ¿Qué se entiende por “espacios de comunicación”?

En el área de matemáticas identificamos una confusa secuenciación de contenidos y de niveles de ejecución, como por ejemplo en temas de geometría del primer ciclo:

Competencia N° 23:

- Utiliza instrumentos no formales que le permiten comparar cuantitativamente formas sencillas y definir sus propiedades geométricas.

No se define exactamente qué instrumentos se pueden utilizar, ni qué formas o propiedades se deben conocer en este ciclo. La Competencia N° 31, por otro lado, se considera lograda si el alumno “construye y utiliza la noción de ángulo, perímetro, superficie y volumen a partir de la composición, descomposición y la comparación de figuras geométricas simples”. Observamos que la Competencia N° 31 ofrece información clave para una apropiada ejecución de la competencia N° 23, o sea las nociones o conceptos específicos a adquirir para poder comparar cuantitativamente las figuras en cuestión. Es decir, una lista de conceptos organizados y bien secuenciados, permitiría que las competencias pudieran desarrollarse en forma más ordenada y articulada.

Se observa en el currículo un fuerte énfasis en la educación bilingüe. Se dispone la provisión de la educación en una segunda lengua (originaria) para todos los alumnos, de manera de garantizar la adquisición de la lectoescritura y de herramientas efectivas de comunicación en ambas lenguas. También se ofrece una extensa explicación sobre la organización de los contenidos para favorecer la adquisición de competencias plurilingüísticas en el nuevo currículo. Sin embargo, no se observa en el diseño una organización interdisciplinaria de los contenidos de manera que las capacidades lingüísticas se desarrollen paralelamente a las cognitivas en las diferentes disciplinas, ni se aprecia una manifestación clara de la necesidad de adquirir o expresar saberes disciplinarios en más de una lengua. En otras palabras, no hay una interacción explícita entre el desarrollo del plurilingüismo y de la adquisición de saberes disciplinarios, tales como las matemáticas o las ciencias naturales. Se ofrecen algunas consideraciones teórico-prácticas sobre la educación bilingüe en la fundamentación del área de lenguaje, así como un cuadro explicativo sobre el uso de las lenguas en los módulos de aprendizaje, pero si consideramos la complejidad de los procesos de asimilación de una segunda lengua podemos anticipar que los alumnos enfrentarían serias dificultades para acceder al currículo oficial en más de un idioma.

Ejemplo: País 06. Ciencias Naturales, Lengua Castellana y Matemática. 2º Ciclo.

El currículo oficial de este país presenta los contenidos como “Indicadores de logros curriculares por conjuntos de grados para los distintos niveles de la educación formal”. Cada sección de indicadores agrupa ciclos de tres años, y cada ciclo está dividido en áreas curriculares. Los indicadores en cada área se podrían considerar como objetivos muy generales, como muestran los siguientes ejemplos:

Ciencias naturales en 2º ciclo:

- Narra sucesos ambientales apoyándose en esquemas explicativos coherentes.
- Planea y realiza experimentos para poner a prueba sus propias hipótesis, las de sus profesores y compañeros.
- Etc.

Lengua castellana en 2º ciclo:

- Establece semejanzas y diferencias entre los tipos de textos y actos comunicativos, en términos de los significados, las estructuras y los contextos.
- Propone planes textuales previos al acto de escritura.
- Etc.

Matemática en 2º ciclo:

- Explora y descubre propiedades interesantes y regularidades de los números, utiliza habitual y críticamente materiales y medios para verificar predicciones, realizar y comprobar cálculos y resolver problemas.
- Reconoce características de sólidos, figuras planas y líneas, los utiliza en su vida cotidiana en trabajos prácticos como mediciones, elaboración de dibujos y construcción de modelos.
- Etc.

Ninguna de las asignaturas en este programa presenta una clasificación de contenidos en bloques o ejes temáticos, ni tampoco se distingue entre conceptos, procesos y actitudes. Uno de los artículos del documento normativo analizado propone ciertos criterios para la elaboración institucional de «indicadores de logros específicos». Los criterios también son muy generales, tales como [tener en cuenta] “Los intereses, necesidades, expectativas y propuestas de la comunidad educativa” o “El devenir del conocimiento, de la ciencia y la tecnología, el ambiente y los cambios individuales, grupales y colectivos que se producen a nivel local, territorial, nacional y mundial”. De acuerdo al artículo 18^o del documento curricular principal, el diseño del “plan de estudios” con indicadores específicos queda en manos de cada centro educativo. Este documento, sin embargo, no proporciona conceptos y procesos específicos para el desarrollo de las competencias propuestas en cada ciclo.

Ejemplo: País 07. Contenidos Básicos Comunes para la Educación General Básica.

Los contenidos del currículo oficial en este caso están planteados como **competencias**. Las competencias educativas a desarrollar se dividen en varios ejes de formación: Ético, sociopolítico-comunitario, del conocimiento científico y tecnológico, y de la expresión y comunicación.

Cada área en educación general básica está organizada en tres ciclos. Cada asignatura está dividida en bloques temáticos regulares que se respetan a lo largo de los tres ciclos. Por ejemplo, en ciencias naturales la asignatura queda siempre dividida en *Estructura de la materia*, *Transformaciones y reacciones químicas* y *Recursos naturales y ambiente*. Los **contenidos** de cada bloque se clasifican sistemáticamente en **conceptuales** y **procedimentales**. Los contenidos **actitudinales** se presentan en la justificación del área al principio, pero se explica que no pueden detallarse en cada bloque. Las actitudes que se espera de los alumnos son válidas para todos los contenidos conceptuales y procedimentales del área.

Una primera descripción de los contenidos en ciencias naturales, por ejemplo, *narra* los contenidos conceptuales generales que deben tratarse en cada ciclo. También se establecen las expectativas de logro al final de la Educación General Básica, pero no se especifican por ciclo. Luego se detallan los contenidos conceptuales y procedimentales para cada ciclo.

Le corresponde a cada escuela desagregar los contenidos por grado, en trabajo de equipo con los docentes de área.

Los indicadores se ofrecen en cuadros de fácil lectura para el docente. Esto permite una visión más clara de la correspondencia entre conceptos y procedimientos, y de la cobertura que se hace de los diferentes temas en todos los ciclos de manera de facilitar una mejor articulación vertical de contenidos. Veamos un ejemplo:

Bloque 3 de Ciencias Naturales en el Segundo Ciclo: Estructura y cambios de la materia.

CONTENIDOS CONCEPTUALES	CONTENIDOS PROCEDIMENTALES
<ul style="list-style-type: none"> - Propiedades y composición del agua, el aire y el suelo. - Estructura atómica de la materia - Modelo atómico. - Elementos. - Noción de compuesto. Modelos sencillos de su conformación a partir de átomos. - Noción de molécula. - Soluciones, dispersiones y suspensiones. Métodos de separación. - La capacidad disolvente del agua. - Efectos de la temperatura en la solubilidad. Concentración, saturación. - Soluciones acuosas en la naturaleza deshielos, ríos, mares, agua potable, aguas duras y blandas. Usos y propiedades. - Acidez, alcalinidad y salinización del suelo. - Cambio de las propiedades del agua por agregado de solutos. - Suspensiones en gases humo, bruma, contaminación atmosférica. Estructura de la materia. 	<ul style="list-style-type: none"> - Elaboración de indicadores a partir de sustancias familiares (jugo de remolacha, repollo colorado). - Uso de indicadores para medir acidez y alcalinidad. - Observación y control de variables en el proceso de cristalización de sales en solución. - Comunicación de resultados mediante informes sencillos u otras técnicas.

Observamos que algunos conceptos son más específicos que otros. Cabe preguntarse, por ejemplo, quién define los usos y propiedades de las soluciones acuosas en la naturaleza. ¿El maestro o maestra, el libro de texto que se elija, o el directivo? También notamos que los procedimientos constituyen actividades; en vez de competencias complejas, y

que no se corresponden exactamente con todos los conceptos propuestos. En este sentido, sin embargo, se observan importantes diferencias entre las definiciones de procedimientos que hacen las distintas áreas curriculares. En lengua, por ejemplo, los procedimientos tienden a expresar las habilidades cognitivas y metacognitivas que se han de desarrollar en relación a los conceptos propuestos, aunque también se incluyan ciertas actividades de aula.

Bloque 2 de Lengua para el Segundo ciclo: Lectura

CONTENIDOS PROCEDIMENTALES

- Selección de estrategias de lectura específicas en función de propósitos determinados: búsqueda de datos, interpretación, análisis de recursos, resumen, etc.
- Escucha y oralización de lecturas. Cotejo de interpretaciones.
- Empleo de distintos soportes textuales en la lectura de textos en diversas situaciones escolares. Identificación de contexto y utilidad.
- Empleo de estrategias de verificación de la lectura: relectura, comentario grupal.
- Etc.

Ejemplo: País 08. Programa de estudios de Matemática. Segundo Ciclo.

Luego de hacer una fundamentación para el área y de algunas consideraciones metodológicas y de evaluación para el segundo ciclo en general, el programa de estudios analizado especifica los indicadores de logro por cada año escolar. El área se halla dividida en bloques temáticos similares a los del resto de la región. Los **objetivos** que se listan, 35 en total, constituyen los indicadores de logros para matemáticas en cuarto año. Luego, para cada eje o bloque temático se presentan los **contenidos**, los **procedimientos**, los **valores** y **actitudes** y los **criterios de evaluación**. Tomamos como ejemplo el bloque de Geometría.

Objetivos: (solo se transcriben los correspondientes a geometría)

1. Identifica los elementos de polígonos y los clasifica según el número de lados; compone y descompone polígonos, insistiendo en las características definitivas de los polígonos a través de su experiencia y trabajo.
2. Reconoce y caracteriza polígonos rectangulares e irregulares en objetos del entorno; los representa mediante dibujos, en el geoplano.
3. Reconoce y discrimina paralelogramos, trapecios y triángulos; compone y descompone figuras, los caracteriza a partir del número de ejes de simetría, lados, ángulos,... en figuras y objetos del entorno, realizando transformaciones de rotación.
4. Clasifica y reconoce diversas clases de triángulos, según la naturaleza de los ángulos internos y el número de lados iguales, aplica las nociones de simetría axial, identifica y construye y mide la altura mediante la composición y descomposición de triángulos.

Contenidos:

- Líneas poligonales abiertas y cerradas.
- Línea poligonal cerrada: borde e interior.
- Polígonos: cuadriláteros y triángulos, reconocimiento, clasificación y construcción.
- Polígonos irregulares y regulares.
- Polígonos regulares: características, clasificación: paralelogramos, no paralelogramos y construcción.
- Ejes de simetría en el rectángulo, en el cuadrado y realización de transformaciones de rotación.
- Etc.

Procedimientos:

- Identificación de líneas poligonales abiertas y cerradas, en representaciones gráficas, nombrando sus semejanzas y diferencias.
- Identificación de polígonos de tres y cuatro lados (triángulos y cuadriláteros), empleando la definición en figuras del entorno.

- Clasificación de polígonos en irregulares y regulares, nombrando sus características e identificándolos en representaciones gráficas.
- Clasificación de los cuadriláteros en paralelogramos y no paralelogramos, analizando sus características y sus diferencias, en figuras construidas para tal efecto (o por los estudiantes).
- Etc.

Valores y actitudes: El estudiante:

- Reflexiona sobre el valor de la exactitud del razonamiento para la obtención de mayores beneficios.
- Confronta las acciones que ejecuta con la realidad.
- Se autoexige y responsabiliza por los actos realizados; es capaz de enfrentar las dificultades y resolver situaciones problemáticas.
- Elige, de diversas estrategias, la que mejor se ajusta a la resolución de problemas de proceso, en forma objetiva.
- Etc.

Criterios de evaluación: El estudiante:

- Identifica y construye figuras geométricas de tres y cuatro lados mediante el doblado del papel, ejercicios en el geoplano y el manejo de reglas, compás y transportador.
- Identifica los ejes de simetría en triángulo, cuadriláteros; reconoce en los cuadriláteros los paralelogramos y no paralelogramos.
- Reconoce las figuras geométricas y experimenta, haciendo rotaciones de las partes de un rompecabezas, para obtener otra configuración o figura geométrica que es simétrica.
- Etc.

Notamos bajo cada título una definición más objetiva y más precisa de los resultados de aprendizaje esperados. Se aprecia una distinción más elaborada entre conceptos y procedimientos. La diferencia entre procedimientos y criterios de evaluación es menos aparente, excepto por el hecho de que los últimos ponen más énfasis en la acción del estudiante que en los contenidos conceptuales.

Ejemplo: País 09. Objetivos Fundamentales y Contenidos Mínimos Obligatorios de la Educación Básica.

Los contenidos mínimos del currículo oficial persiguen dos clases de objetivos:

- a) **Objetivos Fundamentales Transversales:** Aquellos que miran a la formación general del estudiante y que trascienden a un sector o subsector específico del currículo escolar.
- b) **Objetivos Fundamentales Verticales:** Aquellos que se dirigen específicamente al logro de competencias en determinados dominios del saber y del desarrollo personal. Se aplican a determinados cursos y niveles y su logro demanda aprendizajes y experiencias vinculadas a ámbitos disciplinarios específicos del currículo.

Para cada nivel de educación básica, la MCB (Matriz Curricular Básica) presenta la lista de **sectores** y **subsectores de aprendizaje** (que equivalen a asignaturas y áreas dentro de la asignatura), los **objetivos fundamentales** para cada asignatura y los **contenidos mínimos**.

En el sector de Lenguaje y Comunicación en 4° año, por ejemplo, no se distingue claramente la diferencia entre los objetivos y los contenidos. Los objetivos están expresados en forma de competencias complejas, mientras que los contenidos se expresan como actividades o procedimientos, aunque involucren conceptos determinados.

Objetivos fundamentales:

- Escribir textos legibles, en forma manuscrita, respetando aspectos léxicos, semánticos, ortográficos y gramaticales básicos.

Contenidos mínimos:

- Producción de textos: escritura de recados, invitaciones, saludos, canciones, adivinanzas, cuentos, poemas y otros textos breves y sencillos que expresen sus necesidades de interacción, sus vivencias, imaginación, humor y hechos de la vida.

En matemática de 5° año se observa una mayor diferenciación y mejor correspondencia entre objetivos fundamentales y contenidos mínimos.

Objetivos fundamentales:

- Distinguir elementos de un cuerpo geométrico y establecer correspondencias entre un cuerpo y su representación plana.

Contenidos mínimos:

- Figuras Geométricas: Diferenciar cuadrado, rombo, rectángulo y romboide a partir de modelos hechos con varillas articuladas. Identificar lados, vértices y ángulos en figuras poligonales. Distinguir tipos de ángulos, con referencia al ángulo recto.

La mayor parte de las competencias que se mencionan en matemáticas son procesos mentales de orden superior, como la clasificación o la comparación. Sin embargo, las competencias de aplicación a la vida real son escasas y no se mencionan objetos ni situaciones de la vida cotidiana.

En el sector ciencia, el currículo desagrega los subsectores por ciclos. Los objetivos fundamentales y contenidos mínimos en 5° año, al igual que en matemática, guardan una relación coherente.

Objetivos fundamentales:

- Identificar a las especies y poblaciones biológicas como niveles de organización de vegetales, animales y seres humanos.

Contenidos mínimos:

- El cuerpo humano como organización biológica: reconocer, en forma elemental, los aspectos óseos, anatómicos, nerviosos y sensoriales, del ser humano y apreciar la función que estos componentes desempeñan en la vida del hombre y en las interacciones de este con su ambiente.

Los documentos curriculares de ciencias naturales para el Quinto Año Básico proponen un tratamiento exhaustivo de cada unidad. Para ello, incorporan a los objetivos fundamentales y a los contenidos mínimos obligatorios los *aprendizajes esperados*, las *orientaciones didácticas*, *actividades de aprendizaje y ejemplos*, y *actividades de evaluación y ejemplos*. Las actividades de evaluación nos interesan en particular ya que proponen criterios de creciente complejidad:

Ciencias: Unidad 2: El cuerpo humano como un todo organizado: sistema nervioso, sentidos y movimiento. Ejemplos de preguntas para una prueba y criterios a considerar:

Criterio 1:

- Muestra conocimientos sobre los órganos que participan en un movimiento, su papel y las estructuras y procesos nerviosos involucrados.

Criterio 2:

- Razona, a partir de un caso de un movimiento, establece el origen de un estímulo, los órganos involucrados y representa la secuencia de los movimientos en un esquema.

Criterio 3:

- Determina, a partir de un caso relativo al aparato locomotor y la salud, los efectos del peso sobre el esqueleto y las precauciones que es necesario tomar para evitar efectos negativos.

Observamos que el primer criterio demanda solo habilidades de reconocimiento mientras que los siguientes demandan capacidad para establecer relaciones entre conceptos e inclusive entre disciplinas. Esta organización y secuenciación de contenidos y objetivos constituye uno de los pocos ejemplos en la región de indicadores de logros por niveles de ejecución¹⁰ y que por esa misma razón se aproxima al modelo de elaboración de estándares que se está utilizando en otras regiones, y que PREAL recomienda en su informe “El futuro esta en juego” (PREAL, 1998).

En esta sección intentamos ejemplificar algunos de los aspectos más débiles, así como algunas fortalezas, en el diseño de los currículos de la región. Una apropiada elaboración de estándares requiere de una estipu-

¹⁰ Los niveles de ejecución son los grados de complejidad en los contenidos y en las habilidades que se espera que los estudiantes demuestren. Una conducta que demuestre la adquisición de un saber conceptual y procedimental complejo puede especificarse en diferentes niveles de complejidad o ejecución, de acuerdo a los tiempos de exposición a los nuevos conceptos, a las relaciones entre conceptos, a los contextos de aplicación, a las relaciones con otras disciplinas, etc.

lación clara y acabada de los contenidos conceptuales y procedimentales a ser adquiridos por los estudiantes, apuntando a un equilibrio crítico entre unos y otros. No se pueden desarrollar competencias complejas sin garantizar el acceso a conceptos complejos. De la misma manera, los conceptos pierden valor si no se le brinda al estudiante la oportunidad de demostrar su competencia para resolver situaciones complejas. En este juego de palabras, que consideramos debería ser mucho más que un juego, radica el objetivo último de la noción de estándar: apuntar a contenidos complejos, a un equilibrio entre el saber y el saber hacer, y a una definición de niveles de ejecución o complejidad de las competencias que tengan como plataforma mínima un conjunto de saberes comparables a los que proponen los currículos de los países más industrializados del mundo.

3. IMPLEMENTACIÓN CURRICULAR

Por medio del análisis de los documentos curriculares y de las entrevistas realizadas para este estudio intentamos identificar los principales factores que afectan la implementación curricular en América Latina, según el criterio de quienes formulan el currículo intencional. Aunque el presente no es un estudio del currículo implementado (es decir, el currículo que se desarrolla entre docentes y estudiantes en el aula), interesa descubrir los criterios de los agentes encargados por el currículo intencional respecto de la implementación. Tales criterios, sin duda, contribuyen a explicar algunas de las características específicas de los instrumentos curriculares, puesto que se formulan con el propósito de guiar la implementación.

En este sentido, se observa que los técnicos encargados de elaborar los documentos oficiales han tenido en cuenta la necesidad de hacer explícitos los insumos y procesos críticos para una efectiva implementación curricular. En algunos casos se hacen extensas fundamentaciones teóricas para la transformación y que explican los factores a tener en cuenta durante la ejecución de los nuevos currículos. Otros países aportan herramientas de implementación y de gestión curricular más concretas y operacionales.

En primer lugar, existe un consenso a nivel regional sobre la necesidad de promover una mayor autonomía curricular. Esta autonomía viene acompañada de dos nociones comunes a varios países. Una es la de

«adecuación curricular» y la otra es la de «gestión curricular». La primera implica mayor libertad para que instituciones y docentes practiquen el diseño curricular en los centros educativos, respetando los contenidos mínimos de cada país pero adaptándolos a las necesidades regionales, locales y áulicas. La idea de *gestión curricular*, que se explica dentro del marco de relativa descentralización educativa en la región, enfatiza la necesidad de crear mecanismos de gestión dentro de las escuelas para que el currículo sea debatido y diseñado por diferentes actores. La gestión curricular deja de estar a cargo de organismos de control, como supervisores e inspectores, y pasa a ser responsabilidad de cuerpos consultivos institucionales, cuyos participantes varían de país en país, pero que generalmente incluyen a directivos, secretarios, docentes y padres. Se espera que dentro de este modelo de gestión curricular las escuelas consulten a otras organizaciones públicas y privadas sobre las necesidades locales que deben ser atendidas por el currículo. La herramienta más concreta para la efectiva gestión curricular que se menciona es la del Proyecto Educativo Institucional (PEI), también llamado Proyecto Curricular Institucional, o Proyecto Pedagógico Institucional. Los documentos relevados, sin embargo, no ofrecen una definición operacional de este recurso aunque sabemos que hay extensa bibliografía sobre los PEI en la región, y que en algunos países se viene aplicando desde hace varios años. Según se señala en las entrevistas realizadas, además de posibilitar la autonomía y adecuación curricular, el proyecto educativo institucional tiene como objetivo garantizar una mejor articulación horizontal y vertical de contenidos.

Las entrevistas revelan que esta búsqueda de equilibrio entre las normativas curriculares ministeriales que imponen contenidos mínimos y obligatorios para todo el país y las nuevas políticas de adecuación curricular necesariamente genera conflicto entre el poder central y el poder local e institucional. La ausencia de sistemas consolidados de evaluación y medición para todas las escuelas, y en consecuencia la falta de responsabilización institucional por los resultados académicos, implica que los centros educativos hacen uso de sus nuevas facultades para la recontextualización curricular sin priorizar la calidad mínima de los aprendizajes que se espera desde la administración central. Por otro lado, las escuelas manifiestan su disconformidad con las expectativas ministeriales en contextos donde los recursos materiales y humanos son inadecuados para una efectiva ejecución de los currículos. Este conflicto, sin embargo, se percibe en muchos casos como un proceso necesario y positivo dentro

del marco de democratización y descentralización que atraviesan los países de América Latina.

La autonomía en la gestión curricular trae aparejada una segunda dificultad crítica de implementación que se mencionó sistemáticamente en todas las entrevistas y se plantea en la mayor parte de los documentos: la formación y capacitación docente. El nuevo paradigma de aprendizaje por competencias, y que apunta a la formación de un ser humano integral, necesita de docentes preparados para elaborar y transmitir un currículo con definiciones menos acabadas sobre los productos de aprendizaje esperados que los que proponían modelos curriculares anteriores. Efectuar diseños curriculares en base a una lista de contenidos conceptuales y procedimentales es una tarea compleja para la cual gran parte de los docentes de América Latina, según el criterio de muchos expertos, no están preparados. Los sistemas educativos enfrentan entonces, una doble dificultad de implementación. La dificultad técnica del diseño, por ejemplo para la construcción de un PEI, de un proyecto de aula, o de una unidad de aprendizaje integrado, y con una reducción crítica en las *oportunidades de aprendizaje*¹¹ debido a métodos de enseñanza y recursos didácticos inadecuados (Valverde, 1997; Porter, 1993; McDonnell, 1995).

Ante esta situación, que se anticipa explícitamente en la mayor parte de los documentos oficiales, encontramos que todos los países proponen políticas de mejoramiento en la capacitación y formación docente. Estas políticas varían, desde promover el vínculo entre los institutos de formación docente y los centros universitarios e independientes de investigación educativa, hasta la creación de redes de capacitación locales y a distancia destinadas a fortalecer el proceso de implementación. Las entrevistas, sin embargo, revelan que estas políticas no son suficientes o efectivas. Entre los obstáculos específicos que se mencionan en las entrevistas encontramos que:

- a) El docente necesita ser «acompañado» a lo largo de todo el proceso de implementación curricular, pero la infraestructura actual no permite que esto suceda (falta personal que lidere los cambios, se vive «una gran soledad académica», etc.)

¹¹ Traducción del concepto en inglés de Opportunity-to-Learn (OTL).

- b) Los programas de formación docente, aun cuando son universitarios, están desactualizados y no brindan herramientas para educar dentro del nuevo modelo humanista-constructivista-reconstructivista.
- c) Existe oposición gremial e ideológica al cambio de paradigma. Como resultado, los docentes en algunos países se aferran intencionalmente al currículo derogado o solamente a los contenidos de los libros de texto.

Desde la percepción de los curriculistas y líderes de la transformación actual, el factor crítico indispensable para garantizar la oportunidad de aprendizaje reside en el cambio de la postura epistemológica del docente. Un maestro o maestra que reflexiona y comprende su nuevo rol en un modelo de enseñanza-aprendizaje centrado en las necesidades y los conocimientos previos del alumno puede lograr que el currículo oficial alcance a todos los estudiantes.

4. EVALUACIÓN DE DESEMPEÑO Y RENDIMIENTO

Todos los documentos curriculares analizados hacen referencia a la importancia de incorporar la evaluación como herramienta clave para el mejoramiento de la calidad educativa. Sin embargo, se observan diferentes nociones sobre qué significa evaluar y cómo llevarlo a cabo.

La mayor parte de las referencias a la evaluación proponen una evaluación formativa que atienda a los procesos de desarrollo del alumno, tanto en los aspectos cognitivos del aprendizaje como en los valores y actitudes demostradas. Los documentos que favorecen este tipo de evaluación no hablan de medición del rendimiento sino de observación del desempeño; los instrumentos de evaluación que se proponen son la observación de las actitudes de los alumnos, la participación de padres en el proceso, la autoevaluación del alumno, y el uso de “juicios valorativos” en vez de escalas numéricas o alfabéticas.

Los documentos curriculares y normativos que proponen la medición del rendimiento explican la necesidad de medir la calidad en términos de “lo que se sabe y se puede resolver”. Como instrumento principal para estas evaluaciones se mencionan las pruebas nacionales de rendimiento, generalmente aplicadas a dos o tres grados de la educación básica en las áreas de lenguaje, matemáticas y ciencias naturales. Algunos países han incorporado las áreas de estudios sociales y lenguas extranjeras. Un in-

forme reciente del Banco Mundial indica que 14 países de América Latina y el Caribe están aplicando o han aplicado pruebas nacionales muestrales solamente, 3 países aplican pruebas censales solamente, y otros tres han experimentado con sistemas mixtos de pruebas muestrales y censales (Rojas y Esquivel, 1998).

Es importante destacar que en ningún caso se mencionan referentes externos al sistema como indicadores de logro, tales como exámenes internacionales estandarizados o estudios comparados de rendimiento. Los países que actualmente administran pruebas nacionales tampoco ofrecen estas pruebas como estándares o modelos para que puedan ser utilizados por las instituciones que apunten a mejorar la calidad de los aprendizajes en años subsiguientes. Tampoco se hallan referencias a ningún tipo de “responsabilización” de los distintos actores e instituciones del sistema educativo por el rendimiento académico de sus estudiantes. El mejoramiento de la calidad educativa en cada institución se plantea implícitamente como una cuestión de voluntad y ética profesional de docentes y directivos, pero sin parámetros externos de desempeño y rendimiento que le den ímpetu y dirección a la gestión curricular dentro de cada centro educacional.

Cabe agregar a modo de conclusión parcial que, en general, el diseño curricular y la definición de indicadores de logro en América Latina no constituyen instrumentos sólidos sobre los cuales se puedan proyectar sistemas nacionales de evaluación como los que proponen la mayor parte de los países de la región. Si bien esto es natural, ya que los contenidos curriculares en muchos países industrializados se han perfeccionado a partir de la introducción de pruebas estandarizadas, notamos que en América Latina existe cierto desfase entre los objetivos de alcance del currículo y lo que las evaluaciones se proponen medir. Esto puede deberse, según lo expresado por diferentes actores entrevistados, a un quiebre en la comunicación entre los agentes encargados de la evaluación y aquellos que se ocupan del desarrollo y diseño curricular en cada país.

5. ESTÁNDARES EDUCATIVOS

Las entrevistas realizadas para este informe incluyeron una pregunta específica sobre la noción de estándares educativos en cada país, y que apuntaba a revelar el estado de la discusión en la región sobre la necesi-

dad de establecer estos estándares. Los países de América Central, por estar participando en el Proyecto de Establecimiento de Estándares para la Educación Primaria en Centroamérica, centraron sus respuestas en el estado de la elaboración de estos estándares en particular. El proyecto mencionado consiste en un proceso técnico de elaboración y validación de estándares de contenido, desempeño y ejecución en las áreas de matemática, español y ciencias naturales para los grados tercero y sexto del primer nivel en los seis países del istmo. Esta “regionalización” de la educación en Centroamérica es percibida de diferentes maneras por los técnicos curriculares de cada país. En un caso, esta regionalización se percibe como una consecuencia lógica de la globalización, y que tendrá un efecto positivo en virtud de las problemáticas similares que atraviesan los sistemas educativos de la región. En otro caso, se considera que los estándares serán difícilmente implementados debido a las grandes *diferencias* en las problemáticas políticas y económicas que enfrenta cada país. Por otro lado, se remarcó que el proceso de elaboración e implementación de estándares dentro de cada sistema debería ser relativamente sencillo debido al tamaño de los países en Centroamérica. Uno de los técnicos entrevistados atribuye la “desestandarización” de la educación en su país a las diferencias de infraestructura y recursos entre distritos, y no a las diferencias en los contenidos curriculares. Los procesos de elaboración de estándares también varían en cada país dependiendo del grupo de técnicos encargados de la tarea, y el contexto político en que se desenvuelve. En El Salvador, por ejemplo, el equipo de elaboración de estándares es el mismo que tuvo a su cargo el diseño del currículo oficial, y se espera que esto permita mayor continuidad en los procesos de transformación curricular. En Panamá, por otro lado, los equipos técnicos a cargo del diseño curricular oficial y de estándares son diferentes, por lo cual se reporta que los contenidos de los nuevos estándares pasarán por una instancia más de validación. La entrevista con Panamá también indica que el término de «estándar» suele ser rechazado pues al centrarse en los aspectos exclusivamente académicos de la educación se lo percibe en oposición a los principios constructivistas y socio-reconstructivistas que inspiran el currículo oficial.

En América del Sur no tenemos referencias sobre políticas o proyectos concretos de elaboración de estándares, con excepción de Chile, cuya Ley Orgánica Constitucional de Enseñanza asigna al Ministerio de Educación y Cultura cuatro instrumentos de gestión curricular. Uno de estos instrumentos es la elaboración de los “requisitos mínimos de egreso”,

que son considerados estándares por los oficiales ministeriales aunque no se los denomine así en la Ley, y otro es la evaluación periódica de los resultados de aprendizaje. De acuerdo a la entrevista realizada a un funcionario del Ministerio de Educación, en este momento se encuentra en su fase final el desarrollo de planes y programas de estudio, y la siguiente tarea será la de producir estándares. En la actualidad no se han tomado decisiones acerca de cómo se procederá para escribirlos. Uruguay, por su parte, tiene exámenes de egreso para que los alumnos de las escuelas privadas “autorizadas” puedan validar su diploma en base a los contenidos del currículo oficial. Estos exámenes, sin embargo, están restringidos a un uso específico y no se ha discutido la posibilidad de tomarlos como modelo y hacerlos extensivos al resto del sistema educativo.

Argentina está elaborando estándares en el ámbito universitario solamente, y dentro del contexto del Mercosur, muy posiblemente a efectos de compatibilizar programas y requisitos de habilitación profesional. La entrevista con técnicos en Colombia revela que el término de “estándar” no tiene buena recepción por ser vocablo extranjero, aunque se considera que las consultas nacionales sobre las expectativas de la población constituyen un paso importante hacia la definición de estándares. En Venezuela, por último, se están haciendo intentos por establecer estándares que permitan una mejor medición de la calidad al final de cada etapa de la educación. La entrevista indica que la elaboración de estándares “es muy importante para la política nacional, no solamente para la calidad de las instituciones”.

Los documentos curriculares que se analizaron para este informe no hacen referencia alguna al concepto de «estándar» educacional o curricular. No se utiliza el término para designar partes del currículo ni tampoco para explicar los criterios de promoción y graduación de alumnos. Las fundamentaciones del currículo en cada país expresan la necesidad de elevar la calidad de los aprendizajes y se han realizado esfuerzos importantes por mejorar el nivel y la organización de los conceptos y competencias a desarrollar. Los antiguos programas de estudio que ofrecían meros listados de datos que debían adquirirse únicamente como conocimiento declarativo han sido reemplazados por currículos que contemplan la necesidad de educar individuos para una sociedad compleja. Sin embargo, observamos que los sistemas educativos donde se implantan estos currículos no ofrecen oportunidades para medir y comparar resultados en base a parámetros claros de desempeño y rendimiento. Aunque existen contenidos mínimos a alcanzar, los criterios de promoción y graduación

son establecidos por cada escuela. Es decir, la escuela determina los niveles de ejecución y cuáles se consideran apropiados para promover a los estudiantes a un nivel superior. Esto implica grandes diferencias entre escuela y escuela, o entre regiones o provincias, ya que el nivel de exigencia queda a discreción de los docentes y directivos. Por oportunidades de medición y comparación nos referimos tanto a los sistemas de evaluación nacionales como a las instancias de evaluación que regulan el acceso a los niveles superiores de educación. El acceso a las universidades en América Latina, por ejemplo, aun en los casos en que está condicionado a un examen de ingreso, permanece desvinculado al desempeño y rendimiento de los estudiantes durante sus estudios primarios y secundarios. Esto implica que la función propedéutica de la escuela queda delineada por la motivación y por la visión o ideario de cada centro educativo pero que no se valida con ningún parámetro externo a la institución.

Esta necesidad de validación con referencias externas se refleja en la educación privada y de élite, que desde hace años recurre a la utilización de exámenes internacionales, como el International Baccalaureate o el IGCSE de Cambridge University, para determinar metas y orientar los procesos pedagógicos. Estos exámenes ofrecen estándares de desempeño distintos y posiblemente más exigentes que los requeridos por los currículos oficiales de América Latina, y por lo tanto cubren las expectativas de aquellas instituciones que aspiran a ofrecer una educación de excelencia. El mismo fenómeno se observa con los colegios públicos universitarios, que disponen de ciertos estándares de ingreso y cuyos programas de estudio proponen niveles de alcance muy superiores a los de los colegios públicos comunes. Estos centros buscan permanentemente nuevas oportunidades de cotejar sus logros con instituciones internacionales y de participar en eventos competitivos y de intercambio académico. Las fuentes consultadas para la elaboración de este informe no indican ninguna tendencia en la educación pública de América Latina hacia la competitividad y la excelencia académica comparables a la que manifiestan las escuelas de élite y colegios universitarios de la región. La noción misma de 'estándar' educativo ha surgido de contextos políticos europeos y norteamericanos en donde la competitividad (internacional y/o intranacional) y excelencia académica son temas políticos de importancia. La relativa ausencia de estos elementos en el debate político en nuestra región quizás explique la también relativa ausencia de una preocupación mayor por la formulación de estándares educativos.

SUMARIO Y CONCLUSIÓN

Las reformas educativas emprendidas por América Latina en la última década responden a una necesidad imperiosa de brindar a la población las herramientas necesarias para construir naciones fuertes, política y económicamente, y para lograr que el desarrollo se haga sostenible a largo plazo. Estas reformas no solamente plantean la necesidad de garantizar el acceso y la permanencia de la totalidad de la población en edad escolar básica en el sistema educativo, sino que aspiran a mejorar la calidad de los aprendizajes a través de un currículo renovado y dinámico.

Esta voluntad de transformación curricular es un paso importante en el camino hacia la modernización y el mejoramiento de la educación pero que debe reafirmarse en forma permanente con la práctica democrática y de auténtico profesionalismo que los actuales procesos de reforma demandan. Para este informe recurrimos a la noción de estándar educativo como medida de comparación para validar nuestras apreciaciones sobre los currículos de la región. En estas apreciaciones hemos tomado como referencia las características que consideramos esenciales para el desarrollo de currículos que permitan a nuestras naciones alcanzar las grandes metas de transformación social que se han propuesto:

- La validación técnica y política de los contenidos curriculares;
- La calidad de los contenidos y del diseño del currículo como instrumento vital de comunicación entre las agencias de transformación educativa y los diferentes actores a cargo de la implementación;
- La evaluación sistemática y sistémica de la calidad de los aprendizajes.

Los procesos de consulta popular durante la elaboración de los currículos para las diferentes reformas constituyen un avance importante hacia la validación de los contenidos en la región. Consideramos, sin embargo, que esta validación debe consolidarse como proceso permanente, donde la sociedad civil y los diferentes actores del sistema educativo participen en forma dinámica y sostenida para garantizar la renovación y mejoramiento de los contenidos curriculares. Cabe recordar que esta validación es un proceso técnico, en tanto requiere de la participación de especialistas, a la vez que es político y democrático y que por lo tanto requiere de la participación de la sociedad toda en un diálogo donde estén representados los valores e intereses de todos los sectores de la población.

Los documentos curriculares de la región presentan una amplia variedad de modelos, y existen grandes diferencias en la calidad del diseño y de los contenidos propuestos por diferentes países. Encontramos que los países que ponen mayor énfasis en los aspectos axiológicos del currículo tienden a restarle espacio a los contenidos conceptuales y procedimentales o a descuidar la organización y secuenciación de los contenidos. Los países que han priorizado los contenidos conceptuales y procedimentales, por otro lado, proponen los contenidos actitudinales como complementarios a los primeros. En estos casos, los criterios de organización del currículo son más claros para el público y por eso constituyen un instrumento de comunicación más efectivo. Finalmente, podemos concluir que solo una pequeña minoría del total de países considerados en este estudio presenta una definición precisa de todos los conceptos y procedimientos a desarrollar en cada año del sistema escolar, así como de los niveles posibles de complejidad a tener en cuenta durante los procesos de entrega y evaluación del currículo.

Por último, consideramos brevemente el estado actual de la evaluación de los sistemas educativos. Observamos que los sistemas nacionales de evaluación no se han constituido aún en una fuente sólida de información y retroalimentación para el mejoramiento de la calidad educativa, y que no existe un vínculo lo suficientemente estrecho entre los agentes de desarrollo curricular y los agentes de evaluación y medición de la calidad de la educación. La elaboración de estándares educativos apuntaría, precisamente, a que estas dos actividades de desarrollo y evaluación curricular se reforzaran mutuamente.

La Comisión Internacional sobre Educación, Equidad y Competitividad Económica (PREAL, 1998), como mencionamos en nuestra introducción, hace cuatro recomendaciones fundamentales para el mejoramiento de la calidad educativa en América Latina. En este informe nos hemos limitado casi exclusivamente al análisis de estado del currículo prescrito en la región sobre la base de la primera recomendación de PREAL sobre la necesidad de elaborar estándares educativos. Aunque con menor énfasis, también tratamos aspectos de las otras recomendaciones realizadas por PREAL, más precisamente aquellas relacionadas con la responsabilización local e institucional y de la necesidad de fortalecer la formación y la capacitación docente. Debido a restricciones de tiempo, y porque escapa al tratamiento sobre contenidos curriculares, optamos por no discutir en este informe la cuarta recomendación de la Comisión sobre la necesidad de aumentar la inversión por alumno en la educación

básica. Consideramos, sin embargo, que ninguna de estas recomendaciones puede hacerse efectiva sin la provisión de las otras tres, y que habernos concentrado en el tratamiento de la primera recomendación sobre estándares no implica que esta sea más importante o urgente que las demás. Simplemente esperamos haber contribuido a ilustrar unos de los aspectos más ignorados en el debate regional sobre la calidad de la educación: la definición clara y operacional de las competencias que han de desarrollarse en nuestros estudiantes, que permita una evaluación sistémica y efectiva de la calidad educativa y la consecuente apropiación de los resultados y de los mecanismos de ajuste necesarios por parte de la sociedad entera.

APÉNDICE A: Documentos normativos y curriculares analizados

Argentina

- Ministerio de Cultura y Educación (1993). *Ley Federal de Educación N° 24.195*.
- Contenidos Básicos Comunes de Lengua, Matemática y Ciencias Naturales para Educación General Básica (www.mcye.gov.ar)
- Ministerio de Educación, Pcia. de Buenos Aires. *Módulo de Capacitación Docente 7-8*.
- Ministerio de Educación y Cultura de la Nación (1996). *Contenidos Básicos Comunes para la Educación General Básica*, Primera Edición.

Bolivia

- Secretaría Nacional de Educación, Ministerio de Desarrollo Humano (1994). *Ley de la Reforma Educativa*, en En la Hora de la Reforma Educativa, Fascículo VI.
- Ley de Reforma Educativa, Decreto Supremo N 23950, Reglamento sobre Organización Curricular, 1 de febrero de 1995. (www.educa-bol.com/leyes.htm)
- Secretaría Nacional de Educación, Ministerio de Desarrollo Humano (1995). *Nuevos Programas de Estudio de la Reforma Educativa*. Versión Preliminar Completa.

Brasil

- Parámetros Curriculares Nacionais de 1ª a 4ª Séries y Parámetros Curriculares Nacionais de 5ª a 8ª Séries. (www.mec.gov.br/sef/param/param/htm)

Chile

- Ministerio de Educación (1990). *Ley Orgánica Constitucional de Enseñanza N° 18.962*.
- Ministerio de Educación. Unidad de Currículo y Evaluación (1999). *Modificación del Dto. 40: Objetivos Fundamentales y*

Contenidos Mínimos Obligatorios de la Educación Básica. Aprobado por el Consejo Superior de Educación. Decreto en trámite.

- Programa de Estudio Quinto Año Básico INB3: Estudio y Comprensión de la Naturaleza.
(www.mineduc.cl)

Colombia

- Un decálogo para iniciar el debate. Plan Decenal de Educación. (www.icfes.gov.co/mineducacion/p112-3.html)
- Ministerio de Educación Nacional. Dirección General de Educación (1994). *Lineamientos Generales de Procesos Curriculares: Hacia la construcción de comunidades educativas autónomas* - Documento 1.
- Ministerio de Educación Nacional (1996). *Resolución Número 2343 de junio 5 de 1996. Por lo cual se adopta un diseño de lineamientos generales de los procesos curriculares del servicio público educativo, y se establecen los indicadores de logros curriculares para la educación formal.*

Costa Rica

- Ministerio de Educación Pública (1996). *Programa de Estudios de Matemática. 2º Ciclo.*

Ecuador

- Ministerio de Educación y Cultura. Consejo Nacional de Educación (1996). *Reforma Curricular para la Educación Básica 1996.*

El Salvador

- Ministerio de Educación (1995). *Reforma Educativa en Marcha: Documento III Lineamientos del Plan Decenal de Educación 1995-2005.*
- Ministerio de Educación. *Dominios Curriculares Básicos: Educación Parvularia, Básica y Media.*
- Ministerio de Educación. *Programa de Estudios de Tercer Grado.*

Guatemala

- Ministerio de Educación (1995). *Curriculum Educación en Población para el Nivel Preprimario y Primario*. Proyecto GUN92/P03.
- SIMAC (Sistema de Mejoramiento de los Recursos Humanos y Adecuación Curricular) (1989). *Revista de Educación: Guías del CEF (Ciclo de Educación Fundamental. Institucionalizado por Acuerdo Gubernativo 470/89)*.
- *Revista de Educación: Guías Curriculares del CEC (Ciclo de Educación Complementaria)*.

Honduras

- Secretaría de Educación Pública (1993). *Rendimientos Básicos e Indicadores de Evaluación del Nivel Primario*. Acuerdo N 5068-SEP-93.

México

- *Plan y Programas de Estudio para la Educación Primaria*. Primera edición. 1993. (<http://www.sep.gob.mx/rebn.planes/pri/pri2.htm>)

Nicaragua

- *Programas de Español y Matemática de Tercer Grado. Serie: Enseñemos y aprendamos activamente*. 1996.
- *Programas de Español y Matemática: Sexto Grado. Serie: Enseñemos y aprendamos activamente*. 1998.

Panamá

- ME-BID (1999). *Escuela Siglo 21: Revista del Proyecto de Desarrollo Educativo*. Año II, N° 3.
- Ministerio de Educación Oficial (1996). *Programa de Educación Primaria para el Sexto Grado*.
- *Programa de Educación Básica General para las Escuelas de la República. Programa de Sexto Grado*. 1999.

Perú

- Ministerio de Educación (1998). *Nueva Estructura Curricular: Lineamientos generales de planes de estudio.*
- Ministerio de Educación. *Estructura Curricular Básica de Educación Primaria. Segundo Ciclo.* (Sobre Resolución Ministerial N° 201-98-DE. Dirección Nacional de Educación Inicial y Primaria. Ministerio de Educación.)
- Ministerio de Educación. *Programa Curricular Básico de Educación Primaria de Menores. 1er Ciclo.* Ediciones Abedul.

República Dominicana

- Secretaría de Estado de Educación, Bellas Artes y Culto (1994). *Plan Decenal de Educación en Acción. ¿Por qué? ¿Para qué? De la Transformación Curricular.*
- Secretaría de Estado de Educación, Bellas Artes y Culto (1994). *Fundamentos del Currículum: Tomo I. Fundamentación Teórico-Metodológica.*
- Secretaría de Estado de Educación, Bellas Artes y Culto (1994). *Fundamentos del Currículum: Tomo II. Naturaleza de las Áreas y Ejes Transversales.*

Uruguay

- Administración Nacional de Educación Nacional. Consejo de Educación Primaria (1995). *Programa de Educación Primaria para las Escuelas Urbanas.* Revisión 1986. Contenido.

APÉNDICE B: Países en que se realizaron entrevistas

Argentina	Honduras
Chile	Nicaragua
Colombia	México
Costa Rica	Panamá
Ecuador	Uruguay
El Salvador	Venezuela
Guatemala	

REFERENCIAS

- Cariola, L. et al.** (1994). *La educación media en el mundo: Estructura y diseño curricular en diferentes países*. Ministerio de Educación de Chile.
- McDonnell, L.M.** (1995). Opportunity to Learn as a Research Concept and a Policy Instrument. *Educational Evaluation and Policy Analysis*, 17, 3, 305-322.
- Marzano, R. & Kendall J.** (1997). *The Fail and Rise of Standards-Based Education*. Mid-continent Research for Education and Learning (www.mcrel.org)
- Pinto Cueto, L.** (1999). *Currículo por competencias: Necesidad de una nueva escuela*. Alerta: Revista de educación y cultura, 43,10-17.
- Porter, A.C.** (1993). School delivery standards. *Educational Researcher*, June-July 1993, PP. 24-30.
- Gajardo, M.** (1999). *Reformas educativas en América Latina: Balance de una década*. PREAL Documentos, N° 15.
- Rojas, C. & Esquivel, J.M.** (1998). *Los sistemas de medición de logro académico en América Latina*. The World Bank: Latin America and the Caribbean Regional Office, LCSHD series, 25.
- Valverde, G. A.** (1997). Evaluación y Estándares de Currículo en una Era de Reforma Educativa. en B. Álvarez & M. Ruiz-Casares (Eds.), *Evaluación y Reforma Educativa: Opciones de política*. Washington DC: U.S. Agency for International Development.

¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?(*)

PEDRO RAVELA

INTRODUCCIÓN

En 1999 el Grupo de Trabajo sobre Estándares y Evaluación (GTEE) del PREAL produjo un primer documento de análisis sobre la situación de los sistemas nacionales de evaluación de aprendizajes en América Latina, el cual propuso algunos pasos claves para avanzar en el desarrollo de los mismos¹. Aquel trabajo fue producido en el marco de un taller de trabajo realizado en Lima en agosto de 1999, recurriendo básicamente al conocimiento directo de los participantes sobre la experiencia de diversos países en materia de evaluación de aprendizajes a nivel nacional.

El presente documento constituye, en cierto modo, una profundización de aquel análisis a partir del estudio sistemático de un amplio es-

(*) Este documento fue originalmente publicado por PREAL en la serie PREAL documentos N° 22, febrero de 2002.

¹ RAVELA, P., WOLFE, R. *et al.*, (2000). Los próximos pasos: ¿Hacia dónde y cómo vamos en la evaluación...

pectro de Informes Nacionales producidos en la región en el período 1995-2000. Fueron analizados un total de casi 70 informes².

El análisis del estudio se centró en las siguientes preguntas:

- ¿Qué tipo de información producen y difunden los sistemas de evaluación?
- ¿Cómo la presentan?
- ¿A quiénes está dirigida?
- ¿Con qué propósitos?
- ¿Cómo se espera que la información sea empleada?
- ¿Sirve la información efectivamente para esos propósitos?
- ¿Cómo mejorar nuestros modos de reportar los resultados?

El documento ha sido organizado en cinco capítulos. En el primero se analizan los objetivos de las evaluaciones y de los reportes de resultados que aparecen explícitamente señalados en los Informes Nacionales. Se discute la disyuntiva principal en términos de la política de evaluación, que consiste en decidir si las pruebas deben tener algún tipo de consecuencias directas para las escuelas o si deben ser, principalmente, de carácter formativo³.

El segundo capítulo ofrece una visión sistemática sobre las “métricas” o tipos de datos empleados en la región para el reporte de los resultados, tales como porcentaje de respuestas correctas, puntajes o porcentajes de alumnos que dominan ciertos objetivos. Muchas veces el lector no especializado no logra diferenciar el significado de unas y otras. Se ilustra y explica el significado de cada una de ellas y se discute qué tipo de enfoque es más adecuado para la pregunta principal que, en principio, deberían ayudar a responder los sistemas de evaluación: ¿han aprendido nuestros alumnos lo que se supone deberían haber aprendido?

En el tercer capítulo se analiza una de las principales debilidades en los informes de resultados, que es la falta de adecuada atención a las diferencias socioculturales existentes entre las poblaciones atendidas por los distintos sectores del sistema educativo o los distintos tipos de escuelas. Los alumnos llegan a la escuela con fuertes diferencias en cuanto a

² Por razones de espacio no se incluye en esta “versión abreviada” la nómina completa de informes analizados. La misma puede ser consultada en la versión completa, en la página WEB de PREAL.

³ Lo que en la literatura anglosajona se denomina con la expresión *highstakes vs. lowstakes*.

su dominio del lenguaje estándar y capacidades simbólicas, y tienen durante su escolarización diferentes condiciones de vida y niveles de apoyo en el hogar en relación a las tareas, todo lo cual favorece o dificulta la labor de la escuela. Por tanto, reportar resultados ignorando estas diferencias puede dar lugar a conclusiones groseramente inapropiadas. Tal vez la principal de ellas sea que las escuelas privadas son mejores que las públicas.

Entre los principales propósitos explícitos en los informes de resultados se encuentran dos: entregar a los docentes información relevante para mejorar la enseñanza; y aportar a las autoridades información útil para la toma de decisiones y formulación de políticas. El cuarto capítulo analiza cuatro tipos de estrategias desarrolladas para el primero de estos propósitos, en tanto el quinto capítulo está dedicado al segundo de ellos. En ambos casos se analizan las fortalezas y debilidades de la información contenida en los informes nacionales en relación a su potencial uso por los destinatarios definidos.

Cada ser humano es hijo de su circunstancia, del contexto en el que vive, e inevitablemente su visión del mundo está sesgada por esa circunstancia. En este sentido, deseo asumir explícitamente que la perspectiva desde la que este documento ha sido escrito es doblemente parcial. Por un lado, porque veo los informes nacionales desde el trasfondo de la experiencia que hemos realizado en Uruguay y desde las prioridades que aquí asumimos para nuestro sistema de evaluación. Por otro lado, porque veo la experiencia de otros países de manera parcializada, a partir de la lectura de algunos informes. Por tanto, pido disculpas de antemano por las omisiones o errores de perspectiva en que seguramente habré incurrido al analizar los informes. También debo expresar que el trabajo ha sido realizado básicamente sobre informes publicados hasta el año 1999 y que varios países han iniciado cambios de enfoque desde entonces, que no son contemplados en este documento.

De todos modos, deseo enfatizar que las referencias críticas a ciertos aspectos de los informes nacionales no tienen como finalidad descalificar ni enjuiciar el trabajo realizado por las unidades aludidas, sino simplemente ilustrar debilidades y dificultades que en buena medida compartimos todos los países de la región y cuya superación debemos asumir como desafío colectivo. El sentido de este trabajo es aportar a la reflexión y a la acumulación de conocimiento en la región sobre cómo mejorar nuestros sistemas de evaluación de aprendizajes y nuestros modos de reportar los resultados, de modo que tengan un impacto real en el

mejoramiento de los aprendizajes de los niños, en las prácticas de enseñanza de los maestros y en la toma de decisiones de política educativa.

Finalmente, deseo agradecer especialmente a Santiago Cueto, Richard Wolfe, Gilbert Valverde y Patricia Arregui por sus enriquecedores comentarios a la versión preliminar de este documento.

CAPÍTULO I

¿QUÉ SE PROPONEN LOGRAR LOS REPORTES NACIONALES?

Una primera mirada a los informes de resultados de las evaluaciones nacionales consiste en analizar cómo se presentan a sí mismos y qué propósitos dicen perseguir tanto los reportes como, más en general, el sistema de evaluación de aprendizajes.

En la literatura sobre evaluación de las instituciones y de los sistemas educativos se suele distinguir entre:

- Evaluaciones cuyos resultados tendrán consecuencias directas para individuos o instituciones y, por tanto, implican un cierto nivel de riesgo o amenaza porque conllevan algún tipo de sanción positiva o negativa.
- Evaluaciones cuyos resultados tienen una función únicamente informativa y formativa, pero no entrañan consecuencias.

A las primeras se las denomina “high stakes” y a las segundas “low-stakes” (MESSICK, 1999), lo que podría traducirse como alto y bajo riesgo, respectivamente. La figura 1 presenta esquemáticamente una tipología de los modos en que los resultados de las evaluaciones pueden ser utilizados con el propósito de mejorar la calidad de los sistemas educativos.

FIGURA 1
Tipología de usos posibles de las evaluaciones nacionales para mejorar el sistema educativo

BAJO RIESGO	ALTO RIESGO
Información al público sobre el estado del sistema educativo a través de la presentación periódica de resultados generales: nacionales o en grandes desagregados tipo urbano/rural o por regiones o provincias.	Exámenes de acreditación al finalizar cierto nivel educativo o pruebas de selección de los alumnos que desean ingresar al nivel terciario.
Aporte a la definición de las políticas educativas a través de la provisión de información a los tomadores de decisiones en niveles superiores e intermedios.	Establecimiento de incentivos económicos para las escuelas o los docentes, a partir de los resultados de sus alumnos en las pruebas y otros indicadores.
Aporte al mejoramiento de las escuelas y al desarrollo profesional de los docentes a través de la devolución de resultados por escuela y la producción de materiales de carácter didáctico.	Publicación de los resultados de cada establecimiento en la prensa con la finalidad de que las familias puedan juzgar la calidad de las escuelas y elegir la que desean para sus hijos.

El análisis de los objetivos y propósitos incluidos explícitamente en los informes nacionales permite constatar que **la mayoría de los sistemas nacionales de evaluación desarrollados en América Latina durante los 90' enfatizan sus finalidades de carácter formativo y sin consecuencias directas sobre las escuelas.**

De dicho análisis se desprende que existen básicamente dos grandes tipos de roles explícitos para los sistemas nacionales de evaluación:

- constituirse en sistemas de información al público sobre el estado de la educación nacional;
- constituirse en proveedores de información para la toma de decisiones por parte de actores específicos –autoridades, supervisores, directivos, docentes, familias–.

En el recuadro 1 se consigna a título ilustrativo algunas de las definiciones tomadas de los informes, que representan el tono que caracteriza a la gran mayoría de ellos.

RECUADRO 1

Finalidades y propósitos de los sistemas nacionales de evaluación

BRASIL - Sistema de Avaliação da Educação Básica (SAEB)

“Los datos del SAEB permiten:

- inferir los conocimientos y habilidades de los alumnos en las áreas y disciplinas investigadas, mediante la aplicación de tests, no con la intención de “evaluar” al alumno, sino con el objetivo de identificar lo que el alumno sabe o es capaz de hacer en los distintos momentos de su trayectoria escolar (proficiencia), con la finalidad de ponderar la calidad y la equidad de la enseñanza brindada;
- analizar la efectividad de los currículos en función de los resultados de aprendizaje de los alumnos;
- verificar los factores contextuales y escolares que inciden en la calidad de la enseñanza básica ofrecida en el país, utilizando cuestionarios que colectan datos sobre las características infraestructurales y de los recursos de que dispone la unidad escolar, el perfil del director y los mecanismos de gestión escolar, el perfil del profesor, de su práctica pedagógica, de las características socioculturales y de los hábitos de estudio del alumno. El análisis de los resultados de los relevamientos del SAEB posibilitan acompañar la evolución del desempeño de los alumnos y de los diversos factores incidentes en la calidad y en la efectividad de la enseñanza brindada en el sistema educativo, posibilitando la definición de acciones dirigidas a la corrección de las distorsiones identificadas y del perfeccionamiento de las prácticas y de los resultados presentados por las escuelas y por el sistema de enseñanza brasileño. Esas informaciones son utilizadas por gestores y administradores de la educación, investigadores y profesores”.

MEC-INEP Bases para um Ensino de Qualidade. Brasil, 1999.

CHILE - Sistema de Medición de Calidad de la Educación (SIMCE)

“El SIMCE mide para mejorar la calidad de la educación. Con este propósito, entrega información acerca de los resultados de algunos aspectos del proceso educativo, tanto a nivel de establecimiento como provincial, regional y nacional, para su análisis, interpretación y, para la formulación de planes de acciones remediales eficientes, concretas y evaluables (...) El análisis conjunto, unido a la experiencia profesional y al conocimiento que los profesores tienen de su Unidad Educativa, harían de los resultados obtenidos una valiosa fuente de información, que lleve a formular planes de acciones remediales más efectivos”.

Ministerio de Educación, República de Chile. Boletín Informativo N° 5 para Profesores. Educación Básica. Análisis e Interpretación de Resultados 1996.

ECUADOR - Sistema Nacional de Medición de Logros Académicos (APRENDO)

“El informe técnico “APRENDO 97” puede servir a una variedad de usuarios, según el interés y actividad de los mismos: Los líderes y gerentes del sistema educativo pueden emplearlo como una fuente de información confiable para la toma de decisiones a nivel de la gestión financiera, administrativa y curricular. Los maestros y maestras pueden utilizarla información como insumo en los procesos de microplanificación curricular y de evaluación del aprendizaje. Los padres de familia y la sociedad civil en general pueden servirse del informe para conocer y comprender mejor la realidad del sistema educativo ecuatoriano. Finalmente, la comunidad académica puede encontrar en el informe técnico hipótesis y datos valiosos para emprender proyectos de investigación aplicada en el campo de la educación”.

Ministerio de Educación y Cultura del Ecuador - EB/PRODEC, 1998.

I.1. Los sistemas de evaluación como sistemas de información al público en general

La primera gran apuesta compartida por todos los sistemas nacionales de evaluación es dar cuenta a la opinión pública acerca del grado en que el sistema educativo está logrando ciertos objetivos en ciertas áreas del currículo (generalmente Matemática y Lenguaje y, en menor medida, Ciencias Naturales y Ciencias Sociales) y en ciertos grados o niveles (generalmente aquellos que implican el final de un ciclo de la enseñanza).

Desde esta perspectiva, las Unidades de Evaluación tendrían una función comparable a la de los Institutos Nacionales de Estadística: así como estos últimos entregan periódicamente al público información sobre el costo de vida (índice de precios al consumo) o sobre el mercado laboral (índice de desocupación), **la función principal de los informes nacionales sería la de brindar al público información periódica sobre los aprendizajes en el sistema educativo, garantizando la calidad técnica de la misma.**

El uso e interpretación de la misma sería responsabilidad principal de los usuarios, del mismo modo que normalmente los Institutos de Estadísticas no tienen una preocupación especial por garantizar que la población comprenda la información que producen ni por hacer un trabajo de divulgación “didáctica” de la información, lo que generalmente queda en manos de los medios de prensa.

En el caso de la evaluación sobre aprendizajes, es discutible que pueda trabajarse sobre este supuesto. En primer término, porque no existe la acumulación suficiente de masa crítica respecto a las formas de evaluar y a los indicadores a utilizar. A diferencia de los indicadores de población, empleo, evolución económica, etc. –en torno a los cuales ha existido abundante trabajo académico y hay ciertos consensos básicos en cuanto a las formas de medición e interpretación, sus debilidades y limitaciones– en las evaluaciones de aprendizajes aún se está lejos de alcanzar esos acuerdos sobre los enfoques básicos en los países de la región. Existe una gran diversidad de modos de medir, reportar e interpretar los datos y es necesario aún un esfuerzo de discusión y acumulación técnica, que debería incluir necesariamente la apertura de los instrumentos y procedimientos de evaluación al escrutinio académico.

En segundo término, los indicadores de tipo demográfico y económico, en alguna medida, han pasado a formar parte de la cultura general y tienen un significado al menos genérico para los ciudadanos mediana-

mente informados⁴. Esto no ocurre con los indicadores de resultados educativos. Por eso **es necesario realizar un trabajo sistemático de difusión para que el público en general pueda atribuir significado a los datos que los reportes de las evaluaciones ofrecen.** En especial, es imprescindible realizar dicho trabajo con los medios de comunicación, con el fin de crear una cierta cultura básica en torno al significado de la evaluación de aprendizajes a nivel del sistema y sus resultados, de modo de evitar simplismos en el tratamiento de la información que llega al público masivo.

Se puede decir, entonces, que en los países de la región se ha dado un primer paso relevante: periódicamente se está informando a la sociedad acerca de los aprendizajes de los alumnos, lo que contribuye a poner la educación en la agenda pública y a que la preocupación ya no esté centrada exclusivamente en el acceso al sistema educativo, sino también en los aprendizajes que efectivamente se adquieren en él. Pero, simultáneamente, es preciso señalar que **las formas de reportar los resultados que se han desarrollado hasta el momento, en muchos casos pueden dar lugar a serios equívocos,** según se analizará con detalle más adelante en este trabajo.

I.2. Los sistemas de evaluación como sistemas de información dirigidos a públicos específicos

Además de proveer información en forma sistemática a la ciudadanía, **la mayoría de los sistemas de evaluación asumen como misión principal la de aportar información relevante a tres públicos específicos que, supuestamente, la usarán para mejorar la calidad de la educación: las autoridades y decisores de los sistemas educativos, los directivos y docentes de los establecimientos educativos y las familias de los estudiantes.**

⁴ Si bien en muchos casos, como por ejemplo los indicadores adecuados para medir la pobreza, los mismos son objeto de debate. Asimismo, toda interpretación del estado de una sociedad o de una economía a partir de un conjunto de indicadores es siempre materia discutible. Los indicadores no nos entregan la realidad, simplemente nos colocan ciertos puntos de referencia empíricos para su interpretación. Del mismo modo, ningún sistema de indicadores educativos o de resultados de evaluaciones puede eliminar la necesidad de debate ni tener pretensiones de dar cuenta de manera inobjetable y acabada del estado del sistema educativo. Sin embargo, sí es deseable que los indicadores sean pertinentes y su medición técnicamente correcta.

El primer problema que aparece en algunos países es que se proponen informar a esos diferentes destinatarios a través de un informe nacional único. Asimismo, la mayoría de los reportes nacionales se limita a describir los datos incluidos en los cuadros o gráficos, con escaso o nulo análisis e interpretación de los mismos, suponiendo que son autoexplicativos. En muchos reportes tampoco se incluyen advertencias acerca de los usos válidos de la información y sus limitaciones.

En la figura 2 se muestra el tipo de informes que se ha producido en cada país. Chile, Ecuador, Argentina y Uruguay producen informes específicamente dirigidos a los docentes, con el fin de realimentar el análisis de sus prácticas pedagógicas a través del estudio de los objetivos mejor y peor logrados y de sus implicancias pedagógico-didácticas. En varios países se entrega a cada establecimiento los resultados de sus alumnos, con la finalidad de propiciar el análisis y toma de decisiones a partir de sus propias insuficiencias. Este aspecto será revisado en profundidad en el capítulo IV.

FIGURA 2
Tipos de informes de difusión de resultados producidos
en los países de la región

Países	Informe general de resultados	Informes específicos sobre factores asociados	Informes o publicaciones específicas dirigidas a los docentes	Informes o publicaciones específicas dirigidas a las familias	Informe de resultados por establecimiento educativo (censales) ^(*)
ARGENTINA					
BOLIVIA					
BRASIL					
COSTA RICA					
CHILE					
ECUADOR					
PARAGUAY					
PERU					
URUGUAY					

(*) Algunos países, como por ejemplo Ecuador devuelven resultados por establecimiento en el marco de operativos de carácter muestra a los que integraron la muestra.

En cuanto a las autoridades y decisores de los niveles centrales, en muchos países se han producido estudios de los “factores asociados” a los resultados. Sin embargo, por momentos parece predominar una visión excesivamente optimista –tal vez algo ingenua– acerca de lo que la investigación sobre “factores asociados” puede aportar a la definición de las políticas educativas. En la mayoría de los reportes se asume una relación directa, y en cierto modo mecánica, entre la identificación de “factores” internos del sistema educativo asociados con los resultados y la adopción de medidas de mejoramiento educativo, del tipo “*Diseñar acciones directas sobre los factores controlables por el propio sistema educativo que incidan en el logro de las competencias básicas de los alumnos*”. Si bien este último propósito es loable, es preciso tener en cuenta que la relación entre resultados de investigación y toma de decisiones de política nunca es directa e inmediata, sino que está atravesada por múltiples mediaciones de tipo conceptual, social, político y administrativo. En el capítulo V se profundizará en el análisis de este aspecto.

En relación al tercero de los actores privilegiados en las declaraciones de propósitos de los sistemas nacionales –las familias de los estudiantes–, es preciso enfatizar que los reportes nacionales difícilmente son comprensibles aun para una familia de nivel cultural medio. Chile parece ser el único país con una estrategia de difusión dirigida específicamente a las familias, que consiste básicamente en la publicación en la prensa y en Internet de los resultados de cada escuela, y en la distribución a las familias de folletos explicativos con los resultados del establecimiento al que asisten sus hijos⁵.

I.3. Usos de los resultados con consecuencias “fuertes”

Si bien los enunciados explícitos de los sistemas nacionales de evaluación enfatizan su finalidad formativa y un uso de los resultados de “bajo riesgo”, existen en la región algunas experiencias importantes de uso de los resultados de las evaluaciones nacionales con consecuencias más “fuertes” que merecen ser identificados⁶.

⁵ Esto último aparece anunciado en la página WEB del SIMCE pero no fue posible tener acceso a ejemplares de los mismos.

⁶ Asimismo, es preciso mencionar que en varios países coexisten con los sistemas nacionales de evaluación de aprendizajes, Otros sistemas de pruebas de selección para el ingreso a la universidad o de acreditación del final de la educación secundaria.

El caso más destacado es el de Chile donde, si bien no aparece entre los propósitos explicitados en los informes nacionales, desde los comienzos del SIMCE se efectuó una fuerte apuesta a utilizar los resultados de las evaluaciones nacionales como mecanismo para establecer incentivos y sanciones para las escuelas. Los resultados de cada establecimiento se publican en la prensa nacional –y más recientemente en Internet–, con el propósito, entre otros, de que los padres utilicen esa información a la hora de elegir a qué escuela desean enviar a sus hijos, en el marco de una política dirigida a generar un mercado competitivo entre los establecimientos educativos. El supuesto es que de esta forma se obliga a los establecimientos a preocuparse por mejorar sus resultados.

Asimismo, desde 1996 los resultados de las evaluaciones SIMCE están siendo empleados para establecer incentivos económicos para los docentes a través del Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales Subvencionados (SNED). Los establecimientos educativos son “rankeados” en función de un índice que incluye los siguientes seis factores:

- *Efectividad*: promedio de los resultados del establecimiento en las pruebas de Matemática y Castellano aplicadas en los años previos.
- *Superación*: promedio de las diferencias que cada establecimiento tuvo en los puntajes de esas mismas pruebas respecto a la medición anterior.
- *Iniciativa*: incluye diversos indicadores, tales como la realización de actividades formativas complementarias de libre elección para los alumnos o tener un “proyecto educativo institucional”.
- *Mejoramiento de las condiciones laborales*: se construye a partir de una categorización de los establecimientos realizada por el sistema de inspección.
- *Igualdad de oportunidades*: incluye diversos indicadores, tales como la tasa de aprobación de alumnos, la tasa de retención y la incorporación de alumnos con discapacidades.
- *Integración y participación de profesores, padres y apoderados*: se construye a partir de indicadores tales como la existencia de un consejo de profesores y de centros de padres y de alumnos que cuenten con espacios de participación avalados en el proyecto institucional, o la percepción de los padres acerca del establecimiento, entre otros.

Prácticamente las dos terceras partes del valor del índice dependen de los resultados en las pruebas SIMCE: el factor *efectividad* pesa un 37% y el factor *superación* un 28% del valor final del índice SNED⁷.

Otro caso relevante es el de México. Durante los años '90 los esfuerzos de evaluación de aprendizajes a nivel nacional se realizaron principalmente en el marco del programa “Carrera Magisterial”, el cual establece incentivos económicos individuales para los maestros, asociados a cinco grandes factores: antigüedad, grado académico, preparación profesional, acreditación de cursos de actualización y desempeño profesional. Este último factor, que tiene el mayor peso en el índice (35%), se mide a partir de la evaluación del maestro por parte del equipo docente y del denominado “aprovechamiento escolar”, que es el resultado obtenido por los alumnos de cada maestro en pruebas estandarizadas de logro (Comisión Nacional SEPSNTE, 1994). La participación en este programa ha sido de carácter voluntario, pero ha incluido a varios cientos de miles de maestros cada año. Durante muchos años las pruebas aplicadas a varios millones de alumnos fueron utilizadas únicamente con la finalidad de otorgar estos incentivos, sin que sus resultados tuvieran ningún tipo de difusión pública. Desde hace muy poco tiempo la Secretaría de Educación Pública ha comenzado a difundir algunos resultados.

I.4. El propósito central de la difusión de resultados: ¿“consecuencias fuertes” o evaluación “formativa”?

La discusión acerca de cuál debe ser la apuesta principal de los sistemas de evaluación de resultados es de enorme trascendencia y merece especial atención en la construcción de visiones sobre cómo avanzar en el desarrollo de los sistemas nacionales de evaluación de aprendizajes en la región.

Desde algunos sectores, principalmente políticos y economistas, parece existir una creciente presión por ir transformando los sistemas nacionales de evaluación en sistemas de rendición de cuentas públicas y de establecimiento de incentivos al estilo seguido por Chile. Recientemente, Argentina ha comenzado a publicar en la prensa los resultados por establecimiento en sus evaluaciones de carácter censal. En cambio, otros países –como Uruguay y Bolivia– han enfatizado la finalidad exclusiva-

⁷ Para más detalles, véase Ministerio de Educación, República de Chile, 2000.

mente formativa de las evaluaciones y los resultados por escuela tienen carácter estrictamente confidencial.

En dos trabajos recientes, refiriéndose a los Estados Unidos –probablemente el país con mayor acumulación de experiencias, diversidad de enfoques y debates en torno a este tema– Linn & Gronlund (2000) señalan que el **problema central de usar pruebas estandarizadas en el marco de una estrategia de “alto riesgo o consecuencias fuertes” –léase rendición de cuentas públicas o establecimiento de incentivos económicos por resultados– es que exacerban la preocupación de directores y docentes por enseñar la prueba y el tipo de actividades que ellas incluyen**, es decir, a dedicar creciente tiempo de instrucción a ejercicios similares a los que serán planteados en las pruebas y exclusivamente sobre los temas y competencias que serán evaluados. Si las pruebas tienen una fuerte carga de ítems de opción múltiple, esto resulta altamente contraproducente, porque los resultados dejan de ser un indicador de un conjunto de competencias más amplio y pasan a ser estrictamente una medida de la capacidad de los alumnos para responder a las preguntas de las pruebas. Se podría mejorar las pruebas, hacerlas más complejas y abarcativas, pero entonces se vuelve difícil garantizar la corrección confiable de las mismas a gran escala pues, por definición, cualquier sistema de incentivos o rendición de cuentas públicas por escuela debe tener carácter censal. Como señalan Linn & Gronlund⁸:

“El grado en que la creciente presión (por hacer que las evaluaciones tengan ‘consecuencias fuertes’) ayudaron o dañaron a la educación, sigue siendo controversial. Los proponentes de evaluaciones con ‘consecuencias fuertes’ argumentan que las pruebas miden objetivos que es importante que los estudiantes aprendan y que es deseable que los maestros focalicen su atención en ellos. Señalan con orgullo los incrementos en los puntajes de las pruebas observados en los programas de evaluación estatales y distritales a finales de los ’80 y principios de los ’90.

“Los críticos del énfasis creciente en los resultados de las pruebas argumentan que otorgar un exceso de importancia a estos distorsiona la educación. Argumentan que objetivos educativos importantes son ignorados cuando no están incluidos en las pruebas que importan. Más aún, reclaman que el incremento en los puntajes ofrece una visión errada de la realidad, porque los maestros enseñan las cuestiones específicas de las

⁸ LINN, R. & GRONLUND, N., 2000, pp. 8-9, en inglés en el original.

pruebas en lugar de dominios de contenido más generales (...) Como los ítems de una prueba solo constituyen una muestra del dominio que interesa, el puntaje en la prueba y el grado de comprensión no son lo mismo. Se requiere una generalización, y es la generalización, no la prueba en sí misma, lo que importa. Cuando se enseña para responder el tipo de ítems específicos de las pruebas, la validez de la inferencia sobre el nivel de logro de los estudiantes se ve amenazada”.

La evaluación “sin consecuencias” o de “bajo riesgo” minimiza estos problemas, porque directivos y docentes están menos amenazados por la evaluación, con lo cual pueden asumirla con mayor libertad como indicadores de competencias más complejas, sin verse obligados a forzar artificialmente el logro de buenos resultados en las pruebas.

Como contracara, es preciso señalar que **la ausencia de consecuencias hace que nadie en las escuelas esté obligado a hacerse cargo de los resultados y, muchas veces, ni siquiera a enterarse de ellos o a considerarlos.** En este sentido, es posible argumentar que la confidencialidad de los resultados por escuela, en cierto modo, las exime de responsabilidad pública.

Otro de los problemas importantes derivados del hecho de que las pruebas no tengan consecuencias directas para los alumnos, es que ello puede afectar el grado en que estos se esfuerzan por realizar las tareas que se les han propuesto. Esto es particularmente delicado en la educación media, dada la edad de los alumnos, y puede afectar la confiabilidad y validez de las mediciones.

En relación al uso de los resultados para generar competencia entre los establecimientos, Carnoy (2000) señala que la evidencia empírica del caso chileno indica que la creación de un mercado competitivo en el sector educación no ha tenido un impacto de mejora de la calidad del sistema, sino de ampliación de las desigualdades. Cuando se controla el nivel socioeconómico del alumnado de cada escuela, se observa que los colegios católicos, que existían con anterioridad al proceso de privatización de los años '80, tienen resultados algo mejores que las escuelas públicas, pero a un costo mayor por alumno. Simultáneamente, los colegios privados “nuevos” –que “crecieron como hongos” a partir de la implantación del sistema de “vouchers” en los años '80– no están logrando mejores resultados que las escuelas públicas cuando se controla el nivel sociocultural de los alumnos. Sí son económicamente más “eficientes” –tienen un costo por alumno inferior– debido a que pagan salarios inferiores a sus docentes. Al respecto señala:

“Por tanto, no hay evidencia de Chile de que un sistema nacional de ‘vouchers’ que ha estado operando por casi 20 años haya tenido un efecto positivo sobre los resultados de rendimiento de los millones de estudiantes que pasaron por el sistema durante ese tiempo. El plan puede haber ahorrado algún dinero de impuestos en los '80 y principios de los '90, pero lo hizo principalmente porque los colegios privados con fines de lucro pudieron pagar a los docentes menos que las escuelas públicas, no porque hubiese mayor eficiencia en el uso de los recursos. Por el otro lado, existe cierta evidencia de que la privatización aumentó la estratificación durante los '80; en la medida que los colegios privados “descremaron” a los mejores estudiantes –de nivel social más alto y más talentosos– de las escuelas públicas y tendieron a ubicarse en las municipalidades de mejores ingresos... Estudios en otros países con sistema de ‘vouchers; como Nueva Zelanda, también argumentan que los mismos incrementan la estratificación entre los estudiantes (Lauder & Hughes, 1999)”⁹.

Nuevamente se plantea en este punto un dilema. Si bien la evidencia empírica y el sentido común indican que la publicación de “rankings” de escuelas en función de sus resultados tiene como efecto principal una mayor segmentación sociocultural del sistema educativo, simultáneamente es preciso reconocer que, en la medida en que se genera información sobre los resultados alcanzados por los alumnos en un establecimiento, las familias tienen derecho a conocerlos. Se plantea así una contradicción difícil de resolver y que los sistemas deberán asumir en los próximos años: ¿cómo conjugar el derecho de las familias a conocer lo que sus hijos están logrando con el cuidado que debe darse al tratamiento de la información para evitar interpretaciones inapropiadas y la deslegitimación espuria de las escuelas que trabajan con la población más carenciada?

La disyuntiva de fondo es en qué medida los objetivos *formativos* por un lado, y de *establecimiento de incentivos*, por otro, pueden resultar compatibles entre sí. El tema merecería ser profundizado a través de una investigación dirigida a analizar de qué modo las escuelas utilizan los resultados de las evaluaciones y de qué manera el contexto y la estrategia global del sistema de evaluación incide en la aceptación y uso de los resultados por parte de los establecimientos educacionales. Asimismo,

⁹ CARNOY, M., 2000, p. 18, en inglés en el original.

sería relevante comparar el uso de la información entre países en que la difusión de la misma se realiza en un contexto de capacitación –lo que en principio aparece como más efectivo, aunque también más costoso–, y países en que el sistema de evaluación se limita a la distribución de los informes a las escuelas.

En todo caso, el uso de los resultados de los sistemas de evaluación desarrollados hasta el momento con fines de “rendición de cuentas”, requiere de un cuidadoso análisis de los riesgos que implica, como advierte Linn:

“No deberíamos esperar que pruebas poco costosas diseñadas para otros propósitos de ‘bajo riesgo’ se hagan cargo de las presiones que actualmente se ponen sobre ellas por parte de los sistemas de rendición de cuentas de ‘consecuencias fuertes’... Sistemas de evaluación que son monitores sumamente útiles pierden buena parte de su credibilidad para ese propósito cuando se les incorporan ‘consecuencias fuertes’. Los efectos negativos no buscados de los usos de las evaluaciones con fines de rendición de cuentas de ‘consecuencias fuertes’ a menudo superan los efectos positivos buscados”¹⁰.

CAPÍTULO II

PORCENTAJES, MEDIAS, NOTAS Y PUNTAJES: LAS MÉTRICAS EMPLEADAS PARA REPORTAR LOS RESULTADOS

Como una segunda aproximación a los informes de las evaluaciones nacionales, en el presente capítulo se analizarán las “métricas” que estos utilizan, es decir; los tipos de datos que se presentan para describir los resultados a nivel nacional.

Siete han sido las “métricas” empleadas en los diversos reportes nacionales para caracterizar la calidad de los sistemas educativos e informar sobre los resultados de las evaluaciones:

- porcentaje promedio de respuestas correctas
- promedio de los puntajes “brutos”
- puntajes estandarizados
- puntajes de “Teoría de Respuesta al Ítem”
- porcentaje de alumnos que logran cierto nivel de puntaje o de desempeño

¹⁰ LINN 2000, pp.12-14, en inglés en el original.

- porcentaje de alumnos que logran el dominio de objetivos específicos
- una transformación de alguna de las métricas anteriores a una calificación numérica expresada en la misma escala que habitualmente se emplea en el sistema para calificar a los alumnos.

Comprender qué significan las diferentes métricas o tipos de datos que se ofrecen en los informes nacionales es relevante por dos razones principales:

- Porque está en juego la “comprensibilidad” de la información para los destinatarios.
- Porque está en juego la “utilidad” de la información que se entrega en función de las necesidades de los destinatarios. Por ejemplo, si se espera que la información sirva a los docentes para revisar el modo en que encaran la enseñanza, es necesario plantearse si resulta suficiente informar sobre el porcentaje de respuestas correctas en la prueba a nivel nacional o por regiones.

En otro plano, es necesario preguntarse si ofrecer al público el dato de que a nivel nacional hubo un promedio de 58% de respuestas correctas permite responder a la que parece ser la demanda principal del público: saber si los alumnos están logrando o no lo que se espera de ellos al cabo de ciertos niveles de escolarización¹¹.

A lo largo de este capítulo se explica en forma sintética el significado de cada una de las métricas empleadas en los distintos países de la región, incluyendo una tabla de ejemplo y una rápida revisión de sus potencialidades y problemas. El lector interesado en más detalles puede recurrir a la versión completa de este texto en la página WEB de PREAL.

II.1. Porcentaje promedio de respuestas correctas y/o medias de puntajes brutos

El porcentaje promedio de respuestas correctas se construye calculando primero qué porcentaje de los alumnos respondió correctamente cada uno de los ítems de la prueba y luego el promedio de esos porcentajes

¹¹ Este tema involucra la cuestión de si los datos han sido construidos dentro de un enfoque referido a normas o dentro de un enfoque criterial, cuestión que ha sido tratada con más detalle en RAVELA, R (editor) y otros, 2000 y que será analizada al final de este capítulo.

para todos los Ítems de la prueba. En términos matemáticos esta “métrica” es equivalente al promedio de los puntajes brutos¹² de todos los estudiantes.

Argentina, Chile y Paraguay son los países que emplean el porcentaje promedio de respuestas correctas en cada prueba como la métrica principal de sus reportes. En la figura 3 se incluye a título ilustrativo una tabla tomada de uno de los reportes de Argentina. El reporte a través de esta métrica ofrece una visión muy general sobre los logros alcanzados. Entrega información sobre el dominio de ciertas asignaturas en el sistema educativo, pero no sobre lo que los alumnos individualmente considerados son capaces de hacer, pues las unidades de análisis son los ítems, no los alumnos. Un resultado de 55% de respuestas correctas puede obedecer tanto a que la mayoría de los alumnos se ubican en torno dicha media, como a una fuerte dispersión en situaciones extremas: alumnos con alto dominio de los contenidos evaluados, por un lado, y alumnos cuyos puntajes han sido muy bajos, por otro. Como la selección de los ítems que forman parte de las pruebas en general se efectúa privilegiando los que tienen niveles de dificultad media y no a partir de una determinación sustantiva de cuál es el nivel de logro que se espera que alcancen todos los alumnos, en realidad el resultado no dice nada acerca de en qué grado los estudiantes son capaces de hacer lo que se espera de ellos. En otras palabras, los resultados usualmente observados de entre 50% y 60% de respuestas correctas en realidad están determinados *a priori* al seleccionar los ítems, pero no significan que se esté logrando ese porcentaje de lo esperado como logro de los alumnos. Este tipo de métrica puede ser empleada en forma válida para efectuar comparaciones entre distintos sectores del sistema (departamentos, regiones, escuelas, etc.), analizando cuáles están por encima y cuáles por debajo de la media nacional, así como las distancias entre ellos, como se realiza en Argentina (figura 3). Finalmente, el reporte a través de porcentajes promedio de respuestas correctas también puede ser empleado para informar acerca de cambios en los resultados del sistema a lo largo de los años, como se ha hecho en Chile y Argentina¹³.

¹² En una prueba de opción múltiple los puntajes “brutos” son los que se calculan para cada estudiante a partir del número de ítems que respondió correctamente. Si una prueba consta de 40 ítems, el puntaje “bruto” variará entre 0 y 40. Si un alumno respondió correctamente 22 ítems, su puntaje bruto es 22.

¹³ Para que esta información sea válida para apreciar la evolución en el tiempo de los logros en el sistema educativo, es necesario que las pruebas estén adecuadamente equiparadas al nivel de dificultad de la primera medición. Que midan los mismos contenidos y/o compe-

FIGURA 3
Argentina - Valores Medios de Rendimiento en Lengua
por Jurisdicción y Nivel

Nivel Primario - 6º Grado Lengua		
Muestra 1996		
Jurisdicción	Lengua 6º Grado	
	Muestra 1996	Relación con respecto al valor medio
Capital Federal	72,39%	1,24
Mendoza	67,63%	1,16
Río Negro	66,28%	1,14
Neuquén	65,37%	1,12
La Pampa	60,58%	1,04
Chaco	60,37%	1,04
Córdoba	58,77%	1,01
Santa Cruz	58,69%	1,01
MEDIA NACIONAL	58,19%	1,00
Buenos Aires	57,88%	0,99
Santa Fe	57,87%	0,99
Entre Ríos	57,26%	0,98
Gran Buenos Aires	56,98%	0,98
Tierra del Fuego	56,87%	0,98
Misiones	55,83%	0,96
Chubut	55,41%	0,95
Salta	54,71%	0,94
Corrientes	54,12%	0,93
San Juan	53,79%	0,92
Jujuy	52,16%	0,90
La Rioja	50,14%	0,86
Tucumán	49,04%	0,84
Formosa	48,94%	0,84
Santiago del Estero	47,74%	0,82
Catamarca	46,48%	0,80

Fuente: Ministerio de Cultura y Educación de la Nación, Argentina, 1997. Operativo Nacional de Evaluación de Aprendizajes 1996; página 33.

II.2. Porcentajes de alumnos que dominan ciertos objetivos o que logran niveles de desempeño predeterminados

Costa Rica, Ecuador y Uruguay emplean un enfoque sustantivamente diferente al de los tres países anteriores para reportar sus resultados, dado que utilizan como métrica principal el porcentaje de individuos (alumnos) que logran niveles de dominio predefinidos en las competencias, destrezas y/o contenidos evaluados. Las diferencias principales con el enfoque del apartado anterior son que:

- los resultados no son únicamente descriptivos, sino que implican un juicio acerca de si los alumnos han alcanzado o no lo que se espera de ellos
- los resultados, por tanto, son interpretables no solo en términos de comparaciones entre individuos y grupos, sino también en términos del grado en que las metas de enseñanza están siendo alcanzadas.

Costa Rica y Ecuador elaboran sus pruebas dentro del enfoque de evaluaciones referidas a criterios. Para ello, en cada disciplina evaluada se define un conjunto de diez a catorce objetivos o destrezas específicas que se consideran prioritarias. El “dominio” de cada objetivo o destreza por parte de los alumnos es evaluado a través de varios ítems referidos al mismo: tres en el caso de Costa Rica y cuatro en el caso de Ecuador. En el primer país se considera que un alumno domina un objetivo si responde correctamente al menos dos de los tres ítems referidos al mismo. En el caso de Ecuador, cuando responde correctamente al menos tres de los cuatro ítems.

Los reportes de resultados de estos países emplean como “métrica” principal el porcentaje de alumnos que alcanzó el dominio de cada objetivo. En los reportes, los resultados aparecen desagregados por destreza u objetivo, ofreciendo así una visión detallada de lo que los alumnos son capaces de hacer. En la figura 4 se consigna a título ilustrativo una tabla perteneciente a un informe nacional de Costa Rica.

Este modo de reportar presenta una dificultad a la hora de ofrecer una visión global del desempeño de los alumnos en el conjunto de la prueba, o de los objetivos y destrezas de una disciplina. Para resolverlo,

tencias y que tengan la misma estructura en términos de cantidad de ítems y de pesos de las diferentes competencias en el puntaje total. En los informes públicos analizados para estos países no se da cuenta de los procedimientos seguidos para asegurar la equivalencia de los instrumentos de medición utilizados en los diferentes años.

FIGURA 4
COSTA RICA. Porcentaje de estudiantes que dominan cada objetivo
por región. Sexto año.
Español. IIMEC-MER Octubre-noviembre. 1996

OBJETIVO	MUESTRA NACIONAL	CENTRAL	BRUNCA	CHOROTEGA	HUETAR NORTE	PACÍFICO CENTRAL	HUETAR ATLÁNTICA
1	69,9	72,0	67,5	60,6	75,0	65,1	64,2
2	69,9	87,8	87,6	78,8	85,3	80,2	85,5
3	93,7	95,5	91,6	85,0	92,6	92,5	90,5
4	81,4	83,4	76,7	76,3	82,4	61,6	72,6
5	67,5	69,6	71,3	51,3	70,6	64,9	62,6
6	78,7	80,9	76,9	66,3	75,0	77,9	77,1
7	93,6	95,2	91,0	86,3	88,2	93,6	92,7
8	91,7	93,8	88,6	85,6	90,4	89,5	85,5
9	93,9	65,1	92,9	85,6	91,1	91,3	97,1
10	61,0	63,8	58,7	48,1	55,1	56,4	58,1
11	78,4	84,1	66,9	63,1	72,8	75,0	65,4
12	81,4	84,2	77,8	62,5	80,1	81,4	75,4
13	94,5	95,9	91,6	88,8	94,1	90,7	95,0
14	80,1	83,7	80,8	64,4	74,3	77,3	71,5

Fuente: Ministerio de Educación Pública - Universidad de Costa Rica/Instituto de Investigaciones para el Mejoramiento de la Educación Costarricense (IIMEC), Costa Rica, 1997 Informe Nacional sobre el *Desarrollo, Validación y Aplicación de las pruebas de Diagnóstico de conocimientos*. Octubre-noviembre de 1996; pág. 39.

tanto Costa Rica como Ecuador recurren por diversos caminos a una media de puntajes brutos, al igual que los países mencionados en el apartado anterior. En el caso de Costa Rica, se reporta un puntaje promedio del conjunto de la prueba calculado como porcentaje del puntaje total. Ecuador utiliza una calificación sobre 20 puntos –que es la escala que rige en el sistema para la evaluación de los alumnos– que se construye, como en los casos anteriores, como un porcentaje promedio del puntaje total de la prueba.

La modalidad empleada por Uruguay para reportar sus resultados es similar en cuanto la métrica o dato principal es el porcentaje de alumnos que alcanzan un cierto estándar de rendimiento o categoría de logro definido *a priori* y que no es meramente descriptivo sino que implica un

juicio acerca del logro de los alumnos. Sin embargo, en Uruguay esta categoría de logro es denominada “suficiencia” y se establece para el conjunto de la prueba –en este sentido es diferente a lo que hacen Ecuador y Costa Rica– y para cada una de las grandes “competencias” en que está estructurada cada prueba.

Para establecer la “suficiencia” se definen “puntos de corte” en el puntaje bruto de las pruebas, los que no están referidos únicamente a la distribución normal de la población sino al juicio experto de especialistas y docentes acerca de lo que los alumnos deberían ser capaces de lograr al finalizar la educación primaria. Los ítems con alto o bajo grado de dificultad no son descartados, sino que se mantienen en la medida que miden aprendizajes definidos como fundamentales para un alumno que completa la educación primaria. A la luz de los resultados de la prueba piloto y a partir de la opinión experta de jueces, se seleccionan los ítems de modo tal que, como mínimo, se espera que todos los alumnos respondan 14 ítems sobre un total de 24 (lo que equivale a aproximadamente el 60% del puntaje bruto total). Este es el nivel de “suficiencia” para el conjunto de la prueba. Los reportes están centrados en informar sobre el porcentaje de alumnos que alcanzaron ese nivel de desempeño (ver figura 5).

FIGURA 5
URUGUAY. Resultados en Lengua y Matemática, Sexto Año, por tramos de puntaje en 1996 y 1999

Porcentaje de alumnos:	LENGUA		MATEMÁTICA	
	1996	1999	1996	1999
Con desempeño altamente satisfactorio (20-24 puntos)	15,8	14,0	6,8	9,0
Con desempeño satisfactorio (14-19 puntos)	41,3	47,3	27,8	31,8
SUFICIENTES	57,1	61,3	34,6	40,8
Con desempeño insatisfactorio (7-13 puntos)	37,7	35,3	54,5	50,9
Con desempeño muy insatisfactorio (0-6 puntos)	5,2	3,4	10,9	8,3
INSUFICIENTES	42,9	38,7	65,4	59,2
Totales	100,0	100,0	100,0	100,0

Fuente: Administración Nacional de Educación Pública - Uruguay/Unidad de Medición de Resultados Educativos. *Evaluación Nacional de Aprendizajes en Lengua y Matemática. 6º año Enseñanza Primaria - 1999. Primer Informe*; pág. 25.

Al igual que en los casos de Chile y Argentina, Uruguay ha presentado sus resultados de 1999 en tablas comparativas con los de mediciones anteriores¹⁴.

Una de las debilidades de los reportes de Uruguay es que no están explícitamente fundamentadas las razones por las cuales los puntos de corte son esos y no otros, y qué es lo que se supone que marca la diferencia, en términos de conocimientos y competencias, entre un alumno “suficiente” y uno que no lo es¹⁵.

II.3. Puntajes estandarizados

Bolivia utiliza para la presentación de sus resultados puntajes “estandarizados”, que en estadística son denominados como puntajes “T”. Se trata de una transformación matemática del puntaje bruto de cada alumno a partir de la distribución de los puntajes brutos encontrada en la población. El puntaje asignado a cada individuo representa así su distancia respecto a la media de la población, en unidades de desviación estándar. Para ello, se emplea –en Bolivia– una media de 50 puntos y una desviación estándar de 10. Eso significa que los puntajes 40 y 60 son los que corresponden a los alumnos que se ubican a una desviación estándar por debajo o por encima de la media nacional. Los alumnos por debajo de 40 tuvieron resultados significativamente inferiores a la mayoría, en tanto que aquellos que están por encima de 60 son quienes tuvieron resultados superiores a la mayoría del universo evaluado.

Este modo de reportar es similar al de los porcentajes promedios de respuestas correctas empleado en Argentina, Chile y Paraguay, en cuanto su potencialidad principal es la comparación de resultados entre individuos o grupos. La transformación a puntajes T permite afinar estadísticamente esas comparaciones, pero es menos comprensible para el público no especializado.

¹⁴ Las pruebas empleadas en 1999 son formas equivalentes de las utilizadas en 1996 y el procedimiento de construcción de las mismas está descrito en el reporte público ANEP/UMRE, 1999; *Evaluación Nacional de Aprendizajes en Lengua y Matemática. 6º año Enseñanza Primaria- 1999. Primer Informe*; pp. 3 a 9.

¹⁵ El establecimiento de “puntos de corte” para definir categorías de desempeño nunca es una actividad “objetiva”, sino sumamente compleja y fuertemente “atada” a los puntos de vista de los expertos que los determinan.

En el reporte boliviano se establecen puntos de corte en los 45 y 55 puntos de la escala estandarizada, y a partir de los mismos se definen tres grandes categorías de desempeño de los alumnos:

- *Satisfactorio*, cuando el alumno supera los 55 puntos
- *Regular*, cuando el puntaje del alumno se sitúa entre 45 y el 55
- *Riesgo*, cuando el alumno no alcanza los 45 puntos

Los reportes muestran luego el porcentaje de alumnos en cada una de estas tres categorías. Sin embargo, debe tenerse presente que esta clasificación no es absoluta sino relativa: *riesgo* significa que un alumno está por debajo de la media nacional y en la parte inferior de la distribución, pero no dice nada acerca del nivel de logro de los objetivos curriculares. Del mismo modo, *satisfactorio* indica que el alumno está por encima del rendimiento de la mayoría de los alumnos, pero no necesariamente implica que haya alcanzado en forma satisfactoria los objetivos curriculares. Es decir, se trata de un estándar relativo a la distribución poblacional, por lo que es sustantivamente diferente al enfoque adoptado en Costa Rica, Ecuador y Uruguay.

II.4. Teoría de Respuesta al Ítem

Brasil ha sido el primer país en la región en utilizar la Teoría de Respuesta al Ítem (TRI) para la construcción de los puntajes de sus evaluaciones nacionales¹⁶. La TRI es un modelo matemático que permite estimar la capacidad de los individuos en una determinada área o disciplina a partir del supuesto de que la misma es unidimensional. Es decir: se asume, por ejemplo, que existe en los alumnos una capacidad o competencia para la matemática que es la que determina la probabilidad de que un alumno realice adecuadamente las diferentes actividades que integran el banco de ítems.

La explicación detallada de la TRI excede largamente los límites del presente trabajo, pero basta decir que tiene algunas ventajas sobre el enfoque clásico:

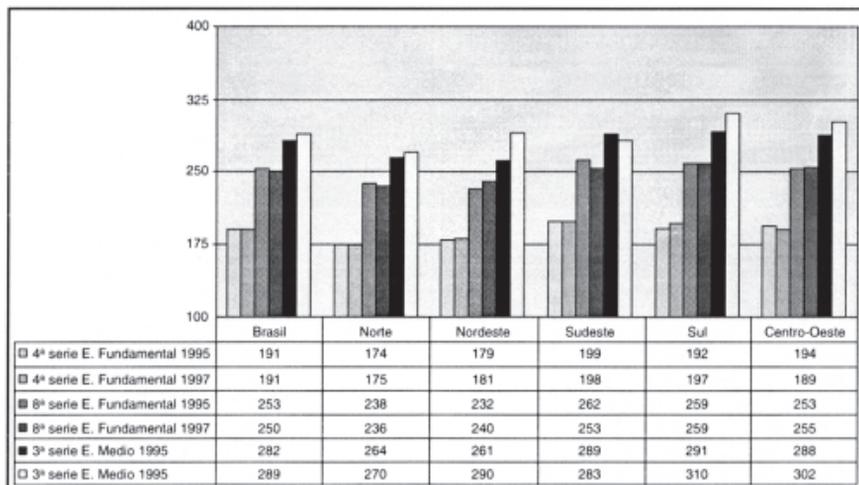
- Permite poner en una misma escala ítems y alumnos.
- Permite construir estimaciones más precisas de los cambios en el tiempo mediante la equiparación de las puntuaciones.

¹⁶ A partir de 1999 Chile y Argentina han comenzado a reportar sus resultados con este modelo que, por otra parte, es el empleado en los estudios internacionales. También lo utiliza Perú, que ha publicado sus primeros resultados en el año 2000.

- Permite estimar una medida de la capacidad de los alumnos que tiene en cuenta la dificultad de los ítems; es decir, los más difíciles tienen un peso mayor en la determinación del puntaje que alcanza un individuo.

Los puntajes de TRI no tienen una métrica específica. Normalmente se toma arbitrariamente una media de 250 puntos y desviación estándar de 50, o una media de 500 puntos y una desviación estándar de 100. En el caso del SAEB brasileño, se emplea una media de 250 puntos que corresponde a la media nacional de 1997 de los alumnos de 8º. Como se puede apreciar en la figura 6, a partir de los supuestos de la TRI es posible construir una única escala de puntajes para poblaciones pertenecientes a diferentes niveles de escolaridad; en el caso de Brasil, 4º y 8º de Educación Básica y 3º de Educación Media. Ello permite comparar las medias de “proficiencia” en cada disciplina entre los diferentes niveles del sistema educativo y regiones del país y para diferentes años, colocando todos los niveles en una misma tabla.

FIGURA 6
Brasil. Proficiencia media por grado en matemática. Brasil y regiones. 1997



Fuente: Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INEP); <http://www.inep.gov.br/saeb/saeb97/saeb97.htm>. SAEB/97. Estudio Comparativo.

Al igual que en el caso de los puntajes estandarizados empleados por Bolivia, uno de los problemas que plantea el uso de la TRI es la dificultad para comunicar al público y a los docentes cómo se llega a la determinación de los puntajes y qué es lo que estos significan. Difícilmente tendrán significado para el lector no especializado, mientras no exista una tradición o cultura en torno a su interpretación.

Un modo de enfrentar esta dificultad, crecientemente empleado en los estudios internacionales y en los reportes del NAEP en los Estados Unidos, es realizar una clara descripción de lo que son capaces de hacer los alumnos que se encuentran en ciertos puntos de la escala de puntajes. En el caso del SAEB, se establecen puntos de corte cada 75 puntos, se explica qué son capaces de hacer los alumnos en cada uno de esos “niveles de proficiencia” (figura 7), al tiempo que se reporta el porcentaje de alumnos que alcanza o supera cada nivel. La descripción de dichos niveles se realiza a partir de los ítems que logran responder los individuos cuyo puntaje se ubica en torno a uno de los puntos de corte y que no logran responder los individuos cuyo puntaje se ubica en torno al punto de corte anterior.

En los estudios internacionales, este modo de reportar es empleado de manera estrictamente descriptiva; es decir, se describe lo que los alumnos son capaces de hacer en cada tramo de la escala, pero no se responde a la pregunta: ¿es eso lo que deberían saber, es más o es menos?, ¿estamos bien, regular o mal? Obviamente, quedarse en un plano descriptivo es adecuado para los estudios internacionales, dado que sería casi imposible acordar criterios de desempeño comunes a todos los países. Sin embargo, esta tarea sí es necesaria a nivel nacional.

FIGURA 7
BRASIL/SAEB 1997 - Escala de Proficiencia en Matemática

Ejemplos de desempeño	NIVEL	Resultados del SAEB/97
<ul style="list-style-type: none"> • Los alumnos reconocen el valor de cédulas y monedas. • Leen la hora en relojes digitales y analógicos y saben que una hora tiene 60 minutos. • Resuelven problemas sencillos de adición y sustracción con números naturales. 	175	<p>En el nivel 175 o arriba de él se encuentran:</p> <ul style="list-style-type: none"> • 56% de los alumnos del 4º grado de la Enseñanza Fundamental; • 95% de los alumnos del 8º grado de la Enseñanza Fundamental; • 100% de los alumnos del 3er. grado de la Enseñanza Media.
<ul style="list-style-type: none"> • Los alumnos reconocen polígonos y cuadriláteros. • Establecen relaciones entre los valores de cédulas y monedas y resuelven situaciones al pagar y recibir cambio, aunque todavía no saben operar con decimales. • Son capaces de multiplicar y dividir, así como identificar unidades, decenas y centenas. • Resuelven problemas que envuelven más de una operación. • Adicionan y sustraen fracciones de un mismo denominador y conocen números naturales en la forma fraccionaria. • Interpretan gráficos de barras y de sector e identifican el gráfico más adecuado para representar una determinada situación. 	250	<p>En el nivel 250 o arriba de él se encuentran:</p> <ul style="list-style-type: none"> • 11% de los alumnos del 4º grado de la Enseñanza Fundamental; • 48% de los alumnos del 8º grado de la Enseñanza Fundamental; • 87% de los alumnos del 3er. grado de la Enseñanza Media.
<ul style="list-style-type: none"> • Los alumnos clasifican sólidos geométricos en cuerpos redondos y poliedros. • Interpretan resultados de medidas de longitud, masa, tiempo y capacidad. • Identifican, comparan y ordenan números racionales (en las formas fraccionaria y decimal) así como números enteros. • Interpretan lenguaje algebraico y resuelven ecuaciones y sistemas de ecuaciones de primer grado. 	325	<p>En el nivel 325 o arriba de él se encuentran:</p> <ul style="list-style-type: none"> • 8% de los alumnos del 8º grado de la Enseñanza Fundamental; • 32% de los alumnos del 3er. grado de la Enseñanza Media.
<ul style="list-style-type: none"> • Los alumnos resuelven problemas que envuelven punto, recta, circunferencia y sus relaciones. • Establecen relaciones y hacen conversiones entre fracciones ordinarias y números decimales. • Resuelven problemas que envuelven ecuaciones e inecuaciones sencillas de primer y segundo grado y sistemas de primer grado. • Conocen los principios básicos de polinomios y efectúan operaciones elementales entre ellos. • Conocen las propiedades básicas de exponentes y logaritmos. 	400	<p>En el nivel 400 o arriba de él se encuentran:</p> <ul style="list-style-type: none"> • 5% de los alumnos del 3er. grado de la Enseñanza Media.
<p>Observación: Los niveles 525 y 400 muestran el dominio de habilidades y contenidos más complejos que no corresponden al 4º grado de la Enseñanza Fundamental. Es por ello que no se presentan resultados para los alumnos de este grado.</p>		

Fuente: Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INER), Brasil, 1998; *¿Cómo está la Educación Básica en Brasil?*

El SAEB ha afrontado este desafío de la siguiente manera: a partir de los resultados descriptivos de lo que son capaces de hacer los alumnos que se encuentran en diferentes puntos de la escala, grupos de especialistas definen en qué momento de la escolaridad los alumnos deberían haber alcanzado cada uno de estos niveles, con lo cual se establece un criterio de valoración contra el cual es posible responder las preguntas especificadas en el párrafo anterior (figura 8).

FIGURA 8
BRASIL. Relación entre niveles de proficiencia y ciclos de los niveles de enseñanza

Nivel de proficiencia escala SAEBI97	Matemáticas	Lengua Portuguesa	Ciencias (Física, Química y Biología)
100	No significativo	Hacia la mitad del 1 ^{er} . ciclo de la Enseñanza Fundamental	Hacia <u>la mitad</u> del 1 ^{er} . ciclo de la Enseñanza Fundamental
175	Hacia <u>la mitad</u> del 1 ^{er} ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> del 1 ^{er} . ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> del 1 ^{er} . ciclo de la Enseñanza Fundamental
250	Hacia <u>el final</u> del 1 ^{er} ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> del 2 ^o . ciclo de la Enseñanza Fundamental	Hacia <u>la mitad</u> del 2 ^o ciclo de la Enseñanza Fundamental
325	Hacia <u>el final</u> del 2 ^o ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> de la Enseñanza Media	Hacia <u>el final</u> del 2 ^o ciclo de la Enseñanza Fundamental
400	Hacia el final de la Enseñanza Media	Después del final de la Enseñanza Media	Hacia <u>el final</u> de la Enseñanza Media

Fuente: Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INEP). *SAEB 97 - Primeros Resultados*.

En otras palabras, el lector puede comparar el nivel de la escolaridad en que se supone que los alumnos deberían haber alcanzado un cierto nivel de puntaje o de “proficiencia”, con la proporción de los alumnos de dicho nivel de enseñanza que efectivamente lo han hecho¹⁷.

¹⁷ Al respecto es preciso observar –y el informe del SAEB lo indica explícitamente– que en el caso de Brasil esta tarea es sumamente compleja en virtud de la diversidad curricular existente en el país. Idealmente este tipo de niveles de “proficiencia” o de desempeño se deberían poder establecer a partir de un currículo específico y no como mínimo común denominador de varias diferentes.

Este modo de reportar es una forma efectiva de brindar información, dado que permite al lector atribuir algún sentido a la escala de puntajes en términos de qué son capaces de hacer los estudiantes que han alcanzado cada nivel de la misma y establecer una valoración acerca de la situación.

II.5. Calificaciones

Muchos países transforman los datos recogidos en las evaluaciones a una escala de calificaciones similar a la que se utiliza usualmente para calificar a los alumnos en el sistema educativo. Esta práctica se justifica con el argumento de que padres y educadores están acostumbrados a las calificaciones:

“Esta calificación tiene un significado genérico y, a la vez, limitado, pues no provee información respecto a qué destreza fue mejor o peor dominada por el alumno de un año, que es el asunto que efectivamente importa en las mediciones de ‘APRENDO’. Este Informe presenta, sin embargo, las mencionadas calificaciones porque son de uso común en la educación nacional” (Ministerio de Educación y Cultura del Ecuador - EB/PRODEC, 1998)

Sin embargo, puede generar algunas confusiones sobre las que conviene alertar. La primera y más obvia es, justamente, la de asimilar los resultados de evaluaciones a las calificaciones de uso habitual en el sistema, cuando su sentido y significado es completamente diferente. Por tanto, más bien habría que trabajar en la dirección de ayudar a comprender que se trata de evaluaciones diferentes, que no tienen fines de acreditación y que no son comparables a las que realizan los maestros en el aula.

Por otro lado, en la mayoría de los países en que no se establecen criterios o estándares de logro, el uso de la calificación favorece la imagen de que se trata de un resultado valorativo acerca de lo que el sistema educativo o una determinada región están logrando (por definición, las calificaciones tienen un punto de corte que implican un aprobado y un desaprobado). Esta interpretación es completamente inadecuada cuando la calificación se establece como una transformación matemática de la media de puntajes brutos o del porcentaje promedio de respuestas correctas, a partir de pruebas en que los ítems han sido seleccionados en función de que su dificultad media se ubique entre .30 y .80. Como ya se indicó, cuando las pruebas han sido diseñadas de esta manera sus resulta-

dos solo deben ser interpretados en términos comparativos, evitando cualquier interpretación en términos de alumnos reprobados o de aprobación/reprobación de la labor del sistema educativo o de determinada gestión.

II.6. Resultados descriptivos o estándares de logro: ¿qué es lo que el público y los docentes quieren saber?

En la figura 9 se indica a modo de resumen qué “métricas” son empleadas en los reportes de los diferentes países de la región.

A lo largo del capítulo se ha podido apreciar que detrás de las métricas o “tipos de datos” con que se presentan los resultados de las evaluaciones nacionales de aprendizajes, hay implicancias y significados diversos en los que es preciso formar a la opinión pública y a los docentes, de modo que puedan hacer usos e interpretaciones adecuadas de los datos.

FIGURA 9
Métricas empleadas por los distintos países para el reporte de resultados

Países	Porc. promedio respuestas correctas	Promedio de puntajes brutos	Puntajes estandarizados	Puntajes “teoría de respuesta al ítem”	Porc. alumnos por niveles de desempeño	Porc. alumnos que dominan objetivos	Calificaciones
ARGENTINA							
BOLIVIA			Puntajes T				
BRASIL							
COSTA RICA							
CHILE	Hasta 1997			Desde 1998			
ECUADOR							
PARAGUAY							
PERU							
URUGUAY							

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

La disyuntiva principal implicada es la siguiente:

- limitarse a describir lo que los alumnos logran hacer en función de las tareas propuestas y a establecer comparaciones entre diferentes sectores del sistema educativo o entre años, para establecer conclusiones del tipo “los logros en lenguaje son mejores en las escuelas urbanas que en las rurales”, “en tal provincia respecto a tal otra”, o “los resultados en matemática han mejorado –o han empeorado– con respecto a dos años atrás”; o
- establecer estándares o criterios acerca de cuáles son los desempeños esperables y exigibles a todos los alumnos cuando culminan cierto grado o nivel de la escolaridad, lo que permite, además del tipo de conclusiones indicadas en ‘a.’, otras del tipo “tal proporción de los estudiantes no alcanzan los niveles de desempeño esperados en lectura en 3^{er} grado”, o “tal proporción de los estudiantes domina el objetivo o contenido ‘x’ de Matemática en 6^o grado”.

No cabe duda de que las conclusiones del primer tipo son útiles, fundamentalmente en términos de obtener una visión de las inequidades existentes al interior del sistema educativo y de la evolución de los resultados a lo largo del tiempo. Sin embargo, las expectativas de la opinión pública y de las familias, y el modo en que tienden a interpretar los resultados, parecen estar demandando conclusiones del segundo tipo. Asimismo, conclusiones de este último tenor probablemente tengan un efecto más fuerte sobre los docentes, en términos de darles una señal acerca de qué es lo que se espera que todos los alumnos sepan y sean capaces de hacer al cabo de cada grado o ciclo de enseñanza.

Esto último requiere un considerable esfuerzo para traducir el currículo –que normalmente en nuestros países está formulado en términos bastante generales– en especificaciones más claras y precisas del tipo de desempeños que se espera de los alumnos, e incorporar metodologías específicas para el establecimiento de criterios sobre los niveles de logro a alcanzar al cabo de ciertos grados o ciclos de aprendizaje. En países en que no existe un currículo nacional común, obviamente la tarea implicará un esfuerzo mayor de construcción de consensos acerca de qué es lo exigible para todos.

CAPÍTULO III

COMPARANDO LO INCOMPARABLE: LA PARÁBOLA DE LAS CARRETERAS

III.1 Modos típicos de desagregar la información en los reportes nacionales

Una de las debilidades principales de los reportes de resultados de las evaluaciones nacionales en los países de la región es la falta de contextualización sociocultural de dichos resultados. En prácticamente todos los países se reporta en función de los siguientes niveles de desagregación:

- Resultados según las divisiones políticas y/o geográficas del país (por estado, provincia, departamento, región).
- Resultados de escuelas urbanas y rurales.
- Resultados según forma de administración (escuelas públicas o fiscales, escuelas subvencionadas, escuelas particulares o privadas).

Estos niveles de desagregación, que responden a las diferencias más “visibles” de los sistemas educativos en subsistemas, son relevantes en la medida en que permiten ver las diferencias en el logro de los alumnos entre dichos subsistemas y presentan a la luz pública ciertas inequidades en la distribución de los conocimientos. Sin embargo, el hecho de que sean prácticamente las únicas formas de desagregación de los datos implica una sobresimplificación de la realidad, en la medida en que ocurren dos fenómenos:

- Algunos de estos agregados incluyen en su interior enormes heterogeneidades. Por ejemplo, el agregado “escuelas urbanas” incluye toda la gama de situaciones socioculturales, desde escuelas de sectores medios y altos hasta escuelas en barrios marginales. Lo mismo ocurre con la mayoría de las “jurisdicciones” políticas o regiones geográficas en que se divide un país.
- En otros casos, los diferentes tipos de escuela tienen un alumnado perteneciente mayoritariamente a ciertos sectores sociales y no una población representativa del conjunto: típicamente las escuelas rurales, que trabajan con un alumnado proveniente de familias con mínima trayectoria en el sistema educativo, y las escuelas privadas, que mayoritariamente reclutan su alumnado en los sectores sociales medios y altos.

El problema se plantea cuando de estas formas de desagregación de los resultados se derivan juicios o conclusiones acerca de la eficacia diferencial de unos y otros tipos de escuelas o jurisdicciones sin tener en cuenta que trabajan con alumnados diferentes.

En general, según lo demuestra la abundante investigación empírica¹⁸, las diferencias de resultados obedecen principalmente a las diferencias socioculturales entre las poblaciones con las que trabajan los diferentes sectores o instituciones del sistema educativo, antes que a diferencias en su eficacia pedagógica.

Lo anterior no implica afirmar que todas las diferencias de resultados se puedan explicar a partir de las diferencias en el contexto social, sino que para hacer afirmaciones válidas respecto a la “eficacia” pedagógica es imprescindible haber controlado el efecto del contexto sociocultural.

III.2. La parábola de las carreteras

Para explicar mejor este problema resulta útil recurrir a la siguiente metáfora. En la República Argentina hay una carretera denominada la “ruta del desierto”. Es la Ruta Nacional N° 20, en la provincia de La Pampa, que cruza la provincia en dirección este-oeste. El paisaje es semidesértico, pedregoso, con arbustos y absolutamente llano. La ruta tiene más de 200 kilómetros en línea recta, sin una curva ni una loma. El mayor peligro en esta ruta es la monotonía del paisaje, que hace que el conductor pueda dormirse en el volante. Por este motivo, a lo largo de la ruta se advierte la necesidad de descansar y se han ubicado cada 60 kilómetros sitios de descanso, pequeños oasis con árboles, sombra y agua.

En Argentina hay otra ruta, la N° 40, que atraviesa el país en dirección sur-norte siguiendo la precordillera. Recorre desde Tierra del Fuego en el extremo sur del país hasta la Quebrada de Humauaca en el norte, en el límite con Bolivia. La ruta 40 sube y baja, atraviesa valles y montañas, diversidad de paisajes y terrenos.

¹⁸ Simplemente por citar un estudio abarcativo a nivel internacional, el TIMSS muestra que *“un factor es consistente a través de todas las pruebas TIMSS en todos los grados. En cada país participante, los hijos de padres con educación universitaria promedialmente superior a los hijos de padres cuya educación terminó con la secundaria, y los hijos de padres cuya educación terminó en la secundaria, a su vez superaron a los chicos cuyos padres no completaron la secundaria. Es también cierto que a los estudiantes que tienen un hogar educativamente rico, con muchos libros, calculadoras y espacio para estudiar les va mejor en la escuela que aquellos que no lo tienen”* BEATON, A., 1999. En inglés en el original.

Suponga el lector que el Ministerio de Obras Públicas ha licitado la construcción de estas dos rutas, ha concedido la construcción a distintas empresas y quiere evaluar la eficiencia de cada empresa, con el fin de utilizar esta información como elemento de decisión en futuras concesiones de obras públicas.

Para ello, la División de Información del Ministerio diseña indicadores de eficacia y eficiencia: la cantidad de kilómetros de carretera por unidad de tiempo –por ejemplo, por día o por semana– que ha avanzado cada empresa; o el costo de cada empresa por kilómetro de carretera construida. Luego recoge la información y entrega los resultados al Ministro.

Supongamos que el Ministro de Obras Públicas toma decisiones a partir de esta información. Probablemente se verá en problemas. Como resulta obvio, en apariencia la empresa A, que construyó la ruta 20, resultó más eficiente que la empresa B, que construyó la ruta 40. Sin embargo, mientras la primera simplemente despejó el terreno, pasó las máquinas niveladoras en línea recta y en horizontal y colocó el material bituminoso, la segunda debió dinamitar partes de montaña, construir terraplenes, hacer puentes para sortear cursos de agua, calcular el trazado de las curvas, construir tramos en subida y en bajada, etc.

Por tanto, en realidad el Ministro no sabe nada sobre cuál de estas empresas es mejor. Por el contrario, la información que posee sirve de poco para ese objetivo. Supóngase que, además, el Ministro tiene información sobre el tipo de tecnología utilizada por cada empresa, por lo cual concluye que la tecnología utilizada por la empresa A es más eficiente que la utilizada por la empresa B. Si todos estos análisis se realizan sin tener en cuenta el terreno en el que cada empresa ha debido trabajar, de nada sirven los indicadores que se construyan.

En educación pasa algo muy similar. **La labor educativa no se desarrolla en el vacío ni con un alumnado homogéneo. Los individuos llegan a la institución educativa en condiciones absolutamente diferentes para recibir y aprovechar el trabajo de los educadores.** Unos llegan luego de una primera infancia vivida con infinidad de estímulos visuales y afectivos, en el marco de un hogar en el que cotidianamente se habla un lenguaje complejo, en el que se lee habitualmente y se le enseñó a reconocer el alfabeto aun antes de conocerlo en la escuela. Otros llegan luego de una primera infancia vivida en hogares en los que existen escasos estímulos visuales, donde el lenguaje es más simple, las prácticas de leer y escribir solo ocurren excepcionalmente, los padres apenas completaron la escuela y no tienen una noción acabada del tipo de ayuda

a brindar a su hijo en sus labores escolares. Unas escuelas trabajan con alumnos que se alimentan adecuadamente, tienen un espacio propio en su vivienda y salen de vacaciones todos los años. Otras trabajan con alumnos que se alimentan mal, que viven en condiciones de hacinamiento y que nunca han viajado más allá de 5 kilómetros de su vivienda. Unos estados o provincias trabajan con población altamente alfabetizada y en un contexto de crecimiento económico, en tanto otros lo hacen con poblaciones con tasas importantes de adultos analfabetos, de familias en condición de pobreza extrema.

Los “terrenos” son absolutamente diferentes. Unas trabajan en la llanura y en línea recta y otras deben sortear infinidad de dificultades. Por tanto, reportar diferencias entre provincias, regiones o tipos de escuela sin analizar las poblaciones con que trabajan, equivale a informar al Ministro de Obras Públicas sobre la “eficacia” de las empresas constructoras de carreteras sin tomar en consideración el tipo de terreno en que cada una de ellas desarrolló su trabajo. ¿Pueden los padres y las autoridades tomar decisiones o formular políticas a partir de los datos así presentados? ¿No sería conveniente advertir al usuario que unos maestros están construyendo en la llanura y en línea recta, mientras otros lo hacen en la precordillera?¹⁹

Cualquier análisis de resultados o indicadores educativos que no tenga en cuenta las características del alumnado, resulta de escaso valor para la toma de decisiones o para extraer algún tipo de conclusión válida que pretenda ir más allá de la mera descripción de que las diferencias existen.

III.3. Algunos ejemplos de conclusiones inapropiadas

A continuación se transcriben algunos párrafos de los reportes nacionales que ilustran el problema de la interpretación inadecuada o sobresimplificada de los resultados, en la medida en que inducen al lector a extraer conclusiones acerca de la efectividad de la enseñanza en los

¹⁹ Algo similar ocurre con el modo que normalmente los economistas interpretan los diferenciales de repetición y extraedad: como ineficiencia del sistema educativo. En realidad indican que en ciertos sectores de la sociedad el terreno para la labor educativa es más complicado, por lo que educar allí lleva más tiempo.

agregados descritos, sin advertir la incidencia de la diferenciación socio-cultural de las poblaciones de dichos agregados.

“En ambos regímenes escolares, en los tres años y en todas las destrezas evaluadas, los planteles particulares urbanos superan en porcentajes de dominio a los fiscales urbanos y estos a los rurales. El análisis de varianza demuestra y puntualiza que, en tercero y séptimo año, la ventaja de los planteles particulares urbanos sobre los fiscales urbanos y de estos sobre los rurales es estadísticamente significativa. En el caso de décimo año, no hay diferencia significativa entre fiscales y rurales; pero los particulares sí tienen una ventaja significativa frente a los otros dos”. Ministerio de Educación y Cultura del Ecuador - EB/PRODEC, (1998).

“...en cualquier grado o disciplina evaluados por el SAEB/97, las medias de proficiencia de la zona urbana son siempre superiores a las de la zona rural. Por otro lado, los alumnos de la zona rural de la región nordeste presentan siempre las menores medias... Los alumnos de la red particular de enseñanza, en todos los grados, disciplinas y regiones, presentan proficiencias más elevadas que los alumnos de la red pública (estatal o municipal)...” Ministerio de Educación - Gobierno Federal INEP 1998.

En realidad, las afirmaciones transcritas en los párrafos anteriores son todas estrictamente ciertas y se limitan a describir datos empíricos. Sin embargo, planteadas así, inducen a concluir que las escuelas privadas son mejores o más efectivas que las públicas urbanas y estas que las rurales. A nadie debiera extrañar que la prensa titulara de esa manera la información sobre los resultados de la evaluación nacional.

A modo de ilustración de cómo cambia la lectura de los datos al introducir la consideración del contexto sociocultural, en la figura 10 se incluyen varios “rankings” de resultados por tipo de escuela construidos a partir de datos de Uruguay. En la columna grisada se consigna el ranking resultante de considerar los resultados globales de cada agregado, sin considerar el tipo de alumnado con que trabajan. Al igual que en el resto de los países de la región, las escuelas privadas tienen resultados notoriamente superiores a las escuelas públicas urbanas y estas respecto a las rurales. Sin embargo, en las columnas siguientes se presentan los “rankings” que resultan de comparar a las escuelas controlando el tipo de alumnado con que trabajan. En la colum-

na correspondiente al “contexto muy favorable”, es decir, aquellas escuelas que atienden a los sectores más favorecidos de la sociedad, las escuelas privadas de la capital del país siguen teniendo mejores resultados que las públicas, pero la diferencia se acortó de 24 puntos porcentuales a solo 5,4. Las escuelas privadas del interior pasaron a estar en el tercer lugar. Las escuelas rurales no aparecen en esta columna porque, como es obvio, no hay escuelas rurales que atiendan a población altamente favorecida.

Cuando se pasa a la siguiente columna, que corresponde ya no a las escuelas de los sectores más favorecidos sino a las de sectores de clase media o media-alta, las escuelas públicas exhiben mejores resultados que las privadas, tanto en la capital como en el interior del país. Tampoco existen escuelas rurales en este sector. Finalmente, cuando se comparan los resultados entre las escuelas que atienden a sectores desfavorecidos de la población, se constata que en realidad las escuelas rurales son “mejores” que las restantes.

Otro ejemplo ilustrativo de conclusiones inapropiadas, en el cual no solo hay un problema de omisión de la consideración del contexto al reportar resultados descriptivos de los agregados, sino que además se formulan explícitamente conclusiones respecto a la eficacia diferencial de los sistemas educativos de las diferentes regiones, aparece en el SAEB/97.

FIGURA 10
URUGUAY. Un “ranking” cambiante en función del contexto sociocultural de las escuelas.
Porcentajes de alumnos suficientes en Matemática

Orden	Global	Contexto Muy Favorable	Contexto Favorable	Contexto Desfavorable
1	Privadas Capital 61,2%	Privadas Capital 71,2 %	Público mt. Urbano 48,0%	Rurales 32,1%
2	Privadas Interior 47,1%	Públicas Capital 65,8%	Públicas Capital 46,4%	Privadas Interior 29,2%
3	Públicas Capital 35,3%	Privadas Interior 61,1%	Privadas Interior 44,4%	Público Int. Urbano 24,3%
4	Público Int. Urbano 27,9%	Público Int. Urbano 45,8%	Privadas Capital 41,6%	Públicas Capital 23,2%
5	Rurales 19,4%			Privadas Capital 15,4%

Fuente: ANEP Unidad de Medición de Resultados Educativos, Uruguay, 1997. *Evaluación Nacional de Aprendizajes en Lengua y Matemática. 6° año de Educación Primaria 1996. Segundo Informe de Difusión de Resultados.*

*“Los mayores porcentajes de alumnos de cuarto grado por encima de ese nivel (175 puntos en Matemática) se encuentran en las regiones Sur y Sureste, **indicando la mayor efectividad de la enseñanza en esas regiones que en las demás** y siendo un claro ejemplo de las desigualdades regionales brasileñas”.*

*“Los resultados del SAEB/97 sobre la proficiencia de los alumnos apuntan, en síntesis, a la existencia de dos tipos de problemas relacionados con la calidad del aprendizaje en nuestro país: el primero está referido a la gran heterogeneidad de resultados de desempeño existente entre los sistemas estaduales de enseñanza, que parece ser un factor de desigualdad más que de simple diferenciación, en la medida en que la dimensión de las diferencias significa, para los alumnos que presentan desempeños más bajos, **una oferta de oportunidades de aprendizaje menos efectiva que aquellas proporcionadas a los alumnos que presentan desempeños más altos**”.* Ministerio de Educación - Gobierno Federal. INEP, 1998.

Cualquier lector medianamente informado sabe que los estados del nordeste brasileño son los más pobres y menos alfabetizados. Por el contrario, los del sur son los que tienen un mayor desarrollo económico y cultural. Por tanto, si bien es relevante mostrar que los niños y jóvenes que viven en los primeros están en clara desventaja frente a los segundos, y que existe una fuerte inequidad, es delicado atribuir esas diferencias a la efectividad de los sistemas de enseñanza, de los currículos o del trabajo de los educadores.

El problema es que los diferentes datos no se conectan entre sí. Se analizan por un lado los resultados por jurisdicción, área geográfica y/o tipo de escuela, y por otro, la relación de los mismos con el origen social de los estudiantes. Falta aún dar el paso de interrelacionar estas informaciones, complejizar el análisis y la presentación de los datos, advertir al usuario de estas complejidades y cuidados en la interpretación, buscar modos para establecer comparaciones entre sectores del sistema educativo que trabajan en “terrenos” similares para poder formular conclusiones válidas respecto a la efectividad de los distintos tipos de escuela, currículos o sistemas de enseñanza.

El tratamiento de los resultados de las evaluaciones nacionales sin una adecuada consideración del contexto social y el tipo de alumnado con que trabajan las escuelas es más grave cuando las evaluaciones tienen implicancias fuertes para las escuelas, como ha ocurrido en el caso de Chile. En el reporte de los resultados nacionales correspondientes a 8º Básico evaluados en el año 1995, se publica la lista de los 20 mejores

establecimientos del país, que son aquellos que tienen los puntajes promedios más altos. De ellos, 19 son colegios privados, 12 están ubicados en los barrios acomodados de Providencia y Las Condes, en la capital del país (ver figura 11).

En el mismo informe se construye una clasificación de los establecimientos por “nivel socioeconómico”, con tres categorías (A, B y C) en función del gasto educacional promedio mensual y del nivel de educación de los padres de los alumnos. La distribución de los tipos de colegios según su nivel socioeconómico muestra claramente que los colegios privados pertenecen a las categorías A y B, en tanto que la enorme mayoría de los colegios municipales pertenece a la categoría C (figura 12). Sin embargo, a pesar de estas constataciones, toda la presentación de los resultados está casi exclusivamente centrada en la comparación entre los puntajes promedio globales de establecimientos municipales, subvencionados y particulares pagados.

FIGURA 11
CHILE. Establecimientos con los veinte promedios nacionales más altos en matemática y castellano

Nº ORDEN	DEPENDENCIA	DEPTO. PROVINCIAL	COMUNA	ESTABLECIMIENTO
1	Particular pagado	Cautín	Villarrica	Colegio Alemán
2	Municipal	Santiago Oriente	Providencia	Liceo Carmela Carvajal de Prat
3	Particular pagado	Talca	Constitución	Colegio Constitución
4	Particular pagado	Santiago Oriente	Las Condes	Anglo American International
5	Particular pagado	Santiago Oriente	Vitacura	Colegio Los Andes
6	Particular pagado	Santiago Oriente	Las Condes	Colegio Cumbres
7	Particular pagado	Santiago Oriente	Las Condes	Colegio La Girouette
8	Particular pagado	Elqui	La Serena	Colegio La Serena
9	Particular pagado	Valparaíso	Valparaíso	Seminario San Rafael
10	Particular pagado	Valparaíso	Viña del Mar	Seminario San Rafael
11	Particular pagado	Santiago Oriente	Las Condes	Colegio La Abadía
12	Particular pagado	Santiago Oriente	Las Condes	Colegio Nuestra Señora del Pilar
13	Particular pagado	Santiago Oriente	Providencia	Saint Gabriel's English School
14	Particular pagado	Santiago Oriente	Las Condes	Villa María Academy
15	Particular pagado	Santiago Oriente	Las Condes	Colegio del Sagrado Corazón
16	Particular pagado	Santiago Oriente	Providencia	Saint Gabriel's English School
17	Particular pagado	Santiago Oriente	La Reina	Nuestra Señora del Camino
18	Particular pagado	Santiago Oriente	Providencia	Colegio Huelén
19	Particular pagado	Osorno	Osorno	Instituto Alemán de Osorno
20	Particular pagado	Santiago Oriente	Providencia	Colegio St. Thomas Morus

Fuente: República de Chile, Ministerio de Educación, SIMCE, 1996. *Resultados 95 - 8º año Básico.*

FIGURA 12
CHILE. Distribución de los establecimientos por nivel socioeconómico según tipo de administración

Nivel socioeconómico	Particular Pagado		Subvencionado		Municipal	
	Cantidad	Porcentaje	Cantidad	Porcentaje	Cantidad	Porcentaje
A	287	60,4	66	4,9	8	0,3
B	188	39,6	517	38,5	342	12,3
C	0	0,0	759	56,6	2.434	87,4
Total	475	100,0	1.342	100,0	2.784	100,0

Fuente: República de Chile, Ministerio de Educación, SIMCE, 1996. *Resultados 95 - 6° año Básico.*

En el caso chileno, esta debilidad en la presentación de los resultados se subsana en parte desde la medición de 1997, en la que se incorpora el dato del avance de cada establecimiento respecto a su propio resultado en 1995 (véase el capítulo IV). Esto resulta mucho más apropiado que la comparación con los resultados brutos de otros establecimientos que, en general, trabajan con poblaciones muy diferentes.

III.4. La consideración del contexto sociocultural

Así como la manera válida de construir indicadores comparativos de la eficiencia de las empresas constructoras de carreteras sería medir su tasa de avance sobre terrenos similares, **la única manera válida de construir indicadores que sirvan para sustentar afirmaciones o decisiones sobre la efectividad de las instituciones o sistemas educativos es a través de algún tipo de categorización de los “terrenos” en que trabajan, es decir, de las características socioculturales de los hogares de los que provienen sus alumnos.**

Un ejemplo que ilustra uno de los caminos posibles para abordar este desafío es la construcción de categorías de “contexto sociocultural de las escuelas”²⁰, con la finalidad de comparar los resultados entre escuelas que tienen alumnados similares. La constatación de que el ordenamiento de los resultados de las pruebas en los diferentes tipos de escuela siguen el mis-

²⁰ Es importante destacar que el “contexto sociocultural” es una propiedad de la escuela o grupo de alumnos, no de los individuos. Lo que explica los resultados no es principalmente el origen social individual de los estudiantes, sino la composición sociocultural del grupo escolar o de la escuela a la que pertenecen. Esta advertencia es de gran importancia, porque en algunos análisis se compara, por ejemplo, los resultados de individuos de nivel socio-

mo ordenamiento que los indicadores sociales dio lugar en Uruguay a una clasificación de los establecimientos educativos en cinco categorías según su “contexto sociocultural”: ‘muy favorable’, ‘favorable’, ‘medio’, ‘desfavorable’ y ‘muy desfavorable’. Esta categorización fue construida a partir de dos índices: uno de carácter cultural, elaborado a partir de los niveles educativos de los hogares de los alumnos, y otro de carácter económico, elaborado a partir de los niveles de equipamiento de dichos hogares.

A partir de esta construcción, los resultados se presentan siempre controlando el contexto sociocultural, como en la figura 10 es decir, mostrando los resultados por tipo de escuela, región o departamento dentro de una misma categoría de contexto sociocultural. Asimismo, en la devolución a las escuelas se entrega a cada una de ellas el resultado de sus alumnos, junto con el correspondiente a las restantes escuelas del mismo contexto sociocultural, de modo que no solo se comparen con la media nacional sino también con escuelas de similar composición social (figura 13).

FIGURA 13
URUGUAY. Cuadro tipo de entrega de resultados a cada escuela

	Escuela	Escuelas del mismo contexto sociocultural del Departamento	Escuelas del mismo contexto socio-cultural del país	Total Nacional
Porcentaje de alumnos suficientes en Lengua Materna				
Porcentaje de alumnos suficientes en comprensión de texto argumentativo				
Porcentaje de alumnos suficientes en comprensión de texto narrativo				
Porcentaje de alumnos suficientes en reflexiones sobre el lenguaje				

cultural “bajo” que pertenecen a escuelas públicas, por un lado, ya privadas, por otro. Normalmente las diferencias no serán significativas y, si lo son, los resultados serán mejores entre quienes pertenecen a escuelas privadas, pero ello como resultado de que en dichas escuelas los individuos de NSE “bajo” son minoría, mientras que en las primeras son mayoría. En otras palabras, cuando los individuos de origen sociocultural desfavorecido son minoría en un grupo constituido mayoritariamente por individuos de mejor “respaldo” cultural familiar, los primeros se verán favorecidos por sus pares. Hay un efecto del nivel del grupo que mejora los aprendizajes. En cambio, cuando el grupo está mayoritariamente constituido por individuos de origen social desfavorecido, la minoría con mejor nivel educativo familiar tenderá a tener resultados inferiores a los esperables. Estas constataciones, por otra parte, indican la importancia de favorecer la heterogeneidad social en las escuelas.

III.5. Los dilemas que la relación entre aprendizajes y contextos sociales plantea a la difusión de resultados

El tratamiento y presentación de los resultados de las evaluaciones nacionales a partir de la categorización de los contextos sociales genera un nuevo problema: se corre el riesgo de acentuar una visión determinista en la sociedad y en los educadores, visión que podría ser ilustrada a través de frases del tipo “esto es lo máximo que se puede lograr en este medio” o “en este contexto no se puede esperar ni exigir más de los niños”. En otras palabras, se corre el riesgo de establecer o fortalecer un sistema de expectativas diferenciadas (“exijamos más en los sectores medios y altos, exijamos menos en los sectores populares, porque lo que pueden alcanzar es diferente”). Muchas personas vinculadas al sector educativo y a los sistemas de evaluación temen, justificadamente, que enfatizar la presentación de resultados diferentes por contexto social opere como “coartada” o justificación de los resultados insuficientes para los maestros que trabajan en esos contextos.

Pero, por el otro lado, ignorar el contexto puede dar lugar a conclusiones espurias y, más aún, puede generar una gran frustración e impotencia en las escuelas y docentes que trabajan en los medios más desfavorecidos, en la medida en que aparezca como que toda la responsabilidad por los pésimos resultados es de la escuela. Incluso puede dar lugar a que ciertas escuelas queden injustamente deslegitimadas ante sus comunidades y tener efectos paralizadores. La evaluación puede ser percibida como injusta, porque responsabiliza al maestro y a la escuela por problemas que, al menos en parte, provienen del contexto. Volviendo a la parábola de las carreteras, es como si la empresa que trabajó en la precordillera quedara ubicada, en un ranking de metros promedio de construcción por día, por debajo de la que trabajó en la llanura. Lo sentiría como fuertemente injusto.

La introducción de la noción de contexto social “alivia” el peso de la responsabilidad, ayuda a percibir la evaluación como equitativa²¹ y a comprender que hay factores externos al sistema educativo que inciden en los logros y que la educación no es omnipotente. En definitiva, ayuda a establecer algo que es estrictamente cierto: parte de las desigualdades que se registran en el sistema educativo son reflejo de las desigualdades que existen en la sociedad.

²¹ véase Benveniste, L., 2000.

Algunas formas de enfrentar el dilema anterior son:

- Mostrar en forma sistemática, a través de la difusión de resultados, que incluso al interior de los diferentes contextos sociales, hay escuelas que logran mejores resultados que otras. En otras palabras, que si bien el contexto social tiene un peso específico sobre los resultados, no hay determinismos absolutos y existe un espacio propio de la acción pedagógica y escolar que es necesario potenciar. Esta puede ser además una forma de motivar para que al interior del sistema educativo se haga todo lo posible por lograr mayor equidad en la distribución social del conocimiento y por mejorar las capacidades de los niños de origen más pobre.
- Colocar fuertemente el énfasis en los datos longitudinales y no en las comparaciones transversales entre escuelas. El esquema ideal sería tener una medición inicial y otra final en un mismo año lectivo. Por esta vía se puede eliminar en buena medida el efecto del contexto, centrando la medición y el reporte de resultados en lo que los alumnos y grupos avanzan a lo largo de un año respecto a su propio punto de partida. Con buenos instrumentos de medición de prácticas y enfoques de enseñanza, una aproximación de este tipo debería mejorar la capacidad explicativa de los “factores escolares” y, por tanto, la acumulación de conocimiento en este terreno²².
- En los países que trabajan con evaluaciones censales, centrar el reporte de resultados en los cambios que se registran en cada escuela con respecto a la medición anterior. En esta alternativa, que se aplica en Chile, de todos modos es necesario tener algún control sobre la composición social del alumnado, porque los cambios en los resultados de una escuela pueden obedecer a cambios en el tipo de alumnado que recluta.

²² Obviamente un dispositivo de este tipo es más costoso y más complejo en cuanto al diseño de las pruebas. Una experiencia de este tipo que debiere ser observada, está iniciándose en la provincia de Buenos Aires, con el apoyo de OREALC.

CAPÍTULO IV

LA UTILIDAD PEDAGÓGICA DE LOS REPORTE DE RESULTADOS

De lo expuesto en el primer capítulo se desprende que uno de los caminos principales a través de los cuales se espera que la información producida por los sistemas nacionales de evaluación tenga un efecto de mejora del sistema educativo, es su utilización por parte de los equipos directivos y docentes en las escuelas para la toma de decisiones de carácter técnico-pedagógico que les permitan mejorar sus prácticas de enseñanza.

La elaboración de reportes de resultados para uso de los equipos docentes ha sido uno de los aspectos más trabajados de la difusión de resultados en los países de la región. Argentina, Chile, Ecuador, Paraguay y Uruguay han producido informes específicos dirigidos a los docentes, en los que se explica con cierto detalle qué fue lo evaluado por las pruebas aplicadas y cuáles fueron los principales problemas y dificultades detectados.

Cuatro han sido las principales estrategias empleadas para este objetivo:

- La producción y distribución a todas las escuelas de materiales y documentos didácticos derivados de los resultados generales de la evaluación, independientemente de que la misma hubiese sido de tipo censal o muestral.
- La entrega a cada establecimiento de los resultados de sus propios alumnos, en el marco de operaciones de carácter censal.
- La realización de operaciones de evaluación muestrales con aplicación y corrección autónoma de las pruebas en el universo de escuelas.
- La implementación de talleres de difusión de resultados y programas de capacitación en servicio a partir de los mismos, dirigidos a docentes, directores y/o supervisores.

IV.1. Materiales de orientación didáctica elaborados a partir de los resultados

Argentina ha sido probablemente el país que mayor énfasis puso, desde el comienzo de sus evaluaciones nacionales, en la elaboración de materiales de análisis de las competencias más y menos logradas en las diferentes áreas evaluadas, a través de los denominados “cuadernos metodológicos” que publicó sistemáticamente desde la primera evaluación nacional realizada en 1993.

Dichos cuadernos, que en cierto modo han servido de modelo a otros países de la región, tienen como finalidad aportar a la reflexión didáctica y disciplinar de los maestros a partir de ejemplos concretos de ítems y sus resultados, enfatizando en las actividades que fueron logradas en mayor proporción, así como en las que resultaron más difíciles.

La figura 14 ilustra el tipo de análisis de los ítems que se entrega a los maestros. Según se puede apreciar, se aporta un análisis detallado de la operación cognitiva requerida para resolver la actividad, así como de

FIGURA 14
ARGENTINA. Análisis de ítems

7° año EGB / Matemática	
<p>18. Si al descarozar 20 kg de ciruelas 15 kg son de pulpa, ¿qué porcentaje de las ciruelas es la pulpa?</p> <p>A) 5 B) 25 C) 35 • D) 75</p>	
Respuesta correcta: D	
<p>Resultado: El 22% de los alumnos lo resolvió correctamente.</p> <p>Contenido: Números y operaciones.</p> <p>Capacidad: Resolver problemas.</p> <p>Operación requerida y evaluada: Aplicar el concepto y cálculo de porcentaje para resolver un problema.</p>	<ul style="list-style-type: none"> • La elección de la respuesta correcta D permite suponer que los alumnos comprendieron el problema, reconocieron los pasos a seguir, plantearon correctamente la situación y calcularon el porcentaje solicitado. • La elección de la respuesta NO correcta A indica una incorrecta lectura del enunciado, los alumnos que la eligieron operan restando los datos, sin advertir que se trata de diferentes magnitudes. • La elección de la respuesta NO correcta B indica que estos alumnos pueden calcular porcentaje, responden al porcentaje de carozos y no de pulpa, marcando así un error de interpretación. • La elección de la respuesta NO correcta C señala que los alumnos parecen no dominar el concepto de porcentaje, sumando todos los datos que aparecen en el enunciado.

Fuente: Ministerio de Cultura y Educación de la Nación - República Argentina. Dirección Nacional de Evaluación, 2000. *Operativo Nacional de Evaluación 1999*; pág. 56.

las dificultades que podían dar lugar a la elección, por parte de los alumnos, de los distintas alternativas no correctas o “distractores”.

Esta manera de reportar los resultados permite ofrecer una visión concreta acerca del tipo de tareas a las que se vieron enfrentados los alumnos en las pruebas. Asimismo, la labor de explicitar el tipo de operación cognitiva involucrada en la resolución de la actividad y el tipo de dificultad que puede haber llevado a un alumno a seleccionar una alternativa no correcta, constituyen información valiosa para el análisis de resultados por parte de los docentes. Simultáneamente, cabe señalar como debilidad una atomización en la visión de las competencias a lograr, en la medida en que el análisis se apoya en ítems individuales, cuando estos en sí mismos no necesariamente dan cuenta de la complejidad de una competencia. Esta dificultad se presenta especialmente en el área de Lengua. Un camino adecuado para afrontar esta debilidad es el adoptado por el SAEB, mencionado en el capítulo II: reportar escalas de proficiencia que describen lo que los alumnos son capaces de hacer en los distintos tramos de puntaje de las pruebas.

Si bien el análisis de la calidad y sustento conceptual desde el cual se formulan las recomendaciones en cada país se encuentra fuera de los alcances del presente trabajo, la impresión general es que, al igual que con las pruebas, existe una fuerte heterogeneidad entre los países de la región en cuanto a la actualización de los enfoques disciplinarios y didácticos desde los cuales son elaborados tanto los instrumentos de evaluación como los cuadernos de carácter didáctico.

En el caso de Ecuador, por ejemplo, el análisis del error y dificultad identificada es de carácter muy general (figura 15). Este último ejemplo es bastante representativo de lo que se hace en la mayoría de los países al proponer ejemplos de ítems, que en general van acompañados por un escaso nivel de conceptualización didáctica y disciplinar.

IV.2 Entrega de resultados por establecimiento

La entrega a cada establecimiento de sus propios resultados, en el marco de evaluaciones nacionales de carácter censal, ha sido llevada adelante en Argentina, Chile y Uruguay, y existen también experiencias a nivel subnacional como en el Estado de Paraná en Brasil, que ha desarrollado su propio sistema estadual de evaluación con esta finalidad, o en algunas provincias argentinas como Mendoza. En ambos casos se trata de evalua-

FIGURA 15
Ecuador. Análisis de ítems

Un automóvil recorre 80 km en una hora. ¿Cuántos kilómetros avanzará en una hora y tres cuartos si mantiene la misma velocidad?			
	Opciones	% de elección	Observaciones
A.	160 km	22.20	
B.	140 km	34.73	
C.	120 km	28.35	no tienen claro lo que se pide.
D.	60 km	6.94	
ERROR: identificación incorrecta de la información.			
DIFICULTAD IDENTIFICADA: insuficiente comprensión de la proporcionalidad y poca capacidad de análisis.			

Un equipo de 4 personas en 8 horas tiende 320 metros de redes eléctricas. ¿Cuántas personas se debe emplear para que tiendan 2.400 metros en 12 horas?			
	Opciones	% de elección	Observaciones
A.	45 personas	12.89	
B.	30 personas	14.43	
C.	20 personas	40.66	
D.	6 personas	21.09	consideran los datos de personas y horas y no tienen los referidos a metros.
ERROR: retención parcial de la información.			
DIFICULTAD IDENTIFICADA: poca capacidad de análisis.			

Fuente: Ministerio de Educación y Cultura - Ecuador. EB/PRODEC - Sistema Nacional de Evaluación de la Educación APRENDO, 1997. *Análisis de las Pruebas Aprendo 1996 y de sus resultados*, pág. 25

ciones de carácter censal, por lo que todos los establecimientos educativos de la provincia o estado reciben sus resultados²³.

La ventaja principal de entregar resultados a cada establecimiento es que ello tiene la potencialidad de involucrarlos directamente en el proceso de evaluación. Es altamente probable que la predisposición a utilizar los resultados sea mayor cuando el establecimiento participó de la evaluación y tiene a la vista los resultados de sus propios alumnos. Por el contrario, es probable que la predisposición a hacer uso de los resultados sea mucho menor cuando simplemente se recibe un informe de resultados nacionales, pero se carece de un punto de referencia acerca del grado en que los propios alumnos dominan los aprendizajes evaluados.

En el caso del Estado de Paraná en Brasil merece especial destaque la pauta que se entrega a cada establecimiento con el fin de que el equipo docente elabore un informe de interpretación de sus resultados (figura 16). La estrategia es interesante, porque en cierto modo “obliga” a cada equipo a analizar sus resultados y realizar por escrito una interpretación de los mismos y un plan de intervención y mejora.

Un matiz importante entre los países en cuanto a su estrategia de devolución de resultados hacia los cuerpos docentes es el relativo al acceso a las pruebas aplicadas. En la mayor parte de los países las pruebas no tienen carácter público y únicamente se informa acerca de los contenidos, competencias u objetivos que fueron evaluados y se divulgan ejemplos del tipo de ítems aplicados.

En los casos de Uruguay y Paraná, las pruebas son públicas (en tanto los resultados de cada establecimiento, según se indicó en el capítulo 1, son confidenciales). Cada escuela recibe ejemplares de las pruebas y manuales de interpretación de las actividades incluidas en las mismas –con una estructura muy similar a la ilustrada para Argentina y Ecuador, pero abarcando la totalidad de los ítems de cada prueba–, así como los resultados de los alumnos del establecimiento en cada una de dichas actividades. En el caso de Chile, a partir de septiembre de 1999 el Ministerio de Educación comenzó a hacer públicas las pruebas que habían sido aplicadas en 1998²⁴.

²³ La devolución de resultados personalizados por establecimiento también ha sido realizada en algunos países como Ecuador o Bolivia, cuyas evaluaciones tienen carácter muestral. En este caso la proporción de establecimientos que pueden recibir sus resultados es bastante reducida, por lo que la estrategia pierde buena parte de su potencialidad de impacto sobre el sistema.

²⁴ De todos modos es preciso reconocer que hacer públicas las pruebas trae aparejadas dificultades a la hora de equiparar nuevas pruebas para mediciones comparables que se desee realizar en el futuro.

FIGURA 16
ESTADO DE PARANÁ/BRASIL. Guía para el análisis de los resultados
en los establecimientos educativos

- *“Examine cada pregunta relativa a un determinado contenido. Generalmente son tres preguntas sobre cada uno de los temas o un poco más. Vea el grado de facilidad de cada una de ellas y observe la dificultad. Si 60% de los alumnos acertaron, se trata de una pregunta de dificultad media, según la observación debajo de la Tabla.*
- *Examine cada ítem, ejemplificando con errores típicos que fueron observados en las pruebas de los alumnos de su escuela para cada conjunto de ítems relativos a un contenido. Haga comentarios sobre cada conjunto e intente explicar las situaciones que fueron identificadas.*
- *Analice, comparativamente, el desempeño de los alumnos en cada uno de los contenidos, teniendo en consideración los datos del Establecimiento, del Municipio, de la Región y del Estado.*
- *A partir de los ítems más difíciles observados en su escuela, vea cómo ocurre la distribución por Municipio, Región y Estado. Haga una reflexión crítica y apunte los problemas revelados a través del desempeño de los alumnos en la prueba que exigirían una mayor intervención de los profesores.*

CONCLUSIONES DEL RELATORIO

Ahora usted dispone de elementos para presentar sus conclusiones en el Relatorio de la Prueba de Contenidos Básicos. Reúname con sus colegas que hicieron el relatorio y elabore, en grupo, las conclusiones del mismo.

- *Haga un comentario general sobre las expectativas de la escuela en relación al desempeño de los alumnos.*
- *Manifiéstese críticamente en relación a las características de la prueba, presentando sugerencias sobre posibles aspectos del Currículo. Haga un paralelismo entre lo que se pretende medir y lo que la escuela realmente abordó en el proceso de enseñanza de las clases evaluadas.*
- *Presente los puntos críticos revelados por el desempeño de los alumnos en las pruebas.*
- *Apóyese en sus observaciones anteriores, pero procure no ser reiterativo.*
- *Dé la opinión consensuada de la escuela sobre la Evaluación, el desempeño de los alumnos y la programación para la Enseñanza Fundamental y Media, presentando sugerencias y teniendo en cuenta su experiencia como Educador(a) y su vivencia de la presente Evaluación.*
- *Comente sobre otros aspectos que juzgue necesario, con base en las informaciones que esta evaluación le proporcionó.*

UTILIZACIÓN DE LOS RESULTADOS EN LA ESCUELA

Haga uso del relatorio y de los resultados presentados para su planeamiento de actividades, con vistas al mejoramiento de los procedimientos didáctico-pedagógicos en la orientación del aprendizaje. Procure divulgar el relatorio final de la Escuela entre los demás profesores y los miembros del Colegio y de la Comunidad”.

Fuente: Gobierno del Estado de Paraná. 1998 *Evaluación del Rendimiento Escolar. Resultados de la Evaluación del Rendimiento Escolar 4ª Serie de la Enseñanza Fundamental.*

Un aspecto interesante del caso chileno, mencionado en el capítulo anterior, es la presentación sistemática de información sobre los avances de cada establecimiento en relación a la medición anterior. Esta información se publica tanto en los medios de prensa como en Internet.

IV.3. Evaluaciones de carácter muestral con aplicación y corrección autónoma de las pruebas en el universo de escuelas

Una tercera estrategia desarrollada en la región con el objetivo de que las operaciones de evaluación afecten positivamente el trabajo docente, ha sido la siguiente:

- Realizar una evaluación en una muestra representativa a nivel nacional.
- Distribuir luego los instrumentos de evaluación a todas las escuelas del país, junto con manuales de aplicación y corrección de las pruebas con orientaciones para que estas tareas fuesen realizadas en forma voluntaria y autónoma en las escuelas.
- Distribuir finalmente los resultados de la muestra nacional para que cada escuela pudiese comparar sus propios resultados con los nacionales.

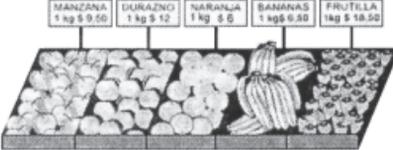
Este tipo de estrategia fue desarrollada en Uruguay en 3^{er} año de educación primaria en 1998, con una prueba de respuesta abierta que integraba en un único instrumento contenidos de Lenguaje, Matemática y Ciencias, y en 1999 con pruebas de selección múltiple en Lenguaje y Matemática en 6^o año de educación primaria.

En la figura 17 se incluye un ejemplo de los modelos de corrección de las pruebas de respuesta abierta empleadas en 1998. Para cada una de las actividades se indicaba a los maestros el proceso de resolución de la actividad, una codificación y sistema de puntajes para los distintos tipos de respuestas posibles, así como ejemplos de respuestas de los niños tomados de la etapa piloto de aplicación de los instrumentos. El manual incluía además los puntos de corte para que cada maestro pudiese establecer los niveles de suficiencia en su grupo, para luego incluirlos en una columna que se dejaba en blanco en los cuadros de resultados nacionales, para que cada escuela pudiese consignar allí los correspondientes a sus alumnos. También ofrecía información sobre los resultados nacionales en cada una de las actividades de la prueba.

La principal fortaleza de esta estrategia es que involucra fuertemente a los maestros, dándoles participación directa en la evaluación y propiciando el desarrollo de una cultura de la evaluación de los aprendizajes

FIGURA 17
URUGUAY - Un Ejemplo de Pautas para la Corrección Autónoma de Pruebas de Respuesta Abierta en 3er Año de Primaria

Otra mamá fue al almacén a comprar frutas para hacer un licuado.



COMPETENCIA - Resolución de problemas.

2. Compró 3 kg de frutas diferentes (1 kg de cada una). Gastó menos de \$25.
 ¿Qué frutas compró?

OBJETIVO - Encontrar como mínimo una respuesta entre dos posibles.

PROCESO POSIBLE PARA LLEGAR A UNA RESPUESTA CORRECTA

- Reconoce la suma como estrategia para relacionar los números de los datos dados.
- Suma los tres precios más bajos.
- Elige una opción cuyo resultado da una cantidad menor a \$25.
- Relaciona los tres precios sumados en dicha suma con el nombre de las frutas.

TABLA DE VALORACION

CODIGO	TIPO DE RESPUESTA	PUNTAJE
1	Durazno, naranja, banana (\$24.5). Manzana, naranja, banana (\$22).	3
2	Nombra 2 frutas, pero respeta el límite de gasto.	1
3	Nombra 3 kilos de una sola fruta (naranja, o banana) respetando el límite de gasto.	1
4	a) Nombra 3 frutas, pero excede los \$25. b) 3 kilos de una sola fruta, pero excede los \$25	0
5	Nombra más de 3 frutas.	0
8	Respuesta inadecuada.	0
9	Respuesta ilegible.	0
0	Ausencia de respuesta.	0

Ejemplos de respuestas dadas por los niños:

Código 2 "no puede comprar 3 kg de frutillas, puede comprar naranjas y duraznos"
 Código 3 "6+6+6=18"
 Código 4 "frutilla, durazno, manzana"
 Código 5 "banana, frutilla, durazno, banana, manzana"
 Código 5 "naranja, banana, frutilla, durazno, manzana, uva, pomelo, sandía, mandarina"
 Código 8 "19.50+12=3070 15+16.50=5700 16.50-1850=100"
 Código 8 "\$34"

Fuente: URUGUAY-ANEP, Unidad de Medición de Resultados Educativos, 1998. Evaluación Nacional de Aprendizajes en 3er año de Educación Primaria 1998. Manual de Corrección de la Prueba y Procesamiento del cuestionario para madres, padres o tutores; pág. 29 - 31.

con instrumentos sistemáticamente validados. Sus debilidades son que no hay garantías de que todas las escuelas realmente se involucren en la evaluación y que con este tipo de instrumentos y procedimientos es prácticamente imposible realizar mediciones comparables en el tiempo.

IV.4. Talleres de difusión de resultados y programas de capacitación en servicio

Un cuarto tipo de estrategia de difusión de resultados hacia las escuelas y docentes ha sido la realización de actividades de capacitación o sensibilización a través de talleres, seminarios y cursillos, en general dirigidos a supervisores, directores y/o maestros y profesores. Sin embargo, no se dispone de información sistemática sobre las características de estas acciones ni sobre sus contenidos y cobertura y, por otra parte, no formaba parte de los objetivos de este trabajo realizar un análisis de las mismas. No obstante, vale la pena mencionar que es probable que la distribución de reportes escritos no sea una vía suficientemente efectiva para garantizar el aprovechamiento de la información por los equipos docentes. El desarrollo de instancias presenciales o a distancia –aprovechando las posibilidades que aportan las nuevas tecnologías– en las que los resultados son explicados directamente, se realizan ejercicios de comprensión e interpretación de los mismos y se presentan análisis de especialistas en didáctica y en las disciplinas evaluadas, sin duda amplía enormemente el aprovechamiento de la información en las escuelas.

Probablemente en el futuro las Unidades de Evaluación deban incorporar este tipo de programas a su misión institucional y no concebirla únicamente en términos de producción de información válida y confiable que otros sabrán utilizar. Este aspecto es relevante, porque normalmente el grueso de la energía de una Unidad de Evaluación está destinado a la realización de los operativos de evaluación, al procesamiento de información y a la producción de informes. Se supone que otras Unidades dentro de los Ministerios de Educación tienen como misión específica la organización de programas de formación o capacitación en servicio. Sin embargo, **no siempre existen las instancias de coordinación y trabajo en común para que los resultados de las evaluaciones se incorporen a los programas de capacitación. Con esto no se pretende decir que las Unidades de Evaluación deban asumir esta tarea –aunque es una posibilidad y algunas lo han hecho–, pero sí que deben preocuparse por establecer negociaciones o estrategias deliberadas para que los**

resultados de las evaluaciones sean incorporados de alguna manera a los programas de formación o capacitación en servicio.

Por otra parte, una tarea a realizar sería la sistematización de las experiencias de capacitación realizadas en la región a partir de resultados de las evaluaciones y la identificación de las modalidades de trabajo más significativas y relevantes para los maestros.

IV.5. ¿Cómo usar las evaluaciones para mejorar las prácticas de enseñanza?

A lo largo de este capítulo se han presentado las diferentes modalidades o estrategias empleadas en la región para que los resultados de las evaluaciones tengan algún impacto en la mejora de las prácticas de enseñanza en las escuelas. De todos modos, sobre este punto existen interrogantes que por el momento no tienen respuesta:

- ¿Qué proporción de los docentes en cada país ha visto los reportes de resultados y otras publicaciones derivadas y qué proporción las ha leído?
- ¿Qué tanto han comprendido los reportes quienes los han leído?
- ¿Cuán útil o enriquecedora les ha resultado la información contenida en ellos?
- ¿Tienen los docentes otras demandas hacia las Unidades de Evaluación?

Responder a estas preguntas requeriría de la realización de estudios de caso que involucren tanto trabajo de encuestas como indagaciones de carácter cualitativo²⁵.

De todos modos es oportuno dejar planteadas algunas hipótesis acerca de cómo hacer más significativos los resultados para la mejora de las prácticas de enseñanza en las escuelas:

- Es sumamente importante que los maestros conozcan al detalle no solo la definición “formal” de las competencias evaluadas, sino también el tipo de actividades que los alumnos debieron responder.
- Se requiere avanzar hacia reportes más elaborados en términos de interpretaciones didácticas de los resultados, evitando los reportes exclusivamente volcados a los resultados numéricos.

²⁵ Sobre este tema, en principio, el único trabajo en la región es el realizado por Luis Benveniste, quien estudió con una finalidad distinta los sistemas nacionales de Chile, Argentina y Uruguay, pero que incluyó una serie de visitas a escuelas y entrevistas en profundidad con docentes, directivos y supervisores respecto a la utilidad en general de los sistemas de evaluación, no específicamente de los reportes (véase Benveniste, L., 2000).

- Es preciso avanzar en el camino de aportar descripciones etnográficas, derivadas de trabajos de investigación de tipo cognitivo –que formarían parte del proceso de diseño y pilotaje de las pruebas– sobre los procesos reales de resolución de las actividades y problemas por parte de diferentes tipos de alumnos, así como el tipo de dificultades y errores que realmente tienen.
- Para que los resultados sean más significativos es importante que los maestros tengan información acerca del desempeño de sus propios alumnos en las pruebas, lo que puede lograrse tanto por la vía de las aplicaciones de carácter censal como con la aplicación autónoma de las mismas pruebas o de pruebas equivalentes.
- Otra línea de trabajo promisoría para que los sistemas de evaluación tengan un impacto directo en la mejora de las prácticas de enseñanza, consistiría en aportar a los docentes instrumentos de evaluación sistemáticamente elaborados y validados, que ellos puedan emplear de manera autónoma de acuerdo a sus necesidades y criterio profesional. Probablemente esta sea una de las principales demandas de los maestros a las Unidades de Evaluación, dado que la construcción de instrumentos de evaluación de aprendizajes es una tarea compleja y demanda una enorme dedicación de tiempo. Poner a disposición de los docentes variadas alternativas de actividades de evaluación para conocimientos y competencias específicas, incluyendo información acerca del desempeño que en cada una de dichas actividades tuvieron los estudiantes de distintos niveles y sectores del sistema educativo, tal vez sería uno de los principales aportes que las Unidades de Evaluación podrían prestar al fortalecimiento de una cultura de la evaluación en las escuelas.

CAPÍTULO V

LOS REPORTES DE RESULTADOS COMO APORTE A LA FORMULACIÓN DE POLÍTICAS EDUCATIVAS

En el capítulo anterior se analizaron las estrategias desarrolladas para que los resultados de las evaluaciones nacionales tengan algún impacto en las prácticas de enseñanza en las escuelas. **En este capítulo el foco está puesto en otra de las finalidades explícitamente enfatizadas en las definiciones de objetivos de los sistemas nacionales de evaluación:**

el aporte de información relevante y oportuna para la toma de decisiones y la formulación de las políticas educativas.

Se supone que un componente central de la misión de las Unidades de Evaluación es aportar información acerca del estado de los aprendizajes, así como sobre los factores o variables que inciden en dichos aprendizajes, a diferentes “tomadores de decisiones” en distintos lugares de los sistemas educativos: responsables a nivel central o regional del diseño de currículos, unidades a cargo de la producción de materiales educativos, instituciones formadoras de docentes, responsables de la toma de decisiones en materia de inversiones y de diseño de programas y políticas.

Al respecto, lo primero que debe ser destacado es que los sistemas de evaluación realizan un aporte central y significativo a la toma de decisiones y a la formulación de las políticas educativas por el mero hecho de producir información sobre lo que está ocurriendo con los aprendizajes de los alumnos. Esta función es trascendente, porque otorga centralidad a los logros educativos y los coloca en la agenda pública.

En ausencia de esta información las discusiones de política educativa corren riesgos tales como centrarse exclusivamente en la evolución de la matrícula y el acceso al sistema, en pugnas por espacios de poder e intereses de tipo corporativo o en debates de carácter puramente teórico. Todo ello, como resulta obvio, de todos modos continúa sucediendo aún en presencia de información sobre los aprendizajes. Pero por lo menos la presencia de esta información permite intentar centrar las discusiones en la razón de ser de los sistemas educativos: garantizar a todos los niños el acceso a ciertos conocimientos y capacidades que son indispensables para su desarrollo individual y social.

Ahora bien, a partir de la lectura de los informes nacionales, es posible identificar dos debilidades principales sobre las que sería preciso trabajar más para que la información producida tenga mayor incidencia en las decisiones de política educativa:

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

V.1. Insuficiente análisis, interpretación y conceptualización

La mayoría de los informes nacionales se caracterizan por su carácter extremadamente descriptivo. En general se limitan a relatar los datos emergentes de los primeros procesamientos, prácticamente sin analizar conexiones con otros datos recogidos en el propio operativo o provenientes de otras investigaciones relevantes. Tampoco suelen hacerse referencias ni interpretaciones conceptuales a partir de la literatura existente en torno a los aprendizajes evaluados, a los debates didácticos relacionados con su enseñanza o a los factores que inciden en ellos.

Como ejemplo puede citarse el informe nacional de Costa Rica respecto a la evaluación realizada en 1997. Este informe está constituido básicamente por 50 cuadros, cada uno acompañado por un par de renglones que indican el dato más sobresaliente, y finaliza con dos páginas de conclusiones muy generales –que resumen los principales datos–, del tipo: *“El rendimiento académico nacional es más alto en tercer año que en sexto... La Región Central supera al resto de las regiones en el número de objetivos dominados en tercer año y en sexto año... El sector privado supera en ambos niveles al sector público. Estas diferencias son estadísticamente significativas...”*²⁶.

En el caso de Argentina es sintomático el hecho de que sus informes prácticamente carecen de texto y se limitan a presentar los datos. Hasta la denominación de los reportes es sintomática: *“Operativo Nacional de Evaluación 1999”*. Se trata de informes sobre los operativos, que describen básicamente cuántas pruebas fueron aplicadas y cuáles fueron los principales resultados, si bien luego existen informes específicos dirigidos a los educadores, con mayor desarrollo sobre las competencias evaluadas y con los ítems que presentaron las mayores dificultades.

En algunos casos es notorio que las conclusiones son las mismas para todos los informes, año tras año. Ello significa que en realidad ha sido escasa la labor de interpretación y análisis de los datos. Esta ausencia de análisis es ostensible en los informes de APRENDO en Ecu-

²⁶ Ministerio de Educación Pública, Programa de Pruebas Nacionales. Informe nacional sobre los resultados de los pruebas de diagnóstico de conocimientos 1997. Costa Rica, 1998. La afirmación de que el rendimiento es más alto en 3° que en 6° está referida a que en el primer curso mencionado los alumnos lograron niveles de dominio de los objetivos evaluados en mayor proporción que los alumnos de 6°. Si bien no se trata de los mismos objetivos, dado el enfoque adoptado en la evaluación de Costa Rica es válido afirmar que hay un mayor o menor nivel de logro de los objetivos esperados para cada curso.

dor, en que las mismas conclusiones se repiten casi textualmente durante tres años (véase el recuadro 2). Esto probablemente es consecuencia, entre otras cosas, de que en muchos países –como en el caso de Ecuador– las Unidades de Evaluación han asumido una dinámica excesivamente ambiciosa de realización de operativos todos los años, en diversos grados y en un número creciente de áreas curriculares, por lo que no existe el tiempo indispensable para un análisis en profundidad de la información producida.

Los ejemplos mencionados muestran que, en general, las Unidades de Evaluación de la región hasta ahora han concebido su labor principalmente en términos de producción de información, la que se supone otros sabrán utilizar y emplear en la toma de decisiones. Por ello, la mayoría de los informes se limita a relatar los datos obtenidos.

RECUADRO 2

ECUADOR. Reiteración de conclusiones en informes de años sucesivos

APRENDO 96: “Resaltan el cuadro de rendimiento en Matemática pues en este se ven dos puntos críticos: que en los tres grados no se observa un dominio promedio de ninguna de las destrezas medidas, y que las destrezas más alcanzadas revelan que los aprendizajes de los estudiantes se concentran en el nivel básicamente operatorio, mientras que las destrezas menos conseguidas manifiestan que la resolución de problemas es el aprendizaje menos consolidado y extendido. Esto puede significar que la población escolar ecuatoriana estaría limitada en cuanto al desarrollo del pensamiento abstracto y del razonamiento lógico”. Ministerio de Educación y Cultura del Ecuador - EB/PRODEC, 1997.

APRENDO 97: “Resaltan el cuadro de rendimiento en Matemática, pues en este hay dos puntos críticos: que el porcentaje de alumnos que dominan las destrezas en cada grado son notoriamente bajos y que las destrezas más alcanzadas revelan que el aprendizaje estudiantil se concentra en el nivel básicamente operatorio, mientras que el aprendizaje menos consolidado y extendido es el relacionado con la resolución de problemas. Esto puede significar que la población escolar ecuatoriana está limitada en el desarrollo de su capacidad de análisis, que es una expresión del pensamiento abstracto y del razonamiento lógico”. Ministerio de Educación y Cultura del Ecuador - EB/PRODEC, 1998.

APRENDO 98: “Resaltan el cuadro de rendimiento en Matemática pues en este se ven dos puntos críticos: que en los tres grados no se observa un dominio promedio de ninguna de las destrezas medidas, y que las destrezas más alcanzadas revelan que los aprendizajes de los estudiantes se concentran en el nivel básicamente operatorio, mientras que las destrezas menos conseguidas manifiestan que la resolución de problemas es el aprendizaje menos consolidado y extendido. Esto puede significar que la población escolar de los CEM está limitada en el desarrollo de su capacidad de análisis, que es una expresión del pensamiento abstracto y del razonamiento lógico”. Ministerio de Educación y Cultura del Ecuador - EB/PRODEC, 1999.

Un modelo diferente e interesante de reporte de resultados ha sido adoptado recientemente por Perú, país en el cual durante muchos años el Ministerio de Educación no autorizó la publicación de resultados. Sin embargo, a partir del año 2000 la Unidad de Medición de Calidad ha comenzado a publicar los boletines denominados *Crecer*. Estos han tomado, en cada número, un tema central a partir de los cuestionarios a alumnos y docentes que normalmente se aplican junto con las pruebas. Por ejemplo, la actitud de los alumnos hacia las clases de Lenguaje y Matemática, o los niveles de formación de los docentes. En cada boletín se hace una revisión de la literatura existente sobre el tema abordado, se explica por qué es relevante, qué dicen las investigaciones sobre el particular y se presentan en forma gráfica los principales datos relevados en el operativo. Del mismo modo, los boletines destinados al reporte de los resultados explican en forma adecuada qué significan los puntajes empleados para reportar, y qué tipo de conclusiones es válido o no formular a partir de la metodología empleada en el diseño de las pruebas. Se trata de boletines breves, gráficamente bien presentados, que tienen un carácter propiamente reflexivo y no meramente de relato de datos²⁷.

Obviamente, es discutible si la función de analizar la información corresponde a los sistemas de evaluación o si estos deben limitarse, como en general ha ocurrido hasta ahora, a la producción de los datos. Asumiendo que sí deberían entregar información más elaborada, uno de los desafíos centrales para los próximos años sería incorporar la función de “analistas de información” de diversas especialidades –didáctica de las distintas disciplinas, ciencias sociales, políticas educativas, por mencionar algunas– que trabajen más los datos, los interpreten y puedan formular hipótesis, recomendaciones e investigaciones.

V.2. Los “factores escolares” asociados a los aprendizajes y las políticas educativas

La insuficiencia en el análisis, interpretación y conceptualización de los datos es ostensible en el tratamiento que en muchos países se ha dado a la indagación sobre los denominados “factores escolares asociados” a los resultados. Dos son los principales problemas que detectan los informes nacionales:

²⁷ Los boletines *Crecer* pueden ser encontrados en: www.minedu.gob.pe/unc/boletines.php

- una debilidad metodológica, que consiste en que algunos reportes se apoyan únicamente en análisis de tipo bivariado;
- una visión en cierto modo ingenua respecto a la relación entre la investigación sobre factores asociados y la toma de decisiones de política educativa.

En relación a lo primero, cabe señalar que formular algún tipo de conclusión en términos de investigación o de recomendación de política educativa, a partir de análisis de tipo bivariado entre variables escolares y resultados, es sumamente riesgoso, por decir lo menos. Para poder hacerlo es necesario, en primer lugar, controlar el efecto de la composición social del grupo o bien trabajar con dos mediciones para la misma población, de modo de medir propiamente el aprendizaje y neutralizar en buena medida el efecto de las desigualdades de origen social de los alumnos. Este aspecto ya fue tratado en el capítulo III. En segundo lugar, es necesario controlar la asociación que los “factores escolares” tienen entre sí.

Para ilustrar este último problema es útil recurrir al Informe del SAE B/97. Por un lado, este reconoce explícitamente la existencia de este tipo de problemas y los cuidados que deben tenerse para no simplificar la interpretación de los datos:

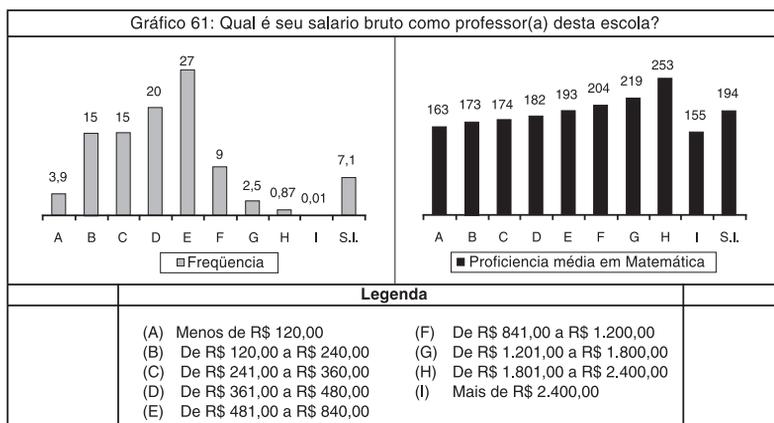
“A los efectos de la discusión de los resultados y de su presentación gráfica, fueron seleccionadas algunas variables de los cuestionarios respondidos por los alumnos y los profesores en el SAEB/97. La selección de estas variables fue orientada por criterios de tasa de respuesta, posibilidades de establecer relaciones con la proficiencia y el potencial para aportar a la discusión de políticas y estrategias de intervención en la realidad escolar.

Es importante llamar la atención para que la lectura de los resultados presentados no conduzca a interpretaciones simplistas o inferencias de causa y efecto, entre una variable considerada aisladamente y el valor que le corresponde. La complejidad del fenómeno de aprendizaje y su multicausalidad deben estar siempre presentes, evitando reduccionismos que conducirían a interpretaciones erróneas de ciertos datos”. (SAEB 1997- Primeros Resultados).

Estas advertencias son absolutamente pertinentes y necesarias. Sin embargo, el informe incluye a continuación un conjunto de cuadros bivariados que conducen precisamente a los errores de interpretación sobre los que se había alertado véanse figuras 18 y 19). Cualquier lector medio no especializado, aún cuando haya sido advertido previamente, probablemente concluya

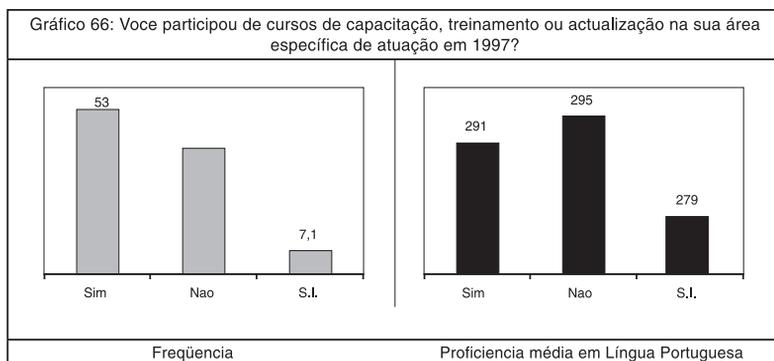
que: a) a mejores salarios docentes, mejores resultados en Matemática; y b) la participación en cursos de actualización no tiene incidencia en la mejora de los aprendizajes (por el contrario, los desmejora).

FIGURA 18
BRASIL - Presentación de Información sobre Factores Asociados



Fuente: Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INEP), 1998. *SAEB 97 - Primeros Resultados*.

FIGURA 19
BRASIL - Presentación de Información sobre Factores Asociados



Fuente: Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INEP), 1998. *SAEB 97 - Primeros Resultados*.

Ahora bien, los datos incluidos en las figuras 18 y 19 pueden ser objeto de múltiples explicaciones o interpretaciones. Probablemente la asociación entre salarios y resultados obedezca a que los salarios más altos corresponden a los colegios privados más caros y, por tanto, a alumnos provenientes de sectores medios y altos de la sociedad. Del mismo modo, la participación en cursos de capacitación no dice nada acerca de la heterogeneidad y calidad de los mismos. También podría ocurrir que quienes participan en mayor medida en dichos cursos sean los profesores más jóvenes y de poca experiencia, y eso explique que los resultados de sus alumnos sean inferiores. En fin, difícilmente pueda llegarse a algún tipo de conclusión válida a partir del análisis bivariado.

Sin embargo, el Informe del SAEB, refiriéndose al gráfico 66 afirma:

“Mientras el 53% de los profesores declaran haber participado de cursos de entrenamiento, capacitación y actualización, 40% no realizaron ninguna de esas actividades. Es interesante notar que esta variable parece no tener ninguna influencia en el nivel de proficiencia de los alumnos” (el destacado es nuestro). Más adelante el informe concluye con una afirmación del siguiente tenor: “En este sentido, el examen de los resultados de proficiencia asociado a algunas características de los alumnos y de los profesores puede dar importantes pistas sobre cuáles pueden ser los nuevos direccionamientos de la práctica de alumnos, profesores, escuelas y familias. Los datos del Sistema Nacional de Evaluación de la Educación Básica son una preciosa fuente de estudios y, principalmente, garantizan el monitoreamiento de aspectos relacionados con la oferta educativa y con los resultados que vienen siendo obtenidos por el sistema educacional brasileño, principalmente aquellos relacionados con la equidad”²⁸.

Nuevamente, lo dicho es estrictamente cierto, en cuanto a que los datos recabados constituyen una preciosa fuente de estudios, pero de otro tipo de estudios más sofisticados que, por otra parte, el propio SAEB ha desarrollado con posterioridad²⁹.

²⁸ Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INEP). SAEB 97 Primeros Resultados.

²⁹ Más recientemente el SAEB ha desarrollado trabajos de investigación empleando técnicas de análisis multinivel, que permiten un tratamiento más apropiado del tema de los factores asociados.

El segundo aspecto que es preciso mencionar es la visión algo simplista e ingenua respecto a la relación directa entre datos y decisiones o políticas sin otro tipo de mediaciones. El tema ya ha sido analizado en una publicación anterior del Grupo de Trabajo sobre Estándares y Evaluación de PREAL³⁰. Muchas veces los informes nacionales dan la impresión de que se espera que a partir de la investigación sobre los “factores asociados” se podrán formular recomendaciones específicas de política educativa. Incluso en informes producidos por los organismos internacionales de crédito era posible encontrar algunos años atrás recomendaciones simplistas del tipo “hay que invertir en libros de texto pero no en formación docente” o “la cantidad de alumnos por grupo no es una variable relevante para mejorar los aprendizajes”, formuladas a partir de revisiones de literatura basadas en análisis sobresimplificados de “factores asociados”.

El caso de Ecuador es interesante e ilustrativo sobre este tema. A través de un análisis de tipo bivariado, el reporte concluye que es difícil establecer políticas o medidas homogéneas para todo el sistema educativo a partir de los datos analizados, porque los factores asociados a los aprendizajes varían entre los diferentes contextos. El tratamiento que se da en el texto a esta constatación parece indicar algún tipo de “culpa” por no haber llegado a conclusiones más contundentes para la toma de decisiones (véase recuadro 3). Sin embargo, el que se haya arribado a esta constatación es sumamente relevante y el informe debería resaltarlo de manera más enfática, advirtiendo a las autoridades acerca de falsas expectativas o visiones simplistas sobre la política educativa. La constatación de que los factores que inciden sobre los aprendizajes no son los mismos en todos los sectores del sistema educativo pone de manifiesto la necesidad de analizar por separado, para cada contexto y sector del sistema educativo, qué factores inciden en los aprendizajes. Este tipo de análisis, a escala más reducida y de carácter más contextualizado, debería ser priorizado por sobre los intentos de construir un modelo explicativo con pretensiones de universalidad o de formular medidas de política educativa de carácter general para todo el sistema derivadas directamente de los “factores asociados”.

³⁰ Véase al respecto Ravela, R., Wolfe, R., Valverde, G. y Esquivel, J.M., 2000.

RECUADRO 4

ECUADOR. Conclusiones del estudio sobre factores asociados

“En general se podría enunciar las siguientes observaciones:

- Las divergencias son mayores que las coincidencias entre los factores de mayor incidencia en el logro académico, lo que dificulta la ejecución de acciones comunes para potenciar a los elementos de mayor impacto. En este sentido se impone tener muy en cuenta las peculiaridades de cada contexto.
- La aplicación de políticas educativas generales se vería muy restringida en razón de que las coincidencias de los factores más importantes en los contextos analizados, que corresponden a los ámbitos de la Gestión Institucional o Escolar, Currículo y Práctica Pedagógica, son muy escasas.
- Se debe destacar la presencia de factores que influyen fuertemente en el logro académico de los estudiantes y que corresponden al ámbito familiar; especialmente la incidencia que en Lenguaje y Comunicación tienen los recursos de lectura en el hogar (aparece en los dos regímenes de escolaridad y en los tres tipos de establecimientos) y, en el caso de Matemática, de igual manera, la educación formal de la pareja del tutor.

De otra parte, cabe anotar que los factores sujetos al ámbito de la política educativa y que no ocupan los primeros lugares porque su condición actual no explica en forma importante los resultados del logro estudiantil, deberían merecer mayor preocupación y estudio para que se pueda establecer, mediante investigaciones pertinentes, qué prácticas pedagógicas, acciones de la gestión escolar o institucional y diseño de currículo son realmente “eficaces” para que produzcan un impacto positivo significativo en el logro académico de los estudiantes”.

Ministerio de Educación, Ecuador, 1999. Factores Asociados al Logro Académico. Resumen de difusión

V.3. Tres vacíos de carácter técnico comunes en la región

Antes de finalizar esta revisión de los reportes de resultados de las evaluaciones nacionales de aprendizajes, es ineludible señalar tres vacíos o debilidades técnicas verificables en la mayoría de los reportes y que son relevantes desde el punto de vista de la información al lector especializado y de la apertura al escrutinio público de la calidad de la información generada.

- Prácticamente en ninguno de los informes analizados se reportan las *tasas de respuesta* a los diferentes instrumentos aplicados, es decir, qué proporción de los alumnos incluidos en el diseño muestral o registrados en las escuelas en el caso de las operaciones censales

efectivamente participaron de la evaluación. Normalmente se informa acerca de la cantidad de alumnos evaluados, pero no acerca de qué proporción representan en relación al total de alumnos que deberían haber sido evaluados. Esta información es crucial dado que en muchos países se registran importantes niveles de ausencia a las pruebas o de falta de respuesta a los instrumentos complementarios.

- Del mismo modo, prácticamente ninguno de los informes analizados incluye la estimación de los *márgenes de error muestral*, que son básicos para analizar la significatividad de la información que se brinda.
- Una tercera debilidad importante es la falta de información acerca de los *procedimientos seguidos para la equiparación de los instrumentos de medición*, en los casos en que se aplican evaluaciones sucesivas en diferentes años en los mismos grados y disciplinas. Este tema es de enorme trascendencia dado que remite a la cuestión de si las diferencias registradas entre años obedecen a cambios reales o a debilidades de nuestros instrumentos.

Estos vacíos en aspectos relevantes de carácter técnico hacen pensar en la necesidad de trabajar conjuntamente en la región en la definición de un conjunto de criterios y requisitos técnicos que todos los informes deberían satisfacer. Para ello puede resultar útil tomar como punto de partida los diferentes estándares para la medición y evaluación producidos en los Estados Unidos por instituciones como la Asociación Americana de Investigación Educativa (AERA) o el Centro Nacional de Estadísticas Educativas (NCES) del gobierno federal.

DIEZ PISTAS PARA PENSAR EN EL FUTURO

Como forma de cerrar este recorrido por los informes nacionales de resultados de pruebas de evaluación de aprendizajes o logros educativos, formularemos a continuación lo que consideramos son las diez principales pistas o desafíos para mejorar nuestros modos de reportar los resultados, así como para mejorar nuestro conocimiento acerca del grado en que nuestros reportes efectivamente están teniendo algún impacto en el sistema educativo.

1. Una primera pista para la reflexión es, en realidad, una advertencia respecto al cuidado que debería tenerse antes de “adosar” a las evaluaciones nacionales algún tipo de consecuencia directa para las escuelas, del tipo incentivos económicos, publicación de rankings u otros. En este sentido, sería pertinente realizar algún tipo de investigación de tipo cualitativo acerca del impacto que realmente tiene sobre las escuelas una estrategia de esta naturaleza en un país como Chile, que ha apostado fuertemente en esta línea. Según se indicó en el primer capítulo, este enfoque genera una fuerte presión sobre las escuelas en la dirección de destinar mucho tiempo a preparar a sus alumnos para responder a las pruebas, lo cual puede dar lugar a una peligrosa reducción del currículo cuando se trabaja exclusivamente con pruebas de selección múltiple o de respuesta sencilla.
2. Si la apuesta principal es a que los sistemas de evaluación aporten a los docentes información relevante para revisar y mejorar su trabajo, entonces deberíamos generar conocimiento específico en torno a las siguientes preguntas: a) ¿están los maestros recibiendo los informes y publicaciones derivadas de las evaluaciones nacionales?; b) ¿en qué proporción efectivamente las leen?; c) ¿en qué medida comprenden el contenido de esas publicaciones?; d) ¿en qué grado la información que les ofrecemos en nuestros reportes les resulta pertinente y enriquecedora para comprender los procesos de aprendizaje de los alumnos y para analizar sus propias prácticas de enseñanza?; e) de los diferentes tipos de estrategias desarrolladas en la región –envío de publicaciones, talleres de capacitación, aplicación y corrección autónoma de pruebas, etc.–, ¿cuáles aparecen como más pertinentes para lograr que las evaluaciones sirvan para mejorar la enseñanza?; y, f) ¿qué otras expectativas o demandas tienen los maestros hacia las unidades de evaluación?
3. Aun cuando no se desee adoptar el camino de la publicación de resultados por escuela en la prensa, queda en pie el desafío de cómo informar a las familias acerca de los resultados de las evaluaciones. Las preguntas a plantearse en este punto son: qué tipo de información brindarles y de qué modo hacerlo. En principio debería tratarse, más que de información numérica –datos de resultados–, de información ilustrativa acerca de qué es lo que se espera que sus hijos sean capaces de hacer; en qué grado lo están logrando y qué pueden hacer las propias familias para apoyar la labor de la escuela.

4. En el mismo sentido, sería útil saber más acerca de cómo otros actores relevantes –autoridades educativas, profesionales de unidades de currículo o formación de docentes, políticos, dirigentes sociales– comprenden e interpretan los resultados y qué demandas de información tendrían ellos hacia los sistemas de evaluación. Un modo de construir respuestas a las interrogantes formuladas en este párrafo y en los dos anteriores sería realizar estudios de casos nacionales mediante encuestas y grupos de discusión con maestros, familias y otros actores relevantes, con el fin de indagar en sus percepciones y demandas sobre los sistemas de evaluación de aprendizajes.
5. Habría que efectuar una revisión sistemática de cómo la prensa ha venido divulgando los resultados de las evaluaciones en los países de la región, el tipo de visión que se deriva hacia la opinión pública y el tipo de errores o interpretaciones inapropiadas más comunes. Eventualmente, debería pensarse en instancias formativas dirigidas a periodistas dedicados a temas educativos (Argentina ha comenzado a hacerlo recientemente, invitando a un especialista español a dictar un seminario sobre evaluación para periodistas).
6. Para dar respuesta a lo que los diferentes actores esperan de los sistemas de evaluación, probablemente será necesario avanzar hacia formar diseños de las pruebas y de los reportes de resultados de carácter “criterial”; es decir, que no se limiten a describir promedios de respuestas correctas y a comparar entre tipos de escuelas o jurisdicciones en que el país está dividido, sino que establezcan niveles de desempeño exigibles a todos los alumnos en cierto grado o nivel, y que reporten el grado en que ello ha sido alcanzado. Asimismo, los reportes numéricos deberían ilustrar ampliamente lo que los puntajes significan en términos de conocimientos y competencias de los alumnos. Estas ilustraciones deberían ir más allá de la publicación de ejemplos de ítems individuales.
7. Es preciso incorporar de alguna manera la consideración de la composición sociocultural del alumnado de los diferentes tipos de escuelas o sectores del sistema educativo en el análisis y presentación de resultados.
8. Habría que dar “valor agregado” a los informes en términos de reflexión sobre los datos y relación de estos con lo que otras investiga-

ciones muestran, entre otras cosas. Asimismo, es preciso reportar sistemáticamente acerca de otras variables sociales e institucionales por sí mismas, de modo de enriquecer la comprensión sobre cómo es el sistema educativo y sus heterogeneidades internas. Esto debería hacerse inclusive antes de hacerlas formar parte de estudios de “factores asociados”, lo que permitiría mejorar la comprensión de lo que significan dichos factores, así como la manera de medirlos. Por otra parte, es preciso llevar adelante o encomendar investigaciones que asuman adecuadamente la complejidad metodológica del tema y moderar las expectativas acerca de lo que las mismas pueden aportar a la formulación de las políticas educativas para evitar caer en simplificaciones. Es conveniente también ampliar el espectro de investigaciones a realizar en materia de “factores asociados” incluyendo abordajes de carácter cualitativo.

9. Un desafío relevante a asumir es el monitoreo en el tiempo de los resultados de nuestros sistemas educativos. Probablemente este sea uno de los tipos de información relevantes tanto para el público en general como para las autoridades y tomadores de decisiones. Esta tarea tiene fuertes implicancias de tipo técnico para garantizar la equivalencia de los instrumentos a lo largo del tiempo.
10. Finalmente, las unidades de evaluación de la región tendrían que trabajar en forma conjunta en la definición de estándares de calidad técnica que deberían satisfacer las evaluaciones y los reportes de resultados. Seguramente este sería un paso enriquecedor para todos que permitiría mejorar la calidad de nuestro servicio al sistema educativo y a la sociedad.

BIBLIOGRAFÍA

- BEATON, A., (1999). *International assessments: the United States TIMSS experience*. En: F Ottobre (ed.), *The role of measurement and evaluation in education policy*. París, UNESCO Publishing, Educational studies and documents N° 69.
- BENVENISTE, L., (2000). *La Evaluación del Rendimiento Académico y la Construcción de Consensos en Uruguay*. Lima: PREAL/GRADE.

- CARNOY, M., (2000). School Choice? Or is it Privatization? En: *Educational Researcher*, Vol. 9, N° 7, pp. 15-20.
- LLACH, J.J., MONTOYA, S. y ROLDÁN, F. (2000). *Educación para Todos*. DISTAL, Buenos Aires.
- LINN, R. & GRONLUND, N. (2000). *Measurement and Assessment in Teaching*. Prentice-Hall, 8ª edición.
- LINN, R., (2000). *Assessments and Accountability*. En: *Educational Researcher*, Vol. 29, N° 2, pp. 4-16.
- MESSICK, S., (1999). *Key Issues*. En: F Ottobre (ed.), *The role of measurement and evaluation in education policy*. París, UNESCO Publishing, Educational studies and documents N° 69.
- MINISTERIO DE EDUCACIÓN, República de Chile (2000). *Evaluación de Desempeño Establecimientos Educativos Subvencionados. SNED 2000-2001*. Santiago.
- OTTOBRE, F., (ed.) (1999), *The role of measurement and evaluation in education policy*. París, UNESCO Publishing, Educational studies and documents N° 69.
- RAVELA, P. (editor), WOLFE, R., VALVERDE, G. Y ESQUIVEL, J.M., (2000). *Los Próximos Pasos: ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?* PREAL/GRADE, Lima.
- SECRETARÍA DE EDUCACIÓN PÚBLICA DE MÉXICO (SEP) - SINDICATO NACIONAL DE TRABAJADORES DE LA EDUCACIÓN (SNTE) (1994). *Carrera Magisterial Folleto Informativo. Primera Versión*. SEP México D.F.
- TUIJNMAN, A. y POSTLETHWAITE, N. (editores), (1994); *Monitoring the Standards of Education*. Pergamon, Oxford.
- UNESCO/OREALC, (1999). *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado*. UNESCO/OREALC, Santiago.

Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: Criterios para guiar futuras aplicaciones^(*)

GUILLERMO FERRER
PATRICIA ARREGUI

INTRODUCCIÓN

En los últimos años, diversos organismos internacionales y regionales han recomendado que los países latinoamericanos participen en pruebas internacionales de logro académico. Esa recomendación suele sustentarse en el argumento de que los datos comparativos sobre el rendimiento estudiantil son útiles para informar a los países sobre la calidad relativa de sus sistemas educativos y, consecuentemente, sobre su competitividad en el mercado global de bienes y servicios. Se argumenta también que las pruebas internacionales garantizan mayor calidad técnica y eficiencia que los sistemas nacionales de medición y que pueden contribuir significativamente al desarrollo de capacidades locales en el campo de la evaluación.

La utilidad efectiva de las comparaciones internacionales de logro académico para informar políticas educativas no es, sin embargo, algo sobre lo cual exista total acuerdo en la literatura actual. Tampoco se ha recogido ni sistematizado suficiente evidencia empírica sobre la difusión que los países han dado a la información recogida en las pruebas interna-

^(*) Este documento fue originalmente publicado por PREAL en la serie PREAL Documentos N° 22, abril de 2003.

cionales ni sobre la manera en que han utilizado esa información para promover el mejoramiento de sus políticas, planes, programas y prácticas educativas.

Ante esta falta de evidencia, y ante la posibilidad de que los países de la región comiencen a privilegiar la participación en pruebas internacionales sin consolidar el desarrollo de sus propios sistemas y capacidades de evaluación y medición, se consideró oportuno iniciar una indagación exploratoria sobre los riesgos y beneficios de esta participación y sobre las condiciones técnicas y políticas en que se han venido realizando.

Para ello se decidió estudiar la experiencia de algunos de los países que han participado en pruebas internacionales de logro académico: Argentina, Chile, Colombia, Cuba, Ecuador, Estados Unidos, México, Perú y Uruguay (que participará próximamente por primera vez en una prueba internacional). La inclusión de Estados Unidos se decidió porque su experiencia arroja cierta luz sobre la forma en que América Latina puede hacer mejor uso de sus oportunidades de participación en pruebas internacionales. Así, aunque esta muestra no se considera representativa de toda la región, permite identificar algunas tendencias y patrones comunes en los procesos de participación de América Latina en pruebas internacionales¹.

El propósito último de este estudio es el de contribuir a que dicha participación, de resultar recomendable o de alguna manera ineludible, pueda servir para el mejoramiento de la gestión y de los resultados de los procesos educativos. A la vez, confiamos en que las recomendaciones finales de este estudio servirán para minimizar los impactos negativos que dichas actividades pudieran eventualmente acarrear.

Antes de concluir esta introducción, deseamos agradecer a Gilbert Valverde, Fernando Reimers, Juan Fernando Vega, José Rodríguez, Leonor Cariola, Carlos Pardo, Silvia Montoya, Shane Hunt, Pedro Ravela y Héctor Fernández por sus valiosos comentarios a los primeros borradores de este informe.

¹ Cabe añadir que la selección de los países estuvo condicionada por las limitaciones financieras de los investigadores, que realizaron las entrevistas personales y telefónicas haciendo uso de tiempo y recursos financieros de otros proyectos de investigación.

I. LAS PRUEBAS INTERNACIONALES DE APRENDIZAJE

I.1. LAS REFORMAS EDUCATIVAS Y LA EVALUACIÓN DEL LOGRO ACADÉMICO EN LOS NOVENTA

Los esfuerzos mundiales por llevar a cabo reformas educativas en las últimas dos décadas han estado motivados, principalmente, por un creciente interés político en asegurar la cobertura y el acceso universal a la educación, así como el aprendizaje efectivo de todos los alumnos de los nuevos contenidos curriculares. Este renovado interés por la calidad de la educación se inscribe en un contexto de fuertes cambios políticos y económicos que son interpretados en formas variadas por las diferentes corrientes académicas y de opinión.

Algunas voces destacan la existencia de condiciones en las que las reglas de libre mercado, la globalización y las nuevas tecnologías de información imponen una redefinición de las metas y objetivos de la educación pública, así como del rol del Estado para la consecución de esas metas. Estos cambios se caracterizan por una serie de objetivos que Carter y O'Neill (en Ball, 1998) definen como constitutivos de una "nueva ortodoxia". Algunos de esos objetivos serían: mejorar las economías nacionales estableciendo vínculos más robustos entre escolarización, empleo, productividad y mercado; mejorar los resultados de aprendizajes relacionados a las competencias y destrezas que requiere el mundo laboral; lograr un mayor y más directo control de los sistemas educativos nacionales sobre los contenidos curriculares y sobre la evaluación; reducir los costos de gobierno en educación; y ampliar la contribución de la comunidad a la educación, mediante su participación en la toma de decisiones escolares y la presión de la libre elección de mercado.

Estas tendencias vienen de la mano de nuevas formas de administración y gestión que plantean la necesidad de mejorar la eficiencia de los sistemas educativos y de fomentar la responsabilización social y profesional por los resultados de la educación. Estas formas de gestión, usualmente denominadas de "gerenciamiento", suponen el uso de estilos y herramientas diferentes a los tradicionalmente utilizados por el sector público. Ejemplo de ello son la delegación de responsabilidad y poder de decisión a las instituciones proveedoras de servicios (escuelas), así como el control "a distancia" de la calidad en la provisión de esos servicios por parte del poder central. De esta manera, el Estado reemplaza la intervención directa y la prescripción por la determinación de objetivos, la com-

paración y la responsabilización (Lyotard, en Ball, 1998). Dentro de este marco de interpretación, sin duda, ha tenido un rol fundamental el uso de mediciones sobre el rendimiento académico de los estudiantes, que permite establecer mecanismos de responsabilización para los alumnos mismos, para los docentes y para las escuelas, en forma de graduación y promoción, acreditación, incentivo a la matrícula, mejoras salariales, y otras tantas formas de estímulo para elevar los resultados educativos.

Paralelamente, muchos países, entre ellos los latinoamericanos, buscan, por medio de las reformas curriculares y de gestión, fomentar y consolidar valores democráticos, de participación ciudadana y de respeto por la diversidad. En este sentido son numerosos los esfuerzos de muchos países de la región por establecer sistemas nacionales de medición de la calidad educativa que permitan, no necesariamente impulsar mecanismos punitivos o de mercado (*school choice*), sino hacer más transparente y participativa la gestión mediante la provisión de información a los actores y usuarios del sistema. Visto de este modo, el Estado no se desentiende de la gestión ni de su responsabilidad por promover mayor equidad y calidad de educación, sino que busca alentar a la población a que utilice la información sobre los resultados educativos para demandar localmente mayores niveles de calidad para sus escuelas.

Simultáneamente, el Estado se reserva para sí la función redistributiva de los ingresos públicos, que se puede hacer más efectiva y equitativa mediante evaluaciones que permitan distinguir cuáles son las subpoblaciones del sistema con mayores dificultades y menores recursos, y así poder intervenir con recursos materiales y pedagógicos para compensar las diferencias de logro académico. Adicionalmente, las evaluaciones han permitido reunir sistemáticamente mayores evidencias sobre las actitudes de los alumnos con relación a los aprendizajes mismos y a los valores personales y sociales, que son contenidos explícitamente delineados en la gran mayoría de los currículos de América Latina donde las reformas educativas de los últimos veinte años se han emprendido en contextos políticos de democratización y pacificación (Ferrer, 1999)².

Es importante notar que la búsqueda del control más directo sobre los contenidos curriculares se ha fortalecido en los países que han tenido

² También en algunos países de Asia, como Japón y Corea del Sur, se ha hecho evidente la necesidad de complementar el enfoque tradicionalmente academicista de la educación con aspectos del desarrollo moral, afectivo y social de los estudiantes (GRADE, 2001).

sistemas educativos tradicionalmente descentralizados, mientras que en aquellos donde la prescripción curricular ha sido históricamente centralizada, los esfuerzos se han concentrado más bien en la flexibilización del currículo en el nivel local y escolar. En cualquiera de los dos escenarios, sin embargo, se percibe la necesidad de realizar evaluaciones centralizadas que permitan medir el logro académico de todos los alumnos bajo un mismo parámetro y con relación a contenidos curriculares a los que todos los estudiantes deberían poder acceder.

En este contexto, una de las herramientas principales de gestión educativa en muchos países del mundo es la evaluación comparativa y cuantitativa de los logros de aprendizaje de los alumnos y la medición del impacto de diversas variables escolares y extraescolares sobre esos logros. Los países que disponían de sistemas nacionales de evaluación desde antes de los años ochenta, reforzaron su aplicación y los vincularon más directamente a sus nuevas metas curriculares. Los países que no disponían de tales sistemas de evaluación, como la mayor parte de los países latinoamericanos, los desarrollaron junto con el diseño e implementación de las reformas educativas de los años ochenta y noventa.

Opiniones especializadas argumentan que el impulso por la creación de estos sistemas de evaluación en los países en desarrollo estuvo influido por la presión de algunas agencias de cooperación internacional que condicionaron los préstamos para educación al desarrollo de sistemas nacionales de medición que permitieran medir los resultados de esas millonarias inversiones en reformas educativas. Sin embargo, esta interpretación es insistentemente desafiada por quienes consideran que, independientemente de las políticas de las últimas dos décadas, comúnmente denominadas “neoliberales”, la evaluación estandarizada del rendimiento académico y el estudio de los factores escolares y extraescolares que afectan ese logro fortalecen los cimientos democráticos de los sistemas educativos, en tanto permiten la opinión informada y la demanda de los beneficiarios, a la vez que el establecimiento de mecanismos de compensación para asistir a las comunidades escolares que más lo necesitan por parte del Estado.

A esta discusión se suma el hecho de que los cambios económicos mundiales de las últimas dos décadas también han dado lugar a una creciente preocupación de los países por su capacidad para competir eficazmente en los mercados internacionales. Para poder competir y prosperar, según se argumenta desde la publicación de *La naturaleza y las causas de la riqueza de las naciones* (Smith, 1776), es menester invertir en capital

humano. Desde entonces, la educación ha sido vista como el principal medio para incrementar ese capital y fomentar así el desarrollo y la posibilidad de competir comercialmente en el escenario global.

La responsabilidad que se le adjudica a la educación por incrementar el capital humano de las naciones se manifiesta actualmente en las múltiples iniciativas de reforma educativa en el mundo. Casi todas estas reformas, si bien muestran grandes diferencias, guardan como principal semejanza los esfuerzos por actualizar permanentemente los currículos y por medir, como se explicó anteriormente, la implementación y aprendizaje efectivo de los contenidos de esos currículos.

Comparaciones internacionales

Junto con el desarrollo y perfeccionamiento de sistemas nacionales de medición de aprendizajes, ha crecido notablemente el interés por las comparaciones internacionales de logro académico. Según Bempechat *et al.* (2000) “la fascinación con el logro académico en diferentes naciones ha sido motivada por los rápidos avances tecnológicos que han transformado la economía mundial [...] Los Estados-nación expresan su preocupación por su capacidad para competir en un mundo que se vuelve más y más pequeño”.

El caso de Estados Unidos es paradigmático en cuanto al impacto que las pruebas internacionales comparadas pueden tener sobre las políticas educativas de una nación. Los resultados en matemáticas y ciencias, de los estudiantes norteamericanos en pruebas internacionales, por debajo del promedio mundial y muy lejos de las potencias asiáticas, alarmaron a los líderes políticos de los inicios de los ochenta, quienes no dudaron en calificar a los Estados Unidos como “una nación en riesgo” (National Commission on Excellence in Education, 1983). La declaración hacía referencia al peligro que corrían los Estados Unidos de perder liderazgo mundial si su sistema educativo público no lograba mejores resultados de aprendizaje en ciencias y matemáticas, y desencadenó una serie de medidas de reforma curricular y de gestión en el país, con el objetivo explícito de ocupar el primer puesto en los *rankings* internacionales de logro académico hacia el año 2000.

Pero a las comparaciones internacionales no solamente se les atribuye un valor competitivo, o de *ranking*. Existen también argumentos conceptuales que justifican las comparaciones entre países como un instrumento metodológico que ofrece mayor información que la que puede ofrecer un

análisis intranacional. Según Joseph Farrell (1979) no se puede realizar ninguna generalización científica sobre educación si no es a través de un estudio comparativo. El autor sostiene que “los datos comparativos son esenciales para establecer la credibilidad de las proposiciones” y que “hasta que no se conoce la relación entre universos particulares y generales, ningún estudio es suficiente para probar una proposición”.

Para muchos investigadores y tomadores de decisiones sobre políticas educativas la participación en pruebas internacionales constituye una “intervención en sí misma”, independientemente de los resultados del estudio, y alerta a educadores e investigadores sobre problemas que podrían no haber considerado anteriormente.

Otras visiones sostienen que las pruebas internacionales pueden cumplir un rol más activo o propositivo. Husén (1987), cuando discute el impacto de la IEA (International Association for the Evaluation of Educational Achievement) sobre las políticas educativas en los países participantes, distingue dos categorías de “clientes”: a) formuladores de políticas (principalmente en el nivel central), y b) ejecutores de esas políticas (técnicos y educadores en el nivel central y local). Según Husén, los primeros suelen estar más interesados en datos y hallazgos de investigación referidos a la estructura del sistema de educación, mientras que los últimos prestan más atención a la información recogida sobre procesos de enseñanza-aprendizaje, prácticas y metodología de aula, materiales educativos y evaluación de aprendizajes en el aula.

Otros autores proponen algunos objetivos específicos que suelen perseguirse mediante la aplicación de pruebas internacionales. Theisen *et al.* (1983), por ejemplo, citan tres objetivos principales para esa participación: comparar el logro académico relativo entre países y por asignatura; identificar implicancias de política educativa para un país a través de los determinantes de logro hallados en otros países; y reevaluar las prioridades de inversión dentro de un país para elevar la calidad de los aprendizajes.

La teoría sobre educación comparada es extensa y controvertida, y los argumentos sobre el valor y validez de las comparaciones internacionales aún son materia de acaloradas discusiones en los medios académicos. Sin embargo, persiste el interés de los investigadores sociales, así como de los formuladores de políticas públicas alrededor del mundo, en realizar comparaciones, cada vez más precisas y sofisticadas, sobre diferentes aspectos de la política social. Este ha sido el caso de las pruebas internacionales de logro académico, que se siguen aplicando en una cre-

ciente cantidad de países cada año y que se van perfeccionando gradualmente en lo que se refiere a la especificación consensuada de sus objetivos, así como a las técnicas de recojo y análisis de la información.

I.2. ¿EN QUÉ CONSISTEN LAS PRUEBAS INTERNACIONALES?

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 se suele estudiar están la inversión per cápita en cada sistema nacional; 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.

Las pruebas pueden ser curriculares o “acurriculares”, dependiendo de su grado de vinculación a los currículos de los países participantes. Cuando las pruebas son curriculares, los instrumentos se diseñan sobre una matriz común que representa los contenidos básicos que se ofrecen en los niveles educativos a ser evaluados en casi todos los países participantes. A estos contenidos comunes se suman contenidos considerados necesarios para una educación de calidad en el siglo XXI, según la opinión de profesionales y académicos internacionales especializados en las áreas disciplinares evaluadas. En el caso de las pruebas “acurriculares”, los diseñadores del estudio establecen estándares de contenido académico disciplinar que se consideran necesarios para una efectiva inserción social y laboral en el mundo de hoy, independientemente de los contenidos de los currículos de los países participantes.

Algunas de estas pruebas se han aplicado una única vez, a modo de diagnóstico o como ejercicio de investigación. Actualmente, se observa una tendencia al diseño de pruebas de aplicación cíclica para seguir el rendimiento de un grupo etario o grado escolar de estudiantes cada cierta cantidad de años, y así establecer relaciones longitudinales con respecto al desempeño académico de esas poblaciones.

Agencias de evaluación

Estas son pruebas diseñadas y coordinadas por agencias internacionales de evaluación, y con cierto grado de participación 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 como es el caso de Educational Testing Services (ETS, ‘Servicios de Evaluación Educacional’) en Estados Unidos, o de la International Association for the Evaluation of Educational Achievement (IEA, ‘Asociación Internacional para la Evaluación del Logro Académico’), que comenzó con un perfil similar pero que gradualmente fue convirtiéndose en una agencia de fuertes vinculaciones con los gobiernos nacionales participantes.

Otras agencias son intergubernamentales desde su concepción y en ellas están representadas las máximas unidades políticas o técnicas de administración o evaluación educativa de los países miembros. Ejemplo de estas últimas son la Organización para la Cooperación y el Desarrollo Económico (OCDE), el Laboratorio Latinoamericano para la Evaluación de la Educación (LLECE) –dependiente de Unesco-OREALC–, o el Southern Africa Consortium for Monitoring Educational Quality (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 diez pruebas internacionales que miden el logro estudiantil en diferentes áreas de conocimiento, tal como se resume en el Cuadro 1.

Difusión y uso de los resultados

Independiente de su origen, estrictamente académico o de aplicación política, las pruebas internacionales buscan el desarrollo de nuevas metodologías de investigación comparada y enfoques analíticos (Forster, 2000); actualmente, sin embargo, se tiende a desarrollar estas pruebas como un instrumento para conseguir insumos de información para el diseño y mejoramiento de políticas educativas orientadas a elevar la calidad de los aprendizajes. Para ese fin, algunas pruebas ofrecen lo que en este informe llamaremos “opciones nacionales”, es decir, la posibilidad de medir subpoblaciones específicas que por alguna razón requieran una medición más enfocada, o de áreas de conocimiento particulares y diferentes de las

CUADRO 1
Resumen de las principales pruebas internacionales
aplicadas en América Latina

Agencia	Prueba	Años de aplicación	Cobertura	Población objetivo
OECD; Statistics Canada; NCES; ETS	IALS (International Adult Literacy Survey)	1994-1998	Global	16-65 años.
OECD; Statistics Canada; NCES	ALL (Adult Literacy and Lifeskills)	2002	Global	16 a 65 años.
OECD	PISA (Programme for International Student)	2000 2001	Global	15 años
LLECE (Unesco OREALC)	Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados	1997	A. Latina	Grados 3 y 4
IEA	PIRLS (Progress in International Reading Literacy Study)	(RLS-1991) 2001+	Global	Grado 4 (9 años)
IEA	TIMSS (Third International Mathematics and Science Study) TIMSS-R	1995 1999	Global	Grados 3-4, 7-8, y último de educación secundaria.
IEA	Educación Cívica	1999 y 1999-2000 (dos fases)	Global	14 y 17-18 años

1/ Se indica el número de países latinoamericanos participantes sobre el número total de países participantes.

Áreas curriculares y competencias	Tipo de prueba	Principales factores asociados	Nº de países latinoamericanos 1/
Lectura: prosa, documentación e información cuantitativa.	Preguntas cerradas y abiertas.	VARIABLES culturales, sociales y demográficas	1/22
Lectura; alfabetización numérica; razonamiento lógico; trabajo en equipo.	Preguntas cerradas y abiertas.	VARIABLES sociales, educacionales, y económicas. Acceso y uso de NTIC.	3/22 [Argentina, Bolivia y Costa Rica]
“Destrezas para la vida”: Matemáticas, Lectura y Ciencias.	Opción múltiple y respuesta abierta.	VARIABLES demográficas, sociales, económicas y educacionales.	6/33 [Argentina, Brasil, Chile, México, Perú y Uruguay]
Lenguaje y Matemática.	Opción múltiple y preguntas abiertas en Lenguaje.	VARIABLES escolares, familiares y sociales.	12 [Argentina, Bolivia, Brasil, Chile, Colombia, Cuba, Honduras, México, y Venezuela]
Lectura	Opción múltiple y respuestas abiertas.	Hogar; escuela; prácticas pedagógicas; currículo; políticas.	2/35 [Argentina y Colombia]
Ciencias y Matemáticas.	Opción Múltiple, Respuesta abierta y demostración práctica.	Currículo; Prácticas pedagógicas; políticas; actitudes; hogar; escuela.	2/45 [Colombia y México] 2/38 [Argentina y Chile]
Ciudadanía y democracia; identidad nacional y relaciones internacionales; cohesión social y diversidad; economía y medio ambiente	Preguntas cerradas y abiertas; escalas de actitud en el hogar.	Clima escolar, actividades extraescolares, recursos culturales	2/32 [Colombia y Chile]

que evalúa la prueba general. Así, un país podría hacer uso del operativo internacional para evaluar, por ejemplo, un grupo etario diferente o agregar ítems de contenidos curriculares nacionales no contemplados en la prueba internacional.

Los resultados de las pruebas internacionales pueden ser reportados y difundidos por las agencias que dirigen los estudios, por las agencias nacionales que coordinan los estudios en los países participantes, por investigadores independientes o por la prensa (Forster, 2000; Macnab, 2000). Los datos que las pruebas arrojan suelen ser cuantiosos y por esa razón se eligen diferentes formas de reportarlos, de acuerdo a las audiencias a quienes están dirigidos. Algunos informes presentan un *ranking* de acuerdo al logro relativo de los países participantes, con relación al total del examen o a grupos de ítems seleccionados para el informe; también suele mostrarse la distribución de los resultados en relación a la media aritmética de cada país (dispersión), el logro de diferentes subpoblaciones, o las relaciones entre logro académico y diferentes variables contextuales relevadas por el estudio (Forster, 2000).

Algunas de las formas en las que se suele usar la información recogida son: la motivación para mejorar políticas a partir de la simple comparación con el logro de otros países; la promoción de políticas de mejoramiento a través de la difusión masiva de los resultados; el establecimiento de nuevos estándares académicos de acuerdo al desempeño de alumnos en otros países; la identificación de los factores escolares y extraescolares que muestran un efecto positivo en el logro de aprendizajes y el estímulo para las investigaciones dentro del país mediante la profundización de los resultados del estudio internacional (Forster, 2000; Macnab, 2000).

I.3. EL DEBATE ACERCA DE LA PARTICIPACIÓN EN PRUEBAS INTERNACIONALES

El valor de las comparaciones de logro académico entre países o entre regiones es materia de discusión permanente entre investigadores. Los argumentos que se esgrimen a favor y en contra de esas comparaciones suelen ser de carácter eminentemente técnico, aunque también entran en juego factores de orden político e ideológico. A continuación se resumen los beneficios y problemas más frecuentemente citados por la literatura especializada con relación a la participación en pruebas internacionales.

Beneficios

- Los resultados comparados internacionalmente atraen la atención de los medios de comunicación, movilizan fuerzas políticas (Greany y Kellaghan, 1996) y suscitan debates sobre las implicancias prácticas de los resultados obtenidos (Husén, 1987; Husén y Tuijnman, 1994, en Kellaghan, 1996).
- La construcción de pruebas requiere de los países participantes una revisión cuidadosa de sus currículos y su comparación con los de otros países, lo cual puede conducir a su actualización y mejoramiento (Beaton *et al.*, 1999). También motiva a educadores e investigadores a estudiar problemas del sector que no se habían considerado antes (Husén, 1987).
- Se desarrolla la capacidad local para la medición de aprendizajes (Greany y Kellaghan, 1996; Torney-Purta, 1990; Bloom, 1974), en la medida en que educadores e investigadores de la educación se ven llamados a interactuar con especialistas internacionales y así se mantienen actualizados (Beaton *et al.*, 1999; Mislevy, 1995).
- Se conoce la efectividad de las políticas implementadas por otros países (Beaton *et al.*, 1999).
- Además del logro académico, permiten comparar otra gran variedad de variables educativas, tales como el clima escolar o los factores familiares y sociales que afectan el rendimiento académico (Husén, 1987; Torney-Purta, 1990). Si bien algunos de estos factores pueden caer fuera de las posibilidades sectoriales de acción, todos contribuyen a una comprensión más integral del estado actual del sistema educativo y sus condicionantes.
- Los gobiernos pueden incluir en estos estudios preguntas e ítemes que de otra manera deberían evaluar por su propia cuenta, a través de sus sistemas nacionales (Husén, 1987).
- La dispersión de los estudios comparativos internacionales suele ser mucho más fuerte y evidente que la de los estudios nacionales, por lo que permiten indagar más profundamente en los efectos de diferentes variables sobre los resultados de aprendizaje (Husén, 1987).

- La construcción y la aplicación colectiva de pruebas puede promover prácticas profesionales democráticas debido al tipo de procedimientos y metodologías utilizadas (Torney-Purta, 1990).
- Algunas pruebas, como TIMSS, utilizan diseños jerárquicos para la recolección de datos, lo cual permitiría, asumiendo que exista suficiente varianza en los datos totales, analizar cuánto de la varianza se debe a diferencias entre países o a diferencias entre estudiantes dentro de cada país (Bempechat *et al.*, 2000). Estos análisis pueden contribuir a revelar cómo algunas políticas similares pueden tener efectos diferenciales en los países, dependiendo de otras variables complementarias³.
- Si bien en el proceso participan técnicos internacionales, los informes de resultados son producidos por investigadores nacionales que conocen a fondo los problemas de su país; por esta razón se espera que exista una interacción entre investigadores y educadores por lo menos igual a la que puede existir en el caso de pruebas nacionales, para que la información recabada y difundida provoque respuestas constructivas a los problemas detectados por el estudio (Bloom, 1974).

Problemas

- Los resultados de los estudios, en particular cuando se presentan como *ranking* de logro entre países, suelen desatar comentarios y críticas por parte de los políticos, los medios de comunicación y el público, en los que solo importa la posición relativa de logro en el escenario internacional y no la abundante información sobre factores sociales, escolares y personales asociados al rendimiento (Greaney y Kellaghan, 1996). La prensa, particularmente, tiende a desconocer los aspectos más complejos y auténticamente informativos de los resultados y solo destaca las posiciones en el *ranking*, a menudo con títulos dramáticos y escandalosos (Husén, 1987; Kellaghan, 1996).
- Si bien las pruebas internacionales utilizan diferentes formatos, algunos de ellos innovadores, los ítems más numerosos continúan siendo

³ Un ejemplo de ello es el efecto positivo que la tarea en el hogar puede tener en el logro académico de los estudiantes chinos, mientras que en Perú el efecto puede ser mucho menor debido al alto índice de analfabetismo de los padres.

diseñados en forma de opción múltiple. Aun cuando estos ítems, si están debidamente contruidos, pueden evaluar habilidades cognitivas complejas, persisten críticas y dudas sobre su validez para ese fin (Greaney y Kellaghan, 1996).

- Se ha criticado a los estudios internacionales la falta de metodologías que fomenten la construcción cooperativa y democrática de las pruebas, lo cual pone en riesgo la participación efectiva y el beneficio que puedan derivar de esa participación de los países con escaso desarrollo técnico (Husén, 1987)⁴.
- La complejidad de los estudios, sumada a la escasa capacidad que suele tener el público para interpretar los numerosos datos que derivan de los estudios comparativos, determina que la información que finalmente se difunde resulte ambigua o contradictoria, que un mismo dato pueda ser utilizado para argumentos antagónicos (dependiendo de la posición política), o que sencillamente se malinterpreten los datos y se llegue a conclusiones erróneas (Husén, 1987). Los indicadores, interpretados aisladamente de su contexto cultural y estructural, pueden conducir a interpretaciones y acciones de política equivocadas (Burstein *et al.*, 1992, en Schmidt y McKnight, 1995).
- La validez de los estudios se ve amenazada por las grandes diferencias culturales entre los alumnos de diferentes países, que hacen que las “condiciones” en que se toman las pruebas no sean iguales en cada país. Entre estas diferencias destacan los diversos grados de familiaridad que los alumnos puedan tener con las pruebas estandarizadas, los estilos de respuesta (como la tendencia a omitir o a adivinar respuestas), y las motivaciones que puedan tener los alumnos para hacer un esfuerzo significativo por resolver las preguntas (Mislevy, 1995).
- A menudo se toman decisiones de política sobre la base de interpretaciones de causalidad entre las variables estudiadas, cuando en reali-

⁴ Debe considerarse que esta construcción cooperativa tiene que ver tanto con la participación efectiva de los técnicos de todos los países involucrados, como con la posible participación de diferentes grupos profesionales en los ámbitos educativos nacionales para la validación de los contenidos de las pruebas.

dad esa causalidad solo podría confirmarse mediante estudios experimentales controlados (Platt, 1975, en Mislevy, 1995).

- La información cualitativa recogida por los estudios comparativos no siempre resulta suficiente para establecer las variaciones que existen al interior de cada cultura y así comprender mejor cómo las creencias y percepciones de los alumnos afectan diferencialmente su rendimiento (Bempechat *et al.*, 2000).
- Si bien se ha propuesto el uso de las pruebas internacionales para fines de responsabilización de los gobiernos y funcionarios, aún no existen definiciones claras sobre quién será responsable, o quién tendrá el poder o la autoridad para modificar y mejorar el rendimiento académico de los alumnos. En general, y como sucede con muchos sistemas de evaluación nacionales, se ha dado poco espacio a la discusión sobre el uso exacto que se dará a las pruebas. Puede pretenderse, erróneamente, que la misma prueba sirva múltiples propósitos cuando en realidad esos propósitos pueden ser incompatibles (Kellaghan, 1996).
- Hasta la fecha, ni las agencias evaluadoras ni académicos especializados han llevado a cabo un registro sistemático de los usos que se ha dado en diferentes países a los resultados de las evaluaciones. Kellaghan (1996) resalta que los informes que existen actualmente presentan información limitada y relativamente superficial, y por lo tanto no se pueden identificar los mecanismos exactos por medio de los cuales la información [de la IEA] se “traduce” en acciones de política. Aunque este no sea un problema inherente a las pruebas mismas, históricamente demuestra que las evaluaciones internacionales no siempre logran informar eficazmente a las audiencias o clientes principales.

El balance entre las ventajas y limitaciones que presentan las pruebas internacionales pareciera indicar que la participación en estos proyectos puede ser valiosa para los países, inclusive para los países en desarrollo (Greaney y Kellaghan, 1996), siempre y cuando se tomen ciertas previsiones con respecto a las características técnicas de las pruebas y a las condiciones políticas e institucionales para su aplicación, incluyendo mecanismos para la correcta interpretación y difusión de los resultados. Kellaghan (1996) propone una variedad de “condiciones” bajo las cuales

debiera concretarse la participación de un país en pruebas internacionales comparadas. Esas condiciones son tomadas en cuenta para sustentar algunas de las recomendaciones que se efectúan hacia el final de este documento.

II. MARCO CONCEPTUAL Y CASOS DE ESTUDIO

El presente estudio fue estructurado sobre un marco conceptual que encuadra la participación en pruebas internacionales de aprendizaje en una secuencia de cuatro fases:

Iniciativa y adopción. Los resultados de aprendizaje obtenidos por un país, según muestran las mediciones nacionales de rendimiento u otros estudios académicos de investigación educativa, estimulan la opinión y la investigación sobre la necesidad de elevar la calidad de los aprendizajes y sobre las formas de lograrlo. La unidad central de administración educativa, o alguna entidad académica con el aval oficial, resuelve que es necesario medirse con respecto al logro académico de otros países y, de acuerdo a las prioridades curriculares y de política educativa en cada país, se opta por uno o más tipos de proyecto de evaluación internacional. Se espera de esta manera realizar un seguimiento de los alcances y efectos de los programas curriculares nacionales.

Implementación. En cada país se ponen en marcha los procesos técnicos de evaluación internacional, desde las primeras reuniones de discusión entre expertos sobre el marco conceptual a ser utilizado, hasta la selección y validación de ítems, la selección de la muestra poblacional y la aplicación final de los instrumentos. Si la relación entre los equipos técnicos de medición y los profesionales del desarrollo curricular en cada país es fluida y cooperativa, es probable que el proceso de diseño de las pruebas contribuya a las primeras revisiones críticas de los contenidos curriculares nacionales, ya sean antiguos o de reciente diseño e implementación. Entran en juego en esta fase los aspectos financieros, así como las capacidades organizacionales y profesionales que afectan la viabilidad y calidad de los operativos de diseño y aplicación de las pruebas. En el contexto latinoamericano juegan un rol particularmente importante los aprendizajes técnicos, de tipo estadístico y psicométrico, que se desarrollan durante la experiencia.

Interpretación y diseminación de resultados. Se lleva a cabo la lectura de los resultados de evaluación por parte de los técnicos de medición así como de los profesionales del desarrollo curricular convocados durante el proceso de implementación. Esa lectura supone el análisis comprensivo y la valoración inicial del nivel de logro obtenido. Se preparan informes de resultados para la prensa, para la opinión pública y para diferentes audiencias profesionales del sector; se ofrecen talleres de difusión y desarrollo de capacidades de interpretación y análisis de los datos, y se liberan las bases de datos para uso de los medios académicos con fines de investigación. Los esfuerzos de interpretación se centran en las inferencias de tipo curricular que se pueden realizar con relación a los contenidos prescritos y enseñados de las áreas de conocimiento evaluadas.

Decisiones, impacto político y nuevas líneas de investigación. Los resultados de las evaluaciones influyen directamente o indirectamente mediante la presión de la opinión pública y profesional, sobre las decisiones de funcionarios y profesionales del sector para elevar la calidad de los aprendizajes.

Casos de estudio

Del total de pruebas internacionales de aprendizaje en las que América Latina ha participado o está participando actualmente, se decidió en este estudio reportar las siguientes experiencias:

- **Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados del LLECE** (Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación), realizado bajo la dirección de la Oficina Regional para América Latina y el Caribe (OREALC) de la Unesco. Resulta interesante por tratarse del único estudio comparado de cobertura estrictamente regional, el cual contó con la participación de un número importante de países latinoamericanos.
- **TIMSS (Third International Mathematics and Science Study); TIMSS-R (repetición de TIMSS); y Educación Cívica.** Todas estas pruebas son administradas por la International Association for the Evaluation of Educational Achievement (IEA). Fueron incluidas por tratarse de pruebas que han convocado la participación de algunos

países de la región y porque la experiencia de su aplicación puede ser contrastada con la de Estados Unidos, donde el impacto de esta evaluación ha sido mucho más significativo que en América Latina.

- **Pruebas PISA (Programme for International Student Assessment)** de la Organización para la Cooperación y el Desarrollo Económico (OCDE). Fue elegida debido a que es una prueba “acurricular” que propone estándares internacionales y ha concitado el interés de varios países de la región.

III. RESULTADOS DEL TRABAJO DE CAMPO

III.1. AGENCIA: LLECE (LABORATORIO LATINOAMERICANO PARA LA EVALUACIÓN DE LA CALIDAD EDUCATIVA, DE UNESCO-OREALC) - PRIMER ESTUDIO INTERNACIONAL COMPARATIVO SOBRE LENGUAJE, MATEMÁTICAS Y FACTORES ASOCIADOS

Cuando se aprobó la idea de crear el Laboratorio Latinoamericano para la Evaluación de la Calidad Educativa - LLECE (en adelante, el Laboratorio), prácticamente todos los países de la región optaron por adherirse. Fue concebido como una red de los sistemas nacionales de evaluación de principios de los noventa, en un momento en que la responsabilidad de crear y sostener un sistema de evaluación era complejo desde el punto de vista técnico y político. El Laboratorio ofrecía un espacio de diálogo y participación para los funcionarios y técnicos responsables de esos sistemas y brindó apoyo para la creación de varios sistemas nacionales de evaluación.

Las principales motivaciones para realizar el estudio comparado fueron la necesidad de establecer una comparación “de raíz latinoamericana” que contemplara las particularidades culturales y educativas de la región y obtener información para apoyar las decisiones de política educativa. Otra razón que motivó a los países a escoger la prueba del Laboratorio fue que era mucho más económica que otros estudios internacionales.

El Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados fue aplicado en 1997 en 13 países latinoamericanos, en los grados 3º y 4º de educación primaria, en las áreas de lenguaje y matemáticas. La prueba incluyó ítems de opción múltiple y también de respuesta abierta, y recogió información sobre variables aso-

ciadas al logro académico, entre ellas de contexto socioeconómico, familiar, escolar y áulico. Los principales objetivos del Laboratorio fueron entregar una comparación de rendimiento académico entre países y analizar los factores asociados a ese rendimiento. Objetivos complementarios fueron: fortalecer los sistemas nacionales de medición y evaluación por la vía de la práctica a través del primer estudio comparado; establecer estándares educativos, y generar una línea de investigación a partir de los resultados.

Los resultados confirmaron muchos de los datos que algunos sistemas nacionales ya habían comenzado a producir, como por ejemplo las diferencias de logro a favor de las escuelas urbanas y privadas, escasas diferencias de rendimiento entre géneros, y una fuerte asociación positiva entre bajo nivel socioeconómico y bajo rendimiento académico. En cuanto a los resultados comparados internacionales, fue particularmente llamativa la diferencia entre Cuba y el resto de los países de la región, con más de dos desviaciones estándar a favor de Cuba. Los resultados fueron publicados por LLECE en dos informes. El primero presenta los resultados comparativos en forma de *ranking* y desagregados por estratos. El segundo informe, que salió tiempo después, ofrece un análisis de factores asociados al rendimiento académico en el conjunto de países participantes, y se concibió con la idea de ofrecer mejores insumos de información para la toma de decisiones de política educativa y para estudios académicos en la región.

El estudio fue financiado con aportes del Banco Interamericano de Desarrollo (BID), sumados a algunos fondos de la Fundación Ford y dinero de los países participantes el cual se les devolvía en forma de servicios de capacitación para la medición.

Para garantizar que los ítems de las pruebas tuvieran una fuerte vinculación con los currículos de los países participantes, se invirtieron dos años y medio en el diseño de las mismas, con reuniones para la selección de ítems, verificando currículos nacionales de matemáticas y lenguaje. Este fue uno de los aspectos más beneficiosos del proceso, pues obligó a los países participantes a efectuar una importante revisión de sus currículos, lo que les permitió descubrir sus debilidades y fortalezas relativas.

Los representantes ante el Laboratorio, llamados “coordinadores nacionales”, fueron en su mayoría los directores de los sistemas de evaluación de cada país, o de la oficina ministerial de la cual esos sistemas dependían. En algunos casos, el diseño de la matriz curricular se hizo con

la participación de los técnicos en currículo de los diferentes ministerios. Sin embargo, en algunos países hubo tensiones importantes entre las oficinas del currículo y los sistemas nacionales de evaluación, lo cual determinó que los departamentos de currículo no estuvieran representados durante el proceso. Debido ello, se reconoció que la prueba final no tendría una base curricular de igual significado para todos los países, y que ello repercutiría en las inferencias de orden curricular que se hicieran más adelante.

La entrega de los resultados finales por parte de la Unesco-OREALC se hizo directamente a los ministerios. Al mismo tiempo, se realizaron conferencias de prensa en Santiago de Chile y se editó un compacto para televisión que fue difundido en América Latina por la Red de TV Española. Aunque todos los países participantes de la prueba del Laboratorio se habían comprometido a publicar los resultados internacionalmente, dos de los países participantes, México y Perú, cuestionaron la validez de los resultados y desautorizaron su difusión.

Era potestad de los gobiernos decidir su difusión en cada país. De ahí que, en muchos casos, los resultados solo circularan en reuniones de cúpula, acaso compartiendo un resumen con la prensa, pero sin la voluntad de generar debate público sobre el tema.

Los responsables por la ejecución del estudio en el Laboratorio coinciden con algunos de los coordinadores nacionales en que la falta de una estrategia fuerte de difusión determinó que los resultados de la prueba no tuvieran un gran impacto político, a través de la opinión pública, ni en las decisiones de política educativa de la región. Las reacciones por parte de los gobiernos fueron variadas, dependiendo de las expectativas y de la coyuntura política de cada país, aunque se reconoce que en general, exceptuando Cuba, fueron de “tono bajo”. Sin embargo, se esperaba que con la difusión del segundo informe sobre factores asociados y con el paso del tiempo, los datos comiencen a ser utilizados por investigadores nacionales e internacionales para análisis secundarios y proyectos de investigación, probablemente comisionados por agentes no gubernamentales, así como por académicos y organizaciones sindicales.

III.2. AGENCIA: IEA - TIMSS, TIMSS-R Y EDUCACIÓN CÍVICA

La prueba **TIMSS** administrada por la IEA fue aplicada en 1995 en unos 45 países de todo el mundo. En algunos países solo participaron algunas provincias y no todos evaluaron todas las poblaciones objetivo del estu-

dio: los grados 3° ó 4° de primaria, 7° u 8° grados, y último año de la educación básica. Las pruebas se centraron en las áreas de matemáticas y ciencias y la matriz curricular se definió mediante un relevamiento previo de los contenidos curriculares prescritos tanto por documentos oficiales como por libros de texto y materiales didácticos de uso masivo en un número significativo de países. La prueba incluyó ítems de opción múltiple, de respuesta abierta y de demostraciones prácticas (por ejemplo, de laboratorio). También se realizó un estudio cualitativo, mediante videograbaciones, sobre procesos pedagógicos de aula en seis países de la muestra.

En cuanto a los resultados comparados internacionales, quedó confirmado que los países del sudeste asiático, como Japón, Singapur y Corea del Sur, logran resultados muy superiores a la mayoría de los países industrializados del mundo. El único país latinoamericano que participó y permitió la publicación de los resultados, Colombia, obtuvo uno de los últimos lugares en el *ranking* internacional.

En 1998 se aplicó **TIMSS-R**, una repetición de la prueba, en 38 países, incluyendo tres latinoamericanos. TIMSS será aplicado nuevamente en el año 2003.

La prueba de **Educación Cívica**, también administrada por la IEA, se aplica regularmente en países de todo el mundo. Los contenidos de la prueba se centran en temas de ciudadanía, gobierno y democracia. También se recogen datos sobre factores sociales, familiares y actitudinales asociados al rendimiento académico.

III.3. AGENCIA: OCDE - PRUEBA PISA

PISA, la prueba comparativa de la OCDE, se aplica regularmente en países de todo el mundo a estudiantes de 15 años de edad, cualquiera sea su grado de escolarización. Los países miembros de la OCDE ingresan a la prueba automáticamente, mientras que los demás países son invitados a sumarse al proyecto mediante el pago de una cuota de membresía. Las pruebas de PISA se diseñan a partir de un conjunto de estándares de contenido y desempeño, elaborados por un grupo de especialistas en diferentes áreas de conocimiento, que se consideran imprescindibles para la adecuada inserción de los jóvenes en la sociedad y en el mundo laboral de hoy. PISA volverá a aplicarse en el año 2003 y luego en el 2006.

Al igual que en el caso de las pruebas de la IEA, la decisión de participar en la prueba PISA la tomó cada país individualmente y por lo

tanto no es posible presentar testimonios de representantes regionales para esta experiencia de evaluación. En consecuencia, se han tomado en cuenta las voces de los representantes nacionales de evaluación de Argentina, Chile, México, Perú y Uruguay.

IV. RESUMEN CRÍTICO DE LOS RESULTADOS

Como se ha podido observar, existe ya un conjunto importante de experiencias de participación en pruebas internacionales en América Latina. Esas experiencias muestran diferentes grados de inserción e impacto efectivos en el escenario de las políticas y la práctica educativa de cada país, dependiendo de los contextos políticos e institucionales en que se inscriben. Aunque el tamaño de la muestra de países en este estudio, y el número limitado de actores entrevistados, limitan las posibilidades de hacer generalizaciones amplias a la región, la cantidad de información obtenida merece un resumen crítico que reúna los aspectos más importantes de los casos analizados, acuerdo a las fases de evaluación comparada internacional utilizadas como marco conceptual en este estudio.

En cuanto al momento previo a la iniciativa y la adopción de pruebas, en la que perciben los primeros “descontentos” con respecto a los resultados de aprendizaje obtenidos por los sistemas educativos nacionales, en América Latina esa percepción se basa, al menos parcialmente, en algunos referentes empíricos, tales como los resultados de pruebas nacionales de logro académico, pruebas exploratorias comparadas en la región, o resultados de pruebas internacionales anteriores. Parte del “descontento” sobre la eficacia de los sistemas educativos, que conduce a la búsqueda de información comparada sobre rendimiento académico, deviene del convencimiento, no del todo fundamentado, de que los grandes problemas económicos de la región tienen origen, en parte, en las deficiencias de aprendizaje de los estudiantes, quienes se gradúan sin estar preparados para las demandas de conocimiento del mundo actual.

Ese uso de referentes empíricos es decididamente más fuerte en países como Estados Unidos, donde se han empleado en forma mucho más extensiva los datos recogidos en pruebas nacionales y pruebas internacionales anteriores, y donde el temor a perder competitividad en el mercado internacional es –o al menos lo fue durante la última década– más marcado.

CUADRO 2

Resumen de la experiencia de cuatro países en la Prueba del LLECE

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Argentina	<p>Se valora la oportunidad de compartir experiencias con América Latina y de discutir con otros países en situación de desigualdad técnica –para lo cual se requiere mucha tolerancia y buena voluntad– y de llegar a acuerdos para la toma de decisiones.</p> <p>Crítica: se tiene la percepción que el Laboratorio no supo aprovechar adecuadamente los conocimientos y la experiencia que ya tenían algunos países.</p>	<p>La principal receptora de los resultados fue la ministra de Educación, quien los compartió con el gabinete ejecutivo de la Presidencia. La lectura dada a los datos fue mínima, en parte porque el primer informe solo fue comparativo y no permitía realizar inferencias sobre las razones que explicaban esos resultados. No se previeron estrategias de difusión y a la prensa solo se le entregó una copia del informe oficial del Laboratorio, sin interpretaciones por parte del Ministerio.</p>	<p>Escasa repercusión a nivel del Ministerio de Educación, pues los resultados llegaron poco tiempo antes de una renuncia de todo el equipo de alta dirección. Sin embargo, muchos de estos ex funcionarios que tuvieron acceso a la información circulan actualmente en medios académicos, desde donde tienen cierta influencia en la formación de profesionales y especialistas en educación.</p>
Colombia	<p>Producto del proceso, se valora el fortalecimiento de la relación entre el Ministerio de Educación y el ICFES (Instituto Colombiano de Fomento a la Educación Superior) que actuó como co-representante ante el LLECE.</p> <p>Críticas: Poca participación efectiva del país en el diseño de ítems, falta de acceso al marco conceptual con el cual se construyeron los instrumentos, poca comunicación entre los países participantes.</p>	<p>El gobierno no tenía recursos para publicar los documentos, y solo se entregaron copias del informe de OREALC a gobernadores o secretarios de educación de departamentos. Se piensa que los resultados no están “contextualizados” a la realidad regional, debido a la ausencia de estándares regionales y a que el proyecto no ofreció un marco interpretativo de los resultados con relación a los objetivos y alcances de las reformas y políticas educativas implementadas en las dos últimas décadas.</p>	<p>Débil impacto sobre la educación. Se considera que las conclusiones a las que llega el Laboratorio no están planteadas de forma que pueda afectar la toma de decisiones de política.</p> <p>Solo el ICFES tomó la iniciativa de profundizar el análisis de los datos.</p>

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Cuba	<p>La experiencia fue valorada como de amplio aprendizaje para diversos actores del área educativa. La participación en el diseño de las matrices curriculares permitió identificar falencias del currículo cubano. El debate y la confrontación técnica entre diferentes formas de evaluar permitió acceder a nuevos conocimientos sobre evaluación, como los relativos al diseño y calificación de respuestas abiertas y nuevas tecnologías de medición (modelo de Rasch y modelos jerárquicos lineales).</p>	<p>Se dio amplia difusión a los resultados del primer informe, con un marcado y deliberado cariz político, destacándolo como un logro significativo de la política social cubana.</p> <p>Se recibió con beneplácito la información sobre los efectos positivos del clima escolar sobre los aprendizajes, contenidos en el segundo informe.</p>	<p>Atendiendo a los buenos resultados, se decidió premiar a los docentes con un incremento del 30% en sus salarios, medida que también pretendía desalentar el doble empleo de los maestros, pues los resultados del estudio mostraban una relación negativa entre el multiempleo y altos rendimientos estudiantiles. En el terreno curricular, se decidió trabajar con mayor diversidad de textos, por lo que se incentivó la producción de nuevos materiales educativos.</p>
Perú	<p>Se consideró de gran utilidad un curso de capacitación de 15 días que el Laboratorio ofreció en Santiago de Chile.</p> <p>Críticas: limitadas posibilidades de influir sobre el diseño de la prueba, no se percibe que la experiencia haya constituido un espacio de identidad y comunicación regional.</p>	<p>Los representantes peruanos pusieron objeciones técnicas a los resultados y solicitaron postergar su publicación internacional hasta obtener mayores detalles sobre el análisis de los datos. Sin embargo, desde antes de la conclusión del estudio, se presumía que el gobierno no tenía intenciones de publicarlos, por razones políticas.</p> <p>La no publicación de los resultados fue interpretada en el país como una privación al derecho constitucional de acceso a la información pública.</p> <p>Los resultados fueron liberados durante el Gobierno de Transición, mediante una publicación de la Unidad de Medición de la Calidad Educativa del Ministerio.</p>	<p>El impacto más evidente de la participación en el estudio internacional del Laboratorio fue quizá la tardía publicación de los resultados, que es considerada un símbolo importante de la democratización del sector. Además, se observa que diferentes medios y actores sectoriales citan frecuentemente los bajos resultados del Perú con respecto a la región para sustentar argumentos sobre la prioridad de la política educativa.</p>

CUADRO 3

**Experiencia de cuatro países en las pruebas de la EIA:
TIMMS, TIMMS-R, Educación Cívica**

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Chile (TIMMS-R y Educación Cívica)	En el caso de la prueba de Educación Cívica, se desarrollaron capacidades relacionadas con las estrategias operativas en el trabajo de campo, con diseños de manuales, aplicación de instrumentos y análisis estadístico. Importante aprendizaje sobre el modelo psicométrico IRT, el cual pudo ser aplicado luego a los operativos de evaluación nacionales.	Los resultados de TIMSS-R fueron publicados solo en forma de ranking por un diario nacional, mientras que la Unidad de Currículum y Evaluación preparó un documento de difusión resaltando los aspectos más sustantivos de los datos recogidos. Para el estudio de Educación Cívica, que incluyó una opción nacional, los resultados incluyeron todos los datos comparativos internacionales de base. Los datos de la aplicación diferenciada (ítemes de opción nacional) fueron difundidos y analizados posteriormente. Los resultados de esta prueba se presentaron en una capacitación para 40 mil profesores.	Es notoria la cantidad de artículos escritos por los académicos de uno de los centros de investigación sobre políticas públicas de Chile, que utilizan los datos para hacer críticas y sugerencias concretas sobre la educación pública en el país. Desde el Ministerio de Educación se está intentando que la información llegue a los profesores, a través de instancias de perfeccionamiento docente, para que haya impacto en el aula y en los contenidos.
Colombia (TIMMS y Educación Cívica)	Fue una excelente oportunidad de aprendizaje para los académicos participantes, especialmente en la etapa de investigación del marco contextual de la educación en Colombia y en el desarrollo de criterios fundamentados para la calificación objetiva de las preguntas abiertas. En el caso de la prueba de Educación Cívica, se valora la oportunidad de participación en el establecimiento del marco conceptual que guiará la construcción de las pruebas.	Difusión de los resultados de Cívica, a través de talleres y seminarios para las escuelas participantes. Hubo importante interés del Ministerio de Educación en la divulgación de los resultados de TIMSS, que incluyó un seminario previo a la publicación de los resultados para funcionarios del Ministerio y del ICFES y publicaciones especiales para maestros de matemáticas con los resultados comparativos con los demás países participantes.	Sobre la base de los resultados, el Ministerio de Educación hizo publicaciones que ofrecían recursos pedagógicos para que los profesores logran mejores resultados de aprendizaje. Uno de los grandes impactos del TIMSS fue que alentó a los técnicos del Servicio Nacional de Pruebas del ICFES a rediseñar el examen que regula el ingreso a la educación superior, con un marco conceptual de referencia a criterios en vez de normas, y con análisis de respuesta al ítem.

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Estados Unidos (TIMSS, TIMSS-R en el caso de algunos estados)	<p>En el nivel local, el aprendizaje más valioso obtenido a través de la participación en TIMSS fue el relacionado al análisis e interpretación de datos.</p> <p>En el nivel nacional, se destacan los avances teóricos logrados mediante el desarrollo de nuevos marcos conceptuales y metodologías de investigación y análisis.</p>	<p>Se publicaron primero los resultados del análisis curricular y varios meses después el ranking internacional. Esto contribuyó a que, cuando se supo que Estados Unidos ocupaba un lugar modesto en ese ranking, las audiencias más especializadas ya tuvieran un conocimiento importante sobre los aspectos sustantivos del estudio.</p> <p>Se contrató una prestigiosa consultora de prensa de Nueva York para las relaciones públicas y las estrategias de difusión de TIMSS. La consultora proveyó entrenamiento a los coordinadores nacionales y asignó personal especializado a la “colocación” de noticias sobre los análisis de TIMSS en diferentes medios de comunicación del país.</p> <p>También en el nivel local se realizaron importantes esfuerzos para lograr una mejor interpretación y difusión de los resultados.</p>	<p>Los resultados de TIMSS pusieron en marcha una serie de medidas gubernamentales, no gubernamentales y profesionales para establecer estándares curriculares más altos, y para alinear los diferentes elementos de la gestión y la práctica educativa desde una perspectiva sistémica.</p> <p>En el nivel local o distrital, los esfuerzos se centraron en la alineación curricular local con los estándares estadales y en mejorar las condiciones pedagógicas para la entrega de mejores oportunidades de aprendizaje.</p> <p>Un efecto singular de TIMSS en el nivel local fue la creación de SMART, consorcio que ha decidido proveer extensas oportunidades de capacitación a docentes sobre currículo, métodos instruccionales y, especialmente, sobre cómo analizar datos de pruebas estandarizadas.</p>
México	Se lograron aprendizajes técnicos valiosos vinculados a técnicas de muestreo y de organización operativa para las aplicaciones.	El país invocó razones técnicas para desautorizar la publicación internacional de los datos.	Se conservaron las bases de datos con los resultados dentro del ámbito de la SEP, pero no fueron analizadas posteriormente.

CUADRO 4

Experiencia y/o expectativas de cuatro países en las pruebas PISA

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Chile (En proceso para participación en PISA 2003)	Se percibe que, en la medida en que se gana experiencia en medición internacional, y a partir de la unificación de los equipos de evaluación y currículo bajo una misma unidad administrativa, la relación entre los procesos de evaluación y desarrollo curricular se va fortaleciendo, especialmente a través de los procesos de revisión curricular sistemática que exigen las pruebas internacionales.	Aun cuando Chile no participó PISA 2000, el Ministerio de Educación difundió públicamente los resultados y anunció que Chile participaría en la siguiente aplicación del proyecto. Se espera difundir los resultados de PISA en forma más beneficiosa y efectiva que con las pruebas internacionales anteriores en que el país ha participado, porque se contará con más tiempo para la planificación de una estrategia adecuada y con mayor apoyo por parte de las organizaciones internacionales.	El gran desafío es lograr que los resultados de las pruebas no causen solo reacciones críticas contra la administración educativa nacional sino que, además, sean utilizados para implementar mejoras pedagógicas y curriculares. Sin embargo, dado que los nuevos marcos curriculares recién están terminando de implementarse en los años superiores de la educación media, se calcula que los resultados de PISA no promoverán cambios en el currículo en el corto plazo.
México (Participación en PISA Plus, PISA 2001 y en preparación para PISA 2003)	La experiencia de evaluación en el proyecto PISA ha permitido un mayor grado de participación técnica que en experiencias anteriores de evaluación internacional La sección de preguntas abiertas de la prueba ha demandado de las autoridades de la SEP una convocatoria más amplia a actores del medio educativo en el proyecto, específicamente para llevar a cabo la calificación de respuestas.	Los primeros destinatarios de los resultados serán los equipos técnicos y autoridades políticas de la próxima administración gubernamental nacional, lo cual aumentará las posibilidades de que los datos sean utilizados y que el proyecto evaluativo se continúe en el tiempo. Respecto a los resultados preliminares de PISA, que consistieron en una aplicación específica de PISA Plus para que México se comparase con otros países de la región antes de ingresar en la aplicación general de 2001, la Subsecretaría de Planeación y Coordinación Institucional ya ha difundido un documento interno. Los datos y las bases mismas también se han entregado a las administraciones estatales.	Ya se están realizando, en función de los datos preliminares obtenidos, algunas propuestas para orientar mejoras en el Programa Nacional de Lectura y en el Programa de Escuelas de Calidad. También se han realizado algunas propuestas para el desarrollo curricular de formación docente inicial, en el currículo de Matemáticas de Educación Básica, y se han facilitado las bases de datos para su análisis a la Academia Mexicana de Ciencias y al Centro de Investigación y Docencia Económica de México.

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Perú (Al momento del estudio, en proceso para participación en PISA 2003)	Se considera que esta experiencia brinda a los técnicos la posibilidad de ensayar con aspectos técnicos y conceptuales diferentes a los de las pruebas nacionales.	Se considera de suma importancia lograr que se difunda información sustantiva, no solamente el ranking internacional. Para un óptimo aprovechamiento de las información, la coordinación del estudio en Perú pretende realizar un taller donde se explique a actores clave del sector educativo cómo interpretar los resultados de PISA.	Se sabe que las noticias de prensa serán escandalosas, pero se cree que en la medida en que el país siga participando y el público se acostumbre a que hay otras informaciones rescatables además del ranking, los medios de comunicación deberán buscar otro tipo de noticias para vender a la población. Un aspecto positivo de la prueba es que el país será propietario de los instrumentos y podrá aplicarlos nuevamente si quiere profundizar las investigaciones.
Uruguay (En proceso para participación en PISA 2003)	Se perciben tres importantes posibilidades de aprendizaje con el programa PISA: 1) generará un mayor conocimiento para el diseño de ítems de las áreas que actualmente no son evaluadas por las pruebas nacionales; 2) permitirá desarrollar capacidades específicas con relación al diseño, calificación y procesamiento de la información de pruebas de respuestas abiertas, y mejor conocimientos de la teoría de respuesta al ítem; y 3) se aprenderá más sobre la medición de variables escolares, curriculares e instruccionales.	Se planea difundir ampliamente los resultados de la prueba, así como lo que pudiera revelarse del instrumento mismo, tal como se ha hecho con las pruebas nacionales. De esta manera, se podría explicar a cada equipo de profesores, en cada establecimiento, qué es lo que las diferentes actividades de la prueba miden, para que ellos puedan aplicar la prueba, o ítems seleccionados de la misma, y así comparar el desempeño de sus alumnos con el de los estudiantes en otros países.	Los técnicos nacionales de medición reconocen que, a pesar del buen uso que se pretende dar a las pruebas, la medición contra estándares internacionales como los de PISA conlleva un riesgo político considerable. Por lo tanto, antes de ingresar en una de estas pruebas, consideraron necesario contar con el aval de las autoridades políticas del sector para asumir el costo de salir en los últimos lugares del ranking internacional.

Las **iniciativas de participación**, así como la elección del proyecto de evaluación comparada más conveniente para cada país en América Latina, han estado en la práctica definidas por tres objetivos principales:

- a) La oportunidad de acceder a capacitación técnica en medición de aprendizajes y diseño de instrumentos. Existe, sobre la base de experiencias anteriores, una expectativa y percepción compartidas entre los países de la región en cuanto a los beneficios de participar en pruebas lideradas por agencias de alta capacidad técnica y organizacional como la IEA o la OCDE. Se valora especialmente lo que se puede aprender de la organización operativa de los proyectos, de la calidad técnica de los instrumentos de medición, de los procedimientos para garantizar la calidad y validez de la información y de los mecanismos que esas instituciones implementan para brindar apoyo y orientación a los países participantes. En cuanto a la vinculación de los proyectos comparativos con los sistemas nacionales de medición, se espera desarrollar capacidades técnicas que fortalezcan a estos últimos, pero no se ha diseñado generalmente un “cruce” sistemático entre los datos obtenidos en ambos tipos de mediciones.
- b) La posibilidad de obtener información objetiva sobre el logro académico de los estudiantes con relación a estándares internacionales de contenido, que impacte la opinión pública y facilite una toma de decisiones mejor informada para el mejoramiento del sistema educativo.
- c) La expectativa, al menos por parte de algunos actores, de que la participación en una prueba internacional estimule la propuesta de estándares regionales contra los cuales medir el rendimiento de los países latinoamericanos. Si esto último no es posible, se espera al menos participar en pruebas internacionales de cobertura mundial que ofrezcan un marco conceptual sólido que permita interpretaciones válidas para el contexto latinoamericano. Cabe destacar que, si bien muchas de las decisiones de participación en pruebas internacionales en América Latina han estado inspiradas en la idea de establecer estándares académicos más claros y exigentes y compararse con ellos, lo que demuestra un interés por los aspectos curriculares de las reformas, la mayor parte de las decisiones se ha tomado sin la participación directa de los responsables de las unidades de currículo de los organismos de gestión educativa nacionales, ni de otros actores que puedan ejercer influencia sobre el desarrollo curricular en cada país. En este sentido, se podría afirmar que las iniciativas y decisiones de participación en

pruebas internacionales están solo parcialmente alineadas con los objetivos nacionales de reforma y logro curricular.

Si bien existen algunas opiniones “supranacionales” en la región que entienden la evaluación internacional dentro de un contexto de política educativa basada en la noción de “gerenciamiento”, en parte motivada por el condicionamiento de los organismos internacionales, la mayor parte de las decisiones de participación no parece haberse tomado en función de ese modelo de gestión. Es decir, los beneficios que se esperan de la información recogida no están relacionados necesariamente a las nociones de descentralización y responsabilización por los resultados, sino más bien al deseo de proveer mayores herramientas de gestión para un adecuado desarrollo e implementación curricular. Ello, pese a lo señalado líneas arriba respecto a la débil vinculación de las iniciativas internacionales con las unidades responsables del currículo.

En cuanto a la fase de **implementación** de las pruebas internacionales, se aprecian grandes diferencias entre la experiencia de participación en la prueba del Laboratorio, por un lado, y en las pruebas de la IEA y de la OCDE, por el otro. Esas diferencias, están relacionadas principalmente con la capacidad técnica, organizacional, financiera y operativa de las agencias que administran las pruebas. La IEA y la OCDE aventajan al Laboratorio en todas esas características, a pesar de que la prueba regional reportó también varios beneficios a los países participantes.

En lo que se refiere a los aspectos estrictamente curriculares, la participación en pruebas de la IEA se ha dado en condiciones de desigualdad con respecto a los países desarrollados, pues América Latina ha ingresado tarde a los proyectos de evaluación, muchas veces con escasas capacidades técnicas para la representación, o con un débil aval financiero e institucional por parte de sus gobiernos nacionales. Por estas razones, las posibilidades de que los países de la región influyan en las decisiones con respecto a los marcos conceptuales y curriculares han sido extremadamente reducidas.

En el caso de la prueba del Laboratorio, si bien la matriz curricular fue discutida con los países, la falta de un mecanismo de comunicación multidireccional no permitió lograr un consenso con respecto a un conjunto de estándares deseables ni a un marco conceptual común que permitiera una interpretación más sustantiva y contextualizada de los resultados y, por lo tanto, la pertinencia curricular del proyecto se vio fuertemente debilitada. Recordemos que en las pruebas de la IEA o de la

OCDE en América Latina se valora precisamente que los países se midan con referentes o estándares externos a la región, mientras que en la prueba del Laboratorio se esperaba llegar a acuerdos regionales sobre medidas deseables de logro.

Una última diferencia que vale la pena resaltar es que, mientras las técnicas de muestreo y los instrumentos de las pruebas de la IEA y de la OCDE son apreciados por su rigor y validez, el Laboratorio enfrentó serios cuestionamientos técnicos por parte de los coordinadores nacionales. Por otro lado, la participación en las pruebas de ambas agencias ha favorecido la reflexión sobre los contenidos curriculares de cada país, en la medida en que las comparaciones revelan diferencias importantes en los enfoques epistemológicos y en la distribución de contenidos a lo largo del currículo de educación básica. Sin embargo, esa reflexión se ha dado principalmente entre técnicos y expertos de las unidades de medición, pero con escasa o ninguna participación de los profesionales de las áreas de desarrollo curricular. Desde el punto de vista de la política curricular, ello implica que la medición internacional en cada país tiene escasa vinculación con los procesos de implementación de los programas curriculares que, según se expresa, se desean mejorar mediante las evaluaciones.

Sin embargo, los aspectos más débiles de los proyectos de evaluación, según se observa en la experiencia de la mayor parte de los países latinoamericanos, han sido el de la **interpretación de los resultados** y su posterior **diseminación**. Además de las insuficiencias en el planeamiento estratégico, la escasa difusión de resultados estuvo marcada por la falta de comunicación y coordinación entre los equipos técnicos de medición y los de desarrollo curricular en cada país. Estos dos factores –falta de una estrategia de difusión y falta de trabajo conjunto entre actores clave– determinaron que se redujeran las posibilidades de impacto en el mejoramiento de la implementación de reformas y en la reformulación de objetivos curriculares.

Durante la prueba del Laboratorio, la debilidad en la interpretación y diseminación se debió, principalmente, a la falta de una estrategia preconcebida, así como a la falta de un marco conceptual sólido y de estándares que permitieran una adecuada interpretación sustantiva de los resultados, a la vez que el segundo informe oficial, que se esperaba tuviera más impacto de política, llegó a los ministerios demasiado tarde, cuando las administraciones nacionales habían cambiado. Cuba fue, de acuerdo a las entrevistas, una excepción en este sentido, dado que allí los buenos resultados obtenidos fueron ampliamente difundidos desde la Presidencia.

En cuanto a los efectos de los resultados de pruebas de la IEA, es evidente que las desfavorables posiciones de América Latina en los *rankings* internacionales han causado un efecto de *shock* en algunas audiencias, especialmente por la forma de reporte que ha utilizado la prensa. En la actualidad, los países que han decidido continuar participando en pruebas de la IEA y de la OCDE están conscientes de la necesidad de “educar” a la prensa y otras audiencias influyentes, de volverlas más especializadas, de manera que puedan difundir informaciones más relevantes que la temida tabla de posiciones. Los reportes que redactan las oficinas de evaluación o currículo de los ministerios, por su parte, se consideran particularmente útiles cuando pueden tener impacto en el ámbito local, escolar y pedagógico y cuando contribuyen a desarrollar capacidades de análisis e interpretación de datos. El caso de Estados Unidos durante la prueba TIMSS fue singular, por el alto involucramiento de diversos grupos de interés y la consulta sistemática a actores del sector, así como también por la fuerte estrategia de difusión utilizada.

Con respecto al **impacto** de la participación en pruebas internacionales en las decisiones sobre política educativa, de repercusiones políticas, o de investigaciones realizadas sobre los datos, nuevamente se hace evidente la diferencia entre la experiencia de América Latina y lo ocurrido en Estados Unidos. En nuestra región, tanto la evaluación del Laboratorio como las pruebas de la IEA han tenido un impacto político y de política relativamente bajo. Las excepciones, según nuestros datos, son Cuba, que tomó algunas decisiones concretas de políticas sobre la base de los resultados del Laboratorio, y Chile, que ha debido enfrentar el impacto sobre la opinión pública de los resultados en las pruebas de la IEA. En el caso del Laboratorio se han realizado algunas investigaciones posteriores, en los países y por parte de la OREALC misma, pero hasta el momento casi no han sido difundidas.

Por otra parte, las implicancias curriculares de las pruebas internacionales en América Latina, han sido casi nulas puesto que no se registra evidencia de que los resultados hayan alertado a los profesionales e investigadores de la educación sobre la pertinencia o adecuación de los contenidos curriculares nacionales. La excepción, hasta donde pudo verificar este estudio, es también Cuba dado que allí la experiencia de evaluación comparada permitió alertar a los representantes nacionales sobre la necesidad de introducir cambios en la secuenciación de algunos conceptos matemáticos, así como sobre el enfoque teórico con que se plantean los contenidos en el área de lenguaje.

En cuanto a la participación en pruebas de la IEA, la experiencia ha repercutido casi exclusivamente en el desarrollo técnico de las unidades de medición, lo cual es significativo, pero no se reportan medidas importantes de política curricular, o educativa en general, en función de los resultados.

Con respecto a PISA, una ventaja anticipada del proyecto que puede repercutir positivamente en el medio educativo es que los instrumentos mismos de medición, serán propiedad de los países participantes, y se podrán utilizar en aplicaciones nuevas e independientes para medir otras muestras y así favorecer investigaciones más profundas en el ámbito nacional.

En Estados Unidos, donde existen más de cien años de trayectoria en medición y evaluación educacional, la experiencia ha sido muy diferente, particularmente con relación a las pruebas TIMSS y TIMSS-R de la IEA. La creación de consorcios educativos, el involucramiento de organizaciones sindicales docentes y asociaciones profesionales de las diferentes disciplinas académicas, así como las múltiples líneas de investigación surgidas a partir de los resultados, confirman que el impacto de la participación en pruebas internacionales depende no solamente del potencial informativo de esas pruebas, sino también de las estrategias de sensibilización previas a la difusión final de los resultados⁵.

Estas diferencias entre las experiencias de Estados Unidos y América Latina confirman que **la ausencia de un diálogo informado entre los funcionarios y usuarios de la información** sistemática, resulta en un escaso impacto de las evaluaciones en las decisiones sobre políticas educativas, y sobre políticas curriculares en particular. (Ver recuadro 1).

Adicionalmente, dado que no se monitorean los resultados sobre las pocas decisiones que se toman en función de las informaciones de estudios comparados, es difícil conocer si se ha avanzado o no en la dirección deseada. Pero hasta el momento la contribución de los resultados de pruebas internacionales a nuevos ciclos de política curricular, o de mejoramiento de los existentes, es mínima y en muchos casos inexistente. En otras palabras, según indican los datos de las pruebas internacionales, es muy probable que las iniciativas y decisiones centrales de política curricular que se tomen en el futuro no consideren la evidencia sobre los resultados de políticas y programas anteriores.

⁵ Vale recordar que para cuando la IEA liberó los resultados de TIMSS, diferentes audiencias y medios de información ya estaban expectantes y deseosos de conocer los resultados.

V. RECOMENDACIONES PARA FUTURAS OPORTUNIDADES DE PARTICIPACIÓN EN PRUEBAS INTERNACIONALES

La evidencia recogida en este trabajo, si bien no se puede considerar representativa de toda la región, ha permitido arribar a algunas conclusiones iniciales sobre la conveniencia de la participación de países latinoamericanos en futuras pruebas internacionales, así como de las condiciones en que esa participación debería darse.

1. Dado que muchos de los países de la región ya han participado en alguna prueba internacional con anterioridad, se recomienda utilizar la información de esas pruebas para generar opinión y sensibilizar a audiencias de interés antes de iniciar nuevos programas de reforma o mejoramiento curricular, y antes de adscribirse a un nuevo proyecto de evaluación comparada internacional. Esta medida tendría el valor adicional de permitir el pilotaje de posibles estrategias de difusión de los resultados.

Una forma viable de sensibilizar a los representantes de la prensa, a las organizaciones sindicales y a los profesionales de la educación es invitarlos a participar en las discusiones iniciales del proyecto, así como en la discusión sobre resultados anteriores, sobre la importancia de iniciar nuevas evaluaciones y sobre el vínculo concreto que se percibe entre las evaluaciones y los contenidos curriculares. Es importante que los coordinadores en el ámbito nacional calculen con anticipación, y en colaboración con otras oficinas ministeriales, el costo material y el tiempo de trabajo que requerirá la fase de discusión y sensibilización, para que esta sea un aspecto formal y viable del proyecto de evaluación.

Durante las discusiones, y mediante la información que se haga circular sobre resultados anteriores, se sugiere alertar sobre los posibles bajos resultados comparativos del país, pero simultáneamente ejemplificar la manera en que la próxima prueba comparada puede informar sobre aspectos específicos de desarrollo o implementación curricular. Paralelamente, esta fase puede comenzar a “educar” a las audiencias convocadas en algunos de los tecnicismos propios de la psicometría y de la sociometría, para que comiencen a desarrollar algunas capacidades de interpretación y análisis de datos más allá de los *rankings*.

Si el proyecto de evaluación está liderado por una institución académica o no gubernamental, es especialmente importante reunir a las

autoridades ministeriales de medición y currículo, no solo para obtener el aval oficial, sino para establecer una especie de “contrato” sobre las necesidades de información percibidas por el ministerio, u organización equivalente, y el tipo de información que puede derivarse de la evaluación comparada.

2. Al momento de decidir sobre el tipo de prueba internacional que se va aplicar, cada país debería considerar los siguientes **aspectos técnicos y contextuales** que afectarán la mayor o menor utilidad de las experiencias de evaluación comparada:
 - a. El **momento histórico** de las reformas curriculares en curso, la duración prevista del gobierno en funciones, y el grado de afianzamiento del sistema nacional de medición de la calidad educativa.
 - b. Las **necesidades reales de la información** que se va a recoger, y las posibilidades concretas de aprovecharla, de acuerdo a los momentos de desarrollo e implementación curricular del país y de acuerdo a las dificultades percibidas o comprobadas anteriormente en esos procesos.
 - c. Las **características de las pruebas**, es decir, el marco conceptual utilizado, los dominios disciplinarios o cognitivos evaluados, y los factores escolares y socioeconómicos relevados con relación al rendimiento. Si la prueba es de cobertura regional y se presenta como un proyecto de elaboración conjunta entre los países participantes, es importante definir un **marco conceptual** que considere el contexto de política educativa y curricular de la región y permita una interpretación contextualizada de los resultados del estudio. Un estudio de esta naturaleza debería representar también una oportunidad para definir con los técnicos de los demás países un conjunto de **estándares** mínimos de contenido para realizar una revisión de los currículos nacionales y comparar, mediante la evaluación, los alcances relativos de los currículos prescritos, implementados y logrados.
 - d. Las **oportunidades de participación efectiva** que el proyecto ofrece con relación a la construcción o adecuación de instrumentos de medición, o a la inclusión de “opciones nacionales” de evaluación; y la capacidad técnica de los coordinadores nacionales para hacer efectiva y viable esa participación.
 - e. La **utilidad y pertinencia del informe** de resultados utilizado por la agencia evaluadora, de acuerdo al tipo de audiencias que se espera informar.

3. Durante la etapa de implementación del proyecto se recomienda motivar y asegurar el vínculo y la **comunicación permanente** entre las unidades de medición y los equipos de desarrollo curricular en los niveles central y regional en cada país. Si parte de estos procesos es eminentemente técnica y requiere poca intervención de personal de currículo, se recomienda mantenerlos informados permanentemente sobre los avances del proyecto y sobre las fechas en que serán convocados para discutir y analizar los resultados del estudio. Simultáneamente, debería mantenerse una comunicación permanente con los representantes de la prensa y de las asociaciones sindicales y profesionales convocadas inicialmente (recomendación 1).

El mismo mecanismo debería operar entre los coordinadores y los actores gubernamentales y no gubernamentales si el operativo de evaluación estuviera en manos de una institución académica independiente. Estos mecanismos de comunicación permanente son una forma de mantener vivo el “contrato” entre los técnicos y los consumidores de información que más pueden influir en la toma de decisiones para cambios de política y práctica curricular. Como existen altas posibilidades de que los usuarios gubernamentales (ministeriales) de la información cambien súbitamente, es más importante aún reforzar los canales de comunicación permanente con la prensa y con representantes de instituciones sindicales, académicas, profesionales y de la sociedad civil.

4. Hacia el momento en que se reciben los datos ya procesados de la agencia evaluadora, y de acuerdo a las recomendaciones anteriores, los posibles usuarios de la información deberían estar informados de las fechas en que serán convocados a **analizar e interpretar** los datos de la prueba. Aquí se sugiere realizar primero una sesión de análisis y discusión entre los técnicos de evaluación y los representantes de desarrollo curricular, para planificar en conjunto una agenda de discusión organizada para el momento de intercambio con representantes de la prensa, del mundo académico y de organizaciones no gubernamentales. Las discusiones con estos actores deberían conducir a una serie de recomendaciones puntuales sobre el tipo de informes que sería deseable difundir, concretamente con relación al formato, al tipo de información que se debe incluir, a las audiencias prioritarias y secundarias que se quiere alcanzar, al número sugerido de publicaciones y a los diferentes objetivos que los informes debieran perseguir

(por ejemplo, si deben brindar insumos de información para medidas de política o para decisiones de tipo curricular y pedagógico en el nivel local y escolar).

Según indica la evidencia, la difusión de la información por medio de informes no es suficiente para lograr un impacto real en las políticas y prácticas educativas. Por ello, es necesario organizar eventos locales con amplia participación de directores de escuela y docentes, en los cuales se explique cómo interpretar los datos y se discutan las diferentes maneras de mejorar el desarrollo curricular en las escuelas y su implementación en las aulas. Este es también el momento de **revisar el contexto de política curricular**: en qué momento de implementación de los programas curriculares se encuentran el país y sus regiones; cuáles son las dificultades que evidencian las pruebas nacionales, los informes de supervisión regionales, las investigaciones académicas y la prueba internacional misma para decidir qué tipo de mejoras (si no cuáles exactamente) se pueden recomendar a las autoridades nacionales y subnacionales de educación.

En este sentido, el análisis de los datos debería conducir a preguntas clave sobre la forma en que otros países logran mejores rendimientos estudiantiles y sobre las condiciones bajo las cuales esos logros son posibles. Vale recordar que una de las ventajas de las pruebas internacionales es que pueden revelar información sobre los efectos diferenciales de algunas variables sobre el logro de aprendizajes que no es posible detectar en un estudio intranacional. También es importante recordar que las asociaciones entre variables dependientes e independientes que muestran los estudios internacionales deberían tomarse con cautela y ser analizadas dentro del contexto particular de cada país.

5. Se recomienda hacer un seguimiento del tipo de decisiones de política y práctica educativa, y curricular en particular, tomadas en función de los resultados difundidos. Una vez identificadas, se sugiere comisionar o alentar a instituciones académicas el estudio de los resultados que sobre el aprendizaje de los estudiantes han tenido, o no, esas decisiones y cambios de prácticas. La información recogida mediante estos estudios debería canalizarse y difundirse adecuadamente para ofrecer evidencia más concreta para la toma de decisiones sobre nuevas políticas y programas curriculares, así como para la decisión sobre nuevas evaluaciones internacionales y el tipo de estrategias que son más efectivas para la difusión y uso de sus resultados.

6. Una vez que se ha participado en una experiencia internacional de evaluación, y si la información obtenida de esa experiencia está a la altura de las expectativas de los coordinadores nacionales y de las autoridades políticas del sector, se aconseja continuar participando en nuevas mediciones, especialmente en las siguientes aplicaciones de la misma prueba. Esto permitiría la medición del progreso académico de las poblaciones evaluadas y, además, que el público y los usuarios de la información se acostumbren a contar con esta periódicamente, y de esta manera maximizar su uso técnico y político.
7. Se recomienda la participación de unidades subnacionales de gestión educativa (por ejemplo, estados, departamentos o grupos de escuelas asociadas). Si la participación no puede ser financiada por esas unidades, como probablemente ocurriría en América Latina, la coordinación central de la evaluación (ministerios u otras agencias a cargo) debería considerar la posibilidad de conseguir recursos para promover esa participación. Es deseable, también, que la muestra nacional, en vez de estar conformada por alumnos de todo el país, sea construida con casos de un departamento o estado que presente la suficiente heterogeneidad demográfica como para asegurar variabilidad en los resultados y, por ende, la posibilidad de analizar estadísticamente su relación con las variables predictivas del logro. Ello permitiría diseñar más fácilmente estrategias de difusión y uso de los datos en un contexto más reducido, para luego replicar esas estrategias en el ámbito nacional.

Esperamos con este trabajo haber contribuido a la comprensión de las condiciones técnicas y políticas bajo las cuales los países de América Latina deberían participar en pruebas internacionales. Es nuestro deseo que los resultados de este informe de investigación alcancen y se difundan en los espacios de decisión sobre evaluación educativa en la región, así como en el ámbito académico, y así coadyuvar a una mejor utilización de los recursos disponibles para la medición de rendimiento académico y, lo que es más importante aún, para su impacto en la calidad de los aprendizajes.

RECUADRO 1

DÍALOGO ENTRE INVESTIGADORES, EVALUADORES Y USUARIOS DE LA INFORMACIÓN

Uno de los aspectos que representa el mayor obstáculo para el uso efectivo de la evaluación internacional en el mejoramiento de la calidad de los aprendizajes, según se ha detectado en este estudio, es la ausencia de diálogo entre investigadores o evaluadores y los usuarios potenciales de la información.

Una explicación al respecto se puede encontrar en Reimers (1999), quien sostiene que en las altas esferas de administración educativa en muchos países del mundo, la mayor parte de las decisiones que se toman para incrementar las oportunidades de aprendizaje se realizan intuitivamente, especulando con la articulación de algunos de los elementos del sistema (currículo, capacitación, supervisión, etc.), en vez de tomarse como resultado del análisis cuidadoso de la evidencia que muestran las investigaciones y los análisis de política. El autor argumenta que esto se debe a la forma en que el conocimiento derivado de la investigación es acumulado y procesado, y a que no siempre se reconoce la diferencia entre la investigación que pretende simplemente explicar la realidad, y la investigación de política que, además, busca describir cómo podría cambiarse esa realidad, que se asume es lo que los funcionarios necesitan saber.

Esa función propositiva de la investigación debería darse en forma de **diálogo entre los investigadores, funcionarios y otros usuarios de la información**, de tal manera que se logre un acuerdo entre las necesidades de información y las posibilidades reales de los investigadores de proveerla.

El modelo racional de toma de decisiones, que asume que los funcionarios seleccionarán opciones de política sobre la base de un análisis informado, está siendo cuestionado. Se reconoce que la mayor parte de los problemas no están, ni pueden ser, perfectamente definidos y que los funcionarios no siempre establecen las prioridades de acción racionalmente sino en función de variables de oportunidad política y de poder. Por ello, para lograr decisiones informadas, es preciso ejercer cierto grado de persuasión en los decisores de políticas mediante un proceso de negociación, y no de diseño unilateral, del proceso de reforma. Ese proceso debe ser interdisciplinario, propositivo –y no solamente descriptivo–, y debe tener en cuenta que el análisis se debe realizar y comunicar dentro de los límites de tiempo que imponen los ciclos de política.

Adicionalmente, Reimers propone el uso de un “mediador de conocimiento” entre el mundo de la investigación y el mundo de la política. La evidencia recogida para este estudio muestra que, en ciertas ocasiones, esa mediación se logra a través de la prensa y de otras instituciones académicas y no gubernamentales formalmente convocadas durante el proceso de evaluación internacional. Sostiene Reimers que “el problema del uso de la información no puede ser considerado recién después de concluida la investigación o el análisis [...] la investigación debe comenzar con la consideración de las necesidades del cliente si se intenta evitar el riesgo de que los resultados del análisis resulten irrelevantes para los funcionarios”.

BIBLIOGRAFÍA

- ATKIN, Myron y Paul BLACK (1997). *Policy perils of international comparisons: The TIMSS case*. Phi Delta Kappan, 79,1, 22-28.
- BALL, Stephen (1998). "Big policies/small world: An introduction to international perspectives in education policy". *Comparative Education*, 34, 2, 119-130.
- BEATON, Albert *et al* (1999). *The benefits and limitations of international educational achievement studies*. International Institute for Educational Planning, International Academy of Education. Unesco, París.
- BEMPECHAT, Janine *et al.* (Borrador, 2000). "Cultural-cognitive issues in academic achievement: New directions for cross-national research". *Board on International Comparative Studies in Education (BICSE)*, National Research Council.
- BLOOM, Benjamin (1974). "Implications of the IEA studies for curriculum and instruction". *School Review*, 82, 3, 413-435.
- BURSTEIN, L. (1989). "Conceptual considerations in instructionally sensitive assessment". Documento presentado en American Educational Research Annual Meeting, San Francisco.
- BURSTEIN, L. *et al.* (1992). "Education indicators", en M. Atkin (Ed.), *Encyclopedia of educational research* (6ª edición). Macmillan, Nueva York.
- CARTER, D.S.G y M.H. O'NEILL (1995). *International perspectives on educational reform and policy implementation*. Falmer, Brighton.
- GRADE (2001). "Mejoramiento de los procesos de desarrollo curricular y de materiales educativos". Informe final de la consultoría N° 010-2000-ED/MECEP para el Ministerio de Educación del Perú.
- FARRELL, Joseph (1979). *The necessity of comparisons in the study of education: The salience of science and the problem of comparability*. Presidential address, annual meeting of CIES, Ciudad de México. Reimpreso en *Comparative Education Review*, 23, 1, 1978.
- FERRER, Guillermo (1999). "Aspectos del curriculum prescrito en América Latina: Revisión de tendencias contemporáneas en curriculum, indicadores de logro, estándares y otros instrumentos". GRADE-PREAL. <http://www.grade.org.pe/gtee-preal/docs/Ferrer.pdf>
- FORSTER, Margaret (2000). *A policy maker's guide to international achievement studies*. ACER Press, Australia.

- GREANEY, Vincent y Thomas KELLAGHAN (1996). *Monitoring the learning outcomes of education systems*. Banco Mundial, Washington DC.
- HUSÉN, Torsten (1994) "Educational research and policy-making", en T. Husén y T. Postlethwaite (Eds), *International Eyclopedia of Education*. Pergamon Press, Oxford.
- HUSÉN, Torsten (1987). "Policy impact of IEA research". *Comparative Education Review*, 31, 1, 29-46.
- HUSÉN, T. y A. TUIJNMAN (1994). "Monitoring standards in education: Why and how it came about", en A.Tuijnman y T. Postlethwaite (Eds), *Monitoring the Standards of Education*. Pergamon Press, Oxford.
- JONES, P.W (1998). "Globalisation and internationalism: Democratic prospects for world education". *Comparative Education*, 34, 133-155.
- KELLAGHAN, Thomas (1996). "IEA studies and educational policy". *Assessment in Education*, 3, 2, 133-160.
- KOHN, Melvin (1987). "Cross-national research as an analytic strategy" (American Sociological Association, 1987 Presidential Address). *American Sociological Review*, 52, diciembre, 713-731.
- LEVIN, Benjamin (2001). "Conceptualizing the process of education reform from an international perspective". *Education Policy Analysis Archives*, 9, 13. <http://epaa.asu.edu/epaa/v9n13.html>
- LYOTARD, J.F (1984). *The postmodern condition: A report on knowledge*. Manchester University Press, Manchester.
- MACNAB, Donald (2000). "Forces of change in mathematics education: the case of TIMSS". *Educational Policy Analysis Archives* 8,15.
- MCKNIGHT, Curtis (1987) *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. A National Report on the Second International Mathematics Study.
- MISLEVY, Robert (1995). "What can we learn from international assessments?" *Educational Evaluation and Policy Analysis*, invierno 1995, 17, 4, 419-437.
- National Commission on Excellence in Education (1983). *A nation at risk*. Government Printing Office, Washington DC.
- PLATT, W.J (1975). "Policy making and international studies in educational evaluation", en A.C. Purves y D.U. Levine (Eds.) *Educational policy and international assessment*. McCutchen, Berkeley, CA.

- REIMERS, Fernando y Noel MC GINN (1997). *Informed Dialogue: Using Research to Shape Education Policy Around the World*. Praeger, Nueva York y Londres.
- SCHMIDT, William y Curtis MCKNIGHT (1995). "Surveying educational opportunity in mathematics and science: An international perspective", *Educational Evaluation and Policy Analysis*, otoño 1995, 17, 3, 337-353.
- SMITH, Adam (1776). *An inquiry into the nature and causes of the wealth of nations*. Dublin, Whitestone, 1776. 3 v. 22 cm. The Library of Congress, EEUU.
- THEISEN, Gary *et al* (1983). "The underachievement of cross-national studies of achievement". *Comparative Education Review*, 27, 1, 46-68.
- TORNEY-PURTA, Judith (1990). "International comparative research in education: Its role in educational improvement in the US". *Educational Researcher*, (volumen, S/D) 32-35.
- TORRES, Rosa María (1999). "*Reformadores y docentes: El cambio educativo atrapado entre dos lógicas*", Documento de discusión. IIPE Unesco, Buenos Aires.
- WIRT, Frederick (1980) "*Comparing educational policies: Theories, units of analysis, and research strategies*". Reimpreso en *Comparative Education Review*, vol.24, 2.
- WOLF, R (1979). "*Sampling*". Boletín # 4: *Secondary Study of Mathematics*. Second International Mathematics Study, Urbana, IL.

ENTIDADES ENTREVISTADAS

(Por razones de confidencialidad, solamente se reportan los nombres de las agencias o unidades ministeriales a la que pertenecen los 25 entrevistados de este estudio)

Argentina

- IDECE (Instituto para el Desarrollo de la Calidad Educativa). Ministerio de Educación.
- Ex Subsecretaría de Evaluación de la Calidad Educativa. Ministerio de Educación.

Chile

- Programa Formación Inicial Docente. División Educación Superior.
- Unidad de Curriculum y Evaluación/ SIMCE. Ministerio de Educación.

Colombia

- ICFES (Instituto Colombiano para el Fomento de la Educación Superior)
- Universidad del Valle. Ciudad de Cali.

Cuba

- Instituto Central de Ciencias Pedagógicas.

Ecuador

- APRENDO. EB/PRODEC. Sistema semiautónomo oficial de medición de la calidad de la educación.

Estados Unidos

- Universidad de Wisconsin-Madison.
- Universidad Estadual de Michigan/ Universidad Estadual de Nueva York, Albany.
- Federación Estadounidense de Profesores.
- Departamento de Educación. (American Federation of Teachers. Educational Issues Department.)
- Consorcio SMART (Science and Mathematics Required for Tomorrow).

México

- Subsecretaría de Planeación y Coordinación Institucional, Secretaría de Educación Pública.

Organización de Estados Iberoamericanos (OEI)

- Programa Evaluación de la Calidad de la Educación.

Perú

- Radio Programas del Perú.
- Unidad de Medición de la Calidad. Ministerio de Educación.
- UNESCO-Perú.

Unesco-OREALC

- Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación (LLECE).

Uruguay

- Unidad de Medición de Resultados Educativos (UMRE) de la Administración Nacional de Educación Pública.

SIGLAS Y ABREVIACIONES

AFT: American Federation of Teachers, ‘Federación Estadounidense de Maestros’.

ALL: Adult Literacy and Lifeskills, ‘Prueba para Adultos sobre Alfabetización y Destrezas para la Vida’.

BID: Banco Interamericano de Desarrollo

ETS: Educational Testing Services, ‘Servicios de Evaluación Educativa’.

ICFES: Instituto Colombiano para el Fomento de la Educación Superior.

IDECE: Instituto para el Desarrollo de la Calidad Educativa.

IEA: International Association for the Evaluation of Educational Achievement, ‘Asociación Internacional para la Evaluación del Logro Académico’.

IRT: Item Response Theory, ‘Teoría de Respuesta al Item’.

LLECE: Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación.

NCES: National Council for Education Statistics, ‘Consejo Nacional de Estadística de la Educación’.

NCTM: National Council of Teachers of Mathematics, ‘Consejo Nacional de Profesores de Matemáticas’.

NSF: National Science Foundation, ‘Fundación Nacional de Ciencias’.

OCDE: Organización para la Cooperación y el Desarrollo Económico.

OEI: Organización de Estados Iberoamericanos.

ONG: Organización no-gubernamental.

OREALC (Unesco): Oficina Regional para América Latina y el Caribe.

PIRLS: Progress in International Reading Literacy Study, ‘Estudio Internacional sobre los Progresos en Alfabetización y Habilidades de Lectura’.

PISA: Programme for International Student Assessment, ‘Programa Internacional de Evaluación de Estudiantes’.

PNUD: Programa de las Naciones Unidas para el Desarrollo.

PROMEDLAC: Comité Regional Intergubernamental del Proyecto Principal de Educación en América Latina y el Caribe.

REDUC: Red Latinoamericana de Información y Documentación en Educación.

SIMS: Second International Mathematics Study, ‘Segundo Estudio Internacional de Matemáticas’.

SACMEQ: Southern Africa Consortium for Monitoring Educational Quality, ‘Consortio de África del Sur para la Evaluación de la Calidad Educacional’.

SEP: Secretaría de Educación Pública.

SIMCE: Sistema de Medición de la Calidad Educativa.

SMART: Science and Mathematics Achievement Required for Tomorrow.

TIMSS: Third International Mathematics and Science Study, Tercer Estudio Internacional en Matemáticas y Ciencias’.

TIMSS-R: Repetición de TIMSS.

UMRE: Unidad de Medición de Resultados Educativos.

WEI: World Education Indicators, ‘Indicadores Mundiales de Educación’.

Los resultados de las evaluaciones educativas en la prensa*

PEDRO RAVELA

PRESENTACIÓN

El Grupo de Trabajo sobre Estándares y Evaluación (GTHE) de PREAL ha trabajado durante los últimos años en torno a cómo mejorar el uso y difusión de los resultados de las evaluaciones nacionales de logros educativos en los países de la región.

Los sistemas nacionales de evaluación educativa nacieron en casi toda la región durante la década de los 90, respondiendo a una necesidad crucial para los sistemas y políticas educativas: contar con información apropiada acerca de los conocimientos y competencias que los niños y jóvenes adquieren a partir de su participación en los centros de enseñanza.

La mayoría de los sistemas educativos latinoamericanos ha justificado la creación de unidades de evaluación de logros de aprendizaje con argumentos similares: contribuir a la mejora de la calidad y equidad del sistema educativo; entregar información relevante, válida y confiable para la toma de decisiones de política educativa; promover la mejora de las prácticas de enseñanza en las escuelas, etc.

Si bien en la práctica las modalidades de implementación y las estrategias de inserción del sistema de evaluación en el sistema educativo han sido diver-

* Este documento fue elaborado en el marco de las actividades del Grupo de Trabajo sobre estándares y evaluación. Un corto resumen fue publicado en la serie Formas y Reformas de la Educación, Políticas N° 17, marzo de 2004.

sas –y algunos fines han recibido más énfasis que otros en distintos países–, casi todos coinciden en sus definiciones de propósitos en cuanto a la necesidad de difundir y utilizar la información recogida para rediseñar o ajustar políticas, planes, programas y prácticas pedagógicas y de gestión escolar.

Sin embargo, la concreción de estos propósitos ha sido más bien limitada. Ello se refleja en el escaso desarrollo –común a toda la región– de mecanismos específicos y efectivos de divulgación y aprovechamiento de resultados, cuestión que es claramente percibida por muchos actores sociales que, aún así, siguen confiando en el potencial de las pruebas nacionales e internacionales de logros como instrumento de política educativa. Tal confianza podría ir disminuyendo, sin embargo, si no se emprenden acciones específicas para fortalecer significativamente esta dimensión central de las evaluaciones: la divulgación y uso de los resultados.

En general se puede afirmar que queda aún un largo camino por recorrer, tanto en relación a la mejora de la calidad técnica de los sistemas de evaluación como en cuanto a la difusión y uso de los resultados para la mejora de la acción educativa. En parte ello obedece a que muchos sistemas de evaluación han tenido una historia discontinua como consecuencia de los vaivenes políticos; varios sistemas nacionales están en estos momentos en proceso de revisión.

Al mismo tiempo, cada vez adquieren mayor fuerza y presencia en el escenario educativo las evaluaciones de carácter internacional que comparan los conocimientos y capacidades logrados por los estudiantes en los distintos países. También en este caso parece haber mucho por aprender respecto al modo de aprovechar efectivamente los resultados de este tipo de estudios.

El GTEE/PREAL, cuya misión es contribuir a consolidar y mejorar los emergentes sistemas de evaluación de América Latina, se ha propuesto examinar las prácticas actualmente vigentes en cuanto a la difusión y uso de los resultados de las evaluaciones, para proponer medidas concretas que podrían tomarse en el corto, mediano y largo plazo para acercar lo real a lo deseable.

En este sentido, con anterioridad, el GTEE ha publicado un estudio sobre los informes de resultados producidos por los países, denominado “*¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?*”¹. Más recientemente, se ha publica-

¹ RAVELA, P., 2002. Documentos PREAL N° 22. Este documento puede ser consultado en la página WEB del Grupo de Trabajo de Estándares y Evaluación de PREAL: <http://www.preal.org>

do el documento *“Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: criterios para guiar futuras aplicaciones”*².

El presente documento focaliza su atención en cómo los resultados son recogidos y entregados a la opinión pública a través de los medios de prensa.

La prensa es uno de los medios principales de difusión de información en la sociedad en todos los temas, y también en materia de resultados de las evaluaciones educativas. El público en general, las familias, e inclusive muchas veces los políticos y hasta los propios profesores, se informan acerca de dichos resultados principalmente a través de las versiones que aparecen en la prensa acerca de los mismos. No es descabellado suponer que son muchas más las personas que leen información sobre resultados de las evaluaciones en algún medio de prensa que aquellas que lo hacen directamente de los reportes producidos por las Unidades de Evaluación de los Ministerios de Educación.

Sin embargo, el tema de los resultados de las evaluaciones educativas es relativamente nuevo y suficientemente complejo desde el punto de vista técnico como para que se cometan errores y existan problemas en los modos en que el tema es abordado y transmitido a la opinión pública por los medios de prensa, tanto por desconocimiento por parte de los medios, como por carencias en los modos en que los Ministerios entregan los resultados.

Usualmente los sistemas de evaluación comunican los resultados a los medios llamando a una rueda de prensa, emitiendo un comunicado que resume los resultados más importantes o haciendo ambas cosas. En muchos casos los medios de prensa se limitan a publicar los comunicados, sin hacer muchas modificaciones. En otros, hacen interpretaciones de los datos presentados tratando de enfatizar aspectos que puedan llamar la atención del público, aunque no necesariamente dichas interpretaciones se deriven de los datos. En algunos casos se realizan entrevistas a actores sociales relevantes o a funcionarios del gobierno, para que comenten los resultados.

En términos generales, en los países de la región, la prensa tiende a destacar lo negativo cuando se presentan los resultados de un levanta-

² FERRER, G., y ARREGUI, P., 2003. Documentos PREAL N° 26. Este documento puede ser consultado en la página WEB del Grupo de Trabajo de Estándares y Evaluación de PREAL: <http://www.preal.org>.

miento de información. Posteriormente, es muy poco lo que se publica sobre el tema, aunque se siguen utilizando algunos datos esporádicamente en diferentes artículos y contextos.

Lo anterior no tiene en principio nada de extraordinario. De hecho, parece la manera normal en que se presenta también la información producida en otros campos. Sin embargo, la complejidad de los resultados de las evaluaciones educativas, las limitaciones técnicas que algunos sistemas tienen y la falta de referentes por parte de periodistas y del público para interpretar correctamente la información, muchas veces impide que estos resultados tengan el impacto esperado: concientizar y movilizar a los diferentes sectores para que se involucren en la mejora de la calidad de la educación. Por el contrario, los efectos que generalmente está teniendo la difusión de resultados son la búsqueda de posibles culpables, el descrédito del sistema educativo en general y el sentimiento de que por mucho que se está haciendo no se está llegando a ningún lado.

La queja más común que se escucha de parte de los ministerios de Educación y, más específicamente, desde los sistemas de evaluación, es que el público no tiene la formación necesaria para entender los datos y que, lo que es peor, reciben una visión distorsionada a través de la prensa, que generalmente se concentra en lo negativo.

Pero, al mismo tiempo, debe señalarse que muchos de los problemas se originan en que los propios ministerios de Educación y unidades de evaluación no tienen una visión clara de qué pretenden de los resultados, qué quieren comunicar y cómo hacerlo adecuadamente. O, también, que lo que pretenden comunicar responde más a los intereses y perspectiva de los especialistas en evaluación que a las preguntas que el público se hace y espera ver respondidas.

En el marco de este panorama, sin duda complejo, el propósito de este trabajo es ofrecer una primera mirada sobre el tratamiento que la prensa en la región ha venido dando a los resultados de las evaluaciones. El objetivo es doble:

- por un lado, aportar a los periodistas elementos que les ayuden a informar mejor sobre resultados de las evaluaciones educativas,
- por otro, desarrollar un conjunto de reflexiones acerca de cómo trabajar mejor con los medios de prensa desde los ministerios de Educación.

El trabajo está escrito desde un doble supuesto.

En primer lugar, que la producción de información a partir de las evaluaciones es abundante, rica y compleja. Los informes oficiales que se publican probablemente son leídos de manera directa por un público muy minoritario. Desde esta perspectiva, establecer una suerte de alianza con al menos una parte del mundo periodístico, con el fin de desarrollar una labor de difusión pública didáctica y apropiada, es un camino importante a recorrer para que una mayor cantidad de ciudadanos se informen sobre el estado de la educación y se interesen por ella.

En segundo lugar, no se puede ser ingenuo respecto al papel de los medios de comunicación. Mientras algunos de ellos de seguro estarán genuinamente interesados en contribuir a la formación del ciudadano y al desarrollo de debates informados sobre la educación, no se puede ignorar que en muchos medios tienen mayor peso otro tipo de intereses: o bien relacionados con la venta del propio medio, lo que en muchos casos lleva al sensacionalismo fácil en el tratamiento de la información, o bien relacionados con posturas político-partidarias, lo que en muchos casos lleva a usar los datos con la finalidad exclusiva de estigmatizar ante la opinión pública las políticas educativas del gobierno de turno, independientemente de que estas sean más o menos buenas, regulares o malas.

Lo afirmado en el párrafo anterior podría sugerir que la empresa que acometemos en parte está destinada de antemano al fracaso: solo sería posible contribuir a mejorar la forma de reportar los resultados con aquellos medios de prensa que realmente están interesados en desempeñar un papel docente en la formación de la conciencia ciudadana y a hacerlo desde una perspectiva seria, amplia, abierta y tolerante. Sin embargo, creemos que en la medida en que al menos una parte de los medios desarrollen capacidades técnicas mayores para cumplir un rol de información seria de los debates educativos, lograrán imponer un estándar a ser emulado por los otros y a ser contrastado por los lectores con lo que producen los diarios meramente sensacionalista. Es desde esta convicción que hemos producido este documento: para apoyar y reforzar la labor positiva que ya vienen realizando algunos medios e incentivar la mejoría de los otros, así como para sugerir estrategias y acciones a las agencias de evaluación para que incrementen el impacto de sus esfuerzos de difusión.

El trabajo fue realizado a partir del análisis de varios cientos de notas de prensa sobre resultados de las evaluaciones, la mayor parte de ellas de países de la región. Los artículos han sido tomados de dos fuentes. Por un lado, la sección “Recortes” de la página web del GTEE,

que en el momento en que se realizó en relevamiento incluía alrededor de 150 artículos correspondientes a países de la región. Por otro, la sección “PISA en la prensa” de la página web de OCDE/PISA, que recoge artículos relacionados con los resultados de PISA en los diferentes países participantes (se trabajó con los artículos correspondientes a Brasil, México, España y Australia).

A partir del análisis de este material se produjo un primer documento que fue objeto de estudio y debate en un taller de tres días realizado en la ciudad de Santo Domingo en marzo de 2003. Participaron en dicho taller, además del autor de este documento:

- Tacyana Arce, periodista del diario *Estado de Minas* de Minas Gerais, Brasil;
- Claudio Sánchez, periodista, corresponsal en Educación de *National Public Radio* de EE.UU.;
- Gilbert Valverde, especialista en evaluación educativa, Universidad del estado de Nueva York en Albany;
- Martha Grijalva, ex responsable del sistema nacional de evaluación de Ecuador -APRENDO;
- Renán Rápalo, actual responsable de la Unidad de Medición de la Calidad Educativa de Honduras;
- Julio Valeirón, actual responsable del Sistema de Pruebas Nacionales de República Dominicana;
- Sarah González de Lora, Decana de Educación de la Pontificia Universidad Católica de República Dominicana;
- Antonio Sancho, oficial de proyectos de PREAL;
- Patricia Arregui, coordinadora del Grupo de Trabajo de Estándares y Evaluación del PREAL.

Buena parte del contenido de este documento es fruto de las discusiones desarrolladas durante este taller y recoge las reflexiones y aportes de los participantes. Deseo además agradecer los comentarios y sugerencias realizadas al borrador final por parte de Renán Rápalo, Martha Grijalva y Tacyana Arce, y muy especialmente a Patricia Arregui por su meticulosa revisión de la última versión y sus agudas observaciones y reflexiones.

El documento está estructurado en cinco grandes capítulos. En el primero se plantea la difícil relación entre medios de prensa y resultados de las evaluaciones, a partir del análisis de la forma en que los medios titulan las noticias.

El segundo capítulo está destinado a analizar e ilustrar los errores en que incurren con frecuencia los medios al informar sobre los resultados de las evaluaciones. Estos errores en muchos casos obedecen al desconocimiento del tema por parte de los periodistas, pero también a tratamientos inapropiados de la información por parte de los propios ministerios de Educación.

El tercer capítulo fue escrito durante el seminario mencionado, por Tacyana Arce. Este capítulo ofrece una mirada sobre el tema desde la perspectiva de los periodistas.

El cuarto capítulo propone algunos modelos de muy buenos informes sobre resultados educativos –encontrados en la revisión realizada– y sistematiza las características de los mismos. A partir de allí se formulan algunas sugerencias para los periodistas en el tratamiento de la información sobre resultados de evaluaciones educativas.

Finalmente, en el quinto capítulo se analizan los problemas de los propios ministerios de Educación para la difusión apropiada de los resultados y se plantea un conjunto de reflexiones acerca de qué hacer para mejorarla.

El documento se propone promover la reflexión e invitar al diálogo. En modo alguno pretende atacar a la prensa, sino simplemente mostrar la complejidad del reporte de resultados de evaluaciones y de la relación con los ministerios de Educación. Pretende también ayudar a la reflexión sobre la debilidad de las propias unidades de evaluación, tanto en su estrategia general de difusión de resultados como en el relacionamiento específico con la prensa. El objetivo final es proponer pistas de trabajo que ayuden a una mejor comprensión de los problemas del sector educativo por parte de la población y a promover un debate ciudadano más informado en torno a los temas educativos.

CAPÍTULO I

LOS TITULARES DE PRENSA: CATÁSTROFE EN LA EDUCACIÓN

Hemos elegido como puerta de entrada natural al tema que nos ocupa a los titulares empleados para presentar las noticias sobre los resultados de las evaluaciones educativas. El titular normalmente es lo que más impacta de una noticia y en general condiciona fuertemente la lectura del contenido de la misma. Muchas veces es lo único que se lee o lo único

que muchos lectores registran en su mente. Los titulares de prensa son, además, motivo de preocupación y potencial fuente de problemas para las autoridades educativas.

Además, como veremos a lo largo de este capítulo, los titulares reflejan una actitud hacia las instituciones educativas y un modo de encarar los debates educativos que le hacen daño al sistema educativo.

A partir de la revisión de los titulares de prensa sobre evaluaciones educativas en la región de América Latina (tanto los incluidos en la sección “Recortes de prensa” de la página web del GTEE/PREAL como en las colecciones de información de prensa sobre los resultados de PISA para México, Brasil y España incluidos en la página web de la OCDE), fue posible constatar un muy fuerte predominio del tono “catástrofe” o “desastre” en el tratamiento de la información.

La lista de titulares consignados en el Recuadro 1 es una selección que pretende ser ilustrativa de este fenómeno. La característica común a todos estos titulares es el uso de adjetivaciones fuertes –del tipo *“educación mediocre”, “país de burros”, “país reprobado”, “los peores del mundo”, “colapso de la educación”, “gran desastre”*– que revelan una intención de crear una sensación de cataclismo que va más allá de los serios problemas que, sin duda, los datos ponen de manifiesto.

Por oposición a lo anterior, otros modos –menos frecuentes– de titular las noticias evitan recurrir a la adjetivación y se focalizan en lo sustantivo de la información (ver Recuadro 2), es decir, adoptan un tono que evita adjetivar y se limita a presentar la información –que en general es suficientemente contundente por sí misma– o a plantear una reflexión u opinión a partir de los datos. Ello no significa que los artículos no pongan de manifiesto importantes carencias o problemas, pero evitan el empleo de adjetivaciones exageradas y llamativas. En otros casos, el tono del titular invita a la reflexión: *“Lo que enseña APRENDO”, “Un puzzle que encaja piezas de 32 países”, “A lição de PISA”*.

RECUADRO 1 TITULARES “CATÁSTROFE”

Titular	País	Medio	Fecha
En cinco años, ninguna mejora	Argentina	<i>La Nación</i>	18/3/99
Chubut tiene el promedio educativo más bajo del país	Argentina	<i>Jornada</i>	8/9/01
Alunos brasileiros ficam em último lugar em ranking de educação	Brasil	<i>Folha Online</i>	4/12/01
O último da fila	Brasil	<i>Diário Catarinense</i>	6/12/01
Brasileiro é último em leitura	Brasil	<i>Diário da Manhã</i>	5/12/01
Troféu inglório	Brasil	<i>Diário de São Paulo</i>	1/12/01
Aluno brasileiro é o pior em leitura	Brasil	<i>Jornal do Brasil</i>	5/12/01
País Reprovado	Brasil	<i>Jornal do Brasil</i>	9/12/01
Evaluación de calidad aplicada a alumnos en cinco departamentos los mostró graves en matemáticas y lenguaje	Colombia	<i>El Tiempo</i>	18/9/02
Informe de Fedesarrollo ‘raja’ a la educación privada en Cali	Colombia	<i>El Tiempo</i>	14/10/02
Simce muestra un gran desastre en la educación	Chile	<i>La Tercera</i>	1/1/01
Ecuador pierde el año	Ecuador	<i>El Comercio</i>	14/1/98
La educación está en su peor crisis	Ecuador	<i>El Diario</i>	14/9/99
El país reprueba matemáticas	Ecuador	<i>Hoy</i>	16/7/00
Una libreta que asusta	Ecuador	<i>El Universo</i>	22/8/99
Los alumnos españoles de secundaria, entre los peores de los países desarrollados	España	<i>El País</i>	5/12/01
Los estudiantes españoles suspenden y se sitúan entre los peores del mundo	España	<i>EFE - Terra</i>	11/1/02
Educación es mediocre	Guatemala	<i>Prensa Libre</i>	8/8/02
La educación no encuentra rumbo	Guatemala	<i>Prensa Libre</i>	28/4/02
Reprueba México calidad educativa	México	<i>Reforma</i>	15/10/01
Según la OCDE. Reprobados en educación	México	<i>Diario de México</i>	5/12/01
Reprueba la OCDE la calidad de la educación impartida en México	México	<i>El Financiero</i>	5/12/01
Contexto: orejas de burro	México	<i>El Financiero</i>	5/12/01
Nos truena la OCDE. Prometen reducir número de burros	México	<i>Metro</i>	5/12/01
¡Le ganamos a Brasil!	México	<i>Metro</i>	5/12/01
Un país de reprobados	México	<i>Reforma</i>	8/12/01
Perú penúltimo entre 15 países de Latinoamérica en nivel educativo	Perú	<i>La República</i>	25/3/00
Escolares peruanos en último lugar de Latinoamérica en matemáticas	Perú	<i>El Comercio</i>	28/2/01
El colapso de la educación	Perú	<i>El Comercio</i>	26/3/01
Somos los peores en matemáticas en el continente	Perú	<i>Tiempos del Mundo</i>	3/5/01

Fuentes: *Recortes de prensa* (<http://grade.org.pe/gte-preal/evalua.htm>) y *PISA en la prensa* (<http://ocde.org/pisa/press.htm>)

RECUADRO 2
TITULARES “INFORMATIVOS” o “REFLEXIVOS”

Titular	País	Medio	Fecha
Persisten las dificultades al finalizar el secundario	Argentina	<i>La Nación</i>	29/5/98
Familia influencia desempenho	Brasil	<i>Jornal do Comercio</i>	3/12/00
Estudantes sabem muito mais do que demonstram na sala de aula	Brasil	<i>Estado de São Paulo</i>	14/12/00
Aluno brasileiro é mal avaliado em pesquisa	Brasil	<i>Folha de São Paulo</i>	5/12/01
Distorção de série e idade prejudica aprendizado	Brasil	<i>O Estado de São Paulo</i>	5/12/01
A lição de PISA	Brasil	<i>O Globo</i>	12/12/01
Pruebas revelan fallas educativas	Costa Rica	<i>La Nación</i>	7/3/02
Los pro y los contra del SIMCE y la reforma educativa	Chile	<i>Sociedad Civil</i>	17/12/01
Lo que enseña “aprendo”	Ecuador	<i>Expreso</i>	13/6/98
El rendimiento estudiantil en el país es insuficiente: Aprendo 97	Ecuador	<i>El comercio</i>	16/9/99
Un puzzle que encaja piezas de 32 países	España	<i>El País</i>	5/12/01
Un 20% de los alumnos españoles no alcanza el nivel mínimo en matemáticas	España	<i>El País</i>	6/12/01
El 60% de alumnos de 3º tienen un nivel satisfactorio	Uruguay	<i>El País Digital</i>	12/12/99

Fuentes: Recortes de prensa (<http://grade.org.pe/gtee-preal/evalua.htm>) y PISA en la prensa (<http://ocde.org/pisa/press.htm>)

El conjunto de titulares incluidos en los Recuadros 1 y 2 responden a una variedad de tipos de artículos: notas informativas, columnas de opinión, entrevistas, etc. Sin embargo, los artículos de información elaborados por un periodista a partir de una presentación pública de los resultados en rueda de prensa y/o a partir de un informe escrito distribuido por el Ministerio de Educación, son los que predominan. Apenas una pequeña parte de los artículos recogidos en las colecciones indicadas son entrevistas a especialistas o de columnas de opinión.

Es por ello significativo que, a pesar de que la gran mayoría de los artículos son notas de carácter informativo, el titular por lo general está fuertemente adjetivado. Esto ocurre tanto con los resultados de las evaluaciones internacionales como con los resultados de las evaluaciones nacionales. En el Recuadro 1 los titulares de periódicos de Perú, México, Brasil y España corresponden a resultados de evaluaciones internacionales, en tanto que los de Argentina, Colombia, Chile, Ecuador y Guatemala corresponden a resultados de evaluaciones nacionales.

¿Por qué los medios de prensa deciden titular las noticias de esta manera? ¿Qué razones, lógicas y motivaciones hay detrás de esta manera de encarar las noticias sobre las evaluaciones educativas?

En principio, cuatro tipos de factores se entrecruzan para generar este fenómeno:

- i. una dosis importante de “sensacionalismo” al servicio de la venta del medio de prensa;
- ii. el desconocimiento del tema de la evaluación educativa por parte de los periodistas;
- iii. el inadecuado tratamiento de la información sobre resultados educativos por parte de los propios Ministerios de Educación;
- iv. la dimensión política que atraviesa la actividad educativa –para bien y para mal– lo que incluye las posturas políticas y/o visiones sobre la educación del periodista, del medio de prensa y de otros actores que actúan como columnistas o tienen espacios en el medio, así como también el uso político de los resultados por parte de las propias autoridades educativas.

1. El “sensacionalismo educativo”

El primer factor que sin lugar a dudas incide decisivamente en el enfoque de los titulares es el hecho de que los medios de prensa son ante todo empresas que venden un producto.

Un medio de prensa puede, además, tener otras motivaciones o encuadres para encarar su labor. Por ejemplo, puede tener opciones de carácter político –o incluso partidarias–, puede asumir una misión de servicio a la formación de la conciencia ciudadana, puede desarrollar su labor en el marco de ciertos principios de ética periodística o puede considerar que su misión es favorecer el control ciudadano sobre el poder político a través de la información. Pero, en primer término, necesita vender su producto, para lo cual debe atraer y retener a algún sector de la opinión pública. De lo contrario, desaparece como medio periodístico.

Algunos medios logran conformar su público a través de estilos reflexivos y analíticos. Pero otros, los dirigidos al “gran público”, parecen lograr su objetivo a través de la sobresimplificación de la realidad y de la incorporación a la información de una cierta dosis de morbosidad –lo cual, por cierto, no es responsabilidad exclusiva del medio, sino también del público–. En este contexto, los titulares son un medio de atraer al público. Dar a los mismos un tono sensacionalista es una manera de construir una noticia que “venda”.

La educación es un sector de la realidad social en el que no es fácil construir noticias que atraigan al público. En la región de América Latina las noticias educativas suelen limitarse a la información brindada por las autoridades respecto a sus planes de acción –construir nuevas escuelas, comprar computadoras, etc.–, a los conflictos sindicales por mejoras salariales, a los reclamos ocasionales de padres por el mal estado de un edificio escolar. Por el momento la prensa en general parece no haber descubierto cómo hacer una noticia acerca de lo que se enseña día a día, de los esfuerzos de escuelas o profesores por mejorar su trabajo y motivar a los niños y jóvenes con el aprendizaje, de los esfuerzos de las familias por educar a sus hijos y del compromiso de maestros y directores que sostienen proyectos educativos valiosos en condiciones muy desfavorables³.

Los resultados de las evaluaciones constituyen un modo relativamente fácil de hacer “noticia educativa”: hay datos que pueden mostrarse y los resultados normalmente son malos, lo cual se presta para el titular escandaloso o catastrófico del tipo de los incluidos en el Recuadro 1, que sin duda pueden atraer la atención del público. En este sentido, los resultados de las evaluaciones se prestan especialmente para crear algo así como el equivalente a la llamada “crónica roja”, que haga al medio más vendible.

Finalmente, cuando los resultados no son tan malos o son buenos, la prensa suele desentenderse de ellos: ya no son noticia. Un ejemplo ilustrativo de ellos es lo ocurrido recientemente en los Estados Unidos en relación a la cobertura de los resultados de un Estudio Internacional sobre Lectura (denominado PIRLS en inglés)⁴.

“Alguna vez se han preguntado por qué el público piensa que las escuelas norteamericanas no son tan buenas? Aquí va un indicio: Ayer se publicaron los resultados de un estudio internacional sobre lectura que muestra que los chicos americanos están en 9º lugar entre 35 naciones y muy cerca en puntaje a los que se desempeñaron mejor.

El Washington Post le dio a la noticia 1/12 de la última página de su sección A. Avisos comerciales de la tienda Hecht’s ocuparon los

³ Lo mismo suele ocurrir con el trabajo de muchos trabajadores y ciudadanos anónimos que día a día sostienen con enorme esfuerzo y generosidad diversos servicios de carácter social en áreas como la salud, la alimentación, la atención a sectores pobres, etc.

⁴ Lo que se transcribe a continuación fue publicado por Gerald Bracey en su página web (www.america-tomorrow.com/bracey/EDDRA/) y ha sido traducido por el GTEE con autorización del autor.

otros 11/12. Las pocas palabras que el Post dedica a la historia son un cable de AP, es decir, algo que no vale lo suficiente como para ocupar el tiempo del reportero nacional de educación del Post, Mike Fletcher, o, en su ausencia, Jay Mathews o Valerie Strauss.

El New York Times le dio cobertura cero. Ni una palabra. Al menos, no en la edición que se imprime en Washington para distribuir a domicilio.

Uno no puede sino imaginarse en qué página de ambos periódicos hubiera aparecido la nota si los estudiantes Americanos hubieran estado bajo el promedio.

El estudio fue difundido aquí vía el Boston College, y el Boston Globe sí colocó un artículo, en la sección B. El enfoque que usó la reportera Megan Tench revela que los reporteros nunca, nunca, nunca escriben una historia simple positiva sobre la educación:

‘Cuarenta años después de que por primera vez se estableció que los estudiantes norteamericanos son peores en matemáticas que sus pares en otros lugares del mundo (gracias, Megan, por este sesudo comentario) un estudio de Boston College muestra que están mejor –pero aun así, no destacablemente mejor– en lectura’...

‘En la primera comparación internacional de importancia sobre capacidades en lenguaje (se equivoca. En 1992 el Estudio “How in the World Do Students Read” analizó las destrezas en lectura de niños de 9 años en 27 países y de niños de 14 años en 31 naciones) los estudiantes estadounidenses quedaron novenos entre 35 países, detrás de Letonia, Bulgaria y Suecia (¡qué tragedia!), aunque por encima del promedio internacional’. (Pero Megan, el puntaje norteamericano está 42 puntos por encima del promedio internacional solo 12 puntos por debajo de Holanda que quedó en segundo lugar, y 18 puntos debajo de Suecia, que quedó primero. ¿No merecería eso algo más que “aunque encima del promedio internacional”?)

No sé si los medios tuvieron el estudio “bajo embargo “por un tiempo hasta ayer a las 11 de la mañana, pero aún si no fue así, sí tuvieron tiempo suficiente como para preparar un artículo extenso para la siguiente mañana.

Jerry.

2. El desconocimiento del tema por parte de los periodistas

A la necesidad del medio de atraer lectores se agrega que la educación tiene, en los hechos, importancia secundaria en la agenda pública y en las prioridades de los medios de prensa en relación a otros sectores de actividad tales como la política o la economía. Por tanto, no suele haber periodistas especializados en educación, como suele haberlos para otros sectores⁵.

Pero, simultáneamente, el tratamiento de la información sobre resultados de evaluaciones educativas es bastante complejo y requiere de cierto conocimiento técnico y de ciertos cuidados en relación a las interpretaciones que pueden derivarse válidamente de los datos y a las limitaciones de los mismos.

La falta de especialización y acumulación de conocimiento y “cultura” en un campo relativamente nuevo en la región –el de la evaluación de sistemas educativos–, lleva a que en los artículos periodísticos revisados se encuentre una gran cantidad de errores, falacias o interpretaciones inapropiadas de los datos. Al análisis detallado de los mismos está dedicado el capítulo II.

3. Los defectos en el uso de los resultados y la comunicación de los mismos por parte de los ministerios de Educación

Junto con los factores atribuibles a los medios de prensa, es preciso reconocer que muchas veces el tratamiento inadecuado de los resultados de las evaluaciones es propiciado por las maneras en que los propios ministerios presentan la información a la prensa.

Según se podrá apreciar en el próximo capítulo, muchos de los errores y falacias en que incurren las notas de prensa sobre resultados tienen su origen en defectos en el modo en que las propias autoridades manejan la información. Por ejemplo, entregando rankings de establecimientos construidos en forma inapropiada, interpretando como “eficacia” educativa lo que probablemente no es más que el efecto de la selección social del alumnado o traduciendo a calificaciones los resultados de pruebas de carácter diagnóstico que, por la forma en que fueron construidas, no

5 El análisis en torno a este tema será profundizado en el capítulo III desde la visión de un periodista.

permiten emitir juicios de valor en términos de si un alumno aprobaría o reprobaría el grado.

Al mismo tiempo, los ministerios de Educación –en particular las unidades de evaluación– muchas veces carecen de una estrategia explícita de relacionamiento con los medios de prensa que defina claramente los mensajes que se desea comunicar a partir de los datos y que permita ir construyendo una cultura de la evaluación entre los propios periodistas. De esta manera la interpretación de la información y la construcción de los mensajes queda absolutamente librada a la buena voluntad e inventiva de personas que, según se indicó antes, normalmente no poseen conocimientos mínimos sobre evaluación.

4. Los usos políticos de los resultados

Un cuarto tipo de factores que contribuye al tono catástrofe que suele tener la difusión de resultados de las evaluaciones de aprendizajes está vinculado con la dimensión política que atraviesa a la educación al igual que a todas las actividades sociales. Este aspecto involucra el uso político de los resultados tanto por parte de los ministerios de Educación como por parte de quienes se oponen a la administración a cargo o desean impulsar otros proyectos o visiones para el sector educativo. El punto es complejo y amerita un tratamiento detenido.

4.1. Los usos políticos de los resultados por parte de las autoridades

Las autoridades pueden utilizar políticamente los resultados de diversas maneras. Una primera, obvia y legítima, es convalidar su gestión cuando los resultados son buenos o mejoran –lo que no suele ser frecuente, dados los problemas históricos de la educación en nuestros países, así como el hecho de que los cambios en los resultados educativos suelen tomar bastante más tiempo que el de una administración–. Sin embargo, aun en este caso, el modo en que esto se haga puede ser contraproducente: la gama de posibilidades va desde entregar información objetiva sobre avances en ciertos indicadores o resultados hasta, en el otro extremo, hacer campaña político electoral a partir de los resultados (a veces en forma subrepticia o disimulada, a veces en forma abierta y explícita). Los usos cercanos a este último extremo, inevitablemente tendrán como consecuencia intentos por descalificar los buenos resultados desde las posturas políticas opuestas y una razonable actitud de sospecha de parte de los medios de prensa.

Cuando se trata de lidiar con malos resultados, es interesante notar que los ministerios los manejan de maneras diferentes e incluso contradictorias: en algunos casos intentan ocultarlos, en otros simplemente explicarlos y en otros los destacan enfáticamente para impulsar reformas o para atacar al sector docente. La primera –cuando los resultados finalmente son filtrados a la prensa o dados a conocer luego de un período de ocultamiento– y la última de estas estrategias, contribuyen fuertemente a generar los titulares de tipo catastrófico.

Por ejemplo, en algunos países de la región, típicamente Perú y México, el tono catástrofe de los titulares de prensa se explica en buena medida por el hecho de que durante muchos años los ministerios de Educación en estos países ocultaron los resultados, tanto de las evaluaciones nacionales, como de los correspondientes a su participación en evaluaciones internacionales –la evaluación latinoamericana de educación primaria realizada por la UNESCO en el caso de Perú, la prueba internacional TIMSS realizada por la Asociación Internacional de Evaluación (IEA) en el caso de México–. Ello dio lugar a que, cuando los resultados se conocieron, el ambiente estuviera predisposto a un tratamiento especialmente escandaloso: por algo los resultados se habían mantenido en secreto durante tanto tiempo. Los titulares fueron además un modo de castigar a las autoridades responsables del ocultamiento de la información.

Por otra parte, en el extremo opuesto, muchas veces los propios ministerios construyen el argumento de la catástrofe educativa a partir de los datos de las evaluaciones, como forma de legitimar sus propios proyectos de cambio –normalmente esto dependerá de si se está al inicio o al final del período de gobierno, pero también de la coyuntura educativa–. Un ejemplo interesante, si bien extrarregional, es el caso de España, en que la visión catastrófica a partir de los datos de PISA es propiciada por funcionarios del gobierno de Partido Popular, en el marco de la discusión de una nueva reforma educativa que dio marcha atrás a la reforma educativa anterior del PSOE:

“Los datos que arroja este informe confirman, según la Secretaria General de Educación, XX, la “necesidad de un cambio de rumbo” y subrayó que en España –donde el estudio se ha realizado entre 2.600 alumnos de todas las comunidades– los resultados en matemáticas están “muy bajos”, seguidos de comprensión escrita y ciencias, también por debajo de la media de la OCDE. “Esto confirma las evidencias y avala la necesidad de cambios en el contenido de lo que estudian los alumnos y el diseño del sistema

educativo, porque estos resultados preocupantes no se corresponden con los recursos y la dedicación de los docentes”, manifestó XX”. (Los adolescentes españoles destacan por su bajo nivel en comprensión escrita, ABC, España, 05/12/01).

“En España, donde el estudio se ha realizado entre 2.600 alumnos de todas las comunidades autónomas, los resultados en matemáticas están “muy bajos” seguidos de comprensión escrita y ciencias, también por debajo de la media de la OCDE. La OCDE estableció cinco niveles de rendimiento, y España se sitúa entre el 2 y el 3, al tiempo que mientras la media de los países desarrollados en el nivel cinco (de excelencia) es del 9 por ciento, en España “solo llegamos al cuatro por ciento”. Por ello la Secretaria General de Educación, XX ha advertido de que los resultados en España del informe de la OCDE sobre los conocimientos escolares “confirma las evidencias y avala la necesidad de un cambio de rumbo” (Los estudiantes españoles suspenden y se sitúan entre los peores del mundo, EFE-Terra, España, 11/1/02).

Como veremos un poco más adelante, cuando se toman en consideración variables socioculturales, los resultados de España no parecen tan malos como lo sugieren las declaraciones de la funcionaria. Por otra parte, resulta algo absurda la categorización de “peor del mundo” cuando la evaluación solo involucró a algo más de 30 países, en un mundo que tiene algunos más...

4.2. Los usos políticos de los resultados por parte de otros actores

En muchos casos los titulares de tono “catástrofe” no responden simplemente a la búsqueda de mayor venta del diario sino que obedecen a una voluntad deliberada de confrontar la gestión gubernamental en educación o de impulsar proyectos alternativos.

El análisis de este aspecto es bastante complejo porque implica conocer en profundidad la situación interna en cada país y las líneas editoriales de los periódicos. Sin embargo, asumiendo el riesgo de equivocarnos, pondremos algunos ejemplos.

El primero de ellos es el caso del diario *La Tercera* en Chile. En el recuadro 3 se incluye un par de artículos de este periódico. El primero es de carácter informativo y es del año 2001, en tanto el segundo forma parte de la página editorial del diario y está fechado en marzo de 2003.

RECUADRO 3 USOS POLÍTICOS DE LOS RESULTADOS (1)

Cifras revelan mínimo avance tras la reforma de 1996

Simce muestra un gran desastre en la educación

La reformulación de currículo y el aumento de recursos destinados a mejorar la calidad de la enseñanza no han dado resultados. Las pruebas a los cuartos básicos arrojaron que un 19% presenta un rendimiento deficiente en lenguaje, un 32% en matemática y un 27% en ciencias. “No podemos estar conformes”, admitió la ministra de Educación, Mariana Aylwin.

(Publicado el 6 de julio).- Un preocupante diagnóstico que reafirma la desastrosa realidad de la educación chilena arrojaron los resultados de las pruebas del Sistema de Medición de la Calidad de la Educación (Simce, aplicadas a los cuartos básicos el año pasado, constituyéndose en el primer traspie de la reforma educacional implementada a partir de 1997.

Pese a que estos escolares están trabajando desde esa fecha con un currículo y metodologías pedagógicas reformuladas, los resultados promedio de los establecimientos municipalizados, subvencionados y particulares pagados son similares a los de 1996, año de la anterior medición a los cuartos básicos. Además, al medir el desempeño en la última prueba, se constató que el 19% de los 285.094 alumnos analizados tienen un rendimiento deficiente en lenguaje y comunicación; el 32% en educación matemática y el 27% en comprensión del medio natural, social y cultural”.

(La Tercera, Chile, 1/1/01) (los subrayados son nuestros).

Nuevo año escolar

“La mayoría de los establecimientos educacionales del país comenzó ayer un nuevo año escolar. Con este comienzan también las tradicionales discusiones sobre el estado, falencias y desafíos pendientes del sistema educacional. El debate ha sido largo y pese a que en los últimos años se han observado ciertas mejorías, la comentada reforma que impulsó el Ministerio de Educación (Mineduc) hace casi ocho años no ha traído consigo los resultados esperados, según estudios nacionales e internacionales.

Si bien en la última década los recursos destinados a educación se han triplicado y alcanzan el 7.2% del PIB nacional... este aumento de fondos no ha ido aparejado de una mejora sustancial de la calidad de la enseñanza, lo que resulta en extremo grave... La centralización que existe por parte del Mineduc⁶ en la distribución de recursos, metas y planes de estudio aparece, entonces, como contraproducente... la eficiencia en la gestión es otro tema que despierta reparos. Para favorecer la competencia y transparencia en el ámbito educacional, el Mineduc anunció que los directores de escuelas comenzarán a ser elegidos a través del concurso público... Pero la competencia debe ir más lejos, dejando atrás antiguas trabas –como las que impone el estatuto docente–...”.

(La Tercera, Chile, 6/3/03) Los subrayados son nuestros).

6 El salto conceptual que hace el editorialista al tema de la centralización/descentralización es, al menos, un tanto simplista. Parece ignorar que una buena parte de los países con excelentes resultados en las evaluaciones internacionales tienen sistemas educativos fuertemente centralizados, en tanto otros con sistemas muy descentralizados, como los Estados Unidos, obtienen resultados bastante magros en la comparación internacional.

Del análisis de los artículos transcritos se desprende, en primer lugar, que los datos ofrecidos, aunque muestran problemas, no parecen sustentar suficientemente el calificativo de “gran desastre”. Se incluyen algunos porcentajes de alumnos que se encontrarían en niveles deficientes en relación a los aprendizajes esperados, que se ubican en el entorno del 20% al 30% lo cual, si bien es materia opinable, en principio no aparece como desastroso. Al mismo tiempo se informa que los avances registrados luego de 5 años son “mínimos” (¿es posible esperar avances espectaculares y rápidos en la educación?, ¿el que haya avances mínimos no podría ser considerado una buena noticia?).

La decisión de encarar la noticia en tono de desastre parece explicarse a partir del editorial transcrito en segundo término. De la lectura del mismo se desprende que la línea del diario sería cuestionar el aumento del gasto en educación y su propuesta pasaría por incrementar la autonomía y la competencia entre los centros educativos, incluyendo la remoción de las regulaciones sobre contratación de docentes que incluye el Estatuto Docente en vigencia⁷.

En otros casos, los titulares obedecen a columnas de opinión de personalidades individuales a las que el medio de prensa abre su espacio, pero que no implican que el medio comparta sus puntos de vista. Un ejemplo de ello es el artículo titulado **“El colapso de la educación”** (Recuadro 4), escrito en Perú en relación con los resultados de la evaluación de la UNESCO cuya publicación no fue autorizada durante el gobierno de Fujimori. Como se puede apreciar claramente, la intención del autor de la nota es confrontar la gestión de Fujimori y las afirmaciones realizadas por su gobierno acerca del estado de la educación en el Perú.

Pero por otra parte el autor, connotado director de un colegio privado limeño, utiliza también la información para promover una visión acerca de la mejor calidad de la educación privada. Para ello alude a los mejores resultados de los colegios privados en relación a los públicos y de los urbanos en relación a los rurales. Sin embargo, omite considerar la cuestión de la composición social del alumnado de los establecimientos educativos, que según todas las investigaciones, incluida la de la propia UNESCO, juega un papel central en la determinación de los logros por

⁷ Por cierto nuestra intención no es cuestionar que el periódico impulse ciertas opciones de política educativa, estas u otras, simplemente buscamos la lógica que puede existir por detrás de los titulares “catástrofe”.

establecimiento, lo que obliga a realizar análisis más complejos antes de formular afirmaciones acerca de la calidad relativa de unos y otros tipos de establecimientos. Por otra parte, el argumento de que si se excluyera a los colegios privados de la muestra los resultados serían peores se aplica a todos los países no solo a Perú. No puede afirmarse que la situación de Perú en la “tabla” latinoamericana empeoraría, también podría mejorar, dependiendo de cómo hayan sido los resultados de los colegios privados peruanos en relación a los colegios privados del resto de los países.

RECUADRO 4 USOS POLÍTICOS DE LOS RESULTADOS (2)

El colapso de la educación

“Finalmente, luego de muchas presiones de los interesados en conocer la realidad de la educación peruana comparada con la latinoamericana, la publicación de los resultados de mediciones internacionales evidencia inequívocamente que nuestra educación es la peor del continente, exactamente lo opuesto a la promesa fujimorista que sostenía que la estaba convirtiendo en “la mejor de América Latina”...

Los resultados muestran que los mejores estudiantes en todos los ítems y niveles son los de Cuba. Más atrás se ubican los estudiantes de Argentina, Brasil y Chile. Un poco más atrás Colombia, México y Paraguay. Los alumnos del tercer grado del Perú quedaron ubicados en último lugar en Matemáticas y antepenúltimo lugar en Lenguaje. Este estudio también ratificó que, salvo Cuba, donde no hay educación privada, **en el conjunto de países los colegios privados muestran mejores resultados que los públicos**; asimismo, que **la educación estatal urbana supera a la urbano-marginal y esta a la rural**, que sigue postrada en las peores condiciones imaginables con la única excepción de Colombia, donde las escuelas rurales tuvieron mejores resultados que las urbano-marginales, gracias a su exitosa modalidad educativa conocida como Escuela Nueva. **Si la muestra peruana hubiera excluido las escuelas privadas, los resultados de la escuela pública hubieran sido peores aun.**

Los peruanos hemos sido triplemente estafados. Primero, porque se nos han escondido los resultados. Segundo, porque se nos ha querido hacer creer que estamos próximos a ser los líderes de América Latina. Tercero, porque el gobierno de Alberto Fujimori gastó cientos de millones de dólares en reformar un sistema educativo con pésimos resultados. Llama la atención que, habiéndose montado en el Ministerio de Educación una Unidad de Medición de la Calidad de la Educación, no se haya permitido que la opinión pública peruana conozca los resultados. ¿Cómo vamos a hacer que nuestra educación sea competitiva sin partir de un conocimiento de su situación real?

El sector Educación requiere urgentemente una “comisión de la verdad” para conocer a fondo sus problemas y posibilidades. Esta es una tarea pendiente que está al alcance de la actual gestión gubernamental...

(El Comercio, Perú, 26/3/01) (los subrayados son nuestros).

Debo afirmar enfáticamente que en ambos ejemplos, tanto el de Chile como el de Perú, no está en cuestión la legitimidad o conveniencia de las posturas o visiones educativas que se busca impulsar ni la utilización para ello de los resultados de las evaluaciones de logro educativo. Ello forma parte de los debates propios de un sistema democrático y es deseable que la información resultante de las evaluaciones sea utilizada para alimentar estos debates.

Sin embargo, es preciso señalar al menos dos requisitos que deberían ser respetados al hacer uso de los resultados de evaluaciones de aprendizaje para impulsar visiones o propuestas educativas. En primer término, las interpretaciones de los resultados no pueden ser antojadizas y arbitrarias sino que deben ajustarse con algún grado de rigurosidad a lo que la información permite afirmar con propiedad. A la profundización de este aspecto está destinado el próximo capítulo.

En segundo término, es importante evitar que el tratamiento de la información alimente una espiral de desconfianza en relación a las instituciones educativas. Este aspecto merece un tratamiento especial en el apartado que sigue.

La confianza en el sistema educativo como valor a construir y preservar

El presente apartado pretende aportar una reflexión acerca del daño que a la educación hacen los enfoques de tipo “catástrofe” analizados a lo largo de este capítulo, tanto los que se derivan del “sensacionalismo” periodístico como los que son utilizados como recurso retórico en el debate público de los temas educativos.

Sin duda los datos normalmente dan sustento a una legítima preocupación por el estado de la educación. Se puede pensar que una función importante de la prensa es generar cierto grado de preocupación e interés de la opinión pública por la situación del sistema educativo. Sin embargo, cabe preguntarse en qué medida el recurso a las adjetivaciones de tono catástrofe ayuda a movilizar a la opinión pública y a comprender los problemas y desafíos del sector educativo. Más bien parece contribuir a generar una imagen crecientemente negativa sobre las instituciones, a desvalorar la educación pública e, inclusive en algunos casos, a la propia nación.

No está en cuestión el derecho de la prensa a informar del modo que le parezca más adecuado, ni el derecho de cualquier ciudadano a formular los juicios de valor que considere apropiados acerca de la situación

del sistema educativo. Tampoco está en cuestión la necesidad de generar cierto grado de preocupación en la opinión pública por la situación de deterioro de los sistemas educativos en nuestra región y por las deficiencias en los logros educativos de los estudiantes. Sin embargo, todo lo anterior no debiera llevar a la creación de una desconfianza radical sobre las instituciones educativas.

En el sector educativo la confianza es tan importante como en el sector económico y financiero. En el sector económico, cuando los actores pierden la confianza, la economía se desmorona. En el sector educativo ocurre algo similar. No es posible educar en un contexto en el que las autoridades desconfían radicalmente de los docentes; los docentes desconfían de las autoridades educativas, sean del partido que sean; las familias desconfían de las escuelas a las que envían a sus hijos y de los docentes que están a su cargo; la opinión pública desconfía de las instituciones educativas en general.

Los enfoques de tipo sensacionalista o de tono “catástrofe” contribuyen a destruir la confianza mínima necesaria para poder construir instituciones educativas fuertes.

Por cierto confianza no significa ingenuidad. Hay políticas educativas buenas, regulares, malas y pésimas. Hay administraciones buenas, regulares, malas y pésimas. Hay docentes buenos, regulares, malos y pésimos. Hay escuelas buenas, regulares, malas y pésimas.

Lo primero y más importante es asumir como punto de partida y presupuesto fundamental que la educación es una tarea compleja y de largo aliento, que requiere del aporte constructivo de múltiples actores y de esfuerzos sostenidos en el tiempo. No hay responsables únicos de los malos o buenos resultados, ni debiera esperarse que alguna política, sea del tipo que sea, obtenga logros espectaculares en tiempos reducidos. Por estos motivos, la educación no debiera ser parte de las discusiones políticas menores –me refiero a los intercambios de declaraciones entre actores políticos que tienen por finalidad principal diferenciarse del adversario para captar votos– y, aún los grandes debates nacionales sobre la educación, debieran darse en un clima de tolerancia y respeto por la diversidad, intentando aprender con humildad de los datos antes que usarlos para defender posturas preconcebidas o para crear percepciones apocalípticas.

En este último sentido, lo segundo es evitar las descalificaciones generalizadas: ‘todos los maestros son malos e incumplidores’, ‘la escuela pública es mala y siempre peor que la privada’, ‘la reforma educa-

tiva del gobierno ‘x’ es integralmente un desastre y ninguna de sus acciones tiene valor alguno’, ‘el sistema educativo de nuestro país es el peor del continente o del mundo’.

Por el contrario, es necesaria una mirada suficientemente amplia como para comprender que en educación las cosas nunca son blancas o negras, más bien grises y llenas de matices. Los factores que inciden en los resultados son múltiples, por lo cual la actitud más apropiada para analizarlos es la ponderación⁸.

Los enfoques que llevan la interpretación de los resultados de las evaluaciones educativas al plano de lo “catastrófico” como recurso retórico, probablemente le hacen más mal que bien a la educación, independientemente del proyecto que busquen sustentar. En especial porque impiden acumular en el tiempo el efecto de las acciones. Cada administración tiene sus nuevos proyectos que salvarán a la educación del caos, mientras poco cambia en la mayor parte de las escuelas.

Por el contrario, lo que se requiere es un enfoque de las noticias y los debates educativos que contribuya a construir apoyos de la sociedad a las instituciones educativas y a los educadores.

Para ello es preciso aprender a utilizar la información de manera apropiada en la formulación de valoraciones matizadas sobre la situación educativa. Es preciso construir las bases que permitan una mejor comprensión de los problemas y desafíos que es necesario encarar en el sector educativo.

En el capítulo IV se analizan algunos ejemplos de reportes de resultados que se caracterizan justamente por su enfoque ponderado y reflexivo.

CAPÍTULO II

LOS ERRORES MÁS COMUNES EN LOS INFORMES DE LA PRENSA

Independientemente de las cuestiones relativas al sensacionalismo, a los usos políticos de los resultados de las evaluaciones educativas y a los

⁸ Probablemente lo dicho en estos párrafos se aplica también a otras áreas de la vida social y requiere de otros modos de hacer política, más centrados en **construir** socialmente que en **combatir** al adversario. Pero bueno, la educación es suficientemente importante en el mundo contemporáneo como para comenzar a experimentar con ella nuevos modos de hacer política.

aspectos éticos conexos, la revisión de las colecciones de artículos de prensa permitió identificar un conjunto muy importante de errores en el tratamiento de la información.

El recurso al sensacionalismo puede ser comprensible cuando no es exagerado. El uso de la información con fines políticos es en general legítimo, o en todo caso, materia opinable en casos extremos. En cambio, un mínimo de rigor en el tratamiento de la información es indiscutiblemente exigible a todos los medios, independiente del modo en que elijan titular las noticias y de los proyectos políticos que busquen impulsar.

El presente capítulo está destinado a mostrar y explicar los errores más comunes en que incurren las notas de prensa cuando informan sobre los resultados de las evaluaciones de logros educativos. El propósito es arrojar algo de luz sobre la complejidad del tema y la necesidad de crear una cultura y un conocimiento sobre el tema entre los periodistas, de modo que la información que se entrega al público y que se utiliza como base para los debates de política educativa, sea lo más apropiada y rigurosa posible.

En algunos casos los errores constatados se deben al desconocimiento del significado de los datos por parte del periodista –como, por ejemplo, interpretar puntuaciones abstractas como cantidad de aciertos en las pruebas–. En otros, se trata de interpretaciones que van más allá de lo que la información permite afirmar –como, por ejemplo, interpretar como “eficacia” escolar lo que en principio puede obedecer simplemente a la composición social del alumnado de una escuela o provincia–. Estos ejemplos se comprenderán mejor a lo largo del capítulo.

Lamentablemente es preciso reconocer que una parte muy importante de las interpretaciones erróneas de los datos son generadas por los propios ministerios de Educación, dado que ellos entregan la información de manera inapropiada y equívoca.

A los efectos de ordenar la exposición hemos organizado los errores encontrados en tres grandes categorías que no pretenden ser exhaustivas:

- a. errores de interpretación del significado de los datos y métricas que se emplean, derivados de la carencia de conocimientos básicos sobre estadística, evaluación y reporte de resultados;
- b. interpretación equivocada de los resultados como calificaciones y de los alumnos como aprobados o reprobados, error derivado de interpretar como “absolutos” (o “criteriales”) resultados que son “relativos”;

- c. construcción inapropiada de rankings e interpretación de los mismos en términos de calidad de las escuelas o provincias sin consideración de las variables relativas a la composición sociocultural del alumnado.

A través de los ejemplos que analizaremos a lo largo de este capítulo se podrá apreciar que en buena parte de los casos la interpretación errónea es inducida por los propios ministerios de Educación y que en general los periodistas no tienen los conocimientos indispensables para detectar lo que muchas veces son groseras contradicciones o interpretaciones falaces o carentes de sentido.

1. Errores de interpretación derivados del desconocimiento estadístico

Informar sobre resultados requiere inevitablemente contar con un conjunto básico de conocimientos generales de estadística, en primer lugar, y del significado específico de las “métricas” o tipos de datos que se emplean para reportar resultados en las evaluaciones educativas. Para ponerlo en otros términos, ningún periodista debería cubrir las noticias económicas si no sabe distinguir un porcentaje de un promedio y si no conoce el significado de términos como Producto Bruto Interno o Índice de Precios al Consumidor. Del mismo modo, para cubrir la información del sector educativo y, en especial, los resultados de las evaluaciones, es imprescindible poseer un conjunto mínimo de conocimientos y asesorarse con personas especializadas, para no cometer errores gruesos: qué diferencia hay entre un porcentaje promedio de respuestas correctas, una calificación –o nota– y un porcentaje de alumnos suficientes; qué diferencia a una prueba normativa de una prueba criterial, etc.

Ejemplo 1.1.

Interpretación de los puntajes como cantidad de respuestas correctas o “aciertos”

Este es un error que aparece en forma reiterada en varios artículos de la prensa mexicana. Todas las evaluaciones internacionales y algunos sistemas nacionales presentan sus resultados a partir de un puntaje cuya media se fija arbitrariamente, por lo general en 500 puntos. Esto se hace a través del tratamiento de los datos mediante un modelo matemático complejo denominado “teoría de respuesta al ítem” –cuyo tratamiento excede los alcances de este documento–. Obviamente un periodista no tiene por

qué conocer los fundamentos matemáticos de la misma, pero sí debiera saber que las medias que se presentan en los informes son puntajes abstractos y no cantidad de aciertos. Es decir, un puntaje de 500 no significa que un alumno haya respondido correctamente 500 preguntas – lo que además parece algo descabellado–. Sin embargo, este es el modo en que se han reportado los resultados del TIMSS en México (véase el Recuadro 5).

RECUADRO 5 ERRORES POR DESCONOCIMIENTO ESTADÍSTICO (1)

“Reprueba México Calidad Educativa”

“El promedio de México quedó atrás en más de 100 puntos con respecto a la media mundial. Por ejemplo, los estudiantes mexicanos de primero de secundaria obtuvieron en matemáticas el último lugar entre 28 países participantes: México sacó 375 aciertos frente a 483, que fue el promedio internacional. Ante los alumnos de las naciones más aplicadas, el contraste es mayor. Singapur, el primer lugar, obtuvo en matemáticas 604 aciertos, Corea del Sur quedó en segundo con 583 y Japón, en tercero con 574. En el examen de matemáticas para segundo de secundaria, los mexicanos se quedaron a 115 aciertos de distancia del promedio internacional... **La media internacional reportó que los niños evaluados lograban resolver acertadamente en promedio 470 problemas, pero los mexicanos sólo cubrieron un total de 362**”.

(Reforma.com, 15/10/01) (los subrayados son nuestros)

“Son mexicanos los burros”

“**Estos países asiáticos alcanzaron calificaciones que superan hasta en más de 100 aciertos el promedio internacional**. Por ejemplo, mientras en el examen de matemáticas de primero de secundaria la media global fue de 483 aciertos, los estudiantes mejor calificados obtuvieron 604 puntos. Después de Singapur, **los coreanos acumularon 583 cuestionamientos resueltos correctamente**, los japoneses 574 y los alumnos de Hong Kong 566”.

(Reforma.com) (los subrayados son nuestros).

Los autores de las notas debieran informarse de que los puntajes no refieren a cantidad de problemas o preguntas respondidas correctamente, sino que se trata de un puntaje arbitrario que forma parte de una escala que se construye a partir de un modelo matemático. Justamente para resolver el problema de la abstracción de estos puntajes es que normalmente los informes incluyen una descripción de lo que se denomina “niveles de desempeño”. Estos niveles describen qué tareas son capaces de realizar los alumnos en distintos tramos de la escala de puntajes. Por

ejemplo indican que los alumnos que obtienen entre 283 y 420 puntos mayoritariamente son capaces de resolver problemas matemáticos de tal tipo pero no de tal otro. Por ejemplo, en la prueba brasileña SAEB-97, el puntaje 325 en la escala empleada para matemática significa que⁹:

- Los alumnos clasifican sólidos geométricos en cuerpos redondos y poliedros.
- Interpretan resultados de medidas de longitud, masa, tiempo y capacidad.
- Identifican, comparan y ordenan números racionales (en las formas fraccionaria y decimal) así como números enteros.
- Interpretan lenguaje algebraico y resuelven ecuaciones y sistemas de ecuaciones de primer grado.

En parte el error en que se incurrió es explicable porque se trata de los resultados del TIMSS, que en México estuvieron ocultos durante muchos años, y que el periódico logró obtener de algún modo no convencional.

Ejemplo 1.2.

Atribuir el nivel de desempeño al país y, por lo tanto, a la totalidad de los alumnos del país

Según se explicó en el ejemplo anterior, es común cuando se reporta con puntajes abstractos describir el significado de la escala en términos de los que son capaces de hacer los alumnos con distintas puntuaciones. Para ello se trabaja con los denominados niveles de desempeño. A título ilustrativo en el Recuadro 6 se incluye un ejemplo de cómo en el Informe de Resultados de PISA 2000 se presentan y ejemplifican los niveles de desempeño en lectura.

9 Véase al respecto el documento **¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?** (RAVELA, P., 2001) disponible en la página web del Grupo de Trabajo sobre Estándares y Evaluación, sección Biblioteca (http://www.preal.org/docs-trabajo/ravela_n22.pdf).

RECUADRO 6 “EJEMPLO DE NIVELES DE DESEMPEÑO EN LECTURA EN LA EVALUACIÓN PISA”

Nivel 5	por encima de 625 puntos	727	<i>Por ejemplo</i> Se mostró a los estudiantes un diagrama de árbol de la población en edad de trabajar en un país, junto con descripciones del estatus laboral de trabajadores individuales. Los estudiantes tuvieron que decidir a qué categoría del diagrama pertenecía cada trabajador. Para ello, tuvieron que deducir qué criterio utilizar para clasificar a los trabajadores a partir de la estructura y contenido del diagrama, extrayendo también información de las notas a pie de página que, por su situación, no era fácil de advertir. Puntuación asociada: 727 puntos en la escala de interpretación.
Nivel 4	de 553 a 625 puntos		
Nivel 3	de 481 a 552 puntos		
Nivel 2	de 408 a 480 puntos	655	Se mostró a los estudiantes una comunicación de un departamento de personal sobre un servicio de ayuda en el área de la movilidad laboral. Los estudiantes tenían que deducir de los dos modos en que este servicio podía ayudar a las personas que habían perdido su trabajo a partir de la información presentada indirectamente, la cual debía distinguirse de otra información distractora que podía confundirse fácilmente con la información requerida. Puntuación asociada: 655 puntos en la escala de obtención de información.
Nivel 1	de 335 a 407 puntos		
Por debajo del Nivel 1	menos de 335 puntos	652	Después de leer una historia de tres páginas sobre la aventura de una mujer, se preguntó a los estudiantes si consideraban que la historia tenía un final apropiado, explicando el porqué de su respuesta. Para obtener la calificación máxima tenían que evaluar el final del relato en términos de su unidad temática relacionando la última frase con las relaciones, los temas y las metáforas centrales de la historia. Puntuación asociada: 652 puntos en la escala de reflexión y valoración.

Fuente: Primeros resultados del Proyecto PISA 2000. Traducción al castellano publicada en la web del INCE de España.

El nivel de desempeño es una característica del alumno: describe lo que un alumno es capaz de hacer, según el puntaje que obtuvo en la prueba. Un error común en los informes de prensa sobre PISA consiste en atribuir el nivel de desempeño al país, según cuál sea el puntaje promedio del país. Es decir que, si el promedio de un país se encuentra dentro del rango de puntajes correspondiente al nivel 2, se afirma que el país está en el nivel 2. Y luego se concluye que dicho nivel describe la capacidad de todos los alumnos del país. Por ejemplo:

“De cinco níveis possíveis de Classificação da média geral, os brasileiros foram os únicos a ficar no nível um. Isso significa que os alunos souberam apenas localizar informações explícitas nos textos” (Estado de Minas, 05/12/2001).

Como el promedio obtenido por Brasil en PISA se ubica dentro del rango correspondiente al nivel 1, la nota concluye que la totalidad de los alumnos brasileños apenas son capaces de realizar lo que corresponde a este nivel. En realidad el promedio es un promedio, hay alumnos con puntuaciones por debajo y por encima del mismo, lo que significa que hay alumnos que hacen más que lo correspondiente al nivel 1. En el propio informe público de PISA presentan varios gráficos con la información acerca de la proporción de alumnos de cada país que quedó ubicado en cada nivel. Para Brasil, el 1% estuvo en el nivel 5, el 3% en el nivel 4, el 13% en el nivel 3, el 28% en el nivel 2, el 33% en el nivel 1 y el 23% por debajo del nivel 1.

El mismo error se puede apreciar en el Recuadro 8, incluido algunas páginas más adelante en este capítulo, en que se coloca a los países dentro de los niveles de desempeño, como si estos caracterizaran al país.

Los niveles de desempeño son un atributo de los individuos y justamente lo que se pretende con ellos es romper con las limitaciones que tiene el puntaje promedio como medida resumen del resultado de un país. Los niveles de desempeño buscan informar acerca de la distribución de los alumnos de cada país, mostrando que hay alumnos con mayores capacidades y otros con menos. Al atribuir al país el nivel correspondiente a su promedio se vuelve a sobresimplificar la realidad en forma equivocada.

Ejemplo 1.3.***En “cuidados intensivos” por estar “por debajo de la media”***

“El Secretario de Educación en Nuevo León prefirió no hacer públicos los resultados del REALE para no “satanizar” a las escuelas con pobre rendimiento.

Indicó que existen cerca de 300 primarias en “cuidados intensivos”, ya que se ubican por abajo de la media estatal en español y matemáticas de acuerdo con el promedio de aciertos de los estudiantes de cada plantel.

“Tenemos escuelas con ‘cuidados intensivos’ que hay que cuidar muy bien. Desde hace dos años les hemos entregado sus resultados, y ellos han trabajado al respecto; hay algunas que se han mantenido en ese nivel, por lo que hablaremos con los directores; maestros y padres de familia para exhortarlos a cambiar esto’; dijo”.

(“Evalúan escuelas, ocultan nombres”, *El Norte*, México, 23/10/01).

Otro problema serio de interpretación tiene que ver con la interpretación del hecho de “estar por debajo de la media”. En principio estar por debajo de una media no constituye ningún demérito especial ya que, por definición, siempre habrá unas escuelas por debajo y otras por encima de la media. El único caso teórico en que esto no ocurriría sería una evaluación en que todas las escuelas tuviesen exactamente el mismo promedio. Salvo esta situación excepcional y muy poco probable, por definición una media implica que alrededor de la mitad de los alumnos o centros están por debajo, en tanto los restantes estarán por encima. El problema no es este, sino cuán lejos de la media se encuentran y cuán lejos está la media de un mínimo exigible o esperable –o, en otras palabras, qué significa la media en términos de capacidades y conocimientos de los alumnos que llegan a ella–.

En este sentido, la calificación de “cuidados intensivos” parece por lo menos exagerada, salvo que se tratase de escuelas muy por debajo de la media. De la nota no surge con claridad si la información fue mal brindada por el funcionario o mal interpretada por el periodista.

Un poco más adelante, en el mismo artículo, se indica que:

“En matemáticas, las escuelas secundarias con índice de aciertos mayores a 110, cifra superior a la media estatal, eran

31 en 1999, mientras que este año se ubicaron en ese rango solo 17. En español, las escuelas llamadas de “alto rendimiento” eran 29 en 1999. Hoy son solo 18”.

“Matemática. En 1999, 31 escuelas tuvieron los mayores aciertos. La baja a 18 en el 2000 y a 17 este año, preocupa a las autoridades educativas”.

(“Evalúan escuelas, ocultan nombres”, *El Norte*, México, 23/10/01)
(los subrayados son nuestros).

La información brindada hace suponer que se trata de una media estandarizada en 100 puntos (ya que de lo contrario sería necesario saber cuál era el puntaje máximo posible). Es decir, nuevamente se trata de un puntaje abstracto y no una cantidad de “aciertos”, como interpreta el periodista. Aparentemente se ha definido como “aceptables” los promedios entre 90 y 110 (probablemente 10 sea la desviación estándar).

El problema es que los juicios de valor y las interpretaciones que se formulan están sustentadas sobre la distribución normal de las escuelas y una distribución normal siempre es así: la mayoría de los casos –alrededor del 70%– se ubican dentro de los márgenes de una desviación estándar por encima o por debajo de la media y una minoría de casos se ubica en los extremos.

Por lo tanto, de este tipo de análisis –que casi seguramente sea responsabilidad de los funcionarios y no de los periodistas– no se pueden derivar conclusiones absolutas (escuelas buenas y malas) sino únicamente relativas (escuelas mejores y escuelas peores en relación al conjunto). Las escuelas que se ubican en los extremos sin duda son mejores o peores en relación al resto –en los logros educativos medidos por las pruebas–. Pero los cambios en la cantidad de escuelas en esos extremos no deberían preocupar al periodista. Este tipo de cambios es altamente probable que obedezca a caprichos de la distribución estadística, a menos que obtener entre 90 y 110 puntos tenga un significado preciso en términos de logros educativos, lo que debería aparecer explicitado. Pero si se está trabajando con una distribución normal, el hecho de que sean menos las que están en el extremo más alto puede obedecer tanto a que las mejores escuelas empeoraron como a que el conjunto de las escuelas mejoró, y por lo tanto son más homogéneas.

Ejemplo 1.4.***Una extraña comparación entre las distancias sociales y las distancias educativas***

Un último ejemplo de desconocimiento estadístico que merece ser incluido en esta colección son las afirmaciones de un funcionario ministerial dirigidas a transmitir el mensaje de que la escuela tiene un efecto de igualdad social:

“Para XX, el estudio demuestra que, si bien la distribución de los niveles socioeconómicos en el país es bastante parecida a los niveles de rendimiento (a más bajos recursos, peores resultados), la diferencia en términos educativos es muchísimo menor que la brecha económica. Por ejemplo, la región Centro tiene un ‘producto bruto que es más del doble del de la región NEA: en cambio, la diferencia en resultados educativos es de solo 9 puntos’.”

“Los alumnos rinden mejor en matemática que en lengua”, *La Nación*, Argentina, 28/7/01 (los subrayados son nuestros).

La afirmación, que es groseramente errónea, tiene implícito el supuesto de una relación directa y lineal entre el producto bruto de una provincia o región y los aprendizajes de los alumnos de la misma. De acuerdo a la afirmación subrayada, por cada punto de diferencia en el producto bruto entre dos regiones debería haber un punto porcentual de diferencia en los rendimientos educativos. El hecho de que haya más puntos porcentuales de diferencia en el producto bruto que en los aprendizajes demostraría que la brecha educativa es menor que la económica.

Para que el lector entienda mejor, lo afirmado en el párrafo anterior equivale a comparar a dos muchachos, Juan y José, de 12 y 15 años respectivamente, uno que pesa 50 kilos y el otro 65, y decir que hay una desproporción porque la diferencia es de 3 años en edad y de 15 kilos en peso. Se trata de unidades que miden cosas distintas que no pueden ser comparadas entre sí sin más. En todo caso habría que establecer cuál es la relación estadística entre peso y edad, para luego ver si las diferencias entre ambos niños son adecuadas o exageradas.

Del mismo modo en educación hay que establecer primero una correlación entre las variables sociales (producto bruto, pero tal vez mejor nivel de alfabetización o educación promedio de la población) y los

rendimientos, para luego analizar si las diferencias son mayores o menores de lo esperable.

Este error fue recogido sin cuestionamientos también por otros medios:

“De acuerdo a la evaluación realizada por el Ministerio de Educación de la Nación las diferencias de resultados entre regiones y niveles sociales son significativamente menores a las existentes en el resto de la sociedad. ‘Esto permite afirmar que la escuela sigue teniendo una notable capacidad de reducción de diferencias sociales’.

(“Chubut tiene el promedio educativo más bajo del país”, Jornada, Argentina, 8/9/01) (los subrayados son nuestros).

2. Aprobados y reprobados

Otra importante fuente de errores en la información al público es la traducción de los resultados a calificaciones y/o su interpretación en términos de aprobación/reprobación. Este recurso, si bien parece facilitar a los periodistas la comunicación con el público, en la mayor parte de los casos es inapropiado, con la única excepción de los casos de exámenes que efectivamente tienen como finalidad aprobar o reprobar alumnos, o de pruebas en las que se establece explícitamente un “punto de corte”, es decir, un puntaje que deben alcanzar los alumnos para que su desempeño sea considerado satisfactorio.

Sin embargo, en la mayor parte de los casos esto último no se hace, sino que las pruebas simplemente buscan describir todo el espectro de capacidades de los alumnos y en todo caso mostrar qué proporción de ellos se encuentra en cada estadio o “nivel” de competencia¹⁰.

Ejemplo 2.1.

“Un país de reprobados”

Un buena cantidad de notas periodísticas referidas a los resultados de TIMSS y PISA en México, Brasil y España recurren a este tipo de inter-

¹⁰ Véase al respecto el capítulo II del Documento ¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina? (RAVELA, P., 2001), disponible en la página WEB del Grupo de Trabajo sobre Estándares y Evaluación, sección Biblioteca (http://www.preal.org/docs-trabajo/ravela_n22.pdf).

pretación o metáfora de la reprobación. Ejemplo de ello son los titulares incluidos en el Recuadro 1: *“Estudiantes brasileños reprobados en examen”*, *“Los estudiantes españoles suspenden...”*, *“Un país de reprobados”*. En todos los casos se trata de modos inapropiados de presentar la información, en tanto refieren a pruebas que no tenían como finalidad aprobar o reprobar alumnos y que no establecieron un criterio respecto a lo deseable o esperable de los alumnos, sino que simplemente pretenden describir a través de cinco niveles de desempeño lo que los alumnos son capaces de hacer en lectura.

Por ejemplo en el caso de PISA, es discutible si el nivel 5, que refleja una capacidad avanzada de lectura e interpretación de textos, es el nivel esperable y exigible para todos y/o la mayoría de los jóvenes de 15 años de edad. PISA no pretende afirmar esto, sino simplemente informarnos que alrededor del 10% de los jóvenes de 15 años de los países miembros de la OCDE logran alcanzar este nivel. Pero es bastante complejo decidir a cuál de los cinco niveles deberían llegar a los 15 años los jóvenes de todo el mundo o de un determinado país o región. En todo caso, lo que los países deberían hacer a partir de la información de PISA y sobre la base de la opinión experta —e inevitablemente subjetiva— de docentes y especialistas, es proponerse metas en términos de los niveles a los que razonablemente deberían llegar la mayoría de sus estudiantes en el mediano plazo.

Más inapropiado aún es considerar al país entero como aprobado o reprobado y, peor aún, atribuir a un organismo como la OCDE la intención de calificar la calidad educativa de un país (véanse los titulares incluidos en el Recuadro 1: *“Reprueba la OCDE la calidad de la educación impartida en México”*; *“País Reprovado”*; *“Reprueba México calidad educativa”*; *“Según la OCDE. Reprobados en educación”*).

La OCDE como institución no dictamina sobre la calidad de los sistemas educativos ni reprueba países. La metáfora del “país reprobado” no parece aportar ni contribuir más que al sensacionalismo. Los alumnos no fueron reprobados, porque no se trataba de un examen sino de un estudio de tipo diagnóstico.

En realidad este tipo de estudios únicamente aporta una información a partir de la cual cada país debe realizar la interpretación y el diagnóstico de lo que está ocurriendo con su sistema educativo, planteándose preguntas tales como: ¿es nuestro currículo el adecuado para los desafíos que los jóvenes encontrarán en el mundo contemporáneo?; ¿estamos enseñando cosas relevantes a los jóvenes?; ¿estamos ense-

ñando bien? (sin descartar la posibilidad de concluir que en realidad las competencias que PISA evalúa no son las que consideramos apropiadas, necesarias ni deseables para nuestros jóvenes, en cuyo caso no nos preocuparán los resultados y probablemente no deberíamos seguir participando en el Programa).

Llevar adelante esta última tarea implica tiempo y estudio –de los académicos, de los técnicos en los Ministerios, de los propios periodistas–. *Atribuir al organismo externo un papel de “certificador de la calidad educativa” es en realidad una forma de eludir el trabajo de análisis, interpretación y reflexión colectiva que debe realizarse a partir de los datos.*

Ejemplo 2.2.

“La nota no llega a 7”

En algunos países, como Ecuador y Argentina, la confusión ha sido promovida por los propios ministerios de Educación, que añaden en la presentación de sus resultados una “traducción” de los mismos a un promedio similar al de la escala de calificaciones vigente en el país, con la intención de facilitar la comprensión por parte del público. Pero a pesar de la buena intención, con ello se induce a la población –y a los medios de prensa– a interpretar las pruebas como si establecieran puntos de corte y de acreditación o certificación, del mismo modo que lo hace la calificación escolar, lo cual en realidad no es así, ya que se trata de pruebas de carácter descriptivo que no suelen incluir una explicitación sistemática de cuál es el nivel mínimo que significa aptitud para aprobar un curso.

En Argentina, por ejemplo, desde el propio Ministerio se dice que el puntaje máximo en las pruebas nacionales –10– solo refleja el dominio de los contenidos mínimos indispensables para aprobar el curso (véase la nota incluida en el recuadro 7). Ello significa que únicamente los alumnos que alcanzaron 10 puntos en la prueba estarían en condiciones de aprobar el curso, es decir, de alcanzar una calificación de ‘7’ –la nota de aprobación en el sistema educativo– (aunque las pruebas nacionales no tienen consecuencias de promoción para los alumnos).

En este contexto, carece de sentido todo el tratamiento que se realizó en la Argentina de los puntajes como calificaciones. Si un puntaje 10 refleja el dominio de lo básico, entonces en términos de calificaciones debiera ser traducido como 7, es decir, la calificación mínima de aprobación. Y todos los alumnos que no alcanzan el puntaje 10 debieran ser

considerados como reprobados. Por tanto carece de sentido transformar mecánica y linealmente los porcentajes promedio de respuestas correctas (que están en una escala de 1 a 100) a la escala de 1 a 10, como si fueran calificaciones, porque 7 significaría en ese caso que un alumno domina solo el 70% de los conocimientos mínimos e indispensables y no la totalidad de ellos. Sin embargo, en los informes del propio Ministerio se traducen los porcentajes promedio de respuestas correctas a calificaciones.

RECUADRO 7 LA METÁFORA DEL APROBADO/REPROBADO

“La nota no llega a 7”

“Un 6,69 en lengua y un 6,70 en matemática son las notas promedio que los alumnos que el año último terminaron el secundario obtuvieron en las pruebas tomadas por el Ministerio de Educación para medir la calidad de la enseñanza.

En la última evaluación, la nota subió a 6,7, tanto en lengua como en matemática.

En séptimo grado, la nota en lengua fue 6,6, un punto por encima del resultado de 1997.

En matemática, los alumnos de séptimo grado obtuvieron un 6, cuando la nota había sido 5,4 en 1997.

A pesar de la mejora, los resultados siguen siendo poco alentadores: en ningún caso el promedio nacional llegó al 7, que por lo general es el mínimo necesario para aprobar una materia en el colegio.

Además, el 10 no significa, en el criterio ministerial, “excelencia”, sino únicamente que el alumno domina los contenidos mínimos y elementales de cada materia”.

(La Nación Line, Argentina, 28/5/99) (los subrayados son nuestros)

Asimismo, hay dos contradicciones flagrantes en las siguientes declaraciones del funcionario ministerial que el periodista no logra detectar:

*“En las calificaciones de las pruebas de los Operativos de Evaluación de la Calidad Educativa, en la escala del 1 al 10, **el puntaje más alto no es sinónimo de la excelencia. El “10” solo representa lo mínimo que un alumno debe saber para el grado o año que esté cursando. Según estos parámetros, solo un tercio de los que el año pasado terminaron la escuela media está cerca de saber lo indispensable...***

“Esto se traduce en una realidad: la escuela pública argentina sigue siendo un factor importante de igualdad social. Puede producir resultados positivos entre los sectores más vulnerables,” aseguró el subsecretario de Educación Básica, XX. Y

también reconoció: “La escuela está más capacitada para llevar a los alumnos hacia la “media”, pero tiene dificultades para que desarrollen las más altas habilidades”. (La escuela secundaria no logra objetivos mínimos, Clarín, Argentina, 28/8/01) (los subrayados son nuestros).

La primera es la afirmación relativa a la capacidad de la escuela a llevar a los alumnos “hacia la media”. ¿Qué significa esto? Como se explicó más arriba en este texto, siempre, por definición, la mayoría de los alumnos se encontrarán en torno a la media, salvo casos excepcionales en que la distribución tenga forma de dos campanas en los extremos.

La segunda y más importante es que, mientras en el primer párrafo citado se indica que los ítems solo reflejan “lo mínimo que un alumno debe saber” –y que eso es lo que refleja un puntaje de 10–, en el segundo párrafo se afirma que la escuela argentina tiene dificultades para desarrollar en los alumnos las más altas habilidades. Esta última afirmación no tiene ningún sustento empírico, dado que estas habilidades no están siendo medidas por las pruebas que, según lo afirmado por el propio funcionario, solo miden las habilidades básicas. Por tanto, no hay forma de saber si se están alcanzando o no “las más altas habilidades”.

En este contexto la “media” no representa un nivel de competencia “medio” en relación a lo que debiera saberse, sino algo muy por debajo. Y un puntaje alto no significa un alto nivel de competencia, sino apenas el dominio de lo básico.

3. El uso exagerado e inapropiado de los “ranking”

Una aproximación que tiene “gancho” periodístico es la del **ranking o tabla de posiciones** estilo campeonato de fútbol. Los periodistas parecen sumamente proclives a seleccionar del conjunto de información que se brinda, aquella relativa a las tablas de posiciones, ya que ofrece datos, es simple y fácil de comprender y resulta atractiva para el lector.

Este enfoque se suele priorizar tanto cuando los resultados son malos como cuando son buenos. En Australia, por ejemplo, cuyos resultados en la evaluación internacional PISA fueron muy positivos, se recurre continuamente al titular *“top of the class”* (los mejores de la clase) –promovi-

do desde el propio Ministerio de Educación en su primera comunicación oficial a la prensa sobre los resultados del mencionado estudio¹¹—.

Lo mismo ocurrió en Brasil y México, con artículos fuertemente focalizados en describir el ranking (como el incluido en el Recuadro 8). Otros titulares le ponen un toque de humor ácido a la metáfora, como el titular mexicano “¡Le ganamos a Brasil!” (que hace referencia a que en el estudio PISA el puntaje promedio de los alumnos mexicanos fue algunos puntos superior al de los brasileños).

Este mismo fenómeno ocurre con los resultados de evaluaciones nacionales en países en los que se publican ranking de establecimientos. Un ejemplo ilustrativo es la nota publicada en Chile por el diario *La Segunda* el 21 de noviembre de 2002, sobre los resultados del SIMCE¹² 2001, que utiliza titulares tales como “*Tops en Región Metropolitana. Ranking de los 20 mejores promedios*” o “*Posiciones. Por área de desempeño. Por comunas*”¹³ acompañados de diferentes rankings de colegios con sus respectivos promedios.

Del mismo modo, en Argentina y Colombia es común encontrar rankings de provincias o información sobre supuestos “mejores colegios”.

“Siguiendo con la tendencia que se ha dado en los últimos años, los colegios privados se llevaron los laureles en lengua con 7 de promedio, y dejó en el segundo puesto a las escuelas técnicas (6,3). Las escuelas estatales comerciales y los bachilleratos estatales quedaron en el tercer puesto, con 6,2”.

(Argentina, *La Nación Line*, 28/5/98) (los subrayados son nuestros).

“Los trece colegios de Cali, rankeados este año entre los primeros 96 del país por parte del Instituto Colombiano para el Fomento de la Educación Superior, ICFES, tienen sus claves. Pese a que algunos bajaron en la clasificación frente a lo alcanzado el año pasado, la calidad educativa tiene sus claves y secretos. El primer centro docente caleño categorizado como ‘muy superior’ por el ICFES es el Liceo Femenino Benalcázar, que aparece en el puesto quince. Luego están el Colegio Co-

¹¹ Véase la sección sobre Australia en la página de OCDE, pág. 5.

¹² La evaluación nacional de aprendizajes chilena.

lombo Británico en el lugar 22, el Berchmans en el 24, el Colegio Bolívar en el puesto 39 el Juanambú Aspen en el 60, Lacordaire en el 61, el Diana Oese en el 71, el Colegio Alemán en el lugar 74, Liceo Tacuri en el 82, el Hebreo Jorge Isaacs en el 91 y el Franciscano Pío XII en el puesto 93".

(*El País*, Colombia, 2/2/02) (los subrayados son nuestros).

Existen dos problemas centrales vinculados al exagerado énfasis en reportar rankings. En primer lugar, se suele dejar de lado lo principal: qué conocen y qué son capaces de hacer los estudiantes, transformando la educación en algo así como un campeonato de fútbol, trivializando lo que hay detrás del esfuerzo educativo y evaluativo. En segundo lugar, en general los ranking tienen muy serias debilidades técnicas, en la medida en que se construyen sin considerar la composición sociocultural del alumnado de las escuelas, provincias o países, con lo cual inducen al lector a conclusiones erróneas acerca de la calidad educativa de las entidades rankeadas¹³. Veamos algunos ejemplos.

Ejemplo 3.1.

Brasil “fica” último en PISA

Parece inevitable que cuando los reportes incluyen algún tipo de ranking la prensa se focalice fuertemente sobre el mismo, aun cuando no sea lo central de la información que se aporta. En el caso de PISA por ejemplo, el informe oficial intenta focalizar la atención más en la descripción de los niveles de lectura de los alumnos y las proporciones de alumnos que alcanzan cada nivel. Sin embargo, esta información en general queda en un segundo plano en relación a las posiciones en el ranking, e incluso esos niveles de desempeño, que se construyen para caracterizar a los alumnos, son mal interpretados por algunos medios de prensa en términos de niveles asignados a los países, según fuera explicado más arriba (véase el Recuadro 8). En el mismo artículo se puede apreciar que la nota está básicamente destinada a mostrar el ordenamiento de los países en la “tabla de posiciones”, incluyendo muy escasa información sobre qué significan los puntajes y qué es lo que están aprendiendo los alumnos.

¹³ Además de lo cual, no suelen considerarse los márgenes de error, que generalmente hacen que una buena parte de las diferencias no sean estadísticamente significativas.

RECUADRO 8 LA TABLA DE POSICIONES

“Brasil queda último en evaluación sobre educación”

El Ministerio de Educación divulgó este martes que Brasil quedó en último lugar en la prueba del Programa Internacional de Evaluación de Alumnos (Pisa), que midió la capacidad lectora de los adolescentes de 15 años de 29 países de la Organización para la Cooperación y el Desarrollo Económico (OCDE) y de otros invitados a participar. La muestra incluye alumnos de escuelas públicas y privadas de zonas urbanas.

El Ministerio de Educación atribuye el mal desempeño al desfase entre grado y edad de los alumnos de la red pública. Los alumnos de 15 años de Brasil están distribuidos desde el 7º grado de enseñanza básica hasta el segundo año de enseñanza media. Lo correcto sería que un estudiante de esa edad estuviese en el ter. grado de enseñanza media. Se calcula que aproximadamente la mitad de los 4.800 estudiantes que rindieron la prueba Pisa están en el 7º u 8vo. grado.

Entre los estudiantes con nueve o más años de escolaridad, o sea, sin atraso escolar, el promedio mejora y el país pasa a ser incluido en el nivel 2, al lado de Grecia, Rusia y Polonia. La evaluación establece cinco niveles de desempeño. El mejor resultado fue el de Finlandia, que obtuvo el índice 4.

La mayoría de países evaluados - 25 de un total de 32 - llegó al nivel 3. Ese bloque incluye principalmente a países desarrollados.

Los estudiantes en el nivel 1 son prácticamente considerados analfabetos funcionales, capaces de identificar letras, palabras y frases, pero no de comprender el sentido de lo que leen (nivel 2). Para poder alcanzar el índice 3 se requiere tener capacidad de abstracción, hacer comparaciones y evaluaciones. Los estudiantes que logran hacer una evaluación crítica de un texto son clasificados en el nivel 4.

La prueba exige que los alumnos comprendan la lectura, a partir de la identificación y recuperación de información, interpretación y reflexión. Presentaba varias situaciones cotidianas y una serie de preguntas para ser respondidas de acuerdo con los textos.

Paulo Renato - “Esperaba un desastre peor”, dijo el ministro de Educación, Paulo Renato Souza, al anunciar el resultado. Para el ministro, ese desempeño está directamente relacionado con el desarrollo socio-económico. Basándose en indicadores como el PIB e o el índice Gini, que mide la concentración de ingresos, afirmó que el desempeño de los estudiantes termina viéndose perjudicado por las condiciones de vida de los brasileros.

Fuente: Diário do Grande ABC - 05/12/2001. Publicado en la sección sobre Brasil de PISA EN LA PRENSA (<http://www.Disa.oecd.Org/News/PISA2000IPISABlaZf1.Ddf>). Traducido por los editores.

Confirma o ranking:

Nivel 4

1º - Finlandia	546
	puntos

Nivel 3

2º - Canadá	534
3º - Holanda	532
4º - Nueva Zelandia	529
5º - Australia	528
6º - Irlanda	527
7º - Corea del Sur	525
8º - Reino Unido	523
9º - Japón	522
10º - Suecia	516
11º - Austria	507
12º - Bélgica	507
13º - Islandia	507
14º - Noruega	505
15º - Francia	505
16º - Estados Unidos	504
17º - Dinamarca	497
18º - Suíza	494
19º - España	493
20º - República Checa	492
21º - Italia	487
22º - Alemania	484
23º - Liechtenstein	483
24º - Hungría	480

Nivel 2

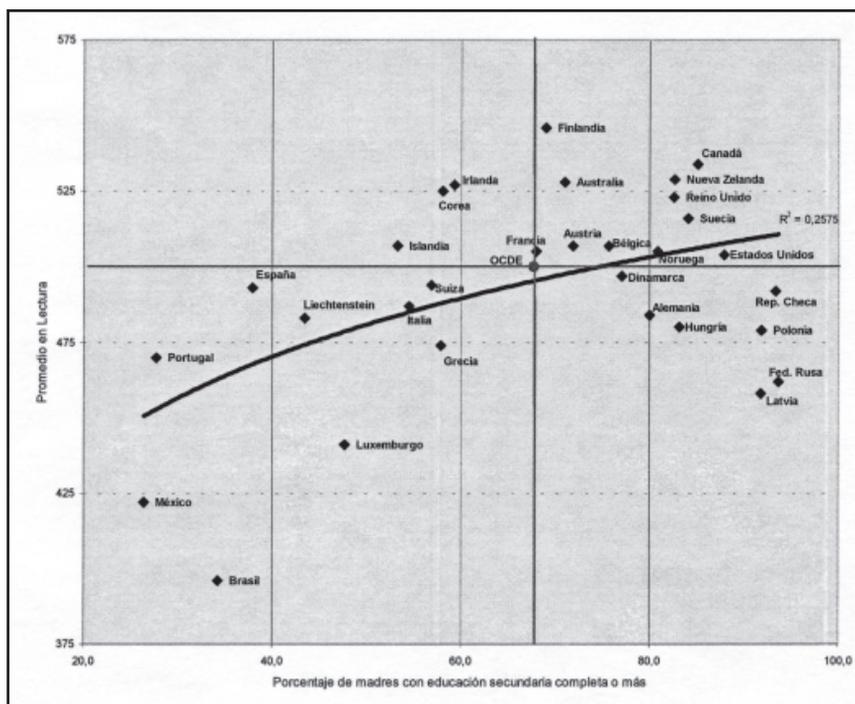
25º - Polonia	479
26º - Grecia	474
27º - Portugal	470
28º - Federación Rusa	462
29º - Letonia	458
30º - Luxemburgo	441
31º - México	422

Nivel 1

32º - Brasil	396
--------------	-----

Pero además de desviar la atención de lo principal, los rankings tienen como defecto principal que no toman en cuenta la composición del alumnado, cuando muchas investigaciones muestran que la principal variable que determina los aprendizajes de los alumnos es la composición social del alumnado de la escuela. En el caso de PISA, si se hace el ejercicio de poner los datos del ranking en un gráfico de dos ejes que tenga en cuenta los niveles educativos de las madres de los alumnos que hicieron las pruebas, se puede arribar a una visión mucho más rica y apropiada que la que arroja la “tabla de posiciones” (véase el Gráfico 1).

GRÁFICO 1
El Ranking de Pisa considerando el contexto sociocultural



Fuente: Elaboración propia a partir de datos del Informe de Resultados de PISA 2000 (Knowledge and Skills for Life. First Results From PISA 2000, OCDE, 2001)

En este gráfico cada país está ubicado según dos valores: el puntaje promedio en lectura (eje y) y el porcentaje de alumnos cuyas madres terminaron toda la educación secundaria o tienen un nivel educativo terciario (eje x). Es decir, se trata de las madres que llegaron más allá que el nivel educativo que actualmente están cursando sus hijos, quienes están terminando en general el primer ciclo de la educación media –8°, 9° o 10° grados–. Estas madres completaron 12° o más. Es importante señalar que la variable –el nivel educativo materno–, si bien es importante por sí misma, más lo es como indicador del nivel educativo y cultural del hogar –mayor nivel educativo materno está correlacionado con mayor cantidad de libros y elementos educativos, mejores ingresos y condiciones de vida, etc., todos elementos que inciden fuertemente en el desarrollo del tipo de capacidades que PISA evalúa en los niños y adolescentes–.

Es interesante notar que el promedio de la OCDE se ubica en el entorno del 70%, es decir, una holgada mayoría de los alumnos son hijos de madres que como mínimo completaron 12 años de escolaridad formal. La relación entre el porcentaje de madres con educación secundaria completa y más y el puntaje promedio en lenguaje se refleja en el R^2 de 0,2575, lo cual significa una correlación más que respetable de 0,51¹⁴.

Brasil, México, España y Portugal se ubican bastante lejos del promedio de la OCDE, por debajo del 40% de madres con secundaria completa. En este contexto, en primer término, la interpretación del resultado de España podría ser bastante diferente de la que fuera planteada en las declaraciones de los propios funcionarios ministeriales en las notas de prensa analizadas anteriormente en este documento. Teniendo en cuenta esta desventaja en los antecedentes educativos de su población, los resultados no son tan “malos” como se los muestra. Esto no quiere decir que no sea importante preocuparse por la brecha que separa a España del resto de los países de la OCDE. Pero no puede atribuirse esa distancia exclusivamente a deficiencias del sistema educativo y/o de los educadores, sino que debe analizársela como un problema más complejo que atañe a toda la sociedad.

¹⁴ En realidad se pueden lograr correlaciones bastante más altas si se construyen índices más sofisticados para el nivel educativo del hogar del estudiante, pero para facilitar la comprensión del lector preferimos trabajar con una variable sencilla, como el porcentaje de madres que completaron la educación secundaria o más.

La situación de Brasil y México también merece un comentario. Si bien quedan muy abajo, aun en el gráfico de doble entrada, debe considerarse que el indicador de educación materna utilizado no muestra la realidad de estos países en su cabalidad. La OCDE y los países del norte en general, cuando quieren definir bajo nivel educativo materno ponen la línea de corte en no haber completado la educación secundaria. Para ellos no haber completado la primaria es algo absolutamente excepcional. Sin embargo, en nuestros países, esto es bastante normal entre las madres de los estudiantes¹⁵.

Algunos países que en el ranking simple aparecen con resultados destacados mantienen esa situación –como Finlandia– mientras que este modo de presentación permite resaltar la calidad de los resultados de otros países que, como Corea e Irlanda, no destacan en el ranking simple. Un último comentario relevante sobre lo que muestra el gráfico así construido es la situación de los países de Europa del Este, los que, a pesar de poseer muy buenos niveles de educación en los hogares, tienen resultados muy por debajo de lo esperable. En principio pueden plantearse dos hipótesis al respecto: o bien la educación que se brinda en estos países está dirigida a aprendizajes muy diferentes de los que PISA evalúa, o bien la crisis posterior a la caída del muro de Berlín ha tenido efectos devastadores sobre los sistemas educativos, lo que hace que estos países estén perdiendo el capital cultural acumulado por las generaciones anteriores.

Más allá de los análisis, el Gráfico 1 tiene como propósito principal ilustrar que es posible, de modo relativamente sencillo, hacer una presentación de resultados mejor que una simple tabla de posiciones que conduce a múltiples interpretaciones inapropiadas.

Ejemplo 3.2.

La diferencia de calidad entre las provincias

Esta traducción automática de resultados de logro educativo a interpretaciones en términos de calidad de los sistemas educativos sin tomar en

¹⁵ En Uruguay por ejemplo, un país con larga trayectoria en la universalización de la educación primaria, el 38% de las madres de los alumnos apenas completaron la primaria. Probablemente en México y Brasil haya cifras de este entorno, pero de madres que no completaron la escuela primaria. Todo esto significa que si se trabajara con más datos el indicador del eje x podría ser más complejo y la correlación más alta. Pero ello escapa a los alcances de este trabajo.

consideración la composición social del alumnado se verifica también en la mayor parte de los reportes de prensa sobre resultados nacionales, obviamente con responsabilidad de los ministerios que ofrecen información insuficiente o que directamente propician este tipo de interpretaciones erróneas.

En el artículo incluido en el Recuadro 9 nótese, en primer lugar, cuáles son las provincias “en la cima” y en el “piso”. Casualmente “en la cima” están las provincias más desarrolladas de la Argentina, y la última –Santiago del Estero– es una de las más pobres. Lo grave del artículo es que la interpretación de este hecho en términos de calidad y eficacia de los sistemas educativos provinciales es propiciada por el funcionario ministerial declarante.

El artículo es un exponente paradigmático de los problemas y falacias que genera no tener en cuenta la composición sociocultural del alumnado de escuelas, provincias o países cuando se presentan y analizan los logros educativos¹⁶. Santiago del Estero es una de las provincias más pobres de la Argentina. En la misma nota se dice que casi el 100% de las escuelas de la provincia formaban parte del “Plan Social”, un programa de apoyo a las escuelas de medios pobres, lo cual sugiere que los bajos resultados no necesariamente se deben a la ineficacia o falta de calidad del sistema provincial –puede ser que así sea, puede ser que no–, sino simplemente a que es más difícil enseñar allí porque los niños viven en condiciones de pobreza material y cultural. Tampoco debe extrañar que un colegio privado, ubicado en la provincia más pobre del país y al que seguramente asisten los hijos de la élite provincial, haya logrado el mejor puntaje nacional. Pero difícilmente haya motivo para sentir “*orgullo*” por algún tipo de “*mérito*”.

¹⁶ Véase al respecto el capítulo III, denominado “Comparando lo incomparable. La parábola de las carreteras” en el documento **¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?** (RAVELA, P., 2001), disponible en la página web del Grupo de Trabajo sobre Estándares y Evaluación, sección Biblioteca (http://www.preal.org/docs-trabajo/ravela_n22.pdf). En dicho capítulo se muestra cómo comparar resultados entre provincias o escuelas sin tener en cuenta su contexto social es como comparar la eficiencia de empresas constructoras de carreteras sin tener en cuenta el tipo de terreno en el que construyen.

RECUADRO 9 UTILIZACIÓN INAPROPIADA DE RANKINGS (1)

“No conforman los resultados de las pruebas del secundario”

“Santa Cruz, Catamarca, La Rioja, Chaco y Misiones están bien por debajo de los promedios nacionales alcanzados en las pruebas de matemática (6,86), lengua (6,57) e historia (6,48) tomadas en julio último a los estudiantes que este año terminan la escuela secundaria. En cambio, la ciudad de Buenos Aires, la provincia homónima y las de Córdoba y Santa Fe se ubicaron en la cima de la lista evaluadora.

‘La diferencia de calidad que se refleja entre las provincias es histórica. Desde que se comenzó a evaluar en 1991 se advirtió esta segmentación en el sistema educativo’, dijo XX a *La Nación*.

Las diferencias de rendimiento son dispares incluso dentro de una misma provincia. En Santiago del Estero se encuentra el colegio privado más exitoso del censo nacional, el Hermas de Bruijn, cuyos alumnos obtuvieron un promedio general de 9,38. Festearon con orgullo el mérito de ser el mejor colegio evaluado.

Sin embargo, las escuelas que rodean al Hermas de Bruijn distan bastante de esa misma excelencia educativa. Así lo demuestran los promedios que ubican a la educación de la provincia de Santiago del Estero entre las menos eficaces de la Argentina: 5,36 en lengua, 5,4 en matemática y 5,42 en historia.

‘Santiago del Estero tiene rendimientos muy heterogéneos. Casi el cien por ciento de las escuelas son del Plan Social Educativo’, dijo XX en referencia a un programa de asistencia y asesoramiento a las escuelas más necesitadas que promueve el Ministerio de Educación.

El promedio de 9,2 le permitió al Instituto de Educación Media N° 288 de la ciudad cordobesa de Río Tercero acceder al primer puesto del ranking de escuelas estatales de la Argentina. Y si bien Córdoba figura entre las provincias que salieron mejor paradas en la evaluación nacional el promedio general de 6,97 ni siguiera se acerca al estupendo desempeño demostrado por su escuela estrella”.

La Nación Line, Argentina, 27/11/99 los subrayados son nuestros.

También resulta algo cómica la calificación de “escuela estrella” que utiliza el periodista, quien además, nuevamente se equivoca al pretender que el promedio de la provincia se acerque al de su “escuela estrella” ya que como vimos anteriormente, por definición nunca el promedio de una provincia se podrá acercar al de su mejor escuela, ya que un promedio indefectiblemente se ubica en una zona central entre las mejores y las peores escuelas de una provincia.

Ejemplo 3.3.

“Los mejores colegios de Cali”

A continuación incluimos otro ejemplo de cómo los periodistas en nuestra región asumen sin dudar que los colegios con mejores resultados

son efectivamente los mejores colegios en términos de la calidad de la educación que imparten (Recuadro 10).

El articulista asume sin más que los colegios privados son buenos en cuanto instituciones educativas, ignorando que por ser privados su población está social y culturalmente seleccionada, por lo que en realidad las “claves” del éxito pueden no tener ninguna relación con los resultados de las pruebas. Del mismo modo, construye un juicio crítico respecto a la calidad del sector público, que tampoco puede sostenerse sin un análisis más complejo de los datos.

RECUADRO 10 UTILIZACIÓN INAPROPIADA DE RANKINGS (2)

“Los mejores colegios de Cali revelan sus secretos para estar entre los primeros del país. Las diez claves de la educación”

La persona es más importante que la tecnología. La clave del éxito está en una formación integral. Los trece colegios de Cali, rankeados este año entre los primeros 96 del país por parte del Instituto Colombiano para el Fomento de la Educación Superior, ICFES, tienen sus claves.

Pese a que algunos bajaron en la clasificación frente a lo alcanzado el año pasado, la calidad educativa tiene sus claves y secretos.

El primer centro docente caleño categorizado como muy superior por el ICFES es el Liceo Femenino Benalcázar, que aparece en el puesto quince. Luego están el Colegio Colombo Británico en el lugar 22, el Berchmans en el 24, el Colegio Bolívar en el puesto 39, el Juanambú Aspen en el 60, Lacordaire en el 61, el Diana Oese en el 71, el Colegio Alemán en el lugar 74, Liceo Tacuri en el 82, el Hebreo Jorge Isaacs en el 91 y el Franciscano Pío XII en el puesto 93.

1- Poner a la persona (el alumno) en primer lugar, en el centro de la educación, es la primera clave para Ann Marie Leroy, coordinadora académica del Liceo Benalcázar. Lo mismo opinan Gloria Stella Ramírez y Juan Carlos Arana, coordinadores académicos del Colegio Berchmans.

2- La formación integral es una segunda clave del éxito. Eso implica concebir una educación no solo como la transmisión de conocimientos, sino con gran contenido ético, estético y espiritual.

La calidad de la educación en Cali, sobre todo la pública, volvió al centro de la discusión, luego de que el ICFES incluyera solo trece colegios caleños en el ranking de los primeros 96 del país. Los trece colegios mejor calificados de la ciudad son privados, no hay ninguno del sector público”.

(El País, Colombia, 2/2/02) (los subrayados son nuestros).

¿Qué hacer con los rankings?

La focalización de la atención sobre los rankings aporta poco a la comprensión de los problemas y desafíos educativos, y constituye una

aproximación más bien simplista al tema. En primer lugar, porque los mismos suelen basarse en una única medida que no da cuenta de la diversidad de aspectos relevantes de la acción educativa de una institución. En segundo lugar, porque por lo general estos rankings no tienen en cuenta la composición social del alumnado de las escuelas o provincias, que es el principal predictor de los resultados. Por lo tanto, los rankings en general, tal como suelen ser construidos, no reflejan calidad educativa sino selección social del alumnado. En tercer lugar, porque en la mayor parte de los casos no va acompañada por otro tipo de información que ayude al lector a comprender más cabalmente lo que fue evaluado y lo que los resultados significan. La gran ausente en los artículos de prensa es la información sobre qué saben hacer y qué no saben hacer los estudiantes, qué es lo que se espera de ellos. Toda la información se centra en los datos numéricos y en los ordenamientos, pero casi no se informa a la sociedad acerca de cuáles son los conocimientos y capacidades que se espera alcancen los alumnos en las diferentes etapas de su escolaridad y en qué grado lo están logrando.

Para algunos medios de prensa esto es meridianamente claro. Véase al respecto el Recuadro 11, que recoge algunos párrafos de un artículo en el que se describe de qué manera muchos medios de prensa en los Estados Unidos están encarando el tema. Los párrafos son suficientemente elocuentes y no requieren de mayores comentarios.

RECUADRO 11 ALGUNOS MEDIOS DE PRENSA BUSCAN NUEVAS MANERAS DE REPORTAR

“Un nuevo agente en la responsabilización: el periódico local”

“Esta primavera, el Detroit Free Press anunció que ya no volvería a categorizar las escuelas y distritos sobre la sola base de sus puntajes en las pruebas estatales de rendimiento escolar.

El periódico llegó a esta decisión luego de conducir durante seis meses un análisis computarizado de los resultados del Programa de Evaluación Educativa de Michigan. Encontró que la pobreza y otros factores que están fuera del control de la escuela están tan fuertemente vinculadas a los resultados de las pruebas de rendimiento que hacer comparaciones directas resulta “inevitablemente fallido” y “mayormente irrelevante”. **Pienso que comprendimos, con algo de vergüenza, que nunca tuvimos ningún derecho a categorizar a los distritos basándonos en los puntajes del MEAP**”, dice Tracy Van Moorlehem, reportera sobre educación escolar del periódico.

Simplemente no es justo ni, en verdad, particularmente preciso”.

En lugar de ello, el Free Press se comprometió a que en adelante produciría una imagen más fina y detallada de cuán bien están funcionando las escuelas de Michigan, dados los retos que confrontan...

“Felicitó a cualquier periódico que intente contextualizar algo de este tema”, dice Linda Leddick, directora de investigación, evaluación y medición del rendimiento de las escuelas públicas de Detroit. **“Por mucho tiempo, se han estado difundiendo resultados de las pruebas como si fueran resultados de carreras de caballos, y eso no ayuda a su comprensión por parte del público”**. Pero ella previene que “este tipo de artículos puede ser particularmente peligroso si los periódicos solo van a comenzar a meter datos de pruebas en una mezcladora y hacer proyecciones sin mayor cuidado”...

Una de las fortalezas de los periódicos es su habilidad para complementar los datos con reportajes más tradicionales. Para su reporte, el World-Herald envió cinco reporteros a una docena de escuelas por dos semanas a visitar las aulas y a entrevistar a los padres, estudiantes y educadores.

Mientras más aprendimos sobre los puntajes en las pruebas de rendimiento y más conocimos sobre sus limitaciones, más sentimos la necesidad de ingresar a las aulas,” dice Mike Reilly, editor de proyectos del periódico....

Lynn Olson

*Extraído y traducido de Education Week on the Web del 17 de junio de 1998, con autorización de sus editores. El artículo completo puede consultarse en la página web del GTEE-PREAL.
(Los subrayados son nuestros).*

El camino no necesariamente es eliminar los rankings, sobre todo cuando ya se han instalado como demanda de la opinión pública, pero en todo caso construirlos de manera más apropiada. Lo que se necesita es utilizar más información y presentarla de un modo más inteligente –a la vez que inteligible para los lectores–.

Si el público quiere comparaciones, lo que hay que hacer es construir las de una manera más apropiada, al estilo de lo que intentamos ejemplificar con el Gráfico 1. No sería difícil, por ejemplo, trabajar con rankings paralelos que permitan establecer comparaciones y distinguir a las mejores y peores escuelas entre las que atienden a los sectores medios y altos de la población, por un lado, entre las rurales, por otro, entre las urbano-marginales, por otro (y al ponerlos en paralelo, comparar a todas entre sí, para no perder la visión de conjunto). O, del mismo modo, entre conjuntos de provincias con similares niveles de desarrollo educativo, cultural y económico. Lo que no tiene sentido es focalizar la atención en un ranking unidimensional que desconoce el conocimiento acumulado en materia de investigación socioeducativa y que conduce inexorablemente a conclusiones erróneas.

CAPÍTULO III

LA VISIÓN DE UNA PERIODISTA¹⁷

Aun cuando no haya sido exactamente una sorpresa, la ausencia de periodistas al encuentro marcado para las 17.00 horas de un día martes en un hotel de Santo Domingo fue emblemático. Establecido para un horario que coincide con el cierre de las ediciones, el encuentro estaba destinado al fracaso desde su nacimiento¹⁸.

El hecho reforzó que, si el objetivo es construir una cobertura más crítica, profunda y contextualizada sobre los sistemas nacionales e internacionales de evaluación, el primer paso es acertar en la relación con la prensa y conocer su lógica, empezando por la lógica operativa.

No se puede simplemente transferir la responsabilidad de un posible fracaso de la cobertura periodística sobre sistemas de evaluación a los periodistas, editores o empresarios de los medios, sin reconocer los límites y deficiencias de las propias agencias operadoras en los primeros contactos con la prensa. Y también sin conocer mejor las potencialidades y determinantes del quehacer periodístico, sus desafíos, límites y condicionantes.

Los problemas clave del periodismo ya los conocemos bien, pero nada nos cuesta citarlos nuevamente. Empezando por la eterna lucha del periodista contra el reloj, lo que lo lleva a la necesidad de tomar decisiones casi instantáneas, a formar juicios de valor en un plazo exiguo de tiempo, a sacar conclusiones apresuradas, lo que muchas veces, aunque no siempre, contribuye a la superficialidad o, peor, al error de información.

Otro gran problema, es la disputa por la “primicia” periodística, por la noticia exclusiva, disputa en la que no es raro que se atropelle a la ética. En la competencia por el liderazgo en el precioso espacio de la

¹⁷ El texto incluido en el presente capítulo fue escrito por la periodista Tacyana Arce – originalmente en portugués– en el marco del Seminario realizado en Santo Domingo y refleja de un modo agudo la visión “desde el otro lado”, es decir, desde alguien que trabaja en los medios de prensa y no pertenece al sector educativo.

¹⁸ La autora se refiere a una reunión de discusión sobre la difusión de resultados de las evaluaciones educativas, de periodistas con los miembros del Grupo de Trabajo de Estándares y Evaluación del PREAL. Dicha reunión era parte de un conjunto de reuniones realizadas con distintos actores –docentes, políticos, padres de familia, empresarios– organizadas como parte del trabajo del taller que se llevó a cabo en la preparación de este documento, según fue explicado antes. La reunión con periodistas fue la única que no pudo realizarse.

opinión pública, en la necesidad de hacerse indispensable para los lectores, no solamente para imponer su primacía sobre los demás, pero, sobre todo, para satisfacer una exigencia que es de los propios consumidores ('si pago un diario, es para que me mantenga informado antes que los demás') los vehículos de comunicación abortan los tiempos de maduración de la noticia, se abstienen de profundizar la información, muchas veces dejando escapar la calidad en beneficio de lo inédito.

Podríamos continuar citando otros determinantes, como la dificultad y los conflictos a la hora de definir los temas relacionando los intereses de los lectores con los intereses de la empresa y los intereses del mercado publicitario (que, finalmente, es el que sustenta un vehículo de comunicación), el constante retorno a temas fetiche, la crisis de identidad de algunos medios, como la prensa escrita, ante los cambios impuestos por la posmodernidad, o por el interés torpe, que no es del periodismo sino de la condición humana, por el sensacionalismo. Pero para no extendernos, citamos otro gran problema: el enfrentamiento entre noticia e intereses económicos y políticos, que solamente en un plano ideal puede dejar de existir, reconociendo que, antes que nada, las empresas periodísticas son grupos comerciales que, como cualquier otro, juegan conforme a las reglas del mercado, con inevitables perjuicios para el periodismo responsable.

Pero la pregunta que nos hacemos ahora es: ante este escenario desfavorable para la calidad de la cobertura periodística, ¿cómo actuar en favor de un comportamiento más adecuado de la prensa ante un tema tan sensible y estratégico como la divulgación de los resultados de las evaluaciones de aprendizaje? ¿Es posible sobreponerse a los intereses políticos, a las disputas simbólicas o a las presiones del mercado? ¿Es posible intervenir en la lógica y en el quehacer periodístico? ¿O esta discusión está tan destinada al fracaso como la reunión con periodistas programada para las 17.00 horas de un martes a la que hiciéramos mención antes?

Si asumimos que las únicas cuestiones que determinan una buena cobertura periodística son las condiciones macro citadas anteriormente, nada nos resta, sino el reconocimiento de que no se puede hacer mucho en la dirección de un cambio de paradigmas en la cobertura periodística. Sin embargo, lo que paso a sustentar ahora es que no son solamente los grandes determinantes los que minimizan las potencialidades del periodismo en Educación, sino también cuestiones menores y por tanto posibles de ser atacadas, las que, aunque no en la misma proporción, también contribuyen a un quehacer periodístico. Enfrentarlas no solo es posible, sino también necesario.

Aunque no conozca a fondo la relación entre prensa y ministerios de educación en los países citados en la colección de notas periodísticas y artículos seleccionados para la muestra de prensa que dio origen a este taller¹⁹, y aunque no esté al tanto de los intereses económicos y políticos de los medios en cuestión, arriesgaría a decir que la mayor parte de los errores y falacias identificados no tienen relación con intenciones explícitas de manipulación, con intenciones deliberadas, sino con la absoluta falta de preparación del periodista para lidiar con números, estadísticas, tablas y otras formas usuales de presentación de los resultados de los sistemas de evaluación. Un buen ejemplo es el artículo '*Reprueba México Calidad Educativa*', donde los puntos obtenidos por los estudiantes fueron confundidos con aciertos. ¿Qué tipo de intención malévolá podría haber por detrás de ese error? Seguramente ninguna. Lo más probable es que el periodista haya pensado que sus lectores no entenderían una escala de puntuación tan diferente como esa, ya que lo habitual es que las escalas varíen de 0 a 100. En la tentativa de hacer su tema más accesible a cualquier lector, regla básica del periodismo, el autor erró.

Entender lo que es una "media", algo muy simple para especialistas, no lo es para periodistas. Simplemente porque para ellos, así como para la población en general, "media" tiene otro significado. En Brasil, por ejemplo, para ser promovido de grado, el alumno precisa lograr una "media" de siete en diez. Quien queda por debajo de eso, es reprobado. Puede parecer estúpido, pero los periodistas, así como los lectores en general, no relacionan inmediatamente "media" con el resultado general de un grupo, en el que muchos logran notas más altas, otros notas más bajas, y que, por eso, obligatoriamente algunos estarán siempre por debajo de la media. Si no son alertados de esto, la primera reacción de los periodistas es creer que sería posible que todos estuviesen por encima de la media, sin percibir que, estadísticamente eso sería imposible.

Es necesario resaltar que tales errores no ocurren solo en la cobertura sobre Educación, sino que se repiten en diversos sectores, siempre que es necesario lidiar con números y similares. Es el caso, por ejemplo, de la cobertura sobre Salud, Índice de Desarrollo Humano, Violencia. Cierta vez un diario en Brasil, estampó un titular diciendo que los casos de homicidio en la Región Metropolitana habían crecido 50%, de 1.800 a

¹⁹ La autora se refiere a los artículos que se analizaron para elaborar una versión preliminar de este informe, que se revisó en el taller al que aquí se hace referencia.

3.600 ocurrencias al año. ¿Cómo pudo el periodista errar un número tan básico? Calculando el crecimiento no sobre el número inicial, sino sobre el número final.

Estoy a punto de argumentar que uno de los principales problemas del periodismo es la formación profesional de sus recursos humanos. En algunos países, el periodismo es una profesión, exigiéndose la formación en el nivel superior. En otros, como es el caso reciente de Brasil, por determinación del Tribunal Superior de Justicia, el periodismo puede ser ejercido por cualquier persona, en nombre de la libertad de prensa. Pero incluso donde se exige la formación de nivel superior, muchas veces la misma se limita a la formación técnica del periodismo: cómo operar cámaras, micrófonos, computadoras; cómo digitar un texto, hacer los títulos y los copetes. Lo que se espera de un buen fotoperiodismo.

Pocas escuelas se proponen formar a sus periodistas en temas de economía, política, ciudadanía. Y otras herramientas básicas para el ejercicio de la profesión, como conocimientos básicos de estadística, indicadores económicos, capacidad de digerir reportes e informes oficiales, también son dejadas de lado.

¿Qué es lo que ocurre en la vida real? El periodista se forma en el ejercicio de su profesión. En Brasil llamamos “foca” a los periodistas con menos de cinco años de formados, porque en ese período serán entrenados y formados. No se manda un foca a cubrir un pronunciamiento presidencial sobre la reforma económica del país, mucho menos a cazar primicias en el Congreso Nacional. Se manda a los focas para cubrir los asuntos que, aunque necesitan figurar en las páginas de los diarios, no tienen el mismo peso en el mercado simbólico del periodismo. En América Latina, el tema Educación es uno de esos temas secundarios. Desafortunadamente, en aquellas ocasiones en que un periodista que cubre el área de Educación (recuérdese, en tanto, que él no se dedica exclusivamente a esta área sino que generalmente la comparte con Salud, Urbanismo y Movimientos Sociales) comienza a destacarse por su buen trabajo, inevitablemente será promovido. Pasará a cubrir Economía y Política, dejando su lugar libre para otro foca.

A partir de aquí identificamos otro problema que puede ser enfrentado: hacer que la agenda educativa sea prioritaria, aunque eso signifique ingresar en un delicado juego de símbolos y sentidos que no tienen respuesta inmediata, menos aún, una fórmula exacta.

¿Por qué los diarios de todo el mundo dedican espacios significativos a la cobertura de Política y Economía? Seguramente porque los lectores de

todo el mundo se identifican con esos dos temas y exigen que los diarios que compran les ofrezcan materiales sobre el tema. ¿Y por qué el tema Violencia gana más espacio en las páginas de los diarios brasileños que en los diarios chilenos? Ciertamente la respuesta no puede ser que los diarios de Brasil sean más sensacionalistas que los de Chile, sino que el asunto es más urgente para los brasileños que para los chilenos. Lo que se pretende decir aquí es que, a diferencia de lo que el sentido común acostumbra repetir, la agenda periodística no es definida por un grupo de periodistas. Quien define la agenda periodística es la sociedad. Los medios se aprovechan comercialmente de deseos y ansiedades generalizados.

Por otro lado, los medios también son bastantes sensibles a ciertas fuentes. Universidades, organismos de desarrollo e investigación, organizaciones multilaterales, empresarios, organizaciones no gubernamentales e institutos de responsabilidad social acostumbran gozar de gran prestigio en las empresas periodísticas. El involucramiento de esos actores en un esfuerzo para despertar el interés de los medios en la agenda educativa puede tener buenos resultados, a través de contactos más frecuentes con periodistas y editores, la utilización de los espacios para editoriales, su reacción ante los buenos o deficientes artículos sobre Educación.

En el caso de los sistemas de evaluación, no se puede dejar de considerar su fuerte potencial periodístico. En sí mismos, ellos son todo lo que los lectores, y por lo tanto, los periodistas, más quieren: contienen información fáctica que puede ser traducida a números (o viceversa), garantizan titulares fuertes y títulos espectaculares, llenan páginas de gráficos coloridos, dan lugar a discusiones y venden diarios. Se debe evitar considerar estas características, que son inherentes al periodismo, y, por lo tanto, difícilmente modificables, como perniciosas. Más bien se las debe explorar. O sea, a partir de ellas, es posible convencer al empresario de los medios que es una agenda que vale la pena explorar, y que, aún más, realizarla de forma competente es asegurar su posición como líder de audiencia y referencia para la opinión pública.

¿Cómo hacerlo?

Llegamos, tal vez, el punto más difícil. Si es fácil identificar dónde están los problemas, menos simple es encontrarles una salida. No es que ellas no existan, pero, ante la complejidad del mercado periodístico, donde cada empresa tiene su realidad, es un poco difícil definir una fórmula, así como no es posible establecer reglas generales para el desarrollo de la

Educación en América Latina. Pero se pueden establecer algunas recomendaciones que intentaremos sean lo menos genéricas posible.

A las fuentes, especialistas y escuelas de comunicación

- Enterarse sobre la lógica de funcionamiento y la línea editorial de los medios de prensa de la región, para prever sus reacciones, expectativas y el tipo de relacionamiento que es posible establecer.
- Tener un buen equipo de prensa capaz de preparar un material con las siguientes características: conciso, claro, sin especulaciones y sin aparente manipulación de los significados. El objetivo es que el periodista pueda confiar en los datos y superar el temor de estar siendo manipulado.
- Convocar ruedas de prensa en un horario que sea conveniente para todos los medios, pero que deje a los periodistas tiempo hábil para revisar los datos. Cuando los periodistas tienen la impresión de que todo el arreglo está siendo hecho para manipular el resultado de su producción, ya se predisponen. Si es posible, poner a disposición los datos con cierta anticipación para sectores interesados tales como consejos de educación, asociaciones de padres y maestros y especialistas de la universidad, para que la repercusión sea de calidad y no solamente basada en impresiones subjetivas (del tipo “me parece que”).
- Nunca ocultar datos a los periodistas, ni siquiera bajo el alegato de que ellos no los podrán comprender. Al contrario, ofrecerse para aclarar los datos o informar cuáles serían los especialistas –no vinculados a los organizadores de la evaluación– capaces de comprender y traducir los datos.
- Intentar establecer una relación ética y respetuosa con posibles periodistas especializados, realizando, de esa manera, una educación continua de tales profesionales. Tener presente, sin embargo, que esto no significa “fidelidad” por parte del periodista. Su función seguirá siendo reportear y denunciar.
- Nunca dejar de devolver los llamados o pedidos de entrevista de los periodistas, por más que se sepa de antemano que la tarea será espionosa.
- Promover, en conjunto con asociaciones de periodistas, asociaciones de dueños de empresas periodísticas, universidades y organizaciones no gubernamentales, programas de capacitación adaptados a las nece-

sidades y la realidad de la profesión. Seminarios prácticos, cortos y puntuales.

- Ocupar espacios editoriales, escribir artículos, manifestarse sobre una buena o mala cobertura de manera constructiva, apuntando las fallas y sus consecuencias, no para el gobierno, sino para el desarrollo de la educación.
- Sin proselitismo, dejar claro dónde están las fallas del sistema de evaluación (si existen), para qué cosas el sistema no sirve o qué tipo de interpretaciones no son válidas, qué otras formas de evaluación complementarias serían necesarias para trazar un perfil más adecuado de los sistemas de educación, etc.
- Nunca intentar interferir directamente en el texto o material producido por el reportero. Nunca pedir para leer, revisar, etc., el material antes de la publicación. Al contrario, colocarse a disposición para cualquier aclaración en el momento en que el reportero esté escribiendo el texto. Esto significa prepararse para dejar por lo menos un técnico libre durante todo el día para posibles nuevas consultas.

A los periodistas

- Si la agenda de la conferencia de prensa está preprogramada, intentar prepararse para la cobertura: leer resultados anteriores, informarse con un especialista independiente sobre los sistemas de evaluación, lo que mide, cómo mide, lo que no mide, dónde están sus problemas, etc. Intentar descubrir qué datos han sido levantados por el sistema y no serán divulgados pero son importantes como contexto para el análisis.
- En el momento de la presentación, asegurarse de que está entendiendo los datos, la forma de presentación, lo que significan. No hacer “traducciones” por cuenta propia, sino más bien confrontar sus interpretaciones con los especialistas.
- No contentarse con la fuente ministerial, aun cuando el plazo sea exiguo. Al mismo tiempo, asegurarse de que la nueva fuente buscada tiene competencia para hacer los comentarios, sin limitarse al “me parece que”.
- Escarbar más allá de los números y las medias. Lo que los datos realmente demuestran y lo que esconden. Cómo se traducen en la realidad de las escuelas y cómo impactan en la vida de los padres y madres y en la vida del país. Cómo esos datos impactan en la economía y en la política.

- Sugerir a los reporteros de investigación y a los redactores de opinión y editoriales detenerse sobre el tema para un estudio más profundo. Aprovechar fechas simbólicas para explorarlos mejor. Por ejemplo: el Día del Estudiante, la Semana del Profesor, o, en el propio período de divulgación de nuevos resultados, es posible desmenuzar mejor lo que fueron los resultados de los años anteriores.
- Mantener contacto permanente con las fuentes, insistiendo en primicias periodísticas, mostrando que una primicia puede ser bien hecha y contribuir a la divulgación más crítica y cualitativa del tema.
- Buscar capacitación si percibe que no domina términos técnicos de estadística, de preferencia con especialistas independientes.
- Enterarse sobre los sistemas de evaluación de otros países, estados o ciudades, para confrontar el sistema de evaluación de su localidad y asegurarse de que sea lo más adecuado para un buen retrato de la educación.
- Recordar que la evaluación no es un fin en sí misma. Se supone que se evalúa para tomar alguna medida o para mejorar algo. Pedir cuenta de las acciones y solicitar metas. Acordarse de volver a ellas cuando el plazo de las metas esté venciendo. Al mismo tiempo, tener en mente que la educación es un proceso que no se resuelve de la noche a la mañana.

CAPÍTULO IV

MODELOS DE BUENOS REPORTES DE PRENSA

El presente capítulo está destinado a proponer diferentes modelos de buenos reportes de prensa sobre resultados de evaluaciones. El propósito es ilustrar las diversas posibilidades de elaboración de informes y notas periodísticas en relación a las evaluaciones de logros educativos y sus resultados, que contribuyan a la información ciudadana y al debate público informado acerca de la educación.

Los ejemplos han sido seleccionados de modo tal que reflejen una diversidad de posibilidades y tipos de artículos. En un caso se trata de una noticia “del día” sobre los resultados de PISA, en otros de notas de tipo “periodismo investigación” realizadas con más tiempo y dedicación, en otro caso, una nota de opinión y, finalmente, se incluyó también una columna elaborada por una ONG mexicana –que se publica regularmente

en la prensa y que ilustra otro modo en que los medios de prensa pueden buscar informar mejor sobre las evaluaciones educativas—.

Ejemplo 1

La cobertura de PISA de “Los Angeles Times”

La nota incluida en el Recuadro 12 es una nota informativa del día aunque probablemente preparada con alguna anticipación. Recoge los resultados de PISA en el momento en que fueron dados a conocer al público.

Se caracteriza claramente por el tono sobrio. No recurre a ningún tipo de efecto sensacionalista ni de creación de alarma pública. No obstante ello, destaca que los resultados se encuentran en torno a la media de la OCDE y que, según el propio Secretario de Educación, eso es malo, porque implica estar por debajo de los países que son los principales competidores de los Estados Unidos en la economía internacional.

El texto incluye de manera clara y concisa, abundante información descriptiva de lo que es PISA, los países que participaron, el tipo de competencias que fueron evaluadas, los países que se destacaron por sus resultados.

Incluye además menciones breves y oportunas a las manifestaciones de tres personalidades diferentes: el Secretario de Educación, un funcionario oficial de la OCDE y un especialista universitario independiente.

RECUADRO 12

UNA BUENA NOTA INFORMATIVA “EN CALIENTE”

“Estudiantes norteamericanos al mismo nivel que países equivalentes”

Por Martha Groves, redactora de Times Education

Según los resultados de una prueba, “los estudiantes norteamericanos de quince años, comparados con sus pares de otros países altamente industrializados respecto a su capacidad de aplicar capacidades lectoras, matemáticas y científicas a situaciones de la vida real, se desempeñan a un nivel promedio.

Los hallazgos, publicados el lunes, se basaron en el desempeño de los estudiantes de 32 de los países más industrializados. Ellos rindieron el examen, el primero de su tipo, en la primavera del 2000.

Los resultados coinciden con otros datos de pruebas recientes que indican que el desempeño en matemáticas y ciencias de los estudiantes norteamericanos va cayendo por debajo del de sus pares de muchos otros países a medida que se hacen mayores. La noticia fue tomada con consternación por el principal funcionario de educación. “Desgraciadamente,

estamos en el promedio en con otras naciones industrializadas”, dijo el Secretario de Educación Rod Paige. “En la economía global, estos países son nuestros competidores-el promedio no es suficiente”.

La prueba, llamada Programa para la Evaluación Internacional de los Estudiantes, es auspiciada por la Organización para la Cooperación y el Desarrollo Económico, una organización de 30 países desarrollados que incluye al Reino Unido, Japón, Corea del Sur y Australia, y que está basada en París. La prueba fue administrada por el National Center for Education Statistics del Departamento Norteamericano de Educación.

Los países desarrollaron la prueba para crear una medición periódica y confiable que ayudase a orientar políticas educacionales.

Estados Unidos fue uno de los 20 países que obtuvieron puntajes “promedio” en capacidad lectora.

Finlandia, Canadá y Nueva Zelanda obtuvieron puntajes bastante más altos que el promedio norteamericano.

Muy por debajo del promedio de Estados Unidos quedaron Grecia, Portugal Luxemburgo y México.

En matemáticas y ciencias, Estados Unidos también obtuvo un puntaje promedio. En matemáticas, ocho países, incluyendo a Japón y Corea del Sur, obtuvieron puntajes bastante más altos que el promedio de Estados Unidos y cinco países quedaron muy por debajo de éste. En ciencias, siete países obtuvieron puntajes bastante más altos que el promedio de Estados Unidos y cuatro países quedaron por debajo.

Funcionarios de educación dijeron que los resultados de Estados Unidos no podían atribuirse únicamente a la diversidad de la población evaluada y sus capacidades lingüísticas.

Finlandia, Canadá y Nueva Zelanda fueron los únicos países que demostraron desempeños por encima del promedio en las tres áreas

La prueba se rendirá cada tres años. Cada ciclo se concentrará en un área temática particular, pero se evaluará a tres cada vez. La lectura fue el foco principal de la versión del 2000; matemáticas será el tema principal en el 2003, ciencias en el 2006.

Los ítems de lectura están elaborados para medir la capacidad que tienen los estudiantes que se acercan al final de su educación escolar formal obligatoria de comprender material que encontrarán en su vida, tal como formularios de bancos, documentos gubernamentales, mapas y artículos periodísticos.

En un ítem se le pide a los estudiantes que elijan una respuesta que mejor identifique el propósito de una carta pugnaz al editor. También se les pide que escriban un texto sobre qué tipo de respuesta esperaba generar el autor de la carta.

En otro, ellos deben analizar las horas en que funcionan diversas bibliotecas, tal como se desprende de un marcador de páginas. Otro pide a los estudiantes completar una ficha de garantía utilizando un recibo de una cámara nueva. Los ítems van de sencillos a complejos.

Los desgloses indicaron que a los estudiantes norteamericanos les iba mejor reflexionando sobre el significado de los textos que encontrando información relevante o interpretando textos.

Un funcionario de la Organización para la Cooperación y el Desarrollo Económico dijo que los resultados ofrecen claras implicancias de políticas para Estados Unidos. “El motivo por el que Estados Unidos en promedio tenga logros promedio, es que a muchos les va mal” dijo Barry McGaw, el director adjunto de educación de la organización. “Lo que Estados Unidos tiene que hacer es levantar el rendimiento de los que están peor. No tiene que sacrificar la calidad para lograr la equidad”.

Otros países, señaló, han demostrado que se puede superar los efectos negativos de un entorno social difícil.

Muchos educadores notaron similitudes entre estos resultados y los del Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS), una mirada continua al desempeño en matemáticas y ciencias de los países industrializados y en vías de desarrollo.

Un desempeño promedio de los estudiantes norteamericanos “no es sorprendente, dado... el hecho de que Estados Unidos se encuentra en el medio de los resultados de octavo grado (en el estudio de matemáticas y ciencias), que serían los más parecidos a estos”, dijo James Hiebert, un profesor de educación matemática de la Universidad de Delaware. A diferencia de esta prueba orientada a la vida real, la prueba de matemáticas y ciencias para los estudiantes de octavo grado explora qué tan bien les va a los estudiantes en los ítemes relacionados con los currículos escolares.

Hiebert encontró extraño que Japón y Corea del Sur, que a menudo salen mejor parados que Estados Unidos en pruebas internacionales, también estuviesen en el bloque “promedio” en lectura.

Entre otros hallazgos importantes:

- En promedio, 10% de los estudiantes de quince años de los países más desarrollados del mundo tienen capacidad lectora excelente y son capaces de comprender textos complejos, evaluar información y elaborar hipótesis. En Australia, Canadá, Finlandia, Nueva Zelanda y el Reino Unido, el porcentaje va de 15% a 19%. En Estados Unidos es 12%.
- Un promedio de 6% de los estudiantes –y en algunos países más del doble de ese porcentaje– están por debajo del nivel más bajo de proficiencia en lectura. Eso significa que no podían ubicar un fragmento sencillo de información o identificar el tema central de un texto. El porcentaje en Estados Unidos es de 6%.
- Japón y Corea del Sur son los que mejor se desempeñan en matemáticas y lectura.
- Algunos países-particularmente Finlandia, Japón y Corea del Sur-mantienen una brecha relativamente reducida entre quienes se desempeñan mejor y peor, a la vez que alcanzan niveles promedio altos.
- En muchos países, los varones están cayendo más en lectura. En todos los países participantes, las mujeres son, en promedio, mejores lectoras que los varones.
- En casi la mitad de los países a los varones les va mejor en matemáticas que a las mujeres. En la mayoría del resto, están casi parejos.

En Estados Unidos, 7,000 estudiantes de 9° y 10° grados participaron en la prueba. Cada uno de ellos rindió una prueba de 90 minutos y resolvió un cuestionario de 25 minutos. Funcionarios de educación señalaron que los estudiantes constituyeron una muestra bastante representativa. En la página web <http://www.nisa.oecd.org> están disponibles ejemplos de los ítemes.

Fuente: Los Ángeles Times - Estados Unidos - 4/12/2001. Traducido por los editores.

Ejemplo 2

Una reflexión sobre PISA en Brasil

El segundo ejemplo, incluido en el Recuadro 13, es un artículo de opinión, de tipo reflexivo, escrito por un académico brasileño. El medio de prensa abre su espacio para una reflexión equilibrada, lo cual es sumamente importante en el contexto de los titulares que vimos anteriormente para PISA en Brasil.

El artículo no rehuye al reconocimiento de los graves problemas educativos que los datos muestran y reconoce, al mismo tiempo, el valor y la insuficiencia de los esfuerzos realizados a través de diversos programas de inversión para la mejora educativa.

Es un buen ejemplo de análisis equilibrado que, si bien señala carencias muy importantes, evita ingresar en el tono catástrofe o en la búsqueda simplista de culpables. Más bien hace un esfuerzo por señalar la complejidad del esfuerzo por mejorar la educación. Señala explícitamente que no hay soluciones mágicas y reconoce la importancia que tiene contar con información sistemática sobre la situación, como uno de los requisitos indispensables –aunque en sí mismo insuficiente– para ir encontrando caminos para enfrentar los desafíos.

RECUADRO 13

UNA COLUMNA DE OPINIÓN DE CARÁCTER REFLEXIVO

La lección de Pisa

Simón Schwartzman

“Los resultados de Brasil en el Programa Internacional de Evaluación de Estudiantes (Pisa), último entre 38 países, justo debajo de México, no llegan a sorprendernos, pero ayudan a recolocar en agenda la cuestión de la pésima calidad de nuestra educación. Los detalles de esta historia ya eran conocidos, y están analizados de forma pormenorizada en el informe del Sistema de Evaluación de la Educación Básica para 1999, realizado y publicado por el INEP. En portugués, en cuarto grado de educación básica, todos los brasileros se encuentran, en promedio, en el nivel inicial de aprendizaje, que va de 150 a 200 puntos, en los cuales los estudiantes son capaces de “lectura con comprensión localizada de textos pequeños, con frases cortas en orden directo, vocabulario y temática próximos a la realidad del alumno”. En esta escala, los promedios de los estados del sureste y del sur se aproximan a 180; y los estados del noreste están debajo de 160 (la media brasilera es 170).

Esto significa que casi la mitad de los estudiantes de cuarto grado del noreste y gran parte de los estudiantes del centro-sur están por debajo de este nivel mínimo. Es decir, son analfabetos funcionales. Lo que es peor: entre 1997 y 1999, la media nacional parece haber disminuido, sobre todo en Minas Gerais, que ya tenía el nivel más alto del país. Nadie sabe exactamente lo que sucedió, la caída puede haberse debido a cambios de la misma prueba, a la mala implementación de la política para combatir la repitencia o, quién sabe, a cierto desánimo nacional que se coló en nuestras escuelas. Pronto sabremos lo que nos dirá el SAEB del 2001. ¿Qué se puede hacer? Las respuestas usuales ya las conocemos: invertir más en educación, pagar mejor a los maestros. Pero nuestro problema más serio no es que gastemos poco, o que nuestros maestros ganen mal: el problema más serio es que la enseñanza continúa siendo mala cuando gastamos más, y cuando los maestros están más formados. Los datos muestran que tampoco gastamos tan poco, y el nivel de los maestros está mejorando. Según la OCDE, Brasil ya gasta cerca de 17% de los recursos públicos en educación, en comparación con el 13% a 14% que gastan Canadá y Francia, y el 16% que gasta Chile. Puede ser que estos datos estén sobreestimados, pero muestran que difícilmente podríamos cambiar el nivel de gasto sin un crecimiento significativo de la economía.

En cuanto a los maestros, 30% de los que enseñan hasta cuarto grado, y 40% a 50% en las regiones Sureste y Sur, tienen educación superior, y reciben salarios compatibles con este nivel para el mercado brasilero. Es cierto que la situación de una escuela en el estado de São

Paulo o en Río Grande do Sul es mucho mejor que la de una escuela del interior de Bahía o de Marañón. Pero ni siquiera así la educación recibida por los estudiantes de Sao Paulo es tanto mejor. Y eso que estamos hablando sólo de la educación más elemental. La situación de la enseñanza media, con casi la totalidad de los alumnos en cursos nocturnos siguiendo currículos obsoletos, es todavía más seria. Por dar un ejemplo extremo, el desempeño de los alumnos de 3er. grado de enseñanza media en portugués en Sao Paulo bajó de 305 a 267 entre 1997 y 1999 (la media brasilera bajó de 290 a 266 en el mismo período).

Tanto el gobierno federal como varios gobiernos estatales han tratado de enfrentar este problema con una serie de programas de estilo y costos variados. Al nivel federal, la lista de proyectos incluye, además de sistemas de evaluación como el SAEB y el ENEM, el Programa Nacional del Libro Didáctico, de distribución de libros; el Proyecto Fundesco-la, de obras y equipos para las regiones más pobres; el Programa Nacional Bibliotecas en la Escuela; el ProInfo, de introducción de tecnologías de información en las escuelas; la TV Escola; el Fondo de Mantenimiento y Desarrollo de la Enseñanza Básica y de Revaloración del Magisterio; los Parámetros Curriculares Nacionales para la Enseñanza Básica y Media; el Programa de Aceleración del Aprendizaje; el Programa Dinero Directo en la Escuela; y el nuevo programa Bolsa-Escuela. Sería absurdo e injusto concluir que, como la educación no está mejorando, estos programas son inútiles y deberían abandonarse. En muchos casos, ellos tienen un tiempo de maduración prolongado, y están dirigidos por equipos competentes y dedicados. Pero también es ingenuo creer que van a lograr todo aquello que se prometió y pensó cuando se iniciaron.

El hecho es que no existe una receta mágica para los problemas de la educación básica. Diez años atrás sabíamos poco sobre lo que estaba sucediendo, y todos decían que era preciso construir escuelas, educar mejor a los maestros y acabar con la deserción. Hoy, gracias sobre todo a la labor del INEP, ya tenemos excelentes datos, casi todos los niños asisten ya a la escuela, los profesores se gradúan de las universidades. Ahora descubrimos que los chicos no aprenden, y tratamos de crear programas de lo más variados para corregir los problemas que vamos encontrando. Pero todavía no aprendemos a evaluar los programas existentes, conocemos poco de las experiencias de otros países, así como de los buenos ejemplos que existen en nuestro medio. Falta una cultura moderna y actualizada de estudios e investigaciones educacionales, que pueda ir dando respuestas y señalando caminos en esta gran jornada que tenemos por delante».

Fuente: O GLOBO - Brasil - 12/12/2001. Traducido por los editores.

Ejemplo 3

Una nota de investigación sobre el SIMCE en Chile

El tercer modelo que hemos seleccionado para este capítulo de buenas prácticas es un trabajo realizado por la revista “Qué Pasa” de Chile. En este caso se trata de una investigación sobre el propio sistema de evaluación, que incluye un análisis tanto de los resultados como del tipo de actividades que los alumnos deben realizar, de los rankings que el sistema produce y de los efectos que el sistema tiene sobre los colegios. El trabajo se apoya fuertemente en entrevistas, tanto con especialistas externos al Ministerio como con funcionarios del mismo.

RECUADRO 14

UNA NOTA DE INVESTIGACIÓN SOBRE EL SISTEMA DE EVALUACIÓN

El ABC del SIMCE

Loreto Aravena | Elizabeth Simonsen

Cada vez que se publican los resultados del Sistema de Medición de la Calidad de la Educación (Simce), se desata una fiebre colectiva. El país queda dividido entre los colegios “top” y los otros y los resultados pueden llegar incluso a provocar renunciaciones, como la que presentó en julio pasado el coordinador nacional de la prueba, Cristián Cox, ante los malos resultados en el test de matemáticas de cuarto básico.

Cox sigue en el cargo, pero su actitud refleja que el Simce ha cobrado una importancia quizás no imaginada cuando en 1988 se instauró para evaluar la educación chilena.

Sus predecesoras fueron la Prueba Nacional (aplicada entre 1968 y 1971), dirigida a los octavos básicos. En 1982 se creó el Programa de Evaluación del Rendimiento Escolar (PER) para los cuartos y octavos básicos. Era una fórmula administrada por un organismo autónomo —la Universidad Católica— cuyos resultados nunca vieron la luz, porque el ministerio consideró que era como ponerle nota a los profesores.

La idea de mantener los resultados en secreto permitía evitar vicios que empezaron a aparecer alrededor del Simce (ver recuadro).

Qué Pasa quiso ir más allá de los resultados globales. Por ello, inicia una serie de reportajes que consideran tanto el contenido del test, como sus proyecciones sobre el nivel de la educación chilena. En esta primera parte, se muestra el nivel real de conocimientos de los alumnos, lo que mide el Simce y si sus resultados son comparables con el otro indicador de la enseñanza: la PAA.

Los tres vicios

La prueba Simce ya muestra ciertas deformaciones que la apartan de su objetivo inicial. Para Jorge Manzi se ha desvirtuado en tres aspectos.

1) Hay colegios que preparan exhaustivamente para el examen a la generación que debe rendir el test, lo que a juicio de Manzi, “genera una especie de acostumbramiento en los niños”. El propio Mineduc envía folletos con preguntas tipo, que sirven de base para que cada establecimiento prepare la prueba.

Estudios norteamericanos dicen que el entrenamiento en una prueba puede conducir a un alza en el rendimiento, que no corresponde necesariamente a lo que ha aprendido el alumno. Esto jugó en contra de los estudiantes en 1999, cuando rindieron un examen con estructura y formato diferentes (ninguno de los niños estaba acostumbrado al prototipo de preguntas abiertas).

2) El incentivo económico que reciben los docentes de un establecimiento educacional por medio de bonos, también distorsiona el sistema. “Como el dinero se reparte entre todo el profesorado, puede suceder que los pedagogos adopten actitudes poco éticas, como aquellas constatadas en un establecimiento de Mulchén (donde un docente ayudó a los alumnos a responder la prueba); o bien, la dirección del establecimiento puede solicitar a aquellos niños cuyo rendimiento no es óptimo, que no asistan a rendir la prueba ese día”, enfatiza Manzi.

3) La calificación de los colegios en un ranking es otro “vicio” del Simce, pues se ha transformado en una especie de “lista de ofertas”. Muchos padres esperan ver los resultados del test nacional para tomar una decisión acertada a la hora de matricular a sus hijos. “Esto ha despertado un conflicto de autoestima entre los diversos establecimientos del país, ya que solo un 10% de ellos logra buenos puntajes. El resto son escuelas pobres, que no tienen los medios para competir con centros educacionales mejor dotados”, dice Manzi.

La prueba es fácil

Las investigadoras del Centro de Estudios Públicos (CEP), Bárbara Eyzaguirre y Loreto Fontaine, analizaron en un exhaustivo estudio las pruebas Simce aplicadas a los segundos años medios y octavos básicos en 1998 y 1997, respectivamente. Y concluyeron que tenían una complejidad equivalente a dos y tres niveles menos que el curso al que se aplicaba la prueba.

Por ejemplo, el 78% de los contenidos de lenguaje del test aplicado a 8° básico pertenecía a los currículos de los cursos de 3° a 6° básico. No se medía ninguno de los temas más difíciles que se enseñan en 7° y 8°, y que reflejan un total dominio del tema. Lo mismo sucedía con la sección de matemáticas: las materias evaluadas correspondían a 6° básico en el antiguo currículo y a un 5° básico avanzado, si se toman en cuenta los nuevos objetivos de la reforma.

En la prueba aplicada el año siguiente a segundo medio, se notó un fuerte cambio, pues incorporaba algunas preguntas acerca de contenidos de la reforma, además de aquellas materias antiguas que no eran consideradas obsoletas.

Sin embargo y pese al avance, las investigadoras del CEP detectaron que el nivel medido en aquella oportunidad igual fue más bajo que los contenidos mínimos que exige el Mineduc.

Sucede lo mismo con la prueba aplicada el año pasado a los 4° básicos. Aunque incorpora definitivamente los nuevos contenidos curriculares, la lectura que puede realizarse a partir de los resultados deja en bastante mal pie a la educación chilena: “El 50% de los niños no está cumpliendo los objetivos mínimos exigidos para 4° año básico en lenguaje, y en matemáticas, el 63% no los logra”, sentencia Eyzaguirre.

Según el informe del Mineduc, un 19% de los niños logró un desempeño considerado deficiente en lenguaje y comunicación y un 31% apenas alcanzó el nivel básico. En rendimiento deficiente caen los que respondieron menos del 65% de las preguntas básicas, que no requerían más esfuerzos que el simple hecho de identificar la información explícita en los textos o diferenciar si un escrito era noticia, cuento, poema, receta o aviso publicitario (pregunta sobre “aventuras del lápiz” en página 46).

En el nivel básico se sitúan los niños que respondieron menos del 65% de las preguntas de la categoría inmediatamente superior, esto es, de desempeño “intermedio”, que requiere, por ejemplo, inferir significados implícitos en textos breves, utilizar sinónimos y reconocer la función que cumplen algunas palabras. Todas estas habilidades se deberían alcanzar como mínimo en 4° básico, según lo que incluyen los contenidos mínimos y objetivos fundamentales establecidos por el Ministerio de Educación.

En matemáticas, la situación es incluso peor: el 32% de los niños está por debajo del nivel básico y el 31% por debajo del intermedio. Ellos no alcanzan objetivos como identificar cuerpos geométricos, utilizar una regla graduada para medir longitudes o resolver problemas cotidianos con sumas, restas, multiplicación y división con números naturales y divisor de una cifra.

“Ningún colegio ni profesor puede sentirse satisfecho con estos resultados, ni siquiera los que alcanzaron los máximos rendimientos, esto es, sobre 290 puntos”, dice Eyzaguirre.

Las respuestas

El tipo de preguntas de las pruebas de 2° medio (1998) y de cuarto básico (1999) son un indicio claro de que el nivel de exigencia planteado a los alumnos es bajo, pero el porcentaje de respuestas correctas no es el esperado.

En el test de matemáticas de 2° medio, la pregunta número 10 (ver figura 5) interroga a los jóvenes sobre cuánto mide el largo del quiosco. La respuesta implica medir cuántas veces cabe el gato en el techo y saber que 100 centímetros equivalen a 1 metro, materia propia de Educación Básica. Sin embargo, apenas el 40% de los estudiantes la respondió bien.

Otro ejemplo es el de Chile y sus medidas (figura 6), materia correspondiente al 8° año básico. Apenas el 31% de los estudiantes marcó la alternativa correcta.

En lenguaje, hay más casos. La pregunta 24 (Pablo Neruda, figura 2) tiene por objetivo reconocer el sentido implícito en el texto. Apenas el 40% de los alumnos respondió bien.

La situación se agrava si se consideran las preguntas abiertas, en las que los jóvenes deben redactar una respuesta a partir de una idea.

En la pregunta número 27 (figura 4) había que escribir un diálogo a partir del título “Sorprendente descubrimiento”. Hay respuestas que revelan alto nivel de creatividad, como “los dinosaurios no se extinguieron, sino que viven escondidos”, “el secreto de la eterna juventud” o “no soy hijo de mis padres”. Sin embargo, apenas el 16% contestó de esta forma y pudo redactar un diálogo coherente con al menos dos personajes distintos. El 35% escribió un diálogo sin personajes o bien los inventó, sin relacionarlo con el descubrimiento. Un 36% omitió la pregunta.

La pregunta referente a la librería El Aleph (figura 1), considerada de nivel alto, requiere de conocimientos de matemáticas, como multiplicar los precios por las unidades adquiridas, sumar el total de la compra y luego restar del dinero disponible para obtener el vuelto. A juicio de Eyzaguirre, se trata de una pregunta adecuada para 4° básico, pero no de las más difíciles, pues la multiplicación se podría fácilmente reemplazar por la suma de los artículos comprados. Más de la mitad de los niños (60%) no la supo responder.

Qué dice el Mineduc

En teoría, la prueba debe incluir algunas preguntas de cursos inferiores para medir si los contenidos aprendidos anteriormente se mantienen en el tiempo. Lorena Meckes, vicecoordinadora del Simce, explica que en la nueva forma de evaluar la prueba –que se aplicó por primera vez el año pasado a los 4° años básicos–, las preguntas de cursos inferiores tienen menor puntaje.

- “Aun cuando se hayan incluido preguntas de primer año básico, esto no significa que la prueba sea más fácil. En las ocasiones anteriores, el puntaje dependía de cuántas preguntas del total habían sido bien respondidas”.

El punto en discusión, entonces, es cuántas preguntas “fáciles” incorpora la nueva prueba y si lo catalogado como “difícil” o de nivel alto corresponde realmente a los conocimientos esperados en el nivel en que se aplica. El Mineduc no dará a conocer esta vez en su totalidad la prueba a los colegios e investigadores y se guardará algunos ítems para aplicarlos en el futuro y así tener una vara de comparación. Por ello, solo se ha liberado una breve muestra de las preguntas.

En todo caso, el método para clasificar a los alumnos en niveles de desempeño “alto”, “intermedio”, “básico” o “deficiente” suscita varias dudas entre los investigadores. Meckes explica que los puntos de corte de las preguntas obedecen a que se tomó el total de estudiantes y se los dividió en cuatro categorías, según su rendimiento. De esta forma, se seleccionó el tipo de pregunta más contestada por cada nivel para establecer los puntos de corte.

El método tiene sus detractores. Según Eyzaguirre, el riesgo es que siempre la prueba se termine adecuando a lo que los alumnos saben, más que a lo se defina oficialmente como necesario.

- “No se saca nada con demostrar aumentos en los puntajes si la prueba (de 8° año) en su mayor parte no sobrepasa el nivel de 6° básico”.

El psicólogo Jorge Manzi, de la Universidad Católica, especialista en temas de educación (el mismo que fue coordinador de la mesa de diálogo), opina parecido: “El Simce es un instrumento de monitoreo que sirve poco, porque no refleja el rendimiento integral de los estudiantes. No se puede constatar cómo reflexionan, cómo se comportan con su entorno social (si son solidarios, si respetan la naturaleza, cómo se desenvuelven como ciudadanos, entre otros). Solo mide las mejorías o retrocesos en materias puntuales”.

El Simce y la PAA

Es por esas características técnicas que los resultados del Simce no son comparables con los de otras evaluaciones como la Prueba de Aptitud Académica (PAA). Es más, ni siquiera el Simce es comparable con su propio pasado, pues hasta ahora, sus preguntas –y por ende, su nivel de dificultad– varían año tras año.

- “El Simce y la PAA tienen objetivos distintos”, explica Cristián Bellej, asesor del Ministerio de Educación. “El primero mide la calidad de la educación y la PAA mide habilidades. El Simce toma una fotografía de la educación en los colegios, y la PAA busca seleccionar y ordenar a los mejores alumnos para el ingreso a la universidad”.

Eyzaguirre sostiene que dado el bajo nivel de exigencia del Simce, un establecimiento con buen puntaje, no garantiza que sus alumnos tendrán el mismo nivel en la PAA. No existe en el país ningún estudio que permita confirmar esas afirmaciones, pero según un análisis hecho por *Qué Pasa* a partir de los rankings de los colegios con máximos puntajes en ambas mediciones, la relación entre ambos es relativa. Ninguno de los colegios que tuvieron alumnos con más de 800 puntos en la PAA figura entre los 10 mejores del Simce de 4° básico del mismo año, pero sí obtuvieron promedios que los sitúan en el quintil más alto del Simce (270 y 348 puntos).

Por ejemplo, el Instituto Alonso de Ercilla, con 15 alumnos entre los puntajes más altos de la PAA, no está dentro de los 10 primeros del Simce, pero tiene promedios sobre 300 puntos en sus tres áreas: lenguaje, matemáticas y comprensión del medio.

La situación es distinta en los segundos medios. En este caso, existe una evidente relación de puntajes: los colegios Alonso de Ercilla, Alemán de Santiago y La Abadía figuran entre los 10 mejores del Simce y entre los mejores de la PAA.

Uno de los aspectos más criticados con respecto a los rankings que se realizan a partir del Simce, es la variable en la cual se basan: el promedio del establecimiento en cada una de las materias evaluadas. Como se trata únicamente de la media, el que un niño ingrese a un colegio “top” no asegura su éxito académico. De hecho, en el test de matemáticas para 4° básico, el 12,8% de los alumnos de los colegios con mejores puntajes (más de 270) está en nivel deficiente, y el 22,4% en nivel básico. Aún más, solo alcanzó desempeño “alto” el 29,4% de los niños, lo suficiente como para favorecer su ubicación en el ranking.

Por lo demás, hay múltiples factores externos que condicionan el rendimiento de un alumno en el Simce, y que van mucho más allá del lugar que un establecimiento ocupe en el ranking, como por ejemplo, el nivel de educación alcanzado por la madre.

Fuente: Revista *Qué Pasa* - Chile -27/8/2000

Para ver el artículo original, ir a:

<http://www.quepasa.cl/revista/2000/08/27/t-27.08.QP.TM.SIMCE.html>

Según se puede apreciar, el trabajo refleja una aguda mirada sobre el SIMCE y sus diferentes aristas, discutiendo aspectos clave de la política de evaluación con el apoyo de múltiples puntos de vista. Describe lo que se requiere a los alumnos en las pruebas a través de ejemplos y figuras –que no están incluidas aquí por razones de espacio–. Además, pone en el tapete un aspecto clave de la política de evaluación: si los niveles de desempeño de los alumnos se establecen en función de lo que la sociedad y el Ministerio de Educación esperan que aquellos sean capaces de hacer, o si por el contrario se establecen en función de la distribución real de las capacidades de los alumnos (*“según Eyzaguirre, el riesgo es que siempre la prueba se termine adecuando a lo que los alumnos saben, más que a lo se defina oficialmente como necesario”*).

Ejemplo 4

Una nota de investigación sobre el aprendizaje de las ciencias en los Estados Unidos

Otro ejemplo de periodismo de investigación y reflexión está referido al debate sobre el aprendizaje de las Ciencias en los Estados Unidos. Se trata de un artículo del *The Washington Times* que plantea la insuficiente habilidad de los estudiantes norteamericanos en ciencias y matemática, principalmente por su limitada capacidad para utilizar los conocimientos en la resolución de problemas nuevos. Esto obliga a las empresas a recurrir a técnicos del extranjero.

Nuevamente, sin ingresar en un tono sensacionalista²⁰, el artículo plantea el problema, resume los datos de diversas evaluaciones y contrasta opiniones disímiles acerca de los modos de enfrentar el problema, como por ejemplo las del economista Milton Friedman y las del presidente de la Asociación Nacional de Profesores de Ciencias, Harold Pratt.

²⁰ A pesar de que el título sí lo es en cierta medida.

RECUADRO 15
DIVERSOS PUNTOS DE VISTA SOBRE EL PROBLEMA DE LA
ENSEÑANZA DE LAS CIENCIAS EN LOS ESTADOS UNIDOS

¿Burros en ciencias? - los bajos puntajes de los estudiantes norteamericanos resultan preocupantes

Por Cheryl Wetzstein

“Las compañías norteamericanas necesitan trabajadores que no solo tengan conocimientos factuales y teóricos sobre ciencias y matemáticas, sino la capacidad para aplicar esos conocimientos científicos a nuevas situaciones”, señaló Edward Donley, ex presidente de la Compañía Air Products and Chemicals.

Sucede simplemente que no hay suficientes personas egresadas de las escuelas secundarias norteamericanas que tengan esta capacidad de razonamiento y de aplicación, dijo, agregando que las compañías norteamericanas han estado importando tantos técnicos extranjeros que sus laboratorios parecen “una mini Naciones Unidas”.

Estudios recientes han confirmado los mediocres puntajes norteamericanos:

- El Tercer Estudio Internacional de Matemáticas y Ciencias, realizado en 1999, mostró que los estudiantes norteamericanos de octavo grado se ubicaban en el puesto 14 entre 38 países. Los puntajes de 1999 no habían casi cambiado desde 1995.
- La Prueba Nacional del Progreso Educacional (NAEP) masiva de carácter en ciencias, publicada en noviembre, demostró que 47 por ciento de los estudiantes de último año de secundaria no poseen capacidades científicas “básicas”.
- Otra prueba estudiantil internacional - el Programa para la Evaluación Internacional de los Estudiantes, desarrollada por la Organización para la Cooperación y el Desarrollo Económico (OCDE)- halló que los estudiantes norteamericanos de quince años tienen puntajes promedio en ciencias, ubicándose en la posición 14 entre 31 países.

“En la economía global, el promedio no es suficientemente bueno para los estudiantes norteamericanos”, dijo el Secretario de Educación Rod Paige cuando se publicaron los resultados de la OCDE el 4 de diciembre.

Los resultados de la NAEP “son preocupantes”, señaló el Sr. Donley, quien es miembro de la junta directiva de NAEP. Según NAEP, los puntajes en ciencias empeoran a medida que los estudiantes avanzan grados en la escuela pública: los estudiantes norteamericanos de cuarto grado clasifican en un puesto bastante alto en comparación a sus pares de otros países. Sin embargo, las notas en ciencias bajan al rango del promedio cuando llegan a octavo grado y bajan aun más, hasta quedar por debajo del promedio al llegar al 120 grado. Los puntajes NAEP muestran “que cuanto más tiempo tienen en la escuela los chicos, peor les va”, dijo Milton Friedman, el premio Nóbel de 89 años, economista y promotor de los vouchers *escolares*. Los puntajes en ciencias son “desastrosos”, pero no inesperados, dado el monopolio cuasi-soviético de la educación pública básica en los Estados Unidos, dijo recientemente el señor Friedman en una videoconferencia para periodistas. “Es fascinante notar que en lo que respecta a educación superior, los Estados Unidos son el número uno en el mundo”, dijo el señor Friedman. “Pero en lo que respecta a la educación básica, estamos muy abajo. La diferencia entre ambos es una sola palabra: *che/ce* [Friedman se refiere a la libre elección de la escuela - N. E.]”.

Indudablemente, las universidades norteamericanas fuertes en ciencia y tecnología son reconocidas mundialmente, según los expertos científicos y educativos consultados. Pero ya en algunas de estas escuelas, los estudiantes extranjeros son más que los norteamericanos, y al paso en que van los estudiantes estadounidenses, la brecha seguramente se agrandará.

La llegada de individuos altamente educados y calificados de todo el mundo ha beneficiado enormemente a los Estados Unidos, dijo Bruce Fuchs, director de la Oficina de Educación Científica de los Institutos Nacionales de Salud, quien recordó que el jefe de la Reserva Federal, Alan Greenspan, estuvo explicando al Congreso este hecho hace ya un año.

“Es una gran cosa, siempre y cuando el mundo quiera seguir enviándonos a sus estudiantes más brillantes y capaces”, dijo el señor Fuchs. Pero que ocurriría si algo interrumpe ese flujo, se preguntó. “Es esencial que de alguna manera aprendamos a no depender de que el mundo nos envíe a los mejores y mas capaces, y crear algunos de esos chicos aquí”.

La visión del Sr. Fuchs consiste en que las escuelas relajen sus intentos de que los estudiantes memoricen vocabularios y datos obtusos, en lugar de que tengan experiencias de aprendizaje significativas. “Lo que realmente importa en la ciencia... es que los chicos se entrenen en el pensamiento crítico que la ciencia enseña”, dijo. Los trabajos fabriles están migrando a otros lugares del mundo, y los puestos de trabajo en los EEUU van a requerir más competencias de sus trabajadores, dijo. “Los chicos de hoy tienen que ser pensadores y aprendices continuos; demostrar que han aprendido a aprender”.

Harold Pratt, presidente de la Asociación Nacional de Profesores de Ciencias, que tiene 53.000 miembros, concuerda en que logros pobres en ciencia en la secundaria pronostican desempeños pobres en el mercado de trabajo y “una escasez de recursos humanos técnicos”. Pero lo que falta es una voluntad nacional de dar prioridad a la ciencia, dijo. Por ejemplo, no es raro que en las secundarias norteamericanas los estudiantes “elijan” no llevar ciencias en los grados superiores, especialmente en el último año escolar, dijo el señor Pratt. Una razón para ello es que los estudiantes están buscando admisión universitaria apenas empiezan ese año escolar, reciben “admisión temprana” ya en diciembre “y entonces descansan el resto del año”, dijo.

Es también un hecho que los estudiantes norteamericanos tienen muchas carreras lucrativas entre las cuales escoger, tales como leyes, ventas, atención a la salud, y altas finanzas. Los puestos vinculados a las ciencias no pagan tan bien como estas otras carreras, mientras que “sí vives en otro país, ser un buen estudiante en matemáticas y ciencias es tu boleto de ingreso” a un puesto bien remunerado, dijo el señor Pratt.

Mas aún, los líderes norteamericanos no parecen valorar la ciencia, al menos en lo que respecta al financiamiento, dijo el señor Pratt. La ley de educación recientemente aprobada por el congreso, por ejemplo, enfatiza el dominio del lenguaje y, en menor grado, la matemática, dijo. Sin embargo, el dictamen final cerró un programa de 485 millones de dólares nombrado en honor al Presidente Eisenhower, que estaba dedicado al desarrollo profesional de profesores de matemáticas y ciencias. Esos fondos pasaron a engrosar un enorme fondo para reducir el tamaño de las clases, desarrollo de lecciones, educación alternativa y tutoría- “hay como 15 usos”, dijo Jodi Peterson, un analista del NSTA.

La nueva ley de educación, que se espera sea aprobada por el Presidente Bush en enero, contiene 12 millones y medio de dólares para un programa de alianzas para matemáticas y ciencias, pero ello es solo una fracción de los 450 millones que se habían solicitado originalmente. “A fin de cuentas, lo claro es que las ciencias y las matemáticas - especialmente las ciencias-tienen tercer o cuarto lugar en las prioridades”, dijo Gerry Wheeler, director ejecutivo del NSTA.

Otros observadores dicen que los puntajes en ciencias son solo un síntoma de problemas escolares mayores, y que sin una reforma escolar sería, la educación norteamericana continuará en declive. A menos que ocurra una real innovación en la educación, “creo que continuaremos simplemente arreglándonos como podamos, como hemos venido haciéndolo a lo largo de 30 años de resultados mediocres”, dijo Chester E. Finn, presidente de la Fundación Thomas B. Fordham”.

Fuente: The Washington Times - Estados Unidos - 27/12/2001 Traducido por los editores

Ejemplo 5

El Observatorio Ciudadano de la Educación en México

El quinto modelo que deseamos proponer ha sido incluido como Anexo en virtud de su extensión. En este caso se trata de un trabajo realizado por una ONU mexicana –El Observatorio Ciudadano de la Educación–, que publica regularmente una columna en el periódico *Jornada*.

La columna seleccionada refiere a los resultados de PISA. Es interesante el enfoque del trabajo, sobre todo si se tiene en cuenta el contexto de los titulares con que la prensa mexicana difundió los resultados de PISA.

Además de realizar un análisis reflexivo y balanceado, que expresa con claridad lo preocupante que los resultados son para México, el trabajo muestra cómo los resultados han impactado en otros países y, en particular, realiza una serie de análisis originales a partir de información sobre variables sociales que muestran que parte de los problemas, aunque por cierto no todos, obedecen a las importantes diferencias entre la realidad social mexicana y la realidad de la mayoría de los países miembros de la OCDE.

Este caso es además un interesante ejemplo de cooperación entre un medio de prensa y una ONG, que los beneficia mutuamente. Al medio, porque puede ofrecer a sus lectores un tipo de producto que probablemente le sería difícil producir con sus propios recursos, y a la ONG, porque le brinda un canal de llegada a la población con el que de otro modo le sería difícil contar.

Sugerencias para periodistas

A partir de los modelos analizados a lo largo del presente capítulo, así como de las reflexiones formuladas por Tacyana Arce en el capítulo anterior y, más en general, de las discusiones realizadas en el Seminario de Santo Domingo, intentaremos ahora formular un conjunto de sugerencias para los periodistas y medios de prensa que deseen llevar adelante una mejor cobertura de los resultados de las evaluaciones nacionales e internacionales de aprendizaje. Por cierto, no pretendemos nosotros orientar la labor periodística, simplemente aportar algunas reflexiones que puedan ayudar a realizar mejor la labor.

En primer término, interesa destacar cinco características principales que están presentes en los reportes de prensa propuestos como modelos a lo largo del presente capítulo:

- a. el **tono** empleado evita ingresar en el terreno de la alarma pública o el escándalo. Si bien todos o casi todos ellos ponen de manifiesto serios problemas en los sistemas educativos, lo hacen sin sensacionalismo;
- b. en todos o casi todos los casos se recurre a la **opinión experta de personas especializadas** en los temas tratados;
- c. se realiza un **tratamiento cuidadoso de la información**, evitando conclusiones rápidas y simplistas y aportando al lector información sobre las características y contexto de las evaluaciones, así como sobre **lo que fue evaluado y lo que los alumnos debieron ser capaces de hacer**;
- d. **se confrontan puntos de vista diferentes** respecto a los problemas educativos detectados y a los caminos para enfrentarlos, o respecto al propio sistema de evaluación y a la interpretación de sus resultados;
- e. se busca **relacionar las evaluaciones** y sus resultados **con otros aspectos de la política educativa** y con las acciones necesarias para enfrentar los problemas que los resultados ponen de manifiesto.

Además de las características que se acaba de mencionar, siguen a continuación otras sugerencias que están agrupadas en tres grandes ejes:

- i. un conjunto de preguntas clave que el periodista debería formular e indagar en torno a las evaluaciones;
- ii. un conjunto de conocimientos y saberes de distinto tipo que debería ir acumulando;
- iii. un conjunto de otros recursos y pistas para la elaboración de notas periodísticas sobre resultados de las evaluaciones.

i. Preguntas clave

- ¿Cuáles fueron los propósitos de la evaluación? ¿Para qué se realizó? ¿Cuál es la estrategia global de evaluación que está siendo seguida?
- ¿Cuál es el significado de los datos o métricas –porcentajes, promedios, etc.– que se utilizan para reportar los resultados?
- Los resultados, ¿son principalmente descriptivos de lo que la población evaluada es capaz de hacer en una disciplina dada, sin que ello implique un juicio de valor acerca de si los alumnos alcanzan lo necesario para aprobar el curso o nivel evaluado? o, por el contrario, ¿establecen las pruebas estándares o expectativas definidas acerca de

lo que los alumnos deberían alcanzar como mínimo para que su resultado sea considerado satisfactorio?

- ¿Cuál es la relación entre lo que las pruebas evalúan y los conocimientos y capacidades que, de acuerdo a lo establecido en el currículo, los alumnos deberían lograr?
- ¿Qué actividades concretas los alumnos debían ser capaces de resolver para alcanzar un resultado satisfactorio en la prueba? ¿Qué tipo de actividades fueron capaces de resolver los alumnos con mejores resultados y cuáles los alumnos con peores resultados?
- ¿A quiénes y bajo qué modalidades van a ser entregados los resultados? ¿Cómo se espera que los mismos sean utilizados y de qué modo se espera que contribuyan a mejorar el sistema educativo?
- ¿Cuáles son los principales problemas que los resultados ponen de manifiesto y qué acciones el Ministerio planea tomar –o está tomando– para hacerles frente? ¿En qué plazos se espera que puedan observarse mejoras?
- ¿Son comparables estos resultados en relación a mediciones anteriores? Los procedimientos técnicos seguidos, ¿permiten garantizar la comparabilidad de las mediciones? ¿Están los procedimientos técnicos utilizados abiertos al escrutinio público?
- ¿Se han tomado los recaudos técnicos necesarios para controlar el efecto de la composición sociocultural del alumnado cuando se efectúan comparaciones de resultados entre tipos de escuelas, sectores del sistema educativo o regiones del país, o entre escuelas individuales? ¿Están los procedimientos técnicos utilizados abiertos al escrutinio público?

ii. Conocimientos y saberes

Para poder formular las preguntas anteriores a los funcionarios ministeriales o a los especialistas, así como para intentar usar las respuestas como insumo en la producción de información periodística, es imprescindible contar con algunos saberes y conocimientos básicos:

- Es preciso poseer un conocimiento básico sobre el sistema educativo, sus estructuras fundamentales y el currículo. Cómo están organizadas las escuelas y qué es lo que se proponen enseñar en los diversos niveles, grados y asignaturas.
- Conviene también estar al tanto de las principales iniciativas que se están llevando adelante (por ejemplo, programas de mejoramiento de

cobertura y/o de calidad) y de cuáles son las principales líneas de debate en el sector. Armar una especie de “mapa” con las principales posturas en pugna en relación a las políticas educativas en el país.

- Es imprescindible poseer un conjunto de conocimientos básicos sobre evaluación educativa y estadística. Saber que hay distintos tipos de evaluaciones, distintos tipos de pruebas y distintas finalidades, y que cada una de ellas permite extraer cierto tipo de conclusiones y no otras. Asimismo, saber distinguir entre los distintos tipos de métricas que se utilizan en la evaluación.

iii. Otros recursos y sugerencias

- Es necesario distinguir datos de interpretaciones, y tener conciencia que un mismo dato puede admitir distintas interpretaciones y que algunas interpretaciones pueden ser válidas y otras no. Distinguir siempre entre suposiciones, hipótesis o propuestas de acción y las conclusiones que están apoyadas en los datos.
- Conviene apoyarse siempre en consultas a especialistas dentro y fuera de los ministerios. Asesorarse y confrontar con ellos interpretaciones y dudas.
- Hay que intentar evitar la tentación de “la” noticia espectacular a partir de los resultados de las evaluaciones. En general la educación no da para eso, aunque sí para producir informes sustanciosos y reflexivos. Si una de las claves de una nota periodística atractiva es crear un conflicto, tal vez el mejor modo de hacerlo es promoviendo el debate público informado, identificando actores clave con posturas diferentes y elaborando notas que confronten interpretaciones de los resultados, así como posturas acerca del valor y significado de la propia evaluación.
- Si va a dar datos de escuelas específicas o a comparar los resultados entre provincias o tipos de escuelas, es indispensable ser muy cuidadoso. Hay que recordar siempre que el principal predictor de los resultados es la composición social del alumnado. Hay que saber distinguir entre tener el puntaje promedio más alto y ser una muy buena escuela –pueden coincidir ambas cosas, pero también pueden ser cosas distintas–. Es necesario distinguir entre una escuela relativamente mejor que otra, de una escuela que logra que todos sus alumnos alcancen los conocimientos y capacidades esperados –porque una escuela puede tener el promedio más alto de todas, pero no estar logrando que la mitad de sus alumnos alcancen los niveles de aprendizaje esperados–.

- Como norma general conviene evitar las adjetivaciones y centrarse en la información, dejando que los datos hablen por sí mismos o que las interpretaciones y valoraciones las formulen los diversos actores involucrados.
- Conviene tener siempre presente la complejidad de la labor educativa y evitar las visiones simplistas y los chivos expiatorios. Por el contrario, es preciso intentar comprometer a todos con el mejoramiento de la labor educativa, asumiendo como punto de partida que la educación debe ser una política de estado y un compromiso de toda la sociedad.
- Finalmente, otra pista para la elaboración de notas atractivas es preocuparse por comprender los efectos de las evaluaciones en la vida de la comunidad. ¿Cómo son las escuelas? ¿Cómo viven los maestros y los niños las pruebas? ¿Cómo analizan y utilizan los maestros los resultados? ¿Qué esfuerzos hacen los maestros por ayudar a aprender a los niños que no logran resultados satisfactorios? ¿A qué atribuyen los resultados? ¿Qué podrían hacer las familias para apoyar a sus hijos en el aprendizaje y para ayudar a la escuela?

CAPÍTULO V

¿CÓMO MEJORAR LA COMUNICACIÓN DE RESULTADOS DESDE LOS MINISTERIOS DE EDUCACIÓN?

El presente capítulo tiene como propósito sistematizar un conjunto de reflexiones y propuestas, formuladas y discutidas durante el Seminario de Santo Domingo, en relación al modo en que los ministerios en general comunican los resultados de las evaluaciones de logro educativo y, más específicamente, a cómo se relacionan con la prensa y con la opinión pública en la comunicación de resultados de las evaluaciones. El capítulo tiene dos partes, una de diagnóstico de situación y otra de propuestas y recomendaciones o sugerencias.

Las dificultades de los ministerios de Educación a la hora de comunicar resultados

1. La falta de visión estratégica sobre el rol de la evaluación

Los problemas y dificultades de los ministerios de Educación para comunicar los resultados de las evaluaciones educativas comienzan en el

hecho más básico de que muchos de ellos carecen de una estrategia definida sobre la evaluación en sí misma y sus propósitos, así como de una visión acerca del modo en que se espera que la evaluación impacte sobre el sistema educativo. Muchas veces se parte de una visión algo ingenua, en el sentido de que el mero hecho de evaluar producirá mejoras o de que todos los actores comprenderán y emplearán los resultados que se divulguen.

También existe una visión particular que traza una suerte de paralelismo entre las unidades de evaluación de logros educativos con los Institutos Nacionales de Estadística: su misión sería únicamente producir datos e información cuyo análisis no les corresponde a ellas y debe ser realizado por terceras partes (tomadores de decisiones, políticos, académicos, etc.). El problema es que como en nuestra región recién se empieza a realizar este tipo de evaluaciones, no existen verdaderas audiencias en condiciones de comprender y utilizar la información (como hemos visto, a veces hasta los propios productores de las evaluaciones incurrir en equívocos). Por lo tanto, es necesario invertir tiempo en formar a las audiencias y en “modelar” análisis sólidos y válidos de los datos.

Lo anterior lleva a que normalmente se carezca de una estrategia de difusión y uso de los resultados que defina con claridad cuáles son las audiencias a las que se quiere llegar, cuáles son los contenidos y mensajes que se quiere hacer llegar a cada audiencia y cuáles son los vehículos o tipos de productos de difusión apropiados a cada audiencia.

Normalmente se produce un único informe de carácter predominantemente técnico, pensado desde la perspectiva de los especialistas de la Unidad de Evaluación, pero no desde la perspectiva de las demandas y preocupaciones de las audiencias. En el mejor de los casos se producen materiales dirigidos a las audiencias internas del sector educativo –las propias instancias de decisión ministerial, materiales pedagógicos para los docentes–, lo cual probablemente es lo prioritario. Pero normalmente hay escasa consideración de la comunicación hacia las audiencias externas: las familias, la opinión pública, el sistema político, el mundo académico, las organizaciones sociales y la prensa en tanto vehículo privilegiado para llegar a todas las anteriores.

Al respecto es preciso señalar que los modos de difusión que se elijan deben ser consistentes con la visión general acerca del rol del sistema de evaluación.

2. Inconsistencias técnicas y desarticulaciones internas

Un segundo tipo de problemas está relacionado con la falta de acumulación de conocimiento sobre evaluación en la región, que ha determinado que en las primeras experiencias no se tomarán decisiones técnicas apropiadas. Un ejemplo de ello es la confusión generada por algunos ministerios –según se analizó en el capítulo II de este texto– en torno a pruebas que son de tipo descriptivo pero terminan siendo interpretadas en términos de aprobación y reprobación de alumnos. La demanda pública y el propósito parecen ir en un sentido, pero las pruebas fueron diseñadas desde otra perspectiva, que no permite satisfacer esa demanda.

Un problema crucial es el de la falta de articulación y comunicación entre las unidades de evaluación y de currículo al interior de los ministerios. Suele ocurrir que las pruebas no estén adecuadamente alineadas con los objetivos establecidos en el currículo y/o que estos últimos estén formulados en términos tan vagos y generales que es imposible evaluar su logro. Esta falta de articulación interna a los ministerios suele conducir a que las evaluaciones no puedan responder a la pregunta principal que legítimamente se hacen –o deberían hacerse– las familias y la opinión pública en general: ¿están los alumnos aprendiendo lo que se espera que aprendan?, ¿qué tan bien o mal están en relación a lo que se espera de ellos?

Asimismo, los problemas de falta de articulación interna le quitan fuerza a la evaluación, en la medida que sus resultados no son conocidos y utilizados por otras unidades al interior de los propios ministerios. Este hecho genera flancos débiles importantes cuando la prensa, con razón, pregunta qué se está haciendo para revertir los problemas que la evaluación pone de manifiesto. En ese momento el resto de las unidades debería tener algo que responder.

A veces ocurre también que los técnicos no tienen suficiente visión de conjunto del sistema y las políticas educativas como para armar un discurso apropiado para la presentación de resultados, que los engarce en el conjunto. La falta de relacionamiento interno con otras unidades, que podría enriquecer la interpretación, incide sin duda en este problema.

3. El uso político de los resultados por los propios ministerios

Otro aspecto clave del problema es el relacionamiento de la alta dirección política de los ministerios con los resultados de las evaluaciones. Al respecto parece haber una dinámica contradictoria de atracción y rechaza-

zo. En principio evaluar resulta atractivo, pero luego no se sabe cómo manejar públicamente los resultados.

En algunos casos la alta dirección política del ministerio tiene una visión ingenua respecto a que los resultados mejorarán rápidamente y suele no estar preparada para manejar malos resultados ante la prensa con cierta solvencia. En algunos países esto ha dado lugar simplemente a que los resultados no se publiquen. En otros casos –algunos han sido ejemplificados en el capítulo II– los funcionarios políticos no dominan el tema y transmiten mensajes que incluyen errores de interpretación importantes.

También suele darse la situación opuesta –normalmente al inicio de un mandato–, en que la alta dirección puede estar interesada en utilizar los malos resultados para desprestigiar a la administración anterior o para presionar sobre el cuerpo docente, pero luego esperará que los mismos mejoren rápidamente antes del término de su gestión.

En cualquiera de las dos situaciones hay una visión simplista respecto al sector educativo y los tiempos de cambio.

El caso del Ministro Paulo Renato de Souza en Brasil es un muy buen ejemplo de manejo adecuado del tema: ante los malos resultados de Brasil en la evaluación PISA, almorzó con los periodistas, su principal declaración fue “pensé que los resultados podían ser peores” y destacó los avances realizados por Brasil en la última década en materia de incorporación de estudiantes de 15 años a la educación formal y las distancias sociales entre Brasil y los países miembros de la OCDE. Incluso escribió un artículo con su firma titulado “Vaso medio lleno, vaso medio vacío” destacando que los resultados pueden mirarse desde diferentes ángulos: ha sido un logro incorporar a estos jóvenes al sistema, pero aún no están aprendiendo lo que deberían aprender.

Sin embargo, este caso es bastante excepcional. Normalmente los funcionarios políticos esperan buenos resultados que les sirvan como propaganda o aval a su gestión y los malos resultados, que suelen ser los más comunes, se tornan una “papa caliente” que no saben cómo manejar.

4. El problema de la independencia técnica y política de las unidades de evaluación

Por detrás de los problemas mencionados en el apartado anterior hay una cuestión más compleja, que es la relativa a la independencia política de las unidades de evaluación. Las unidades de evaluación de los ministe-

rios producen y divulgan la información sobre logros educativos, pero a la vez en cierto modo son parte interesada, lo cual coloca inevitablemente a la información bajo sospecha para los periodistas y para otros actores involucrados.

De acuerdo a la experiencia de los países de la región, esta contradicción permanece aún cuando la unidad de evaluación esté fuera de la estructura ministerial, por ejemplo en una universidad o una entidad evaluadora privada. Aún en estas situaciones, la unidad suele depender de los recursos del Estado para realizar las evaluaciones y de decisiones políticas para difundirlos.

En mi opinión el problema no radica tanto en la solución institucional sino en la cultura política y en la cultura de la evaluación. Es preciso que los ministerios de Educación y las direcciones políticas conciban la labor de evaluación como una labor de información al público –además de producir información para ser utilizada al interior del ministerio y por los docentes– y otorguen a las unidades de evaluación amplia independencia y transparencia en la difusión de los resultados. Y es preciso que la sociedad se acostumbre –y exija– a que exista una unidad de evaluación que periódicamente informa sobre el estado del sistema educativo y de los aprendizajes de los niños y jóvenes, de manera clara, técnicamente apropiada y sin interferencias de tipo político.

Lo anterior implica asumir que la opinión pública tiene el derecho a juzgar la gestión ministerial sobre la base de la información válida y transparente, y de diversidad de interpretaciones a partir de la misma. Implica asumir la complejidad de la tarea educativa y el carácter colectivo de la misma, que es responsabilidad de toda la sociedad. Esto último no implica que el ministerio no asuma su responsabilidad por las decisiones y políticas que impulsa, pero sí implica no asumir una actitud de omnipotencia según la cual el sistema de educación pública podría ser cambiado radicalmente en el término de una administración de cinco años por el solo mérito de las decisiones ministeriales, sean del tipo que sean.

Obviamente, la dirección política del ministerio debe tener y ofrecer su propia visión e interpretación de los resultados, emitir los mensajes que considere apropiados y eventualmente confrontar su opinión con la de otros actores políticos. Puede incluso legítimamente utilizar los resultados para destacar logros de su gestión. Pero ello a condición de que esté completamente garantizada la independencia de la unidad de evaluación, la calidad de su trabajo y la transparencia en el acceso a la información producida. Lo cual implica, entre otras cosas, que la sociedad sepa

qué información se publicará, sobre qué aspectos, en qué tiempos, y que esto ocurra tanto cuando los resultados son buenos como cuando son malos.

Aún hay mucho camino por andar en la región para lograr una cultura de la evaluación que permita que las unidades manejen la información con independencia de los intereses y tiempos políticos.

5. El escaso relacionamiento con los medios de comunicación

Un último conjunto de problemas está constituido por las carencias específicas en el relacionamiento con los medios de comunicación. Como fue planteado en el capítulo III muchas veces desde las unidades de evaluación se desconoce la lógica de los medios de comunicación y no se tiene tiempo ni energía para dedicar a desarrollar una estrategia de relacionamiento con los mismos. Dentro de los ministerios, al menos en países chicos, no suele haber personal especializado en la relación con los medios de prensa. En otros casos el personal de prensa de los ministerios en realidad se especializa en organizar las apariciones públicas del ministro y en el cuidado de su imagen.

Recomendaciones y propuestas

Teniendo como contexto el diagnóstico anterior, a continuación se propone un conjunto de recomendaciones y sugerencias a tener en cuenta a la hora de mejorar la difusión pública de resultados de las evaluaciones a través de los medios de prensa.

En términos de la política de evaluación:

- Concebir una estrategia definida en relación a la evaluación: qué preguntas pretende responder; qué tipo de información va a producir; quiénes y cómo se espera que utilicen la información.
- Concebir una estrategia definida de difusión de resultados. Ello implica identificar a las diferentes audiencias y sus respectivas preguntas, demandas y/o posibles expectativas en relación a los logros educativos, para luego definir el tipo de productos de comunicación apropiados y comprensibles para cada audiencia.
- Asumir la devolución de resultados a la opinión pública como una instancia de rendición de cuentas sobre el estado de la educación,

sujeta a debate público. Esto implica informar siempre y de modo transparente, sin interferencias de tipo político; entregar la información a diversos actores calificados y con diferentes posturas, a los cuales pueda recurrir la prensa para obtener interpretaciones y visiones diferentes a partir de los resultados; evitar la tentación de usar los resultados con fines propagandísticos o de ocultarlos cuando no son satisfactorios.

- Otorgar amplia independencia a la unidad de evaluación, lo que puede incluir el hecho de delegar la explicación principal a funcionarios con conocimiento técnico (siempre que no se “pierdan” en los detalles técnicos) y reservar para el ministro o la alta dirección la expresión de los lineamientos de política educativa que se están siguiendo y lo que los resultados aportan a los mismos²¹.
- Desde la alta dirección, así como desde otras unidades clave del ministerio, preparar un conjunto de mensajes claros en respuesta a lo que los datos revelan, balanceando las “buenas” y “malas” noticias, evitando tanto el tono catástrofe como el triunfalismo, enfatizando las líneas de acción para enfrentar los problemas que los datos ponen de manifiesto. Presentar los resultados como desafíos al sistema –como diagnóstico y no autopsia–. Las responsabilidades de todos los actores deben señalarse, evitando cargar “culpas” sobre solo algunos de ellos.
- Fomentar la apropiación del proceso de evaluación y, por tanto, de sus resultados, generando espacios para la participación de actores de otras dependencias del ministerio así como de actores externos al mismo, en los procesos de diseño y validación de las pruebas y sus referentes conceptuales, así como en la interpretación de los resultados lo que además contribuiría a mejorar la calidad y pertinencia y utilidad de las evaluaciones.
- Promover el mensaje de que el sistema de medición es perfectible, que seguramente tiene debilidades, pero que está abierto a mejoras, para lo

²¹ En Uruguay, por ejemplo, es una práctica habitual desde 1996 que las autoridades conocen los primeros resultados un par de días antes de su presentación pública, cuando la misma ya ha sido convocada y los materiales están impresos. La presentación está a cargo de personal técnico y luego el Director Nacional de Educación Pública hace sus comentarios. En algunos países, por el contrario, los reportes pasan varios meses en el despacho del ministro, esperando que el mismo tenga ocasión de leerlos, consultar a los técnicos para comprenderlos, decidirse a presentarlos a la prensa y prepararse para ello.

cual se quiere recoger la opinión calificada de actores diversos –incluyendo a la prensa, a la cual se quiere convertir en aliado crítico–.

En términos de comunicación:

- Seleccionar cuidadosamente los mensajes –pocos, claros, reiterativos y claramente basados en la evidencia resultante de las pruebas y del análisis de los factores vinculados– que se quiere transmitir en cada instancia de divulgación de resultados.
- Del mismo modo, explicitar de manera clara y sencilla el tipo de conclusiones que no pueden ser derivadas de los datos y las principales limitaciones que la evaluación tiene.
- Preparar bien las explicaciones pertinentes y al alcance de diferentes audiencias acerca de aspectos técnicos clave tales como la equivalencia (o no) de las pruebas con las utilizadas en mediciones anteriores; el significado de las métricas empleadas; la posibilidad (o no) de interpretarlas en términos de “aprobación”; el modo en que se han establecido “líneas de corte” (si se ha hecho) o el significado de los “niveles” (si se han establecido).
- Enfatizar siempre como uno de los mensajes principales qué es lo que los alumnos deberían saber y ser capaces de hacer, qué muestran los resultados al respecto y cómo ayudarlos a mejorar.
- Analizar la posibilidad y conveniencia de hacer la presentación del primer informe oficial de una evaluación nacional en un recinto escolar, con la asistencia de maestros, estudiantes y padres de familia. La escuela puede ser seleccionada con criterios que ayuden a anclar alguno(s) de los mensajes en un ejemplo específico. Por ejemplo, se puede escoger un recinto escolar que sea representativo del promedio nacional de escuelas públicas en términos de la comunidad que atiende, la dotación física y la experiencia del personal y que haya mostrado un alto rendimiento en las pruebas.
- Desarrollar un enfoque comunicativo que intente comprometer a todos con la labor educativa, que asuma la educación como una política de Estado y como un compromiso de toda la sociedad: ministerio de educación, educadores, padres de familia, estudiantes y sociedad civil. Evitar tanto la búsqueda de culpables o responsables únicos –los docentes, las escuelas, los supervisores–, como el uso de las acciones educativas con fines de propaganda política, informando con modestia y cautela sobre lo hecho, los avances y los desafíos pendientes.

- Distribuir siempre un comunicado de prensa adecuadamente escrito, claro, conciso, con los principales datos y mensajes que se desea destacar en la presentación oficial de los resultados.

En términos del relacionamiento con la prensa:

- Asegurar la disponibilidad de recursos humanos idóneos y suficientes para la labor de comunicación con la prensa, que no se limitará a la difusión periódica de resultados, sino también a la aclaración de cuestionamientos surgidos en reacción a la cobertura de prensa, de respuesta a reportes o equívocos, preparación de notas de prensa, conseguir otras instancias de presencia en los medios, otorgar entrevistas cuando la prensa las solicita, etc.
- Desarrollar actividades de capacitación para periodistas en relación a los temas clave de las evaluaciones, desligadas de las instancias de presentación de resultados. Muchas de las explicaciones y conceptos técnicos, así como de las advertencias sobre lo que no puede concluirse válidamente de los datos o el tipo de interpretaciones que no son apropiadas, pueden resultar “sospechosas” para el periodista cuando se hacen en el marco de la difusión de resultados. Asimismo, según lo analizado en el capítulo III, en el momento de la cobertura el periodista no tiene tiempo para comprender a fondo cuestiones técnicas complejas. Por ello parece conveniente ofrecer instancias formativas separadas en el tiempo de los momentos de divulgación de resultados.
- Cultivar especialmente el vínculo con periodistas reconocidos o prometedores o especialmente interesados en temas educativos, a los cuales mantener permanentemente informados sin pretender contar con “incondicionales acrílicos” sino más bien con “incondicionales” acerca de la importancia de la difusión de resultados y del debate público sobre la educación”. Trabajar con ellos acerca de los modos de hacer “noticia educativa” de manera constructiva.
- Explorar la posibilidad de entrega “anticipada” y “embargada” de informes a periodistas, para que tengan tiempo de hacer consultas, verificar validez, contrastar opiniones de expertos, formarse opinión y formular selectivamente sus propios mensajes antes de publicación oficial. De este modo podrían realizar sus notas y comentarios de manera rápida pero informada. Cuando esto no es posible porque no hay garantías del mantenimiento de la confidencialidad –que es lo más probable en nuestra región–, tal vez se pueda hacer algo parecido

entregando el informe en forma anticipada y bajo compromiso de confidencialidad a algunos especialistas y actores relevantes a quienes los periodistas podrán consultar o entrevistar en forma inmediata a la presentación pública de los resultados.

- Facilitar a los periodistas una nómina de informantes calificados e instituciones independientes a las cuales consultar en relación a las evaluaciones.
- Comprometer a personajes públicos influyentes u organizaciones no gubernamentales para que analicen los resultados a través de columnas firmadas, artículos –como el del Observatorio Ciudadano de la Educación– o entrevistas –radiales, televisivas o escritas–.

ANEXO

Observatorio Ciudadano de la Educación

Comunicado 67

La evaluación educativa de la OCDE

**Resultados insatisfactorios: México, penúltimo lugar
Rezagos en comprensión de lectura y aplicación de conocimientos en
matemáticas y ciencias**

La desigualdad económica si afecta resultados

El 4 de diciembre se dieron a conocer en París los primeros resultados del Programme for International Student Assessment (Programa para la Evaluación Internacional de Estudiantes, PISA por sus siglas en inglés) de la Organización para la Cooperación y el Desarrollo Económico (OCDE) de la que México es parte. Estos resultados muestran que los estudiantes mexicanos están muy lejos de alcanzar el nivel educativo del mundo desarrollado y que una proporción no ha logrado el mínimo de comprensión de lectura esperado. Los datos son interesantes porque destacan ángulos de nuestra realidad educativa, detectan debilidades y hacen notar el peso de factores económicos y sociales en el rendimiento escolar. Por ello, Observatorio dedica el presente comunicado al análisis de la información contenida en el reporte inicial del PISA-2000.

El examen y sus resultados

El PISA es parte de un programa más amplio de indicadores de calidad educativa acordado por los países de la OCDE y representa el esfuerzo de evaluación educativa internacional de mayor alcance en la historia. A través del mismo se evaluarán cada tres años los conocimientos de lectura, matemáticas y ciencias enfatizando un campo en cada aplicación. El PISA-2000, correspondiente a lectura, se aplicó a jóvenes de 15 años (inscritos en secundaria o bachillerato) y midió sus capacidades de comprensión y aplicación de textos generales, de matemáticas y ciencias. Participaron los 28 países de la Organización, más Brasil, Letonia, Liechtenstein y Rusia; en total se examinaron 260 mil estudiantes, de los cuales cuatro mil quinientos de México.

El reporte correspondiente, Knowledge and Skills for Life (disponible en el sitio www.pisa.oecd.org) da a conocer resultados comparativos, más tarde se publicarán monografías por país. El año entrante se hará la misma prueba en otros países que lo han solicitado, entre ellos Argentina y Perú, en 2003 se realizará la evaluación de matemáticas y en 2006 la de ciencias. La prueba consiste en un cuestionario con preguntas de opción múltiple o abiertas; en general se presenta al estudiante un texto breve y se mide su comprensión y uso de conceptos “en situaciones de la vida real”. El diseño estima una media teórica de 500 puntos con una desviación estándar de 100.

Las calificaciones globales muestran que los países en los primeros lugares (Finlandia 546 puntos, Canadá 534, Holanda 532, Nueva Zelanda 529 y Australia 528) no se separan demasiado de la media y que en ningún caso el rendimiento podría calificarse de excepcional. En cambio, los últimos lugares (México 422 puntos y Brasil 396) están lejos del promedio general. En todas las calificaciones México quedó en el penúltimo lugar; en matemáticas el primer puesto fue para Japón con 557 puntos y México obtuvo 387; en ciencias ganó Corea con 552 puntos y México quedó en 422.

CUADRO 1
Promedio general de las tres pruebas

Lugar	País	Puntaje	Lugar	País	Puntaje
1	Finlandia	546	17	Dinamarca	497
2	Canadá	534	18	Suiza	494
3	Holanda	532	19	España	493
4	N. Zelanda	529	20	R. Checa	480
5	Australia	528	21	Italia	479
6	Irlanda	527	22	Alemania	474
7	Corea S.	525	23	Liechtenantstein	483
8	Reino Unido	523	24	Hungría	480
9	Japón	522	25	Polonia	479
10	Suecia	516	26	Grecia	474
11	Austria	507	27	Portugal	470
12	Bélgica	507	28	Rusia	462
13	Islandia	507	29	Letonia	458
14	Noruega	505	30	Luxemburgo	441
15	Francia	505	31	México	422
16	EUA	504	32	Brasil	396

Nota: El puntaje medio de la prueba es de 500 puntos con una desviación estándar de 100 puntos.

Fuente: OECD, Knowledge and -Skills for Life. First Results from Pisa 2000 (2001)

Además de esas puntuaciones, el reporte ofrece una escala por niveles de dominio de los textos. La escala maneja cinco grados, que significan: 1) Localizar una información específica; 2) Hacer inferencias simples; 3) Captar relaciones entre componentes; 4) Identificar información implícita; 5) Mostrar dominio de lectura y valoración crítica. Se añade el grado -1 que implica incompreensión de los textos y se deduce de las respuestas incorrectas.

Los resultados de México son preocupantes. Mientras que en los países que ocupan los primeros lugares cerca de 20 por ciento de los jóvenes alcanzaron el nivel más alto y menos de 5 por ciento se situaron en el mínimo, en México menos del uno por ciento obtuvo el nivel superior y cerca de 30 por ciento quedó en el inferior. Peor aún, 16 por ciento se ubicó en el nivel de incompreensión de textos.

Otro dato significativo es que, en promedio, los países de la OCDE alcanzaron resultados satisfactorios (niveles 4 y 5 de dominio) en más de 30 por ciento de los casos, mientras en México ese nivel fue ocupado por menos de 7 por ciento, lo que hace evidente la brecha de calidad que nos separa de aquellos.

CUADRO 2
Porcentaje de estudiantes por nivel de dominio en la escala de lectura combinada (general, matemáticas y ciencias)

Table 2.1a
Percentage of students at each level of proficiency on the combined reading literacy scale

		Proficiency levels											
		Below Level 1 (less than 335 score points)		Level 1 (from 335 to 407 score points)		Level 2 (from 408 to 480 score points)		Level 3 (from 481 to 552 score points)		Level 4 (from 553 to 625 score points)		Level 5 (above 625 score points)	
		%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.
OECD COUNTRIES	Australia	3.3	(0.5)	9.1	(0.8)	19.0	(1.1)	25.7	(1.1)	25.3	(0.9)	17.6	(1.2)
	Austria	4.4	(0.4)	10.2	(0.6)	21.7	(0.9)	29.9	(1.2)	24.9	(1.0)	8.8	(0.8)
	Belgium	7.7	(1.0)	11.3	(0.7)	16.8	(0.7)	25.8	(0.9)	26.3	(0.9)	12.0	(0.7)
	Canada	2.4	(0.3)	7.2	(0.3)	18.0	(0.4)	28.0	(0.5)	27.7	(0.6)	16.8	(0.5)
	Czech Republic	6.1	(0.6)	11.4	(0.7)	24.8	(1.2)	30.9	(1.1)	19.8	(0.8)	7.0	(0.6)
	Denmark	5.9	(0.6)	12.0	(0.7)	22.5	(0.9)	29.5	(1.0)	22.0	(0.9)	8.1	(0.5)
	Finland	1.7	(0.5)	5.2	(0.4)	14.3	(0.7)	28.7	(0.8)	31.6	(0.9)	18.5	(0.9)
	France	4.2	(0.6)	11.0	(0.8)	22.0	(0.8)	30.6	(1.0)	23.7	(0.9)	8.5	(0.6)
	Germany	9.9	(0.7)	12.7	(0.6)	22.3	(0.8)	26.8	(1.0)	19.4	(1.0)	8.8	(0.5)
	Greece	8.7	(1.2)	15.7	(1.4)	25.9	(1.4)	28.1	(1.7)	16.7	(1.4)	5.0	(0.7)
	Hungary	6.9	(0.7)	15.8	(1.2)	25.0	(1.1)	28.8	(1.3)	18.5	(1.1)	5.1	(0.8)
	Iceland	4.0	(0.3)	10.5	(0.6)	22.0	(0.8)	30.8	(0.9)	23.6	(1.1)	9.1	(0.7)
	Ireland	3.1	(0.5)	7.9	(0.8)	17.9	(0.9)	29.7	(1.1)	27.1	(1.1)	14.2	(0.8)
	Italy	5.4	(0.9)	13.5	(0.9)	25.6	(1.0)	30.6	(1.0)	19.5	(1.1)	5.3	(0.5)
	Japan	2.7	(0.6)	7.3	(1.1)	18.0	(1.3)	33.3	(1.3)	28.8	(1.7)	9.9	(1.1)
	Korea	0.9	(0.2)	4.8	(0.6)	18.6	(0.9)	38.8	(1.1)	31.1	(1.2)	5.7	(0.6)
	Luxembourg	14.2	(0.7)	20.9	(0.8)	27.5	(1.3)	24.6	(1.1)	11.2	(0.5)	1.7	(0.3)
	Mexico	16.1	(1.2)	28.1	(1.4)	30.3	(1.1)	18.8	(1.2)	6.0	(0.7)	0.9	(0.2)
	New Zealand	4.8	(0.5)	8.9	(0.5)	17.2	(0.9)	24.6	(1.1)	25.8	(1.1)	18.7	(1.0)
	Norway	6.3	(0.6)	11.2	(0.8)	19.5	(0.8)	28.1	(0.8)	23.7	(0.9)	11.2	(0.7)
Poland	8.7	(1.0)	14.6	(1.0)	24.1	(1.4)	28.2	(1.3)	18.6	(1.3)	5.9	(1.0)	
Portugal	9.6	(1.0)	16.7	(1.2)	25.3	(1.0)	27.5	(1.2)	16.8	(1.1)	4.2	(0.5)	
Spain	4.1	(0.5)	12.2	(0.9)	25.7	(0.7)	32.8	(1.0)	21.1	(0.9)	4.2	(0.5)	
Sweden	3.3	(0.4)	9.3	(0.6)	20.3	(0.7)	30.4	(1.0)	25.6	(1.0)	11.2	(0.7)	
Switzerland	7.0	(0.7)	13.3	(0.9)	21.4	(1.0)	28.0	(1.0)	21.0	(1.0)	9.2	(1.0)	
United Kingdom	3.6	(0.4)	9.2	(0.5)	19.6	(0.7)	27.5	(0.9)	24.4	(0.9)	15.6	(1.0)	
United States	6.4	(1.2)	11.5	(1.2)	21.0	(1.2)	27.4	(1.3)	21.5	(1.4)	12.2	(1.4)	
	OECD total	6.2	(0.4)	12.1	(0.4)	21.8	(0.4)	28.6	(0.4)	21.8	(0.4)	9.4	(0.4)
	OECD average	6.0	(0.1)	11.9	(0.2)	21.7	(0.2)	28.7	(0.2)	22.3	(0.2)	9.5	(0.1)
NON-OECD COUNTRIES	Brazil	23.3	(1.4)	32.5	(1.2)	27.7	(1.3)	12.9	(1.1)	3.1	(0.5)	0.6	(0.2)
	Latvia	12.7	(1.3)	17.9	(1.3)	26.3	(1.1)	25.2	(1.3)	13.8	(1.1)	4.1	(0.6)
	Liechtenstein	7.6	(1.5)	14.5	(2.1)	23.2	(2.9)	30.1	(3.4)	19.5	(2.2)	5.1	(1.6)
	Russian Federation	9.0	(1.0)	18.5	(1.1)	29.2	(0.8)	26.9	(1.1)	13.3	(1.0)	3.2	(0.5)

El factor económico

Por otra parte, el informe analiza algunos factores con incidencia en la calidad educativa. Se demuestra que las condiciones socioeconómicas de los estudiantes y las de la escuela pesan en el rendimiento. Por ejemplo, la correlación entre el PIB per cápita y las calificaciones es de 0.53 y la correlación entre gasto por alumno y calificaciones es 0.47. También es clara la relación entre los niveles de desigualdad económica y las calificaciones: los países en los primeros lugares no son los más ricos, sino los que además de contar con economías sólidas distribuyen con equidad los ingresos (como Finlandia, Canadá, Holanda, Nueva Zelanda y Australia). Países poderosos como Estados Unidos, Francia y Alemania, pero con niveles de desigualdad mayores, obtuvieron lugares intermedios en la clasificación. México y Brasil que además de ser pobres tienen niveles de desigualdad extremos, quedaron al final.

En cuanto a las condiciones económicas de las escuelas se observa que, por regla general, los estudiantes que provienen de entornos escolares bien equipados obtienen mejores resultados. Los países con una dotación escolar satisfactoria y distribuida con equidad lograron los niveles superiores, mejores aún que otros con equipo tecnológico sofisticado, como Estados Unidos (lugar 16) o Alemania (lugar 22). Debe hacerse notar, no obstante, que la relación entre economía y resultados escolares es imperfecta. Si nos comparamos con Polonia, Hungría, Grecia o Portugal es claro que tanto el nivel de desarrollo económico de nuestro país, como las condiciones económicas de los estudiantes, son semejantes a esos países, pero allá se obtienen mejores resultados. Ello expresa la posibilidad de mejorar la calidad educativa aún en contextos económicos y sociales deprimidos, si se tiene la voluntad política necesaria; también indica la necesidad de intensificar y aplicar correctamente los programas encaminados a mejorar la dotación de recursos y las condiciones de operación de las escuelas más pobres. Desde luego, es importante esperar los resultados de la aplicación de la prueba en países de desarrollo económico, social y cultural similares a los de México para contar con un marco de comparación más apropiado.

CUADRO 3
Indicadores seleccionados
Puntuaciones combinadas (lectura, matemáticas y ciencias)
por categoría

	Promedio OCDE	México	Diferencia
Tamaño de la localidad			
Menos de 3000	481	359	-122
3,000 a 15,000	486	396	-90
15,000 a 100,000	500	426	-74
100,000 a 1,000,000	503	451	-52
Más de 1,000,000. Capital	522	491	-31
Más de 1,000,000. Otras Cds.	510	447	-63
Tipo de escuela			
Pública	491	409	-82
Privada	523	494	-29
Admisión mediante examen			
Nunca	496	373	-123
A veces	490	404	-86
Siempre	507	453	-54
Grado escolar			
3ro. Secundaria	485	390	-95
1ro. Bachillerato	519	467	-52
Sexo			
Mujeres	515	432	-83
Hombres	483	411	-72

Fuente: OECD, Knowledge and Skills for Life. First Results from PISA 2000 (2001)

La divulgación de estos resultados generó diferentes reacciones y desde luego no solo en México, en donde ya se pronunciaron las autoridades educativas y otros sectores. En España, que obtuvo el lugar 19, la prensa señaló: "...sin un apoyo decidido del desmotivado profesorado, sin un aumento considerable de la inversión en educación secundaria y sin programas dirigidos a fomentar la implicación de las familias en los centros, será difícil solucionar los problemas" (*El País*, 9/12/01). En Brasil el Ministro de Educación, Paulo Renato, atribuyó los resultados al

rezago escolar, aunque anunció un programa inmediato de refuerzo a la lectura por medio de la dotación de colecciones literarias para 8.4 millones de estudiantes (*Folha de San Paulo*, 8/12/01). En Francia la Dirección General de Programación y Desarrollo del Ministerio de Educación urgió a la toma de iniciativas de reforma escolar que permitan una formación secundaria más adecuada, aunque mostró satisfacción con los resultados en comparación con evaluaciones previas (*Le Monde*, 4/12/01). En Inglaterra, la titular de la oficina encargada del estudio, Estela Morris, dijo que “los resultados obtenidos son una reivindicación de las reformas asumidas en los últimos años, pero los retos continúan... no descansaremos hasta que cada niño obtenga las oportunidades que merece” (DFES News Centre, 5/12/01). Por último, en Estados Unidos el Secretario de Educación, Rod Paige, se mostró francamente insatisfecho con los resultados y declaró que “estos países son nuestros competidores; estar en el promedio no es suficiente para los niños norteamericanos...” (*New York Times*, 5/12/01).

Interrogantes

Desde julio del año pasado en la perspectiva del período de transición que se abría y ante el silencio de la administración saliente, que no respondió a las interrogantes publicadas en este espacio, decidimos cancelar las preguntas que acompañaban cada comunicado. Recientemente se cumplió el primer año de ejercicio de esta administración y ya está en operación el Programa Nacional de Educación. Observatorio, como organización independiente preocupada por el análisis y vigilancia de las políticas educativas, continuará demandando información sobre las acciones y esclarecimiento de las políticas. A partir de este comunicado reanudamos nuestra sección de interrogantes y esperamos que las nuevas autoridades respondan.

Respecto al tema que nos ocupa, suponemos que la participación mexicana en evaluaciones internacionales, que implica un gasto y tiene efectos políticos importantes, se basa en el interés del gobierno de identificar el estado de la calidad educativa para mejorarla. Sería lamentable que los resultados se utilizaran solo para sugerir mayor control de la docencia sin apoyar las condiciones de escuelas, profesores y estudiantes. Los resultados también nos muestran el riesgo de incurrir en la peor combinación posible en el terreno educativo: altas expectativas de cambio y bajo presupuesto.

¿Qué acciones se emprenderán para enfrentar los resultados de México en el PISA y en otras evaluaciones internacionales? ¿Qué estrategias se utilizarán para difundir los resultados de las evaluaciones? ¿De qué manera se permitirá el acceso a las bases de datos?

¿Cuándo iniciará la operación del sistema nacional de evaluación educativa previsto en el Programa Nacional de Educación?

REFERENCIAS

1. The OECD Programme for International Student Assessment, Página web (contiene el informe)
2. The OECD Programme for International Student Assessment, Base de datos (contiene tabulaciones y permite realizar cálculos)



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

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

Esta antología reúne algunos de los principales estudios sobre políticas y prácticas de determinación de expectativas y evaluación de logros de aprendizaje en América Latina realizados por el Grupo de Trabajo sobre Estándares y Evaluación del PREAL en el período 2000-2004.

Este Grupo de Trabajo –conocido como GTEE/PREAL– es una red regional de especialistas y educadores que empezó a constituirse a fines de 1998. Su principal objetivo es promover la clarificación de las expectativas de aprendizaje escolar y la evaluación de logros educativos en los países de la región, así como contribuir al fortalecimiento de los procesos de mejoramiento de la calidad, pertinencia, equidad y eficiencia de los sistemas educativos latinoamericanos.

Las actividades del PREAL, como las del Grupo de Trabajo sobre Estándares y Evaluación, son posibles gracias al apoyo que brindan la United States Agency for International Development (USAID), el Banco Interamericano de Desarrollo (BID), el Banco Mundial, la International Association for the Evaluation of Educational Achievement (IEA), The Tinker Foundation, GE Foundation, entre otros.



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



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