

**N° 22**

## **¿Cómo Presentan sus Resultados los Sistemas Nacionales de Evaluación Educativa en América Latina?**

**Versión Abreviada\***

Pedro Ravela, FEBRERO, 2002

<sup>(\*)</sup> El presente documento es una versión resumida de un trabajo más amplio publicado por PREAL en agosto de 2001. La versión completa puede ser consultada en la página WEB del Grupo de Trabajo de Estándares y Evaluación de PREAL, <http://www.grade.org.pe/gtee-preal/docs/Resultados.pdf>.

Las opiniones vertidas en este trabajo son de responsabilidad del autor y no comprometen al PREAL, ni a las instituciones que lo patrocinan.



## SUMARIO

<b>Introducción</b>	<b>5</b>
<b>Capítulo I. ¿Qué se proponen lograr los reportes nacionales?</b>	<b>7</b>
I.1. Los sistemas de evaluación como sistemas de información al público en general	9
I.2. Los sistemas de evaluación como sistemas de información dirigidos a públicos específicos	9
I.3. Usos de los resultados con consecuencias “fuertes”	10
I.4. El propósito central de la difusión de resultados: ¿“consecuencias fuertes” o “evaluación formativa”?	11
<b>Capítulo II. Porcentajes, medias, notas y puntajes: las métricas empleadas para reportar los resultados</b>	<b>14</b>
II.1. Porcentaje promedio de respuestas correctas y/o medias de puntajes brutos	14
II.2. Porcentajes de alumnos que dominan ciertos objetivos o que logran niveles de desempeño predeterminados	15
II.3. Puntajes estandarizados	17
II.4. Teoría de Respuesta al Ítem	18
II.5. Calificaciones	20
II.6. Resultados descriptivos o estándares de logro: ¿qué es lo que el público y los docentes quieren saber?	21
<b>Capítulo III. Comparando lo incomparable: la parábola de las carreteras</b>	<b>23</b>
III.1. Modos típicos de desagregar la información en los reportes nacionales	23
III.2. La parábola de las carreteras	23
III.3. Algunos ejemplos de conclusiones inapropiadas	24
III.4. La consideración del contexto sociocultural	27
III.5. Los dilemas que plantea la relación entre aprendizajes y contextos sociales a la difusión de resultados	29
<b>Capítulo IV. La utilidad pedagógica de los reportes de resultados</b>	<b>30</b>
IV.1. Materiales de orientación didáctica elaborados a partir de los resultados	30
IV.2. Entrega de resultados por establecimiento	32
IV.3. Evaluaciones de carácter muestral con aplicación y corrección autónoma de las pruebas en el universo de escuelas	34
IV.4. Talleres de difusión de resultados y programas de capacitación en servicio	36
IV.5. ¿Cómo usar las evaluaciones para mejorar las prácticas de enseñanza?	36
<b>Capítulo V. Los reportes de resultados como aporte a la formulación de políticas educativas</b>	<b>38</b>
V.1. Insuficiente análisis, interpretación y conceptualización	38
V.2. Los “factores escolares” asociados a los aprendizajes y las políticas educativas	40
V.3. Tres vacíos de carácter técnico comunes en la región	43
<b>Diez pistas para pensar en el futuro</b>	<b>44</b>
<b>Bibliografía</b>	<b>46</b>



## 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 mismos<sup>1</sup>. 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 espectro de Informes Nacionales producidos en la región en el período 1995-2000. Fueron analizados un total de casi 70 informes<sup>2</sup>.

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 formativo<sup>3</sup>.

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 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 pers-

1 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.

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 GRADE.

3 Lo que en la literatura anglosajona se denomina con la expresión “high-stakes” vs. “low-stakes”.

pectiva 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 conse-

cuencias 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 recolectan 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 brasileiro. 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, harán 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 utilizar la 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 currículum (generalmente Matemática y Lengua y, en menor medida, Ciencias Naturales y Ciencias Sociales) y en ciertos grados o niveles (generalmente aquéllos 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 éstos 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 cuáles 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 acumula-

ción 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 medianamente informados<sup>4</sup>. 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 asume 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.**

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

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.

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.

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.

**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) <sup>(*)</sup>
ARGENTINA					
BOLIVIA					
BRASIL					
COSTA RICA					
CHILE					
ECUADOR					
PARAGUAY					
PERÚ					
URUGUAY					

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

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 aún para una familia de nivel cultural medio. Chile parece ser el único país con una estrategia de difusión dirigida específicamente a las fami-

lias, 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 hijos.<sup>5</sup>

### 1.3. Usos de los resultados con consecuencias “fuertes”

Si bien los enunciados explícitos de los sistemas nacionales de evaluación enfatizan su finalidad formativa

<sup>5</sup> Esto último aparece anunciado en la página WEB del SIMCE pero no fue posible tener acceso a ejemplares de los mismos.

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 identificados<sup>6</sup>.

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 Educativos 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

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.

7 Para más detalles, véase Ministerio de Educación, República de Chile, 2000.

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 SNED<sup>7</sup>.

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 SEP-SNTE, 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.

#### 1.4. 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 exclusivamente 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 pública 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 pública por escuela debe tener carácter censal. Como señalan Linn & Gronlund<sup>8</sup>:

*"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 éstos 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 pruebas en lugar de dominios de contenido más generales (...) Como los ítems de una prueba sólo constituyen una muestra del dominio que intere-*

*sa, 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 éstos 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:

8 "LINN, R. & GRONLUND, N., 2000, pp. 8-9, en inglés en el original.

*"Por tanto, no hay evidencia de Chile de que un siste-*

*ma 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)<sup>9</sup>*

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 espúrea de las escuelas que trabajan con la población más carenciada?

La disyuntiva de fondo es en qué medida los objetivos *formativo* 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, 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"<sup>10</sup>.*

9 CARNOY, M., 2000, p. 18, en inglés en el original.

10 LINN 2000, pp.12-14, en inglés en el original.

## 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 éstos 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
- 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 res-

ponder 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 escolarización<sup>11</sup>.

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 para todos los ítems de la prueba. En términos matemáticos esta “métrica” es equivalente al promedio de los puntajes brutos<sup>12</sup> 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 distin-

11 Este tema involucra la cuestión de si los datos han sido contruidos dentro de un enfoque referido a normas o dentro de un enfoque criterial, cuestión que ha sido tratada con más detalle en RAVELA, P. (editor) y otros, 2000 y que será analizada al final de este capítulo.

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.

tos 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 Argentina<sup>13</sup>.

**Figura 3**  
**Argentina - Valores Medios de Rendimiento en Lengua por Jurisdicción y Nivel**

<b>Nivel Primario - 6º Grado Lengua</b>		
<b>Muestra 1996</b>		
<b>Jurisdicción</b>	<b>Lengua 6º Grado</b>	
	<b>Muestra 1996</b>	<b>Relación con respecto al valor medio</b>
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
<b>MEDIA NACIONAL</b>	<b>58,19%</b>	<b>1,00</b>
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 sólo 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.

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 competencias 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-MEP. 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	81,6	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.

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, 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 nor-

mal 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 ítemes 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 ítemes de modo tal que, como mínimo, se espera que todos los alumnos respondan 14 ítemes 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
<b>SUFICIENTES</b>	<b>57,1</b>	<b>61,3</b>	<b>34,6</b>	<b>40,8</b>
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
<b>INSUFICIENTES</b>	<b>42,9</b>	<b>38,7</b>	<b>65,4</b>	<b>59,2</b>
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 anteriores<sup>14</sup>.

Una de las debilidades de los reportes de Uruguay es que no están explícitamente fundamentadas las razones por las cuáles 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 es<sup>15</sup>.

### 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 aquéllos 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.

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 distribu-

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. 6to. 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.

ción 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 nacionales<sup>16</sup>. 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 ítemes.

La explicación detallada de la TRI excede largamente los límites del presente trabajo, pero basta decir que tiene algunas ventajas sobre el enfoque clásico:

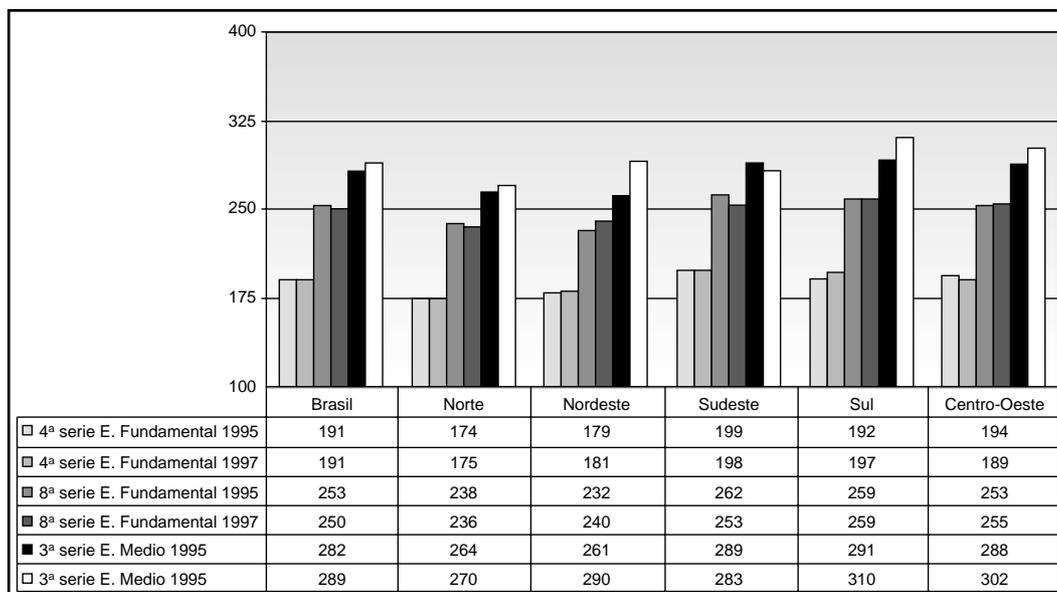
- Permite poner en una misma escala ítemes y alumnos.
- Permite construir estimaciones más precisas de

los cambios en el tiempo mediante la equiparación de las puntuaciones.

- Permite estimar una medida de la capacidad de los alumnos que tiene en cuenta la dificultad de los ítemes; es decir, los ítemes 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 8°. 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, 4° y 8° de Educación Básica y 3° 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 éstos significan. Difícilmente tendrán significado para el lector no especializado, mientras no exista una tradición o cultura en torno a su interpretación.

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.

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 ítemes 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"> <li><input type="checkbox"/> Los alumnos reconocen el valor de cédulas y monedas.</li> <li><input type="checkbox"/> Leen la hora en relojes digitales y analógicos y saben que una hora tiene 60 minutos.</li> <li><input type="checkbox"/> Resuelven problemas sencillos de adición y sustracción con números naturales.</li> </ul>	<b>175</b>	<p><b>En el nivel 175 o arriba de él se encuentran:</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 56% de los alumnos del 4° grado de la Enseñanza Fundamental;</li> <li><input type="checkbox"/> 95% de los alumnos del 8° grado de la Enseñanza Fundamental;</li> <li><input type="checkbox"/> 100% de los alumnos del 3er. grado de la Enseñanza Media.</li> </ul>
<ul style="list-style-type: none"> <li><input type="checkbox"/> Los alumnos reconocen polígonos y cuadriláteros.</li> <li><input type="checkbox"/> Establecen relaciones entre los valores de cédulas y monedas y resuelven situaciones al pagar y recibir cambio, aunque todavía no saben operar con decimales.</li> <li><input type="checkbox"/> Son capaces de multiplicar y dividir, así como identificar unidades, decenas y centenas.</li> <li><input type="checkbox"/> Resuelven problemas que envuelven más de una operación.</li> <li><input type="checkbox"/> Adicionan y substraen fracciones de un mismo denominador y conocen números naturales en la forma fraccionaria.</li> <li><input type="checkbox"/> Interpretan gráficos de barras y de sector e identifican el gráfico más adecuado para representar una determinada situación.</li> </ul>	<b>250</b>	<p><b>En el nivel 250 o arriba de él se encuentran:</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 11% de los alumnos del 4° grado de la Enseñanza Fundamental;</li> <li><input type="checkbox"/> 48% de los alumnos del 8° grado de la Enseñanza Fundamental;</li> <li><input type="checkbox"/> 87% de los alumnos del 3er. grado de la Enseñanza Media.</li> </ul>
<ul style="list-style-type: none"> <li><input type="checkbox"/> Los alumnos clasifican sólidos geométricos en cuerpos redondos y poliedros.</li> <li><input type="checkbox"/> Interpretan resultados de medidas de longitud, masa, tiempo y capacidad.</li> <li><input type="checkbox"/> Identifican, comparan y ordenan números racionales (en las formas fraccionaria y decimal) así como números enteros.</li> <li><input type="checkbox"/> Interpretan lenguaje algebraico y resuelven ecuaciones y sistemas de ecuaciones de primer grado.</li> </ul>	<b>325</b>	<p><b>En el nivel 325 o arriba de él se encuentran:</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 8% de los alumnos del 8° grado de la Enseñanza Fundamental;</li> <li><input type="checkbox"/> 32% de los alumnos del 3er. grado de la Enseñanza Media.</li> </ul>
<ul style="list-style-type: none"> <li><input type="checkbox"/> Los alumnos resuelven problemas que envuelven punto, recta, circunferencia y sus relaciones.</li> <li><input type="checkbox"/> Establecen relaciones y hacen conversiones entre fracciones ordinarias y números decimales.</li> <li><input type="checkbox"/> Resuelven problemas que envuelven ecuaciones e inecuaciones sencillas de primer y segundo grado y sistemas de primer grado.</li> <li><input type="checkbox"/> Conocen los principios básicos de polinomios y efectúan operaciones elementales entre ellos.</li> <li><input type="checkbox"/> Conocen las propiedades básicas de exponentes y logaritmos.</li> </ul>	<b>400</b>	<p><b>En el nivel 400 o arriba de él se encuentran:</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 5% de los alumnos del 3er. grado de la Enseñanza Media.</li> </ul>
<p>Observación: Los niveles 325 y 400 muestran el dominio de habilidades y contenidos más complejos que no corresponden al 4° 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 (INEP), 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 esco-

laridad 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 SAEB/97	Matemáticas	Lengua Portuguesa	Ciencias (Física, Química y Biología)
100	No significativo	Hacia <u>la mitad</u> del 1er. ciclo de la Enseñanza Fundamental	Hacia <u>la mitad</u> del 1er. ciclo de la Enseñanza Fundamental
175	Hacia <u>la mitad</u> del 1er. ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> del 1er. ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> del 1er. ciclo de la Enseñanza Fundamental
250	Hacia <u>el final</u> del 1er. ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> del 2 <sup>do.</sup> ciclo de la Enseñanza Fundamental	Hacia <u>la mitad</u> del 2 <sup>do.</sup> ciclo de la Enseñanza Fundamental
325	Hacia <u>el final</u> del 2 <sup>do.</sup> ciclo de la Enseñanza Fundamental	Hacia <u>el final</u> de la Enseñanza Media	Hacia <u>el final</u> del 2 <sup>do.</sup> ciclo de la Enseñanza Fundamental
400	Hacia <u>el final</u> 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 hecho<sup>17</sup>.

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 simi-

lar 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

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 una curricula específica y no como mínimo común denominador de varias diferentes.

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 ítemes 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 resultados sólo 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.

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							
PERÚ							
URUGUAY							

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

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.

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 3er grado", o "tal proporción de los estudiantes domina el objetivo o contenido 'x' de Matemática en 6° 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 currículum -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 currículum 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 diferenciaciones 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 pro-

veniente 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 empírica<sup>18</sup>, 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.

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.

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 superaron 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 aquéllos que no lo tienen” BEATON, A., 1999. En inglés en el original.

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 aún 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 sólo 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?<sup>19</sup>

**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 sobreesimplificada de los resultados, en la medida en que inducen al lector a extraer conclusiones acerca de la efectividad de la enseñanza en los agregados descritos, sin advertir la incidencia de la diferenciación sociocultural 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 éstos a los rurales. El análisis de varianza demuestra y puntualiza que, en tercero y séptimo años, la ventaja de los planteles particulares urbanos sobre los fiscales urbanos y de éstos sobre los rurales es estadísticamente significativa. En el caso de décimo año, no hay diferencia significativa entre fisca-

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.

les 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 éstas 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 éstas 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 columna 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 sólo 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 una población altamente favorecida.

**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 Int. 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.*

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 sólo hay un problema de omisión de la consideración del contexto al reportar resultados descriptivos de los agregados, sino que además se for-

mulan explícitamente conclusiones respecto a la eficacia diferencial de los sistemas educativos de las diferentes regiones, aparece en el SAEB/97.

*“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 estadales 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 las currículas o del trabajo de los educadores.

**El problema es que los diferentes datos no se conectan entre sí. Se analiza por un lado los resultados por jurisdicción, área geográfica y/o tipo de escuela, y por otro lado 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, currículas 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 8° 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
<b>A</b>	287	<b>60,4</b>	66	<b>4,9</b>	8	<b>0,3</b>
<b>B</b>	188	<b>39,6</b>	517	<b>38,5</b>	342	<b>12,3</b>
<b>C</b>	0	<b>0,0</b>	759	<b>56,6</b>	2434	<b>87,4</b>
<b>Total</b>	475	<b>100,0</b>	1342	<b>100,0</b>	2784	<b>100,0</b>

Fuente: República de Chile, Ministerio de Educación, SIMCE, 1996. Resultados 95 - 8º 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”<sup>20</sup>, 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 mismo 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 sólo 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 sociocultural 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				

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 sociocultural “bajo” que pertenecen a escuelas públicas, por un lado, y a 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 espúreas 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 quedará 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 equitativa<sup>21</sup> 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.**

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 terreno<sup>22</sup>.
- 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.

21 Véase BENVENISTE, L., 2000.

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 debiera ser observada, está iniciándose en la Provincia de Buenos Aires, con el apoyo de OREALC.

## Capítulo IV.

### LA UTILIDAD PEDAGÓGICA DE LOS REPORTES 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 ítemes 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 ítemes 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 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 ítemes individuales, cuando éstos 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.

Figura 14  
ARGENTINA. Análisis de ítemes

**7º año EGB / Matemática**

18 Si al descarozar 20 kg de ciruelas 15 kg son de pulpa, ¿qué porcentaje de las ciruelas es la pulpa?

A) 5%  
B) 25%  
C) 35%  
▶ D) 75%

**Respuesta correcta:** D

---

**Resultado:** El 22% de los alumnos lo resolvió correctamente. ▶ 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.

**Contenido:** Números y operaciones.

**Capacidad:** Resolver problemas. ▶ 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.

**Operación requerida y evaluada:** Aplicar el concepto y cálculo de porcentaje para resolver un problema. ▶ 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.

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 disciplinares y didácticos desde los cuáles son elaborados tanto los instrumentos de evaluación como los cuadernos de carácter didáctico.

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 ítemes, que en general van acompañados por un escaso nivel de conceptualización didáctica y disciplinar.

En el caso de Ecuador, por ejemplo, el análisis del

Figura 15  
ECUADOR. Análisis de ítemes

<b>Un automóvil recorre 80 km. en una hora. ¿Cuántos kilómetros avanzará en una hora y tres cuartos si mantiene la misma velocidad?</b>			
	<b>Opciones</b>	<b>% de elección</b>	<b>Observaciones</b>
<b>A.</b>	<b>160 km.</b>	<b>22.20</b>	
<b>B.</b>	<b>140 km.</b>	<b>34.73</b>	
<b>C.</b>	<b>120 km.</b>	<b>28.35</b>	no tienen claro lo que se pide.
<b>D.</b>	<b>60 km.</b>	<b>6.94</b>	
ERROR: identificación incorrecta de la información.			
DIFICULTAD IDENTIFICADA: insuficiente comprensión de la proporcionalidad y poca capacidad de análisis.			

<b>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?</b>			
	<b>Opciones</b>	<b>% de elección</b>	<b>Observaciones</b>
<b>A.</b>	<b>45 personas</b>	<b>12.89</b>	
<b>B.</b>	<b>30 personas</b>	<b>14.43</b>	
<b>C.</b>	<b>20 personas</b>	<b>40.66</b>	
<b>D.</b>	<b>6 personas</b>	<b>21.09</b>	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.

#### 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 evaluaciones de carácter censal, por lo que todos los establecimientos educativos de la provincia o estado reciben sus resultados<sup>23</sup>.

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

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.

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.

**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. Reúnase 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, 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 Colegiado y de la Comunidad".*

Fuente: Gobierno del Estado de Paraná, 1998. Evaluación del Rendimiento Escolar. Resultados de la Evaluación del Rendimiento Escolar. 4ta. Serie de la Enseñanza Fundamental.

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 ítemes 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 I, 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 ítemes 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 setiembre de 1999 el Ministerio de Educación comenzó a hacer públicas las pruebas que habían sido aplicadas en 1998<sup>24</sup>.

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 3er. 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<sup>to</sup> 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.

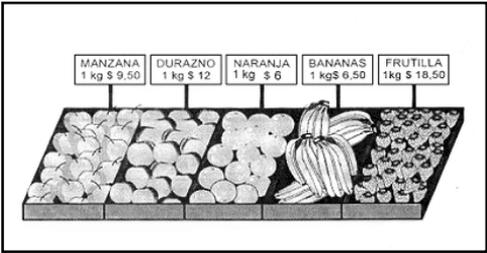
La principal fortaleza de esta estrategia es que involucra fuertemente a los maestros, dándoles participación directa en la evaluación y propiciando el desarrollo de una cultura de la evaluación de los aprendizajes con instrumentos sistemáticamente validados. Sus debilidades son que no hay garantías 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.

---

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 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 limite de gasto.	1
3	Nombra 3 kilos de una sola fruta (naranja, o banana) respetando el limite 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, naranja, manzana"

**Código 5** "naranja, banana, frutilla, durazno, manzana, uva, pomelo, zandia, 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.

#### 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 cualitativo<sup>25</sup>.

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 sólo 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.
- 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 informa-

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).

ción 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 promisorio 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 REPORTE DE RESULTADOS COMO APOORTE 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 currículas, 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 suele 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 saliente, 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...”*<sup>26</sup>.

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 ítemes que presentaron las mayores dificultades.

26 Ministerio de Educación Pública, Programa de Pruebas Nacionales. Informe nacional sobre los resultados de las 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.

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 Ecuador, 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 .

## **Recuadro 2** **ECUADOR. Reiteración de conclusiones en informes de años sucesivos**

APRENDO 96: "Resaltan el cuadro de rendimiento en Matemática pues en éste 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 éste 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 éste 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.

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.

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 *Creceer*. Éstos 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 presenta 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 datos<sup>27</sup>.

Obviamente, es discutible si la función de analizar la información corresponde a los sistemas de evaluación o si éstos deben limitarse, como en general ha ocurrido hasta ahora, a la producción de los datos. Asumiendo que sí deberían entregar información más

27 Los boletines *Creceer* pueden ser encontrados en: [www.minedu.gob.pe/web/eL\\_ministerio](http://www.minedu.gob.pe/web/eL_ministerio)

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 se detectan los informes nacionales:

- 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, o bien 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.

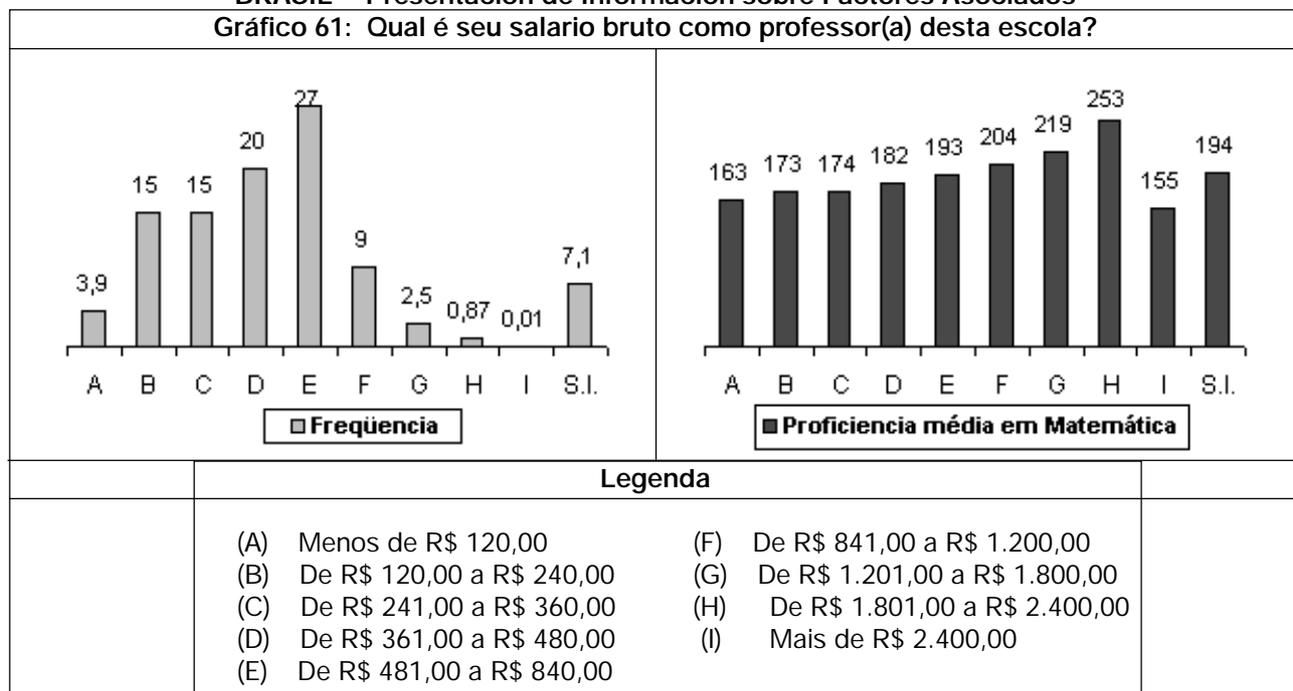
Para ilustrar este último problema es útil recurrir al Informe del SAEB/97. Por un lado, éste reconoce explícitamente la existencia de este tipo de problemas y los cuidados que deben tenerse para no simplificar la interpretación de los datos:

*“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 - Primeiros 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 (véase 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).*

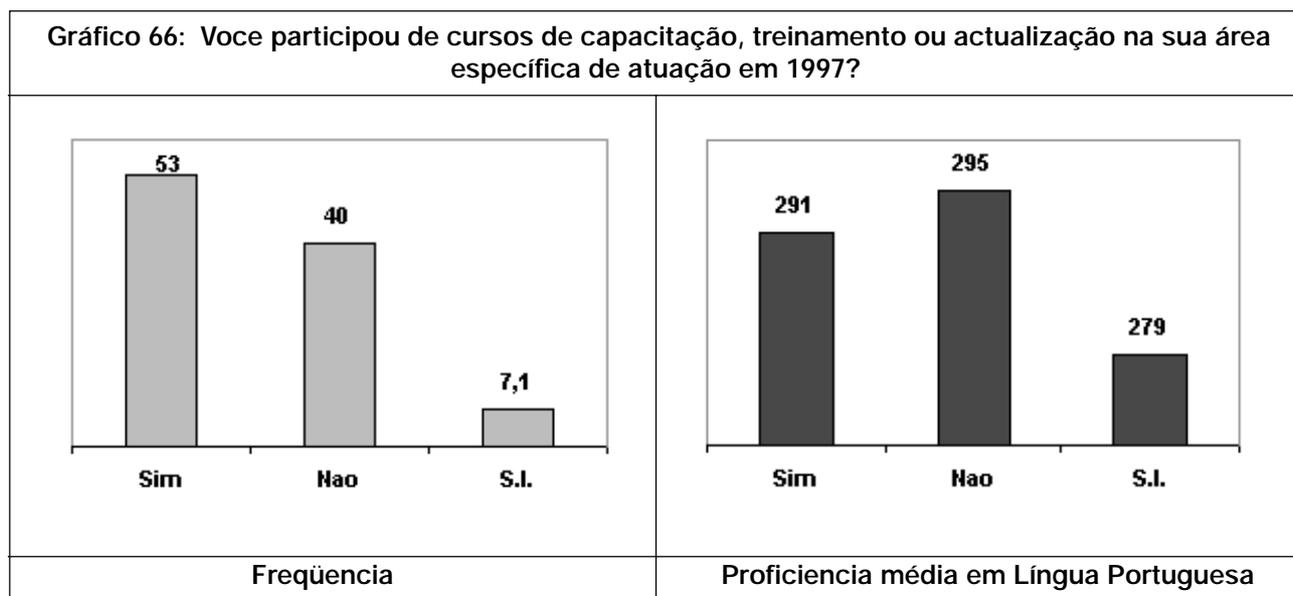
**Figura 18**  
**BRASIL - Apresentação de Informação sobre Factores Associados**  
**Gráfico 61: Qual é seu salário bruto como professor(a) desta escola?**



Fuente: Ministerio de Educación - Gobierno Federal. Instituto Nacional de Estudios e Investigaciones Educativas (INEP), 1998. *SAEB 97 - Primeros Resultados.*

**Figura 19**  
**BRASIL - Apresentação de Informação sobre Factores Associados**

**Gráfico 66: Voce participou de cursos de capacitação, treinamento ou atualização na sua área específica de atuação em 1997?**



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 alumnados 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 brasileiro, principalmente aquellos relacionados con la equidad”<sup>28</sup>.*

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 posterioridad<sup>29</sup>.

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 PREAL<sup>30</sup>. 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”.

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.

30 Véase al respecto Ravela, P., 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 currículum 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. Aún 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 currículum 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 formas 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 ítemes 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 éstos con lo que otras investigaciones 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 tendríamos 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.





**Ve. A. Ce.**

Teléfono: (56-2)2387089 - Fax: (56-2)2389363

E-mail: [info@veace.cl](mailto:info@veace.cl)

Santiago - Chile