

¿QUÉ MIDE REALMENTE EL SIMCE?

Bárbara Eyzaguirre y Loreto Fontaine

Una evaluación estandarizada, masiva y con resultados públicos, como es nuestro SIMCE (Sistema de Medición de la Calidad de la Educación), puede tener efectos positivos o negativos sobre la calidad de la enseñanza. El presente estudio analiza la información disponible sobre el SIMCE, su proceso de construcción y validación, así como los contenidos incluidos en esta prueba y las señales que ellos emiten para el quehacer educativo.

De este estudio se concluye que los progresos (expresados en porcentajes de logro) que se han evidenciado a lo largo de los años de aplicación del SIMCE no son fácilmente interpretables, ya que los niveles de dificultad de la prueba no son absolutos sino variables (es decir, sufren ajustes de acuerdo a la validación experimental que

BÁRBARA EYZAGUIRRE A. Psicóloga educacional especializada en desarrollo cognitivo, con experiencia en programas de mejoramiento de la calidad de la educación en sectores de escasos recursos. Fundadora y asesora pedagógica de la Escuela San Joaquín (Renca), perteneciente a la Fundación Marcelo Astoreca. Investigadora del Centro de Estudios Públicos.

LORETO FONTAINE C. Master en educación (M. S.) Se ha desempeñado como profesora de enseñanza básica, en programas de rehabilitación de problemas de aprendizaje y como coordinadora académica en los colegios Los Andes y Tabancura. Ha asesorado proyectos de biblioteca y lectura de la Sociedad de Instrucción Primaria (Escuelas Matte) y Fundación Los Nogales. Ha publicado antologías de cuentos infantiles. Investigadora del Centro de Estudios Públicos.

recibe la prueba cada año). A su vez, se advierte que la técnica de construcción de la prueba la hace perder validez de contenido, al bajar demasiado el nivel de dificultad. El análisis del contenido de las pruebas revela que, en promedio, su complejidad corresponde a dos y tres grados menores que el cursado por los alumnos. También queda de manifiesto una falta de coordinación entre las exigencias de básica y media, falencias en la formulación de los ítems y en la selección de las materias.

Conforme a las deficiencias encontradas, se sugiere un cambio en la teoría de construcción de la prueba y un mayor celo en su elaboración, de modo que refleje las metas educacionales que el país se ha propuesto y que a la vez permita discriminar entre las escuelas. Se propone, asimismo, avanzar hacia una política de total transparencia en las metodologías y contenidos del SIMCE. Finalmente, se señala que debiera mantenerse la política inicial de que organismos independientes del Estado se encarguen de la prueba.

En 1998, en el afán de actuar con transparencia y de suscitar el análisis y el debate sobre materias educacionales, el Ministerio de Educación dio a conocer parte de las preguntas del último SIMCE administrado a los alumnos de 8° año básico el año 1997. Hasta ese momento, los educadores habían conocido sólo los resultados y contaban con escasos ejemplos de las preguntas que contemplaba la prueba. El paso dado por el Ministerio de Educación de publicar ahora las preguntas es una iniciativa que, sin duda, será fructífera, pues por primera vez se da a padres, profesores y a la comunidad en general la oportunidad de conocer cabalmente este instrumento que no solamente evalúa nuestra educación, sino que tiene una profunda influencia sobre la calidad de ésta. La mayor transparencia así lograda será un paso adelante para el logro de una mejor calidad. La información aportada permite interpretar con mayor claridad el significado de los puntajes obtenidos a nivel del país entero y a nivel de los establecimientos en particular. Por otra parte, debiera ayudar a visualizar mejor cuáles son los objetivos esperados por el sistema educacional y permitir a los colegios identificar con precisión las áreas deficitarias que deben ser corregidas.

Simultáneamente a la divulgación de esta información, y con el fin de adecuarlo a las nuevas orientaciones curriculares, el Ministerio de Educación se abocó al rediseño del SIMCE y distribuyó, a modo de orientación

para los educadores, alguna información sobre lo que sería la nueva generación de pruebas, con ejemplos de las preguntas que se aplicarían ese año en la evaluación de 2º medio. Este estudio analiza los datos disponibles sobre el SIMCE 1997 y 1998 y sus efectos sobre el quehacer educacional.

Influencia de las pruebas estandarizadas en la educación

Si bien Chile es uno de los pioneros en este tipo de mediciones en Latinoamérica¹, la aplicación masiva de pruebas de rendimiento destinadas a medir los logros de aprendizaje obtenidos por el sistema escolar es una práctica común en un gran número de países². Estas mediciones se realizan con diversos objetivos: monitorear la efectividad del sistema y diseñar políticas pertinentes; evaluar la utilización de recursos y la aplicación de medidas; detectar sectores más débiles y diseñar estrategias correctivas. En algunos sistemas educacionales estas evaluaciones cumplen también el papel de seleccionar a los alumnos para las diferentes vías que ofrece el sistema. En otros casos, aunque los resultados no incidan en la suerte académica de los alumnos, la información de los resultados de cada colegio estimula entre ellos la competencia y permite a los usuarios del sistema (padres y alumnos) escoger las alternativas más adecuadas a sus intereses. Estas pruebas permiten también la creación de sistemas de incentivos: existen ciertas iniciativas consistentes en premiar a los colegios con buen desempeño con aumentos en el aporte económico o, incluso, en condicionar parte del financiamiento de las escuelas según los resultados de tales evaluaciones³.

Estas evaluaciones ejercen gran influencia en el proceso de enseñanza-aprendizaje y en ellas se han cifrado altas esperanzas de mejoramiento de la calidad de la educación. Hay también una corte de detractores que sostienen que estas evaluaciones suelen ser injustas y que tienen el poder de degradar la enseñanza.

Para muchos, la evaluación estandarizada y transparente sería una de las herramientas fundamentales para mejorar la calidad de la educación: el que los profesores y los colegios en cierto modo rindan cuentas sobre el

¹ Las primeras evaluaciones de este tipo en Chile fueron en 1968.

² Véase B. Eyzaguirre, "Políticas educacionales comparadas: Consideraciones a partir del Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)" (1999).

³ Véase D. Koretz, "Using Student Assessments for Educational Accountability" (1996).

aprendizaje de los alumnos conduce naturalmente a que concentren e intensifiquen sus esfuerzos en tal sentido. Las pruebas estimularían a los colegios a superarse y a competir por su prestigio o por su financiamiento. También señalarían a los profesores en forma precisa el tipo de competencias que debe exigirse a los estudiantes y las metas de aprendizaje a las cuales se debe llegar. Por otra parte, son instrumentos de costo relativamente bajo, cuya aplicación no requiere perder una gran cantidad del tiempo destinado a la instrucción, y con un inmenso potencial de afectar la calidad de la enseñanza. Además, su utilización extendida por largo tiempo y en forma masiva las hace científicamente confiables⁴. Las investigaciones realizadas en Estados Unidos e Inglaterra, donde este tipo de evaluaciones tienen una larga trayectoria, coinciden en que éstas:

- tienen una influencia directa en la definición de los objetivos de estudio, tanto o más que los currículos diseñados para ello. Dan señales concretas de los logros que el sistema espera (Prophet y Rowell, 1988);
- normalmente focalizan la enseñanza hacia el aprendizaje de contenidos y destrezas básicas. La reducción de objetivos puede implicar una degradación de la enseñanza, pero también puede incidir positivamente. Hay evidencia de que las naciones con currículos focalizados tienen mejores rendimientos⁵ y que los alumnos menos aventajados se benefician con la enseñanza dirigida a objetivos acotados, porque aumentarían el tiempo dedicado a la completación y comprensión de la tarea (Gickling y Thompson, 1985);
- establecen parámetros de comparación que desincentivan el conformismo, siempre y cuando los estándares sean altos;
- estimulan a los alumnos al estudio cuando son utilizadas para discriminar el acceso a las diferentes instancias educativas;
- ayudan al proceso de asimilación de las materias. Las pruebas no sólo evalúan sino que promueven el aprendizaje. Se ha observado que la retención de una materia aumenta sustancialmente en los alumnos que han sido examinados sobre esa materia, aunque no reciban retroalimentación ni vuelvan a estudiar esa materia (Dempster, 1997);

⁴ Véase N. F. Mc. Ginn y A. M. Borden, *Framing Questions, Constructing Answers. Linking Research with Education Policy for Developing Countries* (1995), p. 154.

⁵ Véase B. Eyzaguirre, "Políticas educacionales comparadas: Consideraciones a partir del Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)" (1999).

- proveen retroalimentación objetiva a los establecimientos de los aciertos a mantener y de los errores a corregir;
- son utilizadas por los padres para ejercer fuertes presiones en los equipos de gestión de los colegios. Los directivos presionan a su vez a los profesores para lograr buenos puntajes (Hanushek, 1986)⁶.

A pesar de los efectos positivos que se esperan de este tipo de evaluación masiva estandarizada del desempeño de los alumnos, existen también consecuencias negativas que finalmente se pueden traducir en una degradación de la instrucción⁷:

- La enseñanza se puede reducir, enfocándose sólo a lo que se puede medir con facilidad y dejando de lado aspectos relevantes de más difícil evaluación (Williams, 1998). Para algunos estas pruebas no son un reflejo de las habilidades que exige la vida real, es decir, evalúan aspectos irrelevantes de la formación de los alumnos. Se acusa a los tests de dirigir a los alumnos hacia un tipo de pensamiento correcto-incorrecto, a entregar respuestas simples, en vez de orientarlos hacia el desarrollo del juicio reflexivo y fundamentado o hacia la solución creativa de problemas. Parte de esta crítica se refiere al formato de preguntas prevalente en la mayoría de las pruebas que se usan en la actualidad: preguntas de respuesta breve, del tipo ‘correcto-incorrecto’ o ‘seleccione la alternativa’. Si bien las investigaciones demuestran que es más fácil construir ítems que miden memorización de datos, esto no implica que las pruebas de alternativas no puedan evaluar habilidades de pensamiento superior. Fredericksen, en “The Real Test Bias”, demuestra que las pruebas de razonamiento matemático pueden medir los mismos atributos independientemente del formato que se utilice para hacer las preguntas⁸.
- También se puede limitar la instrucción sólo a lo que se sabe va a ser evaluado. Para entender este efecto es necesario recordar que las pruebas estandarizadas por lo general son sólo una muestra, una selección de todo el contenido que puede abarcar una asignatura o un campo de conocimientos. Se ha detectado que los profesores

⁶ Véase también el caso de Tailandia en N. F. Mc. Ginn y A. M. Borden, *Framing Questions, Constructing Answers. Linking Research with Education Policy for Developing Countries* (1995), p. 227.

⁷ Véase E. A. Hanushek y D. W. Jorgenson, *Improving America's Schools: The Role of Incentives* (1996).

⁸ Véase en E. D. Hirsch, *The Schools We Need: Why We Don't Have Them* (1996), p. 200.

tienden frecuentemente a limitar la enseñanza a aquellos temas dentro de una asignatura que se sabe serán evaluados, disminuyendo así, de hecho, las oportunidades de aprendizaje de sus alumnos. Cuando los profesores están enterados de cuáles son las habilidades que se evaluarán pueden sobreejercitar estas destrezas, en desmedro de otras igualmente importantes. Las asignaturas no evaluadas también son desatendidas: los profesores tienden a destinar las horas que corresponden a esas asignaturas a trabajar en las asignaturas evaluadas⁹. Por otra parte, ocultar o retener la información sobre el contenido de las pruebas resulta contraproducente, pues esa información es condición necesaria para que los profesores puedan mejorar o al menos focalizar mejor su enseñanza. Uno de los axiomas que fundamentan la evaluación como agente de mejoramiento educacional es que los ítems o tareas de la prueba dan a los maestros un modelo acerca del tipo de competencias que debe exigirse a los estudiantes¹⁰. El justo equilibrio se encontraría, entonces, en proporcionar una información amplia sobre el tipo de tareas, habilidades y conocimientos requeridos en las pruebas y luego examinar una fracción significativa de ello.

- Puede también darse una ejercitación excesiva del test mismo en desmedro de la enseñanza de la materia. (Por ejemplo, cuando se asigna tiempo de clases para que los alumnos hagan ensayos con facsímiles.) Smith (1991) constató que el tiempo destinado por los alumnos de EE.UU. a la preparación y ejecución de pruebas estandarizadas tomaba un tiempo promedio de 100 horas de las 1.000 horas lectivas anuales. Esta proporción es más significativa si se considera que el tiempo efectivo de enseñanza, para ese mismo período, es sólo de 300 horas en promedio (Weinstein y Mignano, 1993). El entrenamiento en la lógica de las pruebas puede conducir a aumentos de puntajes que no se condicen con lo realmente aprendido. Williams *et al.* (1996) comprobaron que los alumnos que aprenden estrategias específicas para responder pruebas mejoraban en forma permanente su rendimiento en todo tipo de pruebas. Un sistema adecuado de supervisión de clases y una buena construcción de los instrumentos contribuirían a paliar este problema.
- También se sostiene que las pruebas pueden reflejar mejoramientos espurios de los resultados y, por lo tanto, conducir a un conformis-

⁹ D. Koretz, "Using Student Assessments for Educational Accountability" (1996).

¹⁰ *Ibidem*.

mo innmercido y de efectos perniciosos en la calidad. Esto sucede si los estándares son bajos o si no se atiende a los posibles efectos ya descritos de limitación de la instrucción.

- Otra vertiente de crítica se refiere a que las pruebas estandarizadas tienen un sesgo cultural y utilizan un vocabulario propio del grupo social dominante y, por ello, discriminan negativamente a los niños de niveles socioeconómicos bajos y con carencias culturales (M. Neill, 1994). Esta crítica sería válida solamente cuando se quiere condicionar los incentivos económicos a los diferentes establecimientos según su desempeño en una prueba, o cuando los resultados de las pruebas se traducen en un cierre de oportunidades educacionales para los alumnos con resultados insatisfactorios. En esas circunstancias es importante considerar las ganancias relativas de puntajes tomando en cuenta los distintos niveles de entrada de los niños. Pero cuando se desea tener una imagen exacta y acertada del nivel de desempeño que ha alcanzado la totalidad de los alumnos de un sistema, es necesario reconocer que las diferencias en el nivel de comprensión del lenguaje puestas en evidencia por estas pruebas, son el reflejo real de la brecha que existe entre los diferentes niveles socioeconómicos. Esta distancia tiene que ser reconocida como tal para ser compensada por la vía de la educación.
- Finalmente, hay quienes sostienen que un sistema de evaluación externa crea en la escuela un clima de competencia y ansiedad contrapuesto al placer de aprender por aprender e incluso a un desarrollo emocional sano.

Los efectos perversos de evaluar masivamente con pruebas objetivas y estandarizadas pueden disminuirse con un cuidadoso diseño de estas pruebas, con una adecuada supervisión de clases y con políticas educacionales sensatas. Los inconvenientes de la aplicación masiva de evaluaciones se potencian cuando se asigna a estas evaluaciones un papel que no les corresponde. Por ejemplo, cuando ellas se transforman en el *único* instrumento para tomar decisiones críticas acerca de financiamiento, promoción de profesores o para determinar el futuro académico de los alumnos. También, cuando no están claramente establecidas las metas educacionales de cada etapa, las pruebas pueden llegar a ser la única herramienta para determinar el currículo e incluso el contenido de los libros de texto. Esto da a las pruebas un peso desmesurado en el quehacer educacional. Por último, cabe decir que las pruebas de evaluación alcanzan su máximo potencial de utilidad sólo cuando forman parte de un sistema coherente con metas educacionales claras y bien especificadas, donde hay una adecuada y transparente

comunicación acerca de los objetivos de la medición, las características del instrumento y los resultados obtenidos, y complementación con otros métodos para asegurar la calidad del sistema.

En la actualidad no hay pruebas que eviten todos los inconvenientes y que maximicen los efectos positivos. Elaborar una buena prueba objetiva no es fácil, pero tampoco se trata de una tarea imposible. La reciente aplicación en nuestro país de la prueba internacional TIMSS nos ha dado a conocer ciertamente un ejemplo de evaluación bien construida: por su equilibrio entre la medición de habilidades, conocimientos y procedimientos; por la amplitud del rango de las materias medidas y por la adecuada proporción de ellas; por la claridad en la construcción de las preguntas; por la combinación de preguntas de selección múltiple con preguntas abiertas y preguntas que evalúan desempeño en tareas complejas; por la diferente ponderación dada según el nivel de dificultad de las preguntas¹¹.

Los intentos por encontrar formas alternativas a las clásicas pruebas objetivas de lápiz y papel, como lo son las evaluaciones de desempeño y las de portafolios¹², se han topado con un sinnúmero de desventajas: problemas con los grados de objetividad que pueden alcanzar, altos costos de corrección y, por último, no han tenido los efectos educativos que se les atribuían. Por lo tanto, el debate acerca de los sistemas de evaluación y la búsqueda de mejores instrumentos continuará por un tiempo.

Evaluar pensamiento reflexivo, razonamiento y la capacidad de enfrentar tareas de la vida real es más fácil y posible en la asignatura de matemáticas que en lenguaje. Sin embargo, hay evidencia de que, aun en esta área, se puede alcanzar en las pruebas un alto grado de validez (medir lo que se propone medir) y confiabilidad (distintas mediciones, con el mismo instrumento, entregan los mismos resultados). Al respecto, podemos citar una investigación que demuestra que las pruebas de alternativas pueden medir válida y confiablemente la capacidad de redactar, que es una tarea de gran complejidad, propia de la vida real y que requiere pensamien-

¹¹ Véase B. Eyzaguirre, "Políticas educacionales comparadas: Consideraciones a partir del Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)" (1999).

¹² Las evaluaciones de desempeño se refieren a pruebas que exigen a los alumnos la ejecución de tareas concretas para ser calificados; por ejemplo, diseñar y llevar a cabo un experimento científico. Las evaluaciones de portafolios se refieren a la aplicación de un método tradicionalmente utilizado en la enseñanza del arte, al resto de las asignaturas. Éste consiste en acumular todos los trabajos que el alumno realiza durante un largo período y analizar los avances que ha tenido con respecto a sus primeras producciones. Para una revisión de los problemas de confiabilidad que presenta este tipo de evaluación, véase Koretz *et al.*, "The Vermont Portfolio Assessment Program: Findings and Implications" (1994), pp. 5-16.

to reflexivo. En esta área normalmente se obtiene una baja correlación (0,25) entre las correcciones de distintos jueces sobre un mismo escrito. Para aumentar la confiabilidad de las mediciones se exigió que cinco jueces corrigieran 5 ensayos escritos por cada uno de los estudiantes que participaron en la investigación. Se evaluó más de un ensayo por alumno, porque se sabe que la habilidad para redactar está mejor representada cuando se consideran distintos temas. De este modo, la producción de cada alumno recibía 25 lecturas. Al aumentar el número de personas que revisan las pruebas y el número de ensayos por alumno el coeficiente de correlación entre los evaluadores subió a un 0,84. Esto refleja que entre expertos habría un alto grado de acuerdo sobre lo que se considera un buen escrito. Luego se comparó el resultado de esta prueba, que exigía redactar directamente, con una prueba de alternativas múltiples y se obtuvo una correlación de 0,775. Esto demuestra que la prueba de alternativas múltiples puede ser un buen indicador de la destreza de redacción, aun cuando las tareas que exige no correspondan a la tarea real de escribir. Si bien la prueba de selección múltiple demostró tener un alto grado de eficiencia y bajo costo, los investigadores recomendaron agregar un ítem que exigiera redactar directamente, a pesar de que éste aumentaba la correlación sólo del 0,775 al 0,784. Esto se propone con el fin de dar la señal de que es necesario que los alumnos redacten en clases para aprender a escribir y así evitar que los profesores se dediquen sólo a preparar a los niños para los ítems de alternativas¹³. Citamos esta experiencia porque nos indica cómo la investigación experimental puede orientar la elaboración de instrumentos de evaluación más precisos y eficaces sin los inconvenientes que normalmente se asocian a ellas. Hasta el momento las pruebas que combinan preguntas de alternativas múltiples y preguntas abiertas son las que mejor resolverían el problema de reflejar aprendizajes significativos, a bajo costo y con un alto grado de confiabilidad.

Pese a las críticas enunciadas, pocos se atreverían a descartar la evaluación como un medio para mejorar la educación. H. Gardner, creador de la teoría de las inteligencias múltiples, sostiene que el enigma de la reforma educativa es intrincado y que, de hecho, depende por igual de cuatro núdulos diferentes: evaluación, currículo, educación del educador y apoyo de la comunidad¹⁴.

¹³ En E. D. Hirsch, *The Schools We Need: Why We Don't Have Them* (1996), p. 186.

¹⁴ H. Gardner, *La mente no escolarizada: Cómo piensan los niños y cómo deberían enseñar las escuelas* (1997).

Las evaluaciones estandarizadas en Chile

El interés por las evaluaciones de desempeño escolar a nivel nacional se inicia durante el período de la reforma educacional de los años 60. En 1968 se crea la primera Prueba Nacional dirigida a los 8° básicos, la que se aplicó anualmente hasta el año 1971. El propósito de estas pruebas era aportar información para el proceso de desarrollo curricular y disponer de parámetros para mejorar la asignación de recursos.

Esta política de evaluación se vio interrumpida hasta el año 1982 en que se inicia el Programa de Evaluación del Rendimiento Escolar (PER) para los cursos de 4° y 8° año de enseñanza básica. Uno de los fines del PER era entregar datos sobre el rendimiento escolar del establecimiento y de cada alumno al equipo educativo del colegio y a los padres. Esta información se consideraba una de las herramientas imprescindibles para poner en funcionamiento el sistema de voucher¹⁵, en el cual los padres y las escuelas pasarían a tener un peso importante en la toma de decisiones educacionales. Este sistema sólo tendría efectos sobre el mejoramiento de la calidad de la educación, si la decisión de los padres de cambiar de colegio a los alumnos tenía un costo para el establecimiento. El PER pretendía entregar a los padres información externa y confiable sobre los logros de las escuelas para basar estas decisiones. Este proyecto se realizó mediante un convenio del Ministerio de Educación con la Universidad Católica que estuvo a cargo del diseño y ejecución del proyecto. En ese entonces, se consideró importante que un organismo externo realizara las evaluaciones nacionales de modo que el Ministerio de Educación no hiciera el papel de juez y parte del proceso. La publicación de los resultados en los medios de comunicación y la entrega de los resultados por alumno encontraron fuertes resistencias al interior del Ministerio. El argumento principal se centraba en que se estaba juzgando la labor del profesor sin tomar en cuenta las diferencias de entrada de los alumnos de diversos niveles socio-culturales. Recién el año 1995 los resultados conocieron la luz pública por primera vez.

En el período 1985 y 1986 el Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, dependiente del Ministerio de Educación, se hizo cargo del sistema de evaluación y pasó a llamarse

¹⁵ El sistema de voucher consiste en financiar la educación pública y particular subvencionada a través de un bono que se otorga a los establecimientos por cada alumno matriculado. El monto del bono se determina por el número de días que asiste el alumno a clases. Si el alumno se cambia de establecimiento el bono que le corresponde se traslada con él.

Sistema de Evaluación de la Calidad de la Educación (SECE). En este lapso no se realizaron evaluaciones nacionales y la tarea se restringió a la elaboración de los datos recogidos anteriormente por el PER.

El actual Sistema de Medición de la Calidad de la Educación (SIMCE) empieza a operar en 1988 mediante un nuevo convenio del Ministerio de Educación con la Universidad Católica. En 1992 el Ministerio de Educación asume la responsabilidad total del SIMCE. La Ley Orgánica Constitucional de Enseñanza (LOCE) de marzo de 1990 establece que el Ministerio de Educación Pública es el organismo encargado de diseñar los instrumentos de evaluación periódica. Para cumplir con esta normativa se establece el organismo SIMCE como una estructura funcional, técnica, dependiente de la Subsecretaría de Educación.

Características de la actual prueba SIMCE

El SIMCE es un conjunto de pruebas e indicadores de carácter censal que se aplica anualmente en forma alternada a los alumnos de 4° y 8° año básico. En 1997 en la prueba de 8° fueron evaluados 217.000 estudiantes. Cabe hacer notar que todas aquellas escuelas que tienen menos de cinco alumnos en el nivel a evaluar, no participan en la medición. En educación básica esto representa algo más de tres mil escuelas rurales, las cuales suman alrededor de quince mil alumnos. En educación media se aplica al final del Segundo Año; en este ciclo se han realizado sólo dos pruebas, la primera en 1994 y la segunda en 1998.

Estas evaluaciones mantienen los objetivos del antiguo PER, esto es, obtener información acerca del nivel de logros del sistema educacional, proveer a los directivos de las escuelas de retroalimentación externa sobre su labor docente y entregar a los apoderados datos confiables acerca del desempeño del establecimiento.

Sin embargo, con la incorporación de la Reforma Educacional se definen algunas nuevas funciones como la identificación de las escuelas con rendimientos deficientes para entregar ayuda focalizada y las de rendimiento sobresaliente para instituir un sistema de incentivos. El Programa de las 900 Escuelas consiste en asignar asesoría pedagógica y recursos materiales a las escuelas con peor puntaje en el SIMCE y el Premio de Excelencia se otorga a los profesores de aquellas escuelas que han logrado buenos resultados en un conjunto de indicadores, entre los cuales el buen desempeño en el SIMCE es uno de los más importantes. Otro objetivo que debiera lograr esta evaluación es entregar datos acerca del impacto que ha

producido el conjunto de medidas emprendidas por la reforma. Obtener esta información no es posible aún, pues como veremos a continuación la naturaleza de la prueba no permite hacer comparaciones de un año a otro mientras no se establezca un sistema que las haga comparables.

A los tradicionales índices de eficiencia del sistema que medía el PER, es decir, porcentaje de logro en las pruebas académicas y tasas de repitencia y deserción, se incorporaron indicadores que van más allá de la formación netamente académica. Estos apuntan a la formación general del alumno como es la medición de la autoestima y a factores que muestran el grado de satisfacción de los participantes del proceso educativo (grado de aceptación de la labor educacional de padres, alumnos y profesores). En el futuro, el SIMCE planea integrar a la evaluación los objetivos transversales incluidos en la reciente reforma. Es así que para el futuro se contempla la posibilidad de sondear el pleno desarrollo de la personalidad individual y social del alumno, la formación ciudadana, el ejercicio de la tolerancia y libertad, la cooperación y la solidaridad.

También se está considerando la posibilidad de recopilar información acerca de las variables que tienen relación con los logros escolares y que requieren ser controladas para permitir una comparación justa de la gestión educativa del establecimiento, a saber: capital cultural y económico de la familia; gasto por alumno; participación en programas remediales, etc. La tendencia actual en evaluación es poder determinar el valor agregado de la gestión educacional por parte de cada establecimiento.

En la actualidad, en el nivel académico, las pruebas miden aprendizaje en las áreas de Lenguaje, Matemáticas, Ciencias Naturales y Ciencias Sociales. Cada una de ellas contiene alrededor de 60 preguntas de selección múltiple, las que deben ser respondidas en 90 minutos.

Los resultados se expresan desagregados por establecimiento y no por alumno. Los resultados individualizados se entregaban en el antiguo PER, pero esta medida se descontinuó en parte por costos y en parte porque se argumenta que los directores de establecimientos podrían tomar medidas contra los niños que obtenían bajos puntajes. Sin embargo, el hecho de que la prueba sea anónima quita toda responsabilidad a los alumnos y esto puede tener como consecuencia una falta de compromiso que se puede traducir en una baja en los resultados. El rendimiento final no reflejaría entonces el nivel real de conocimientos de los alumnos. Con esta política además se desperdicia la oportunidad de que los alumnos se apliquen más a sus estudios a partir de la presión de la prueba.

Los informes que se entregan a cada colegio incluyen información acerca del porcentaje promedio de respuestas correctas obtenidas por el

colegio en cada prueba y un desglose de estos resultados en términos de los porcentajes de respuestas correctas alcanzados en cada uno de los objetivos de cada asignatura. Se entregan también índices que permiten comparar al establecimiento con el promedio general del país y con los de los establecimientos de nivel sociocultural similar. A partir de 1997 se está entregando un nuevo indicador, el PALS, que se refiere al porcentaje de alumnos que logra sobre el 70% de respuestas correctas. Persigue entregar, de un modo más comprensible, información acerca de la dispersión de resultados.

Los padres tienen acceso a los puntajes generales a través de los medios de comunicación quedando al criterio de cada colegio el informar con más detalle a los apoderados. El PER entregaba una carta por alumno con la información individual y del establecimiento la cual debía hacerse llegar a cada apoderado. La difusión actual de la naturaleza de la prueba y de sus resultados es adecuada a nivel de establecimientos educacionales, pero no lo es a nivel de apoderados y de público en general. Tampoco lo es a nivel de investigadores: las bases de datos no están disponibles y los aspectos técnicos de la construcción de la prueba no han sido expuestos para su discusión. Una mayor difusión, disponibilidad de los datos y transparencia metodológica es necesaria.

Construcción de la prueba

Los contenidos y destrezas a evaluar se determinan de acuerdo a los planes y programas de estudios vigentes, por lo tanto en el período actual de reforma curricular, el SIMCE está sometido a una serie de ajustes¹⁶. Según la mayor o menor cercanía a los objetivos de la reforma curricular, se habla de instrumentos de primera, segunda o tercera generación:

- Las *pruebas de primera generación* son las que se aplicaron hasta 1997. Su propósito era la medición de los contenidos del currículo definido en los decretos 4002 y 300 para Educación Básica y Media, respectivamente, y vigentes desde 1980. Las preguntas se han compuesto exclusivamente de ítems de selección múltiple.

¹⁶ Los nuevos Objetivos Mínimos y Contenidos Fundamentales datan del año 1996. A partir de éstos se elaboraron los planes y programas de estudio para cada ciclo. Los de 1° a 4° año básico fueron publicados en 1996 y 1997, los de 5° básico en 1998 y los restantes se publicarán en los próximos años. Por lo tanto, una evaluación de la aplicación de estos programas tiene que considerar el tiempo necesario para que su aplicación haya abarcado un ciclo completo.

- Las *pruebas de segunda generación* también miden los contenidos y objetivos definidos en el antiguo currículo, pero excluyen aquellos considerados prescindibles en relación a las nuevas propuestas curriculares. Además de los ítems de selección múltiple se consideró necesario agregar preguntas con respuestas abiertas, orientadas a la medición de aprendizajes más complejos, respondiendo así al cambio de orientación curricular cuyo foco es el desarrollo de destrezas y de habilidades de pensamiento superior más que la adquisición de conocimientos.
- Las *pruebas de tercera generación* medirán solamente los objetivos propuestos en el nuevo currículo (Decreto 40 de Educación Básica de 1980 y Decreto 200 de Educación Media de 1980). Supuestamente se compondrán de ítems de selección múltiple, así como también de preguntas de respuesta abierta, orientados, en una alta proporción, a la medición de aprendizajes complejos.

Interpretar a través de preguntas los objetivos fundamentales y contenidos mínimos que aparecen en nuestro nuevo currículo de estudio, no es tarea fácil, ya que éstos no están redactados como objetivos de logro evaluables. Muchas veces no es posible una interpretación unívoca de los mismos, por lo tanto al SIMCE le cabrá la responsabilidad de fijar qué se entiende por cada uno de los objetivos y determinar los reales niveles de aprendizaje esperados. Por otra parte, el nuevo currículo incluye entre sus objetivos muchos de carácter instrumental, es decir, incorpora objetivos que fuerzan el uso de determinadas metodologías para aprender, con la esperanza de impulsar desde allí la renovación pedagógica. Por ejemplo, el contenido de 4° básico, “conocer monedas y billetes nacionales y establecer equivalencias”, persigue el objetivo de comprender el sistema decimal, la descomposición de números, equivalencias y el valor posicional. La decisión de medir el objetivo instrumental o el objetivo central será del SIMCE, cualquier decisión que tome afectará la enseñanza y el aprendizaje de los alumnos. Es más difícil y mejor enseñar hacia el objetivo central, porque implica un mayor nivel de generalización. Si se evalúa directamente el objetivo instrumental, aquellos niños que hayan sido entrenados específicamente en ellos tendrán mejores resultados que aquellos que tuvieron una educación de carácter más general; por lo tanto, es necesario que el SIMCE elabore las preguntas con la conciencia clara de cuáles son los incentivos que está estableciendo.

Las nuevas pruebas también tendrán que responder, al momento de elegir las preguntas, al desafío de adecuar el sistema de evaluación a la libertad curricular planteada por la reforma. Desde luego tendrán que resol-

ver el problema de evaluar establecimientos que están implementando planes y programas educacionales propios, distintos a los que propone el Ministerio. La dificultad surge por el hecho de que restringirse a los objetivos fundamentales y contenidos mínimos comunes y obligatorios para todos los establecimientos baja el estándar de exigencia nacional. Esto sucede porque ningún colegio se rige sólo por estos objetivos. Los establecimientos tienen que adoptar el programa del Ministerio o uno propio y en cualquiera de los dos casos serán más amplios que los objetivos mínimos. Tampoco sería viable evaluar por alguno de los programas específicos, porque limitaría la libertad de elección de programas propios.

Históricamente, numerosos docentes, seleccionados en concurso público, elaboraban los ítems de selección múltiple de la prueba. Esta medida buscaba aportar a la confección de las preguntas un criterio de realidad, basado en la experiencia en el aula. Las nuevas pruebas no están elaboradas por un grupo de docentes de aula. Este cambio de política se basó en el supuesto de que los profesores no están aún familiarizados con las nuevas orientaciones de la reforma centrada en la enseñanza de habilidades de pensamiento superior; por ello no tendrían el bagaje para construir preguntas dirigidas a la evaluación de este nuevo tipo de objetivos. El desfase entre lo que los profesores y los expertos entienden por preguntas que evalúan pensamiento de orden superior pone de manifiesto la brecha que existe entre lo que se propone el currículo y lo que realmente se implementa. La antigua prueba estaba más cerca del currículo implementado al tomar en cuenta la interpretación de los profesores de aula; la nueva prueba escogió la interpretación de los expertos.

La construcción de la prueba responde a la teoría clásica de evaluación. La selección final de las preguntas se hace, año a año, de la siguiente forma: después de haber elaborado un conjunto grande de preguntas, se ensamblan diez pruebas de formas paralelas de modo que cada ítem sea respondido por 400 niños de una muestra no probabilística. Luego se eliminan todas aquellas preguntas que han obtenido menos de un 30% de respuestas correctas y aquellas que han obtenido más de un 80% de respuestas correctas. Se analiza también el comportamiento de las preguntas y de los distractores, descartándose las preguntas de formulación ambigua y aquellas en que las alternativas sugieren marcadamente la respuesta o inducen a error. También se revisa la correlación de cada pregunta con la prueba total para asegurar que la prueba tenga un grado alto de homogeneidad, de modo que los resultados sean interpretables unívocamente. Por diseño, un pequeño porcentaje de las preguntas debe cubrir contenidos de ciclos anteriores al cual está dirigida la prueba, de modo de sondear la persistencia de los aprendizajes a través de los años y la adquisición de las destrezas y concep-

tos más básicos de la asignatura. La mayor parte de las preguntas restantes debiera estar dirigida a evaluar los objetivos que debieran aprenderse en el año que se aplica la prueba y en el anterior.

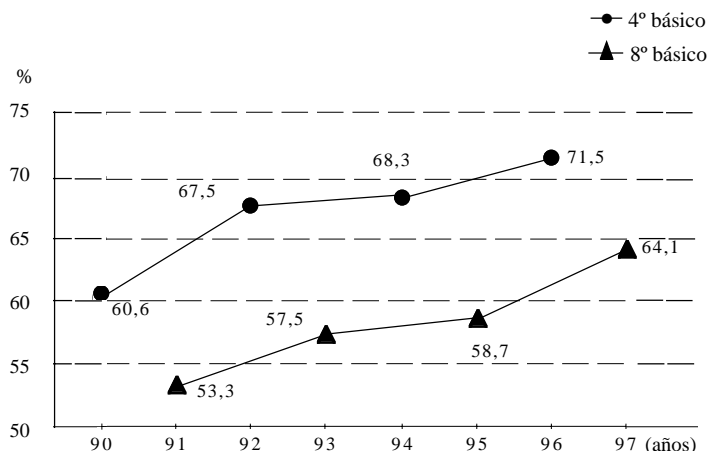
El segundo paso es confeccionar las pruebas definitivas, para lo cual se exige que el conjunto de ítems elegidos hayan obtenido en promedio un 65% de logro en la prueba experimental para el curso de 4° básico y un 60% de logro en 8° básico y II medio.

Este sistema de diseño tiene una serie de consecuencias que normalmente no se conocen. El fijar un promedio de logro esperado según la muestra experimental implica que los niveles de dificultad de la prueba son variables de un año a otro, ya que dependen de la población en la cual se tomó la muestra. Esto significa que es una prueba relativa, por lo tanto no se pueden comparar directamente los porcentajes de logro entre pruebas de distintos años. Un porcentaje de logro determinado en un año puede no ser equivalente, en términos absolutos, al del año subsiguiente. Por otra parte, si el SIMCE hubiera sido construido testeando los ítems en una muestra probabilística y representativa de la población el promedio nacional obtenido debiera haber estado siempre en el margen del 65% en 4° básico y del 60% en 8° y II medio. La representación gráfica de los resultados a lo largo de los años debiera ser una horizontal. Es un error esperar aumentos en ella. Sin embargo, las muestras experimentales no han sido probabilísticas hasta el año 1998, por lo tanto los resultados nacionales podrían haberse posicionado en cualquier valor, dado que las muestras experimentales no eran representativas de la población general.

Los aumentos de puntaje que se aprecian en el Gráfico N° 1 no necesariamente indican aumento de rendimiento, podrían ser reflejo de las diferencias de rendimiento entre el grupo experimental con el cual se construye la prueba y la población total. Si la muestra hubiese sido probabilística y representativa de la población, los resultados debieran estar en los márgenes del 65% para 4° básico y del 60% en 8° básico.

Una prueba relativa, en la cual la distribución de los resultados se decide empíricamente, se hace para lograr una buena dispersión de los resultados y poder así discriminar finamente el rendimiento entre los diversos establecimientos. Es necesario diferenciar este tipo de pruebas de las referidas a criterio, ya que la mayoría de las personas piensa que el SIMCE es una de estas últimas. En ellas se determina el nivel de dificultad según parámetros teóricamente establecidos, por ejemplo, qué porcentaje de los contenidos de 8° básico tendrían que dominar los alumnos después de 7 años y medio de enseñanza. En estas pruebas no se espera una distribución según una curva normal: podría obtenerse que la mayoría de los establecimientos dominaran todos los contenidos si la educación fuese buena o que

GRÁFICO N° 1: PROMEDIOS NACIONALES DE LA EDUCACIÓN BÁSICA 90-97
(Promedio de castellano y matemática (%))



Fuente: Juan Eduardo García Huidobro y Cristián Cox, “La reforma educacional chilena 1990-1998, visión de conjunto” (1999), p. 34.

no dominaran casi ninguno si la educación es deficiente. Este tipo de pruebas indica más directamente el nivel de aprendizaje alcanzado en relación a las metas esperadas por el sistema. Sin embargo, no es óptimo para ‘ranquear’ a los establecimientos por rendimiento.

El SIMCE tiene dos grandes objetivos que son el entregar información acerca del cumplimiento de los objetivos educacionales por parte del sistema general y el de discriminar la calidad de la educación entre los establecimientos. Estos objetivos tienen que compatibilizarse en el diseño de las pruebas, tarea que no resulta del todo fácil. Los requerimientos técnicos para obtener una buena dispersión exigen un porcentaje de logro promedio en el rango 60% al 65%, pero, ¿qué sucede si en la prueba experimental, con los contenidos definidos para ese curso, no se alcanza este porcentaje de logro? ¿Se baja el nivel de exigencia para adaptarse al porcentaje de logro esperado? ¿Se baja tanto el nivel que las preguntas ya no corresponden a los cursos en los cuales están los alumnos? Entonces, ¿cómo se conjugaría la validez de contenido con la necesidad de lograr la dispersión adecuada? El análisis que a continuación se realizará de las preguntas de la prueba de 1997 nos muestra cómo no ha sido posible mantener la validez de contenido en la prueba de 8° básico, esto es: la mayoría de las preguntas no corresponde a este curso sino a cursos inferiores.

Para conseguir la comparabilidad de los resultados en las pruebas relativas se realiza la técnica del *equating*. Este proceso permite comparar los puntajes de una prueba con la de otros años. Consiste en tomar un porcentaje de ítems para repetirlos en años sucesivos. Luego se analiza la correlación de los resultados de esas preguntas en relación a la prueba total lográndose un índice que se utiliza para obtener la equivalencia de puntajes entre una prueba y otra. El *equating* de un año a otro se hizo entre las dos primeras aplicaciones del PER y nuevamente se está realizando desde el año 1997. Por lo tanto, la comparación de pruebas anteriores a 1997 no sería metodológicamente correcta, y sería aventurado formular conclusiones sobre avances en la calidad de educación antes de esa fecha. Recordemos que al probar experimentalmente los ítems cada año se exige un 60% o un 65% de logro promedio a la prueba, por lo que los niveles absolutos de dificultad pueden variar de un año a otro si no se hace *equating*. En 1999 se hará un estudio de comparabilidad de las pruebas de 4° básico repitiendo porciones de pruebas de los 10 años anteriores a muestras representativas de la población. Esto permitirá recuperar los datos de los años en que no se hizo *equating*. Esperamos que este estudio se haga público.

El SIMCE, a diferencia de la Prueba de Aptitud Académica, no corrige puntajes por respuestas al azar. En toda prueba de alternativas hay una probabilidad de responder bien una proporción de ítems por adivinación al azar, por lo tanto se castiga esa actitud descontando una proporción de las respuestas malas a las buenas. No tomar en cuenta este punto puede aumentar el rendimiento espuriamente.

En las nuevas pruebas se utilizará la teoría de Respuesta al Ítem; ésta permite resolver el problema de asignar puntajes equivalentes a preguntas de diferentes niveles de dificultad. Los ítems más difíciles reciben una mayor ponderación que los más fáciles. Las pruebas construidas con esta teoría siguen siendo relativas y no aseguran la comparabilidad si no se hace *equating*.

La prueba SIMCE está en proceso de reforma como también el organismo (llamado con el mismo nombre) responsable de ella. Por una parte, las evaluaciones están sometidas a una serie de cambios y desafíos para adaptarse a la reforma curricular. Por otra parte, el organismo SIMCE amplía sus funciones al pasar a coordinar una serie de evaluaciones internacionales del sistema educacional a las cuales Chile ha decidido incorporarse. Una amplia difusión de las políticas de medición estimularía la discusión en torno al tema, lo cual contribuiría a mejorar la calidad de nuestro sistema nacional de evaluación y la calidad de la educación misma.

El SIMCE de lenguaje y matemáticas de 1997 y orientaciones futuras

Los antecedentes expuestos anteriormente sobre los efectos tanto positivos como negativos de una evaluación estandarizada, masiva, no hacen más que confirmar el efecto profundo que tiene la evaluación sobre lo que sucede adentro de la clase. Aún sin efectos sobre el financiamiento o sobre la destinación académica de los alumnos, una evaluación central, administrada censualmente y cuyos resultados son públicos, emite una señal poderosa a directores y profesores acerca de lo que debe hacerse. Es por lo tanto válido examinar los contenidos del SIMCE de 8° básico de 1997 dados a conocer en 1998 y preguntarse qué señales son las que éste transmite.

Este trabajo pretende hacer un análisis de nuestras pruebas a partir de una doble coyuntura: el hecho de que por primera vez se publiquen las preguntas de las pruebas rendidas en 1997 y el momento de pleno proceso de rediseño de las pruebas futuras. Podría parecer inútil el análisis en período de transición; sin embargo, pensamos que su discusión puede ayudar a interpretar los resultados de las pruebas ya rendidas y aportar antecedentes para la elaboración de las nuevas evaluaciones.

El análisis se basará en los ítems publicados por el Ministerio, que corresponden a una de las formas administradas en 8°. En este curso se busca medir el aprendizaje de los contenidos que debieran haber sido asimilados en el transcurso de enseñanza básica. Una pequeña proporción de los ítems evalúa conocimientos de 4° básico y el resto cubre los del segundo ciclo. Son 48 preguntas de alternativas múltiples para matemáticas y 50 para lenguaje que deben ser respondidas en 90 minutos. La muestra es censal, es decir, se someten a la prueba todos los niños del nivel, lo que equivale a 217.000 estudiantes pertenecientes a 4.755 establecimientos. Basamos nuestro estudio en los documentos publicados por el Ministerio de Educación, *Resultados SIMCE, octavos años, 1997. (Fascículos de castellano y de matemáticas)*, julio 1997, y *Orientaciones para la medición de logros (fascículos de castellano y de matemáticas)*, Segundo año medio, 1998. Para el caso de matemáticas se ha considerado también como punto de referencia la prueba de 8° básico realizada por el Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)¹⁷.

¹⁷ Para más información sobre el estudio del TIMSS, véase B. Eyzaguirre, "Políticas educacionales comparadas: Consideraciones a partir del Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)" (1999).

SIMCE de 8° básico de lenguaje 1997

Como hemos dicho, una prueba implica seleccionar algunos de los elementos que constituyen un campo amplio de conocimientos asumiendo que el dominio de ellos que tenga un alumno refleja o significa un dominio del campo total. En el caso del SIMCE de lenguaje que comentamos (de 1997), se optó por ocupar un 40% de la prueba en materias correspondientes a aspectos normativos, como ortografía, gramática, estructura de palabras, puntuación, etc. Las habilidades de lectura ocupan otro 52% de la prueba. En este grupo incluimos algunas de las preguntas que evalúan contenidos de 4° y de 6°, la agrupación “Diálogo Escrito” y la agrupación “Comprensión Lectora”, pues todas se refieren a comprensión de textos leídos. El 8% restante corresponde a 4 preguntas que supuestamente evalúan razonamiento lógico. (El facsímil publicado, correspondiente a una de las formas de la prueba, consta de 50 preguntas en lugar de las 60 que constituyen la prueba; como se indica en el mismo documento, 10 preguntas han sido mantenidas en reserva para efectos de un análisis histórico.)¹⁸

Nivel de dificultad

El grupo de preguntas de tipo normativo está constituido por un muestreo de *casos* de ortografía, puntuación, gramática, vocabulario y estructura de palabras que supuestamente revelaría el dominio que tiene el alumno de su lengua. Llama la atención en primer lugar el bajo nivel de conocimientos que requieren estas preguntas. El 78% de los contenidos evaluados en este grupo pertenece a los cursos entre 3° y 6° del programa que estaba vigente en ese momento según el decreto 4002. No aparece, entre los contenidos medidos, ninguno de los temas más difíciles que tradicionalmente se enseñan en 7° y 8° y que corresponden a un verdadero y total dominio del idioma. Por ejemplo, la ortografía de algunos verbos irregulares, casos más complejos de sintaxis y de acentuación (acento diacrítico y acento diéresis) que sí estaban contemplados en las evaluaciones anteriores como lo indicaría al menos el informativo entregado a los colegios para la prueba de 8° de 1995. Salvo dos preguntas sobre ‘nexos’ o conjunciones, tampoco hay contenidos gramaticales nuevos propios de 7° y 8°. Sólo algunos ítems de vocabulario corresponden, por el nivel de abstracción de los vocablos evaluados, a un nivel más propio de 8° básico.

¹⁸ La prueba constaría de dos formas paralelas, véase Ministerio de Educación, *SIMCE: Folleto Técnico 1995* (1995).

Todas las preguntas se refieren al dominio más elemental del contenido, nunca a un grado más complejo. Por ejemplo: hay tres preguntas que miden la habilidad de ordenar palabras alfabéticamente. Ésa es una habilidad muy simple que, en general, se puede enseñar en 3° básico y que sirve para ubicar datos en un diccionario, en una enciclopedia o en una guía telefónica. Según el documento, estas preguntas corresponden al 10% de la prueba destinado deliberadamente a verificar lo que los alumnos de 8° saben de los contenidos del subciclo anterior. El grado más simple de complejidad consiste en ordenar palabras que comienzan con letras diferentes; un grado superior sería ordenar palabras que tienen la misma letra inicial, pero difieren en la segunda y así sucesivamente. Pues, de las tres preguntas sobre esta materia tan elemental, sólo una tiene dos palabras que comienzan con la misma letra. En gramática, la única pregunta sobre verbos se refiere al modo indicativo, presente, es decir, el caso más elemental de esta materia. De las posibles respuestas, todas están en modo indicativo y sólo una está en presente. En la pregunta sobre la conjunción 'o' las tres opciones incorrectas son completamente absurdas. Y así sucesivamente.

Entre las preguntas de comprensión de textos escritos, 8 (el 30%) exigen la comprensión de *una sola frase*, 6 (el 23%) se refieren al manejo de documentos muy breves (por ejemplo, un cheque o un aviso económico) y el resto se reparte entre una estrofa poética de cuatro versos y dos fragmentos narrativos de media página cada uno. El grado de dificultad de los textos (salvo, tal vez, en el fragmento poético), su longitud y el tipo de preguntas corresponden en su mayor parte también a un nivel aproximadamente entre 5° y 6° básico. Por ejemplo, se pide: determinar quiénes son los protagonistas de un relato que sólo tiene dos personajes; escoger, de una lista de cuatro, el calificativo que caracteriza a un personaje; el porqué de la acción de un personaje (la opción correcta es una frase textual del párrafo y es la única que lo es), etc. Hay dos preguntas que apuntan a la identificación del tema o motivo principal de la lectura, pero resultan poco convincentes, en un caso, porque es dudoso que se pueda preguntar por el motivo principal de un fragmento tan breve sacado de una novela y, en otro, por lo ambiguo de las alternativas de respuesta. No hay ni una sola pregunta que exija una lectura crítica, ni un aporte personal, ni siquiera las que se refieren a textos de los medios de prensa. (Hay 4 preguntas [19%] con textos de prensa, pero todas ellas referidas sólo a identificar las secciones que conforman un diario. Por ejemplo, se pide determinar en qué sección de un periódico se encontraría determinado aviso.) Dado el bajo nivel de comprensión que requieren estas preguntas es aventurado sacar cualquier tipo conclusiones acerca de cómo comprenden los alumnos el lenguaje escrito.

Es interesante comparar los niveles de exigencia en lectura que evidencia este SIMCE con los descritos para una edad similar en el programa estatal inglés¹⁹: “Nivel 5: Los alumnos demuestran comprensión de una gama de textos seleccionando los puntos esenciales y usando inferencia y deducción cuando sea apropiado. En sus respuestas identifican características principales, temas y personajes, y seleccionan oraciones, frases y cualquier otra información relevante para sostener sus puntos de vista. Son capaces de obtener y relacionar información de diversas fuentes. Nivel 6: Al leer y analizar una gama de textos, los alumnos identifican diferentes niveles de significado y pueden comentar acerca de su importancia y efecto. Dan respuestas personales a textos literarios, refiriéndose a aspectos del lenguaje, la estructura y los temas para justificar su opinión. Son capaces de resumir una variedad de información de diversas fuentes”²⁰.

Redacción elemental

Mención aparte merece la prueba de expresión escrita. Es un avance importante contar con una prueba que evalúe esta habilidad y más aún si, como se sostiene en el documento, desde 1991 se observa un incremento de más de 10 puntos porcentuales a nivel nacional. El aumento de puntaje es especialmente significativo en las dependencias municipal y particular subvencionado. Este avance indica que el hecho de evaluar probablemente ha modificado la conducta de los profesores haciéndoles dedicar más tiempo de clases a la expresión escrita.

Lamentablemente, la información que se nos da sobre esa parte de la prueba es poco clara. Los cuadros y gráficos con que se presenta un estudio del número de líneas escritas por los estudiantes son confusos y, por lo tanto, no es posible comprender el análisis cuantitativo de este punto realizado por los autores del documento. Sí queda claro que los alumnos escribieron un promedio de 21 líneas sobre un tema dado. (No se informa sobre el tiempo que tuvieron los alumnos para ello.)

Tampoco queda completamente clara la forma de corregir esta prueba. Se nos dice que hay una pauta de corrección constituida por siete ‘factores’, pero no se publica la pauta, no se indica cómo se contabilizan los errores del alumno ni qué puntaje tiene cada factor, ni tampoco el peso del puntaje de expresión escrita en el puntaje total del SIMCE. Los siete

¹⁹ Los niveles de desempeño citados corresponden aproximadamente a lo exigido entre 6º y 7º.

²⁰ Department for Education (Inglaterra), *The National Curriculum* (1995), p. 28.

factores son: expresión creativa, expresión léxica, organización interna, ortografía puntual, ortografía literal, ortografía acentual y presentación formal. Una breve descripción de estos factores nos indica que el factor ‘expresión creativa’ verifica que el alumno “título un relato central, donde estén presentes el espacio y el tiempo de la situación narrada o descrita; cree personajes identificables movilizadores de acción, emplee un tono o actitud emotiva, dé opiniones e infiera valores”²¹. Parece discutible acumular en un solo ‘factor’ todo lo que se refiere al contenido del texto escrito y que los otros seis factores apunten a describir el lenguaje utilizado y la aplicación de normas. Esto otorgaría un peso desmesurado a la forma en detrimento del contenido. También llama la atención que se ignoren aspectos tales como la estructura, el propósito, la adecuación del estilo y del lenguaje, etc.

Los autores se felicitan por el logro obtenido en ‘organización interna’ (86,5%). Sin embargo, ese factor no se refiere a la estructura narrativa del escrito, como su nombre hace suponer, sino meramente a “la utilización adecuada de pronombres, ilativos o nexos, la no omisión de términos e ideas; la concordancia de sujeto-predicado, género y número”²². Es decir, al dominio gramatical más elemental. (Aunque cabe la duda sobre cómo se mide la “no omisión de ideas”.) La descripción de éste y de los otros factores que conforman la pauta indican que esta evaluación se refiere a un dominio muy elemental del discurso escrito, limitado al logro de una coherencia básica y de las mínimas condiciones de ortografía y puntuación para obtener claridad. Es decir, un dominio correspondiente, en su mayor parte, a un 5º o 6º básico bastante indulgente.

Ilustra este punto el hecho de que en los países con buen nivel educacional se esperan estos niveles de desempeño en redacción ya en los primeros años de la enseñanza básica²³. Véase, por ejemplo, esta descripción del nivel de desempeño esperado para alrededor de 3º a 4º básico en el currículum nacional británico:

La escritura de los alumnos es a menudo bien organizada, imaginativa y clara. Usan apropiadamente las características principales de los diferentes tipos de textos y empiezan a adaptarse a diferentes lectores. La secuencia de las frases extiende las ideas lógicamente y

²¹ Ministerio de Educación, *Castellano: Resultados SIMCE octavos años, 1997* (1998), p. 40.

²² *Ibíd.*, p. 41.

²³ Véase L. Fontaine, “La asignatura de lenguaje vista en programas escolares extranjeros estatales y privados” (1998).

selecciona las palabras para dar variedad e interés. La estructura gramatical básica de las oraciones es usualmente correcta. La ortografía es usualmente perfecta, incluso la de los polisílabos más comunes. Usan con corrección la puntuación para demarcar las oraciones: puntos, mayúsculas y signos de interrogación. La caligrafía es ligada y legible²⁴.

Igualmente, en el programa de Francia, se espera al finalizar el 5° básico que el alumno sea capaz, entre otras cosas, de escribir un cuento, reescribir un texto, relatar un suceso desde varios puntos de vista, presentar su punto de vista frente a un evento, redactar un resumen, preparar un cuestionario, redactar un informe (y distinguirlo de un relato), tomar notas sobre las lecturas, estructurar un texto para su presentación (párrafos, gráficos), utilizando un procesador de textos. Como puede apreciarse, ya el 6° básico se inicia con un dominio funcional de la escritura mucho mayor que el que se evidencia en nuestro SIMCE de 8°. Un dominio suficiente para desempeñarse en cualquier trabajo y que le asegura al alumno proseguir con éxito las etapas siguientes de su educación.

El grado de dificultad de estas preguntas no corresponde en absoluto a lo que debiera esperarse para 8° y no refleja el conjunto de aprendizajes que debieran estar logrados al terminar la enseñanza básica. ¿Qué pueden deducir los profesores al analizar el nivel de dificultad de la prueba? Si desean obtener un buen puntaje en el SIMCE y se guían por este conjunto de preguntas, les bastará con mantener la materia simple; no enseñar cosas difíciles. Puede ser arriesgado avanzar mucho con la gramática o perder el tiempo en acentuación. Quienes han hecho otra cosa han perdido su tiempo, pues no se reflejará en la evaluación. No harán a los alumnos leer textos avanzados; no vale la pena. Bastará con breves lecturas infantiles y algunos documentos utilitarios. Más que leer la prensa, enseñarán muy bien en qué secciones se divide un diario y se concentrarán en los avisos publicitarios, los cuales, al parecer, son considerados más importantes que las noticias y los artículos de opinión. En cuanto a expresión escrita, podrán contentarse con un dominio elemental de la escritura, limitado al logro de una coherencia básica y de mínimas normas de corrección formal.

Más allá de las señales acerca del qué, cuánto y cómo enseñar que nos entrega esta prueba, interesa que la interpretación de los resultados se ajuste a la realidad. A la luz de las preguntas que se han dado a conocer, parece alarmante sacar conclusiones sobre los rendimientos de 8° y comparar colegios sobre la base de una prueba que en su mayor parte corresponde

²⁴ Department for Education (Inglaterra), *The National Curriculum* (1995), p. 30.

a un fácil 6º básico. Más alarmante es que se use la comparación con años anteriores (ya hemos visto que estrictamente las pruebas de diferentes años hasta 1997 no son comparables) para deducir visiones optimistas que pueden no reflejar progresos verdaderos.

Preguntas ambiguas


Naturalmente surge el interrogante de por qué se dan tan pobres resultados con materias tan fáciles. Hay muchas explicaciones posibles y no es el fin de este documento analizar los múltiples factores desde socioeconómicos hasta metodológicos que pueden dificultar o impedir el aprendizaje. Pero queremos llamar la atención sobre un factor que era para nosotros desconocido antes de que el documento que comentamos cayera en nuestras manos y que quizás explica en parte este bajo rendimiento. Se trata de la formulación de las preguntas: entre los ítems de comprensión de textos se encuentra una cantidad importante de preguntas mal formuladas, de respuesta ambigua. Aparentemente ésta también ha sido una inquietud del Ministerio, pues en los documentos sobre la próxima generación de pruebas se anuncia un nuevo mecanismo de selección y validación de preguntas. Especial mención reciben las tres preguntas de comprensión de textos que, según el documento, corresponden a materia de 4º básico. Consisten en un pequeño dibujo, de muy mala calidad, debajo del cual se leen cuatro oraciones. El alumno debe escoger la oración que mejor representa en el primer caso 'el mensaje', en el segundo el 'contenido' y en el tercer caso 'la idea' que representa la lámina. No sólo es confundidor para el alumno que se le pida de tres formas diferentes realizar una misma tarea sino que, además, los dibujos en sí son tan ambiguos, que permiten responder con casi cualquiera de las alternativas propuestas. Véanse figuras Nº 1 y Nº 2.

No es de extrañar que aunque esta parte de la prueba pretende evaluar habilidades de 4º básico, en ninguna pregunta se superó el 61% de respuestas correctas.

Una revisión de las preguntas con bajo porcentaje de respuestas correctas (menos de 65%) revela una proporción importante de preguntas ambiguas. Hay 14 ítems sobre comprensión de textos que alcanzaron menos de 65% de respuestas correctas; de éstos, al menos 9 son discutibles (las preguntas 4, 5, 6, 8, 33, 37, 38, 43, 45).

Entre las preguntas sobre aspectos normativos hay menos ítems ambiguos. Pero sí hay algunos cuya dificultad no estriba tanto en el contenido

FIGURA N° 1




La oración que representa mejor el **mensaje** del dibujo es:

- A) El trabajo, por humilde que sea, ennoblece.
- B) Debemos trabajar para *ganar* el sustento.
- C) Todo trabajo debe darnos una satisfacción.
- D) El dinero recompensa todo el trabajo.

FIGURA N° 2

¿Cuál es la **idea** que representa la lámina?



- A) El amor lleva a dos personas a conversar.
- B) La amistad también es compartir sin hablar.
- C) Dos personas participan de un mismo pensamiento.
- D) La compañía de dos personas significa amor.

evaluado sino en la manera de preguntar, que exige una lectura muy cuidadosa para no caer en la ‘trampa’. Por ejemplo, la pregunta 23:

¿Cuál de las opciones *no* es *sinónimo* de la palabra subrayada en la siguiente oración?

Todos fuimos invitados a la *boda* de mi prima.

A) matrimonio B) casamiento C) idilio D) enlace

Otro ejemplo es la pregunta 13:

¿En cuál de estas oraciones la *coma* está bien aplicada?

A) La mañana estaba clara, sin embargo, se nubló pronto.

B) La mañana, estaba clara, sin embargo se nubló pronto.

C) La mañana estaba, clara sin embargo, se nubló pronto.

D) La mañana estaba clara sin embargo, se nubló pronto.

Según el documento, la alternativa correcta es A), que fue elegida por un 61,87% de los alumnos. Hubo, sin embargo un importante grupo (18,84%) que escogió D), desorientados por el hecho de que la alternativa correcta tuviera *dos comas* y no una, como se deduciría de la pregunta.

Cuando se interpretan los resultados de una prueba con tales preguntas siempre cabrá la duda de si los alumnos dominan o no el contenido o la destreza evaluada, pues los resultados se ven contaminados por las diversas interpretaciones que han hecho los alumnos acerca de la pregunta. Igualmente, omitiendo el punto (grave de todas maneras) de que algunas preguntas son equívocas en su formulación y que se requiere ciertamente de una mayor prolijidad en su elaboración, cabe hacerse la reflexión acerca del efecto que puede tener sobre la instrucción el hecho de que haya preguntas que aunque estén bien formuladas sean de alguna forma ‘tramposas’. ¿Cuál es el mensaje para profesores y directivos? Para obtener buenos puntajes en el SIMCE conviene gastar el tiempo de clases entrenando específicamente con este tipo de preguntas, para disminuir las posibilidades de error. El degradar la instrucción haciendo más ensayos y menos clases es la consecuencia natural de una prueba mal formulada.

Tareas artificiales

La mayor parte de las tareas que pide esta prueba no corresponde al tipo de tareas que normalmente hacemos cuando usamos el lenguaje. Escribir con buena ortografía no consiste en completar los espacios que faltan en

las palabras; leer, en términos generales, no consiste en escoger la frase que mejor expresa el contenido de una lámina; apreciar la poesía no pasa necesariamente por discriminar entre metáforas, personificaciones, etc. Dominar o comprender los medios informativos no se reduce a conocer las secciones de un diario. Tener un buen vocabulario no consiste en reconocer un diminutivo. Una vez más la asignatura de castellano se ve reducida a una serie de simplificaciones, nomenclaturas y clasificaciones. (Esta tendencia ya fue observada y descrita respecto de los textos escolares de lenguaje de nuestro país.)²⁵ Este punto tiene gran importancia pues el tipo de preguntas incide en lo que se enseña y este tipo de preguntas puede conducir a una enseñanza limitante y que no aporte a los alumnos oportunidades de desarrollar su uso del lenguaje, que es la finalidad de la asignatura. Una vez más, los profesores que deseen trabajar desarrollando verdaderamente el lenguaje de los alumnos con mucho tiempo dedicado a la lectura y a la comunicación oral y escrita pueden verse inhibidos al conocer cómo se les va a evaluar. Por otra parte, es posible que esta forma de evaluar no mida el verdadero desempeño de los alumnos cuando se comunican en la vida real y por lo tanto arroje un diagnóstico erróneo sobre sus capacidades. Afortunadamente, se percibe que la próxima versión de estas pruebas tiende a corregir este aspecto.

El nuevo SIMCE de lenguaje de 2° medio

La prueba aplicada a los 2°s medios en 1998 obedece a una nueva concepción del SIMCE. Ésta surge de la necesidad de ajustar los contenidos de la prueba a los nuevos lineamientos de la reforma curricular. El año 1998 corresponde a una prueba de *segunda generación*, es decir, una prueba en que se miden objetivos y contenidos del programa anterior (Decreto 300, en este caso) pero sólo aquellos que “más se acercan al sentido del nuevo”²⁶. Al momento de escribir este estudio no contamos con información acerca de la totalidad de las preguntas utilizadas y, por lo tanto, nos basamos únicamente en el folleto publicado por el Ministerio de Educación antes de la prueba para orientar a los profesores. En este folleto se incluye la fundamentación teórica en que se basa la prueba, una descripción del instrumento y 18 ítems. Naturalmente no se pueden derivar demasiadas

²⁵ B. Eyzaguirre y L. Fontaine (eds.), *El futuro en riesgo, nuestros textos escolares* (1997).

²⁶ Ministerio de Educación, *Sistema de medición de calidad de la educación: Orientaciones para la medición de logros. Asignatura de castellano, segundo año medio* (1998).

conclusiones sin conocer el conjunto de todos los ítems. Pero intentaremos hacer un comentario sobre la muestra entregada.

El folleto nos ofrece un conjunto de 18 preguntas, referidas a 9 objetivos del programa de castellano según el Decreto 300. Los objetivos que se pretenden evaluar son citados textualmente del decreto, encabezando los ítems correspondientes y, contrariamente a lo que sucede con la prueba de 8º, conciernen en su mayoría a 2º medio. Sin embargo, los objetivos de media en esta asignatura no especifican niveles de dificultad por lo cual el rango de preguntas que se pueden formular a partir de ellos podrían corresponder a cursos inferiores o superiores. Por lo tanto, es la observación detallada de las preguntas la que más nos revelará acerca del nivel de dificultad de esta prueba y de las orientaciones que da para la enseñanza.

Este documento denota un importante progreso respecto de algunas de las deficiencias observadas en la prueba de 8º. Al no conocer el total de las preguntas no podemos formarnos una opinión sobre el nivel de dificultad general, pero se puede apreciar al menos una mayor prolijidad en la confección de las preguntas y una intención de medir destrezas más relevantes. Vista en su conjunto, la muestra ofrece una visión más funcional y más completa de la asignatura. Se busca evaluar el uso de la lengua con ítems más similares a las verdaderas acciones que realizamos cuando nos comunicamos. Éste es, por lo demás, uno de los objetivos que se proponen los autores de esta prueba de ‘segunda generación’ y podemos decir que parece bien logrado. Éste es un punto de gran valor, pues en términos generales influirá a los docentes a realizar prioritariamente esas tareas en la clase y no otras. Revela también una elaboración mucho más cuidadosa de las preguntas. Las instrucciones para responder los ítems son claras y precisas y no se observan preguntas ambiguas. Es decir, el alumno sabe lo que debe hacer. Sólo las preguntas abiertas podrían ofrecer alguna dificultad *en su corrección*, pues las pautas para corregir no aparecen del todo claras. Esta dificultad, sin embargo, podría obviarse al momento de corregir mediante el perfeccionamiento y aclaración de las pautas.

Tal como se indica en la introducción del documento, se ha buscado estimular el uso de destrezas cognitivas superiores y entre éstas se ha dado prioridad a la evaluación de habilidades deductivas e inductivas por sobre las habilidades de memoria y creatividad. Esto, porque “las habilidades deductivas e inductivas son de un nivel cognitivo más alto que las habilidades de memoria, (...) y las habilidades de creatividad, si bien son de alto nivel cognitivo, no se pueden utilizar en exceso, puesto que requieren de

preguntas abiertas”²⁷. Sin duda, este énfasis se debiera traducir en un cambio de las actividades de la clase. Parece especialmente auspicioso respecto a la orientación que da a las actividades de comprensión de lectura, alejándose de la recolección y reproducción textual de los datos leídos y enfatizando más la interpretación y comprensión de aspectos no implícitos.

Lamentablemente este foco en ‘habilidades superiores’ ha llevado a una total prescindencia de los aspectos normativos del ramo. A pesar de que el programa lo requiere, no hay preguntas de ortografía ni gramática, tal vez asumiendo que éstas han sido ya completamente asimiladas por los estudiantes de 2º medio. Éste es un supuesto arriesgado, ya que la experiencia indica que en la actualidad los graduados de enseñanza media en general parecen tener una formación deficitaria en estos aspectos. Transmitir la idea de que estos contenidos no serán evaluados en segundo medio puede conducir a una total falta de atención a los aspectos formales básicos y, por consiguiente, a un deterioro de la expresión escrita de los estudiantes. Este riesgo se hace más real a la luz del bajo nivel exigido en estos aspectos en la prueba de 8º. Sólo asegurando que estos contenidos han sido logrados en los cursos básicos se podría ‘liberar’ a la enseñanza media de cubrirlos.

A pesar de que en esta muestra hay una fuerte proporción de preguntas de lectura y que éstas parecen medir aspectos relevantes de la habilidad de leer, se echan de menos algunas preguntas con trozos más largos. Si se desea medir competencia en habilidades usadas en el mundo real, no puede obviarse que éste nos exige a menudo la lectura de trozos extensos. Éstos, a su vez, exigen habilidades de concentración, retención y captación de la estructura general del texto que no pueden ser evaluados con pocas líneas. Al menos entre los ejemplos dados a conocer no hay trozos que sobrepasen las tres líneas. Podría ser que haya una correlación entre las habilidades requeridas para lecturas largas y el desempeño en preguntas cortas, pero creemos que la señal que emiten las pruebas para determinar lo que pasa con la instrucción es demasiado poderosa y existe el riesgo de que una prueba que sólo se basa en textos breves desincentive la lectura de textos completos en la sala de clases.

Pero tal vez la mayor objeción que se puede hacer a esta muestra de ítems se refiere a lo que las preguntas realmente miden (validez de contenido). Llama la atención que aunque se han escogido objetivos del programa

²⁷ Ministerio de Educación, *Sistema de medición de calidad de la educación: Orientaciones para la medición de logros. Asignatura de castellano, segundo año medio* (1998), p. 13.

antiguo que en las palabras del documento son los que “más se acercan al sentido del nuevo”, los ítems mismos ofrecen a veces una curiosa interpretación de éstos. Por ejemplo, para el objetivo “Conocer obras valiosas de la literatura universal”, el ítem requiere sólo parear una lista de obras con sus autores. Esto es, no examina si verdaderamente se conocen (se han leído) estas obras. Para el objetivo “Conocer y manejar algunas técnicas para escribir sencillas piezas teatrales”, se escogió una pregunta que exige reconocer la *definición* de tragedia. Saber definir tragedia ¿es verdaderamente una técnica para escribir mejor? Para el objetivo “Iniciarse en un encuentro sistemático con la literatura”, se pide reconocer la *definición* de metáfora. Es decir, si bien los objetivos corresponden en algunos casos al ámbito de la valoración, y, en otros, a habilidades de comunicación, las preguntas no recogen estas intenciones sino miden más bien la retención de información. Tampoco lo hace la pregunta 6 que, para el objetivo “Descubrir la propia creatividad y desarrollarla”, propone: “Explica con tus palabras qué es una leyenda”.

El cuidado de la validez de contenido es fundamental para la correcta interpretación de los resultados de un test. En el caso de la pregunta 6 nada serio se podría deducir sobre la creatividad, pues los resultados dependerían del conocimiento que tengan los alumnos acerca del tema de la leyenda. Este tipo de confusiones puede reflejar erróneamente la realidad de los alumnos y conducir a diagnósticos equivocados.

Esta confusión de intenciones se observa al menos en 4 de los 18 ítems. En todos los casos la información exigida tiene que ver con literatura (preguntas 1, 2, 3 y 6). Las materias literarias se prestan magníficamente para evaluar comprensión, apreciación, pensamiento crítico, etc. Sin embargo, no se ha hecho así y se ha optado por preguntar definiciones y clasificaciones. Es sorprendente que, en todos los casos, se ha perdido la oportunidad de “estimular el uso de destrezas cognitivas superiores”, y dar “mayor relevancia a las habilidades deductivas e inductivas que a las de memoria y creatividad”, pese a la intención expresada en el documento de hacerlo así. Esta forma de preguntar tal vez obedece a una visión peculiar sobre la enseñanza de la literatura que percibe a ésta como incompatible con el desarrollo de destrezas comunicativas o con el uso de destrezas cognitivas superiores. Curiosamente, en el documento se anuncia que se otorga “mayor peso a los contenidos no literarios que a los literarios, por considerarse que el espíritu de las nuevas propuestas curriculares así como el del Decreto 300 enfatiza lo comunicacional por sobre lo literario, criterio

que se aplicó en el diseño de la prueba”²⁸. Es discutible que el ‘espíritu’ de las nuevas propuestas curriculares y del Decreto 300 enfaticen lo “comunicacional por sobre lo literario”. En primer lugar, ambos programas de enseñanza media valoran igualmente la literatura y, además, la frase implicaría que literatura y comunicación son cosas distintas. No es así: la literatura no es otra cosa que comunicación. Comunicación del mejor nivel y sobre los temas más importantes para el hombre. En un documento ofrecido como orientación para el profesorado es alarmante que se entregue este tipo de afirmaciones equívocas. Por lo demás, la posición contrasta con lo observado en las propuestas curriculares de otras naciones que logran mejores resultados educacionales para la totalidad de su población y que no trepidan en permear su currículo con una gran cantidad de literatura de la mejor calidad²⁹.

El SIMCE de matemáticas 1997

Como hemos dicho, este análisis se basa en los ítems publicados por el Ministerio de Educación de la prueba del SIMCE de 8° de matemáticas rendida en 1997. Como punto de referencia para este análisis, se utilizará también la prueba de 8° básico realizada por el Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)³⁰. Éste es el estudio más amplio y ambicioso sobre rendimientos comparados jamás realizado. Abarcó más de 40 países y exigió la revisión de los currículos de todas estas naciones para determinar los contenidos cuyo manejo se consideraba indispensable al final de 8° básico. Los contenidos seleccionados en ningún modo representaban las posiciones extremas de los países en cuanto a exigencias, sino más bien el término medio. Las preguntas, cuidadosamente elaboradas, se ocuparon de evaluar tanto conocimientos como destrezas, de que los diferentes contenidos tuvieran una proporción de preguntas de acuerdo a su importancia y de que los niveles de dificultad tuvieran una graduación que permitiera diferenciar los niveles de aprendizaje. Por lo

²⁸ Ministerio de Educación, *Sistema de medición de calidad de la educación: Orientaciones para la medición de logros. Asignatura de castellano, segundo año medio* (1998).

²⁹ Véase L. Fontaine, “La asignatura de lenguaje vista en programas extranjeros estatales y privados” (1998).

³⁰ Para más información sobre el estudio del TIMSS, véase B. Eyzaguirre, “Políticas educacionales comparadas: Consideraciones a partir del Tercer Estudio Internacional de la Enseñanza de Matemáticas y Ciencias (TIMSS)” (1999).

tanto, podemos estar seguros de que esta prueba constituye un buen parámetro de comparación.

El análisis abarcará diferentes aspectos de la prueba intentando dilucidar las señales que el SIMCE envía a nuestra educación.

Nivel de dificultad

En primer lugar, llama la atención el bajo nivel de dificultad de las preguntas del SIMCE. Si partimos del supuesto que debe evaluar lo que se espera que los alumnos aprendan al final de la enseñanza básica, la lectura de los ítems nos indica que el estándar al cual se está aspirando es bajo. Si se hace el ejercicio de tomar ítem por ítem de la prueba y analizar el curso en que debe ser tratado el contenido de cada una de las preguntas, tenemos que las materias evaluadas corresponden a un 6° básico en el antiguo programa de estudio chileno, a un 5° básico avanzado si nos referimos a los nuevos objetivos fundamentales y contenidos mínimos y a un 5° básico si tomamos como estándar el curso en que se trata el tema en la mayoría de los 45 países participantes en el estudio del TIMSS³¹.

A continuación se presenta el Cuadro N° 1, con las preguntas del SIMCE clasificadas según el curso en que dicho contenido aparece en el currículo. Para determinar el curso se consideró el Decreto 4002 (1980), que es el currículo con el cual se hizo el SIMCE de 1997, luego el Decreto 40 (1996), que define los objetivos mínimos y los contenidos fundamentales actualmente vigentes, y en tercer lugar se utiliza el nivel identificado según la mayoría de los países del TIMSS. Se incluye también el puntaje de logro nacional por pregunta en el SIMCE de 1997.

En el Cuadro N° 1 se observa que, en general, en Chile se tratan más tardíamente los contenidos que en la mayoría de los países del estudio del TIMSS. Las naciones que obtienen puntajes sobre y en el promedio en el TIMSS tienen currículos más exigentes. Los alumnos avanzan más rápido de manera que al final de enseñanza básica han tenido la oportunidad de aprender más a fondo los contenidos de álgebra, medición, geometría, probabilidades y recolección, representación y análisis de datos.

Si se analizan los contenidos de la prueba que corresponden a 7° y 8° básico según los programas nacionales, se constata que, en general, se

³¹ Véase W. H. Schmidt, *Many Visions, Many Aims: A Cross National Investigation of Curricular Intentions in School Mathematics* (1997), Vol. 1.

CUADRO N° 1: CLASIFICACIÓN DE LOS ÍTEMS DEL SIMCE DE MATEMÁTICAS DE 1997 SEGÚN EL CURSO EN QUE SE ESTUDIAN LOS CONTENIDOS DE CADA UNA DE LAS PREGUNTAS

Contenido	Pregunta	% logro	Decreto 4002	Decreto 40	Promedio TIMSS
Sentido número	–descubrir si el 1 en 1.234 representa a las unidades, decenas, centenas o unidad de mil	85,00	3°	3°	2°
Propiedades de los números	–reconocer entre 0, 5, 60, 90 un múltiplo de 30	59,44	5°	5°	5°
	–reconocer que los múltiplos de 4 terminan en par	76,31	6°	*	6°
Operatoria	Multiplicar 305 x 123	82,03	5°	5°	4°
	Dividir 18.492:23	56,75	6°		5°
Fracciones	–reconocer la fracción 1/4 en un diagrama	66,13	4°	4°	3°
	–equivalencia de fracciones en dibujo (6/8 y 3/4)	72,31	4°	4°	3°
	–suma y resta de fracciones de igual denominador	74,10	6°	5°	5°
	–suma con distinto de nominador 1/4 + 1/3	50,53	6°	5°	5°
	–simplificar 4/20 por 4	62,96	6°	5°	6°
	–equivalencia distinto denominador	52,98	6°	5°	5°
Fracciones con números negativos	–ordenar fracciones con igual denominador con negativo fracciones negativas	48,06	7°	8°	6°
	–suma y resta combinada de fracciones con número negativo distinto denominador	38,97	7°	8°	6°
	–división de fracciones con números negativos	55,03	7°	8°	6°
	–multiplicación de fracciones con números negativos	57,92	7°	8°	6°
	–incógnita con fracciones con número negativo		7°	8°	6°
Números negativos	–sumar -5 + 3=	55,53	7°	8°	5°
	–multiplicar -150 por 7	75,34	7°	8°	6°
	–dividir -234 por 9	73,27	7°	8°	6°
	–dividir -20 por -4 y sumarlo a 4	66,78	7°	8°	6°

(continúa)

Contenido	Pregunta	% logro	Decreto 4002	Decreto 40	Promedio TIMSS
Decimales	–sumar 0,22+4,08+57,90	90,38	5°	6°	5°
Porcentajes	–calcular el 25% de 60.000	61,09	6°	7°	6°
Proporciones	–proporciones: resolver cuánto es el 100% de un curso si el 10% son 3 alumnos	67,11	8°	7°	6°
	–proporciones: si se tejen 3 chalecos con 48 ovillos, ¿cuántos ovillos se necesitan para 4?	60,44	8°	7°	5°
	–proporción de dinero ganado si se trabajan 60 días en vez de 15 con el mismo sueldo	54,92	8°	7°	6°
Geometría	–reconocer el concepto “cara” en poliedro	79,83	4°	2°	3°
	–construir un ángulo obtuso	64,98	5°	8°	5°
	–reconocer un radio en un dibujo	69,66	6°	7°	5°
	–reconocer ángulo recto como característica triángulo rectángulo	44,02	6°	5°	5°
	–clasificación de un ángulo según sus grados	60,53	7°	5°	5°
	–deducción de un ángulo del triángulo	56,86	8°	8°	6°
	–deducción de un ángulo de un cuadrilátero	46,16	8°	8°	6°
	–ángulo en rectas paralelas con una secante	54,00	8°	8°	6°
Medición	–Cálculo de volumen de un paralelepípedo	51,06	8°	7°	5°
	–Cálculo del área de un rectángulo de 3mts x 2 mts	45,35	6°	5°	5°
Representación de datos, probabilidad y estadística	–lectura de datos en un gráfico de torta con los porcentajes escritos en cada porción	67,05	7°	*	6°
	–sacar promedio de las siguientes notas 5-7-5-6-6	48,87	6°	*	5°
Funciones, relaciones,	–resolver $72.600 = \square \cdot 30$	71,43	7°	6°	6°
	–resolver $40 : \square = (-5)$	64,95	7°	8°	6°

(continúa)

Contenido	Pregunta	% logro	Decreto 4002	Decreto 40	Promedio TIMSS
Ecuaciones	-descubrir la incógnita en una multiplicación de fracciones				
	$-\frac{1}{3} \cdot \boxed{} = \frac{2}{15}$	52,13	7°	8°	6°
	-resolver la ecuación (-50) + X = (-25)	54,06	7°	8°	6°
Problemas de un paso	-resta con reserva (300 cms. - 37 cms.)	75,88	3°	3°	3°
	-multiplicación 138 x 24	82,03	4°	5°	4°
	-división \$18.810:3	56,75	5°	5°	5°
Problemas de dos pasos	-multiplicar 12 por 9 y luego restar al resultado 72	60,52	6°	5°	5°
Problemas de razonamiento	-¿Qué hace falta para resolver el problema?	71,40	5°	*	5°
	-¿Tiene solución el problema?	69,05	6°	*	5°
Curso promedio en que se trata el conjunto de temas del SIMCE		62,7	6°	5,8°	5,1°
		Promedio nacional			

* No figura como contenido en este documento.

refieren al tema de los números negativos, contenido que en la mayoría de los países se trata en cursos más tempranos, ya que no presenta dificultades mayores de razonamiento para los alumnos. Además, en contraste con el grueso de los países del TIMSS, en nuestro país los contenidos de proporciones y de geometría se postergan hacia los cursos finales de enseñanza básica. Entonces, el ‘curso promedio’ en que se trata el conjunto de temas del SIMCE aparece más elevado por el curso en que se ubicaron esos contenidos.

A partir del Cuadro N° 1 y del dato de que sólo el 35,6% de los alumnos logró tener más del 70% de respuestas correctas en la prueba, se puede concluir que existiría un 64,4% de estudiantes que no alcanza el nivel satisfactorio de aprendizaje esperado para 6° básico en el programa chileno, o el de un 5° básico según parámetros internacionales, habiendo cursado ocho años de enseñanza básica.

Este panorama se agrava más aún al analizar la manera de formular las preguntas, porque se comprueba no sólo que los contenidos pertenecen a cursos bajos, sino que el grado de complejidad elegido para cada contenido corresponde a los niveles más elementales con que normalmente se tratan esos temas en los cursos que los trabajan por primera vez. Por ejemplo, se pide sumar los siguientes números decimales: $0,22 + 4,08 + 57,90$, obteniéndose un porcentaje de 90,38% de respuestas correctas. Este buen resultado era de esperarse porque la dificultad real de ordenar los decimales según su valor posicional fue soslayada: todos los números tenían la misma cantidad de dígitos después de la coma y las alternativas no proponían la posibilidad de que los niños alinearán los números de izquierda a derecha. En cambio, la prueba del TIMSS evalúa la comprensión del valor posicional de los decimales pidiendo sumar $3,75 + 5,625$, obteniendo un 66% de logro. La prueba del SIMCE contempla esta única pregunta sobre decimales, en cambio la prueba del TIMSS evalúa además multiplicación y división de decimales y transformación de decimales a fracciones. Por ejemplo, “Escriba 0,28 como una fracción reducida a su mínima expresión” (33% de logro). Esta pregunta refleja el mayor grado de dominio que se espera de los alumnos a nivel internacional. Otro ejemplo, esta vez del tema de medición, nos vuelve a mostrar que el SIMCE evalúa lo más básico de cada tema: la pregunta es:

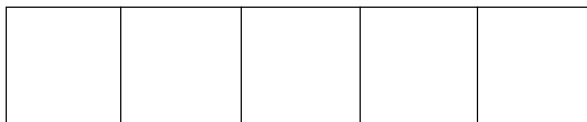
Se desea pintar un letrero rectangular de 3 metros de largo y dos metros de ancho.

¿Cuál es el **área** de la superficie que se desea pintar?

Si comparamos este ítem con uno del TIMSS (véase Figura N° 3), nuevamente comprobamos las diferencias de exigencia en el mismo tema:

FIGURA N° 3

La figura consiste en 5 cuadrados del mismo tamaño. El área total de la figura es de 405 cm².



–Encuentre el área de un cuadrado. Respuesta _____ centímetro cuadrados.

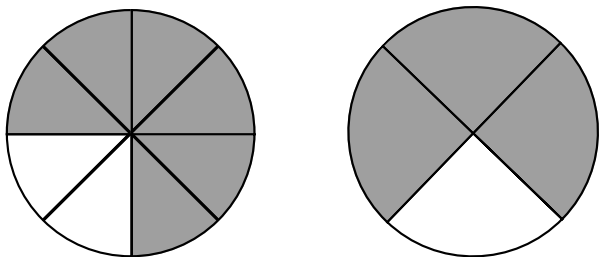
–Encuentre el largo de un lado de los cuadrados. Respuesta _____ centímetros.

–Encuentre el perímetro de la figura total en centímetros. Respuesta _____ centímetros.

El nivel de dificultad también disminuye con la entrega de pistas que facilitan la elección de las respuestas, por ejemplo, al interrogar sobre cuál de las fracciones que aparecen en las alternativas es equivalente a $\frac{1}{4}$, se entrega la ayuda adicional de dos diagramas donde está resuelto el problema gráficamente. En una pregunta sobre interpretación de un gráfico de torta se le entrega los porcentajes inscritos en cada porción, con lo cual la dificultad se reduce a la correcta lectura del achurado. (Véase Figura N° 4)

FIGURA N° 4

Observa atentamente las figuras y responde:



La fracción $\frac{6}{8}$ es equivalente a:

- A) $\frac{1}{3}$
- *B) $\frac{3}{4}$
- C) $\frac{4}{3}$
- D) $\frac{3}{6}$

Opciones	Porcentaje
A	10,24
B	72,31
C	8,75
D	8,70

Las respuestas alternativas también acotan las dificultades, en el problema que pide sumar $1/4 + 1/3$ todos los resultados tienen denominador 12, con lo cual los niños que intuitivamente quieren sumar los denominadores deben revisar su respuesta y adecuarla a las opciones existentes. Otra manera de facilitar los ítems consiste en resaltar las palabras claves de las preguntas en negrita, por ejemplo: “José **desea tener** un promedio 6,0 en Castellano. ¿Qué nota debe obtener en la prueba que le falta si sus calificaciones, hasta el momento, son: 5-7-5-6-6-?” La negrita enfatiza en esta pregunta la idea de que no se está pidiendo un promedio de las notas existentes, sino que la que se debiera obtener para lograr un promedio determinado. De las 48 preguntas 15 tienen este tipo de ayuda.

Esta prueba tampoco exige manejar problemas complejos ni tareas que impliquen un grado moderado de razonamiento para los cuales los alumnos de curso inferiores a 8° básico ya están preparados cognitivamente. Si se analizan las preguntas que requieren resolver problemas formulados verbalmente, se observa que éstas obtienen los más altos porcentajes de logro entre todos los ítems. Estos resultados contrarían las tendencias que normalmente se obtienen cuando se aplican problemas de razonamiento en pruebas de evaluación. En el TIMSS, por ejemplo, el porcentaje de logro más bajo se asoció a la resolución de problemas que implicaban dos o más pasos. Estas diferencias no se explicarían por la calidad de nuestra educación en este aspecto o por las habilidades de razonamiento del alumnado chileno, sino por la menor dificultad en los planteamientos de los problemas. De los doce que aparecen, nueve requieren que los niños realicen un solo cálculo, es decir, son operatoria encubierta. De los tres restantes, dos requieren que los niños determinen si se necesita o no más información y sólo uno exige dos pasos para ser resuelto. Si comparamos este único problema con alguno de los tantos del TIMSS, podremos contrastar inmediatamente la brecha entre los niveles de exigencia esperados entre la prueba chilena y la internacional. El ítem del SIMCE es el siguiente:

Un jardinero debe plantar 12 flores en cada uno de los 9 maceteros del jardín y lleva plantadas 72. ¿Cuántas flores **faltan** por plantar ?

(60,52% de logro.)

Entre los del TIMSS encontramos:

Un grupo de estudiantes tiene un total de 29 lápices y todos tienen al menos un lápiz. Seis estudiantes tienen 1 lápiz cada uno, 5 estudiantes tienen 3, y el resto tiene 2.

¿Cuántos estudiantes tienen sólo 2 lápices?

(47% de logro.)

O consideremos otro más difícil del TIMSS:

Los dos siguientes avisos aparecieron en el diario de un país cuya unidad monetaria son los *zeds*.

Edificio A

Espacio de oficinas disponible
85-95 metros cuadrados
475 *zeds* por mes

100-120 metros cuadrados
800 *zeds* por mes

Edificio B

Espacio de oficinas disponible
35-260 metros cuadrados
90 *zeds* por metro cuadrado
al año

Si una compañía está interesada en arrendar una oficina de 110 metros cuadrados por un año en ese país, ¿en cuál edificio de oficinas, A o B, debiera arrendar oficina para pagar el precio más bajo? Muestre su trabajo.

(19% de logro.)

Con este análisis se concluye que los ítems de nuestra prueba difícilmente corresponden al nivel que debiéramos aspirar para niños que han estudiado 8 años de enseñanza básica; el promedio nacional de respuestas correctas de un 62,7% obtenido no es un resultado satisfactorio si consideramos que la prueba tiene un nivel de dificultad más bajo que el que hemos definido como país, en el currículo, para ese curso; la señal que se envía al sistema es peligrosa por cuanto comunica que se podría aspirar a un menor nivel de exigencia y aun así obtener buenos resultados. Los resultados con esta prueba podrían inducir a un conformismo no merecido por parte del sistema, ya que un 62,7% de logro tiende a interpretarse como un nivel

satisfactorio de aprendizaje (en los colegios la nota 4 corresponde normalmente a un 50% de logro y en los establecimientos exigentes la escala asigna un 4 al 60% de respuestas correctas).

Cobertura de temas

Una comparación de los temas cubiertos por la prueba del SIMCE y la del TIMSS nos muestra que nuestra prueba considera un número mucho menor de temas que la evaluación internacional. La prueba del TIMSS es un buen parámetro de comparación por cuanto fue elaborada por un número significativo de expertos que revisaron los currículos de los 40 países participantes para determinar aquellos temas que se consideraban indispensables en 8° básico. Los temas que la prueba chilena no abarca y que sí están comprendidos en el TIMSS son:

- Estimación y redondeo de cantidades.
- Operatoria: propiedades de las operaciones.
- Fracciones y decimales: transformar de fracciones a decimales.
- Medición: lectura en una línea graduada, medición lineal, cálculo de problema con horas, conversión de unidades métricas, estimación de longitudes, identificación de unidades, márgenes de error.
- Proporciones: escala y proporción inversa.
- Gráficos: representación cartesiana, interpretación de curvas, operación con datos extraídos de gráficos, construcción de gráficos, completación de tablas a partir de gráficos.
- Geometría: simetría y medición de perímetros.
- Álgebra: patrones, completar secuencias, valorizar y ecuaciones, traducir problemas verbales a ecuaciones, utilizar fórmulas.

Una prueba de evaluación siempre implica una selección, pero la muestra debe ser representativa de los contenidos que se consideran más relevantes. De este modo los profesores y alumnos percibirán la correcta ponderación de cada tema. La prueba de 1997 no contempla contenidos que la mayoría de los países del TIMSS considera importantes, probablemente porque no estaban incluidos en el antiguo currículo; sin embargo, algunos temas como álgebra y estadística no figurarán en las evaluaciones futuras porque tampoco están presentes en los nuevos programas de estudio. Cabe preguntarse, por qué, si en la mayoría de los países los consideran parte importante de la formación básica de los alumnos, no forman parte de la de los nuestros.

Jerarquía de los temas

Otro nivel de análisis es la importancia que se le da a cada contenido, y éste se determina por el número de preguntas que se asigna a cada uno de ellos. El Cuadro N° 2 muestra la proporción de preguntas por área encontradas en el SIMCE y en el TIMSS³².

CUADRO N° 2

Contenidos	Proporción de preguntas dedicadas al contenido en el SIMCE	Proporción de preguntas dedicadas al contenido en el TIMSS
Sentido de número, operatoria, decimales y fracciones	71%	34%
Álgebra	2%	18%
Medición	4%	12%
Geometría	17%	15%
Representación de datos, análisis y probabilidades	2%	14%
Proporcionalidad	4%	7%

La importancia adjudicada a los distintos contenidos en el SIMCE de 8° muestra un peso desmedido hacia los temas de aritmética, descuidando la evaluación de los otros contenidos básicos. Nuestra jerarquía contrasta fuertemente con la del TIMSS, en la que encontramos una proporción de preguntas más equilibrada por cada área de contenido.

Si desglosamos las preguntas de aritmética en el SIMCE, tenemos que el 23% del total de ellas incluyen fracciones, que el 25% evalúa números negativos y el 13% está dedicado a operatoria simple de multiplicaciones y divisiones. El resto de las preguntas de aritmética asigna, en la mayoría de los casos, un solo ítem a los siguientes temas: valor posicional, cálculo de promedio, operatoria con decimales y teoría de números. La importancia asignada a los números negativos parece sobredimensionada en relación al espacio asignado a los demás temas. La prueba del TIMSS apenas considera este contenido, dándoles mayor importancia a temas como álgebra, geometría, etc. Las fracciones ocupan un lugar privilegiado

³² La distribución de los ítems de matemáticas según los contenidos se obtuvo utilizando las categorías y criterios de clasificación del TIMSS. Los porcentajes del TIMSS aparecen en: IEA, “TIMSS Mathematics Items: Released Set for Population 2 (Seventh and Eighth Grades)” (1996), p. iv.

tanto en el SIMCE como en la prueba internacional, en la cual el 17% de las preguntas se relacionan con el tema.

La revisión del número de preguntas por temas nos muestra que nuestra prueba tiene un balance precario. Para lograr una buena radiografía de los niveles de aprendizajes de los alumnos, la evaluación debiera equilibrar el número y la dificultad de los ítems de acuerdo a la importancia de cada contenido. De ese modo, el porcentaje final de respuestas correctas sería el reflejo del nivel de dominio de lo que interesa que los niños hayan aprendido en sus años de escolaridad. Dado que las pruebas de evaluación influyen en lo que se enseña, un desequilibrio en la elección de los ítems puede enviar señales equivocadas de las materias a las cuales se debe dedicar más tiempo.

Tipo de tareas exigidas

El TIMSS divide los ítems no sólo por los contenidos evaluados sino también por las destrezas exigidas para resolver cada una de las preguntas. Éstas pueden referirse al dominio de conocimientos, a la realización de procedimientos de cálculos rutinarios o complejos y al razonamiento en resolución de problemas.

La proporción encontrada en nuestra prueba y en el TIMSS es la siguiente³³:

CUADRO Nº 3

Tipo de destreza exigida	SIMCE	TIMSS
Conocimiento	25%	22%
Cálculos rutinarios	52%	25%
Cálculos complejos	15%	21%
Resolución de problemas	8%	32%

El TIMSS aparece como una prueba más balanceada. Da casi la misma importancia a todas las destrezas, poniendo un énfasis ligeramente mayor en las de resolución de problemas por su capacidad de medir comprensión y manejo de conceptos. Llama la atención que esta prueba, que se considera innovativa educacionalmente, les preste importancia a las pre-

³³ La distribución de los ítems de matemáticas según el tipo de tareas exigidas se obtuvo utilizando las categorías y criterios de clasificación del TIMSS. Los porcentajes del TIMSS aparecen en IEA's, "TIMSS Mathematics Items: Released Set for Population 2 (Seventh and Eighth Grades)" (1996), p. iv.

guntas que requieren dominio de conocimientos y a los cálculos rutinarios, ya que los educadores que se consideran de avanzada critican la utilización de la memoria y de los procedimientos que no requieren de un pensamiento crítico y de alto nivel. Sin embargo, aquí aparece validada la opción de asignar un papel importante al conocimiento de datos fundamentales en matemáticas y al dominio de algoritmos de cálculo, incluyendo a la vez un acento correcto en los procesos de razonamiento más complejos.

La prueba del SIMCE, en cambio, tiene un claro desequilibrio en la medición de destrezas, privilegiando las que permiten un manejo más mecánico de los contenidos. La señal que enviaría esta prueba al sistema es que los profesores debieran privilegiar un aprendizaje mecánico de las matemáticas para obtener buenos resultados. En las futuras pruebas este problema no estará presente porque se invierte esta proporción observada, desbalanceándose, esta vez, hacia preguntas de razonamiento complejo.

A partir de los porcentajes de logro generales obtenidos en la prueba de 1997, no se podría deducir cuán bien están razonando, comprendiendo y sabiendo aplicar los conceptos los alumnos de 8° básico, ya que la prueba incluyó un número muy reducido de este tipo de tareas. Por lo tanto, los resultados deben ser interpretados en torno a la capacidad de los alumnos para resolver operaciones y el dominio de datos matemáticos básicos.

Niveles de discriminación de los ítems

Los ítems tienen un rango de dificultad muy estrecho: las preguntas que son muy difíciles, es decir, las que debieran obtener un bajo porcentaje de respuestas correctas, no están presentes. La pregunta que tiene menor porcentaje de logro tiene un 38,97%, la que le sigue tiene el 44,02% de logro y solamente 3 ítems más tienen menos de un 50% de logro. En cambio en el TIMSS encontramos preguntas que obtienen desde un 8% de logro, seguida por una graduación continua hasta el 90%.

La teoría clásica de evaluación utilizada para lograr discriminación de rendimiento entre establecimientos implica seleccionar sólo aquellos ítems que han logrado ser respondidos en promedio por el 60% de los alumnos en la aplicación experimental³⁴. Esta medida dejaría afuera todas

³⁴ Véase Ministerio de Educación, SIMCE, *Folleto Técnico 1995: Directivos y Docentes. Educación Básica* (1995). En la prueba experimental que se realiza todos los años para seleccionar los ítems definitivos, se aplican 10 formas paralelas de modo que cada ítem debe ser respondido por 400 niños de una muestra no probabilística. Luego se seleccionan aquellas preguntas que tienen entre 30% y 80% de respuestas correctas, cuidando que el promedio final sea de un 60%.

aquellas preguntas que podrían indicar que los alumnos no están aprendiendo lo que el currículo nacional ha determinado, que los alumnos debieran aprender. En este sentido la prueba no estaría referida a un criterio fijo de excelencia, sino a estándares relativos fijados por la muestra de los alumnos de la prueba experimental. Esto implica que los resultados no serían indicadores del nivel de aprendizaje alcanzado en relación a los objetivos finales de enseñanza básica. Contar con preguntas que reflejen lo que realmente se espera que los alumnos aprendan, aun cuando sean de un alto nivel de dificultad, permitiría tener indicadores reales de los logros de nuestra educación. La actual prueba nacional establece un techo bajo, por lo que los alumnos de rendimiento mediocre pueden obtener altos porcentajes de respuestas correctas. Ésta es una señal que puede inducir a una conformidad que no debiera existir si no se ha alcanzado un nivel óptimo de aprendizaje en los objetivos considerados mínimos y fundamentales para el curso actual en el que está matriculado el alumno.

El nuevo SIMCE de matemáticas de 2° medio

El rediseño de la prueba SIMCE se ha emprendido con el fin de poner al día nuestra prueba con las nuevas teorías de medición y con los nuevos programas de estudio. Se distinguen tres fases en este proceso: las pruebas de primera, segunda y tercera generación.

Las orientaciones que discutiremos a continuación se refieren a las que se entregaron en 1998 para la preparación de la prueba matemática de II medio a rendir ese año y corresponden a las de segunda generación. Miden los objetivos de los antiguos planes y programas de estudio, pero excluyen aquellos contenidos considerados obsoletos en las nuevas propuestas curriculares. La nueva evaluación muestra cambios importantes que pretenden incorporar los procesos de pensamiento de alto nivel que plantea el reciente cambio curricular. De acuerdo a estos lineamientos, la proporción de ítems destinados al sondeo del aprendizaje de procedimientos operatorios rutinarios y al conocimiento de conceptos básicos disminuye a un 16,6% de la prueba, en cambio aumenta a un 73,3% los ítems destinados a la medición de razonamiento inductivo-deductivo y se agrega un 10% de preguntas dirigidas a la creatividad. Se introduce además la idea de incorporar un 33% de preguntas abiertas y dejar un 66,6% para las de selección de alternativas³⁵.

³⁵ Las cifras aparecen en Ministerio de Educación, *Orientaciones para la medición de logros. Asignatura de matemática: Segundo año medio 1998* (1998).

Las nuevas propuestas son auspiciosas en el sentido de que reorientan la enseñanza de las matemáticas hacia un aprendizaje razonado; sin embargo, tienen componentes que pueden resultar equivocados. En primer lugar, el vuelco en la proporción de las preguntas destinadas al sondeo de las distintas habilidades es demasiado extremo y nuevamente cae en un desequilibrio. Si se observa el Cuadro N° 4, nos damos cuenta de que el TIMSS tiene una posición mucho más equilibrada, asignándole un rol importante tanto al manejo de conocimientos y procedimientos de rutina como a las actividades que implican razonamiento más elevado. En cambio, la nueva prueba casi niega el papel que juega la memorización de datos en matemáticas y el dominio de procedimientos rutinarios.

CUADRO N° 4: PROPORCIÓN DE LOS ÍTEMS SEGÚN LAS HABILIDADES REQUERIDAS PARA ENFRENTARLOS ³⁶

	SIMCE II° 1998	SIMCE 8° 1997	TIMSS 8°
Memoria y procedimiento	16,6%	77%	47%
Razonamiento inductivo-deductivo	73,3%	23%	53%
Creatividad	10%	0%	0%

El marco conceptual de la prueba también puede inducir a los establecimientos a orientar la educación matemática hacia un camino peligroso y que seguramente no es el que el SIMCE promueve. En la sección “Marco Conceptual de la Medición”, aparece un listado con las características del pensamiento superior que fueron consideradas para la elaboración de la prueba³⁷. Éstas corresponden a las categorías formuladas por Resnik³⁸ que sostienen que el pensamiento de alto nivel:

³⁶ El cálculo de las proporciones que aparecen en el cuadro se obtuvieron respectivamente de: SIMCE II en Ministerio de Educación, *Orientaciones para la medición de logros. Asignatura de matemática: Segundo año medio 1998* (1998). SIMCE 8° se calcularon en base a los criterios de tabulación utilizados en la prueba de la población 2 del TIMSS. TIMSS 8° se obtuvieron en IEA, “TIMSS Mathematics Items: Released Set for Population 2 (Seventh and Eighth Grades)” (1996), p. iv.

³⁷ Véase Ministerio de Educación, “Orientaciones para la medición de logros. Asignatura de matemática: Segundo año medio” (1998).

³⁸ *Ibíd*em, p. 7.

- es no algorítmico, es decir, el curso de acción no está completamente determinado desde un principio;
- tiende a ser complejo; el curso para llegar a la solución no es aparente a partir de un punto aislado cualquiera;
- suele producir múltiples soluciones, cada uno de sus costos y beneficios, más que una sola (u óptima solución);
- requiere juicios e interpretaciones sutiles;
- involucra incertidumbre, no todo lo que está implicado en la tarea es conocido;
- involucra la regulación de los propios procesos de pensamiento.
- requiere esfuerzo.

Estos postulados, leídos en el contexto de la educación matemática, pueden inducir a pensar que se pretende introducir a los niños en una matemática mucho más compleja de lo que, en realidad, ha sido el nivel de matemática que normalmente se enseña a los niños. Así, la prueba del TIMSS, los textos de estudio extranjeros y los currículos de países de alto nivel educacional revisados tienden a exigir de sus alumnos un pensamiento que sigue una lógica ordenada, que normalmente busca llegar a una solución y que, por lo general, no involucra incertidumbre, pues los datos son conocidos. Sin embargo, esto no implica que los esfuerzos cognitivos que tienen que realizar los alumnos sean menores, pues está involucrada la construcción mental de un bagaje amplio de conceptos y herramientas abstractas, que deben ser comprendidas y asimiladas para su correcta utilización. El marco conceptual de Resnik aparece en los documentos sobre las nuevas orientaciones para las pruebas de las cuatro asignaturas a evaluar por el SIMCE, por lo tanto no es un marco teórico específicamente pensado para matemáticas. Parece peligroso, entonces, tratar de hacer calzar este esquema a una disciplina que, al menos en toda la etapa de construcción de los conceptos básicos, busca desarrollar un pensamiento lógico-algorítmico. De hecho, las preguntas de ejemplo que se incluyen en este documento de orientación de la nueva prueba de matemáticas no se ajustan a la medición de un pensamiento de alto nivel como el descrito por los postulados. En general, son preguntas que tienen una sola respuesta correcta, siguen un patrón algorítmico de pensamiento y muchas de ellas pueden ser resueltas en forma mecánica si se han aprendido los procedimientos matemáticos necesarios. Estas preguntas requieren sí de un grado mucho mayor de razonamiento que la prueba anterior del SIMCE. La medida de incorporar con mayor fuerza preguntas que exigen razonamiento y aplicación de conceptos a situaciones concretas es adecuada. De este modo se asegura que los

alumnos comprendan los conceptos, conozcan las relaciones entre ellos, establezcan las ligazones que tienen con la realidad y desarrollen la capacidad de seguir secuencias encadenadas de pensamientos. Pero no parece recomendable centrar la prueba en preguntas que impliquen importantes cuotas de ingenio para resolverlas, porque se terminará evaluando el índice de inteligencia de los alumnos en vez de la calidad de su aprendizaje en matemáticas. Justamente, las herramientas matemáticas permiten enfrentar con mayor facilidad problemas que sin ellas requerirían de grandes dosis de ingenio. Si las señales apuntan desbalanceadamente en este sentido, se estaría induciendo a realizar clases de matemáticas centradas en juegos y se descuidaría la enseñanza sistemática de conceptos con su debida ejercitación razonada.

Otro de los problemas que presenta la nueva prueba es que el listado de contenidos a evaluar aparece nuevamente restringido con respecto a la prueba del TIMSS. No se incluyen temas básicos de representación de datos, probabilidades y medición. Es posible que este déficit se aminore en las pruebas de tercera generación cuando entren en vigencia los nuevos programas de estudio. Sin embargo, no creemos que la diferencia desaparezca del todo porque nuestro currículo es de objetivos mínimos y contenidos fundamentales, por lo tanto es esperable que muchos de los contenidos que se exigen internacionalmente no sean requeridos en nuestras evaluaciones.

El nivel de dificultad de estos ítems se ajusta mejor a los niveles de exigencia internacionales que la prueba del SIMCE de octavo de 1997. Si tomamos los ítems de octavo y los de finales de enseñanza secundaria del TIMSS, encontramos que los de la nueva prueba corresponden a un nivel de dificultad intermedia entre estos dos cursos. Ésta es una buena señal; es de esperar que este mayor nivel de dificultad se haya mantenido en la prueba real que rindieron los alumnos de II medio en 1998, cuyas preguntas y resultados conoceremos en 1999.

La inclusión de ítems abiertos debiera estudiarse bien por los altos costos que implica y porque las investigaciones demuestran que las preguntas de alternativas múltiples son igualmente capaces de medir destrezas sofisticadas, que requieren alto nivel de pensamiento. En el documento de orientación aparece un gran número de ítems abiertos, probablemente para familiarizar a los colegios y alumnos con este nuevo formato, pero muchas de ellas podrían reemplazarse por preguntas de alternativa. La revisión del TIMSS, que también incluye preguntas abiertas, muestra que éstas se han dejado solamente para aquellos casos en que se exige que el alumno elaborar un producto como figuras geométricas o gráficos o cuando las respuestas

pueden ser varias y las alternativas reducirían o facilitarían las respuestas. Por ejemplo: “Escriba una fracción mayor que $2/7$ ”.

La orientación general de elevar el nivel de dificultad y de incluir en una proporción mayor preguntas que requieran el desarrollo del razonamiento matemático son auspiciosas. Sin embargo, si se mantiene la política de fijar el promedio en un 60% de logro para la prueba nacional en la prueba experimental, estos aumentos de dificultad pueden no ser sostenibles.

Conclusiones

La idea de publicar las preguntas de las pruebas de evaluación es una medida acertada que contribuirá a que el sistema pueda aprender más acerca de sus fortalezas y deficiencias. Los ítems dan una idea mucho más concreta del nivel de aprendizaje esperado y de los logros, que un listado de contenidos y objetivos que pueden tener referentes muy distintos para cada lector. Para los apoderados y para la opinión pública es más fácil entender lo que significan los resultados si tienen a mano lo que se exige saber en concreto, a través de las preguntas mismas. Una adecuada comprensión de los logros y de las deficiencias es necesaria para generar la discusión profesional al interior de los establecimientos. También contribuye a generar el compromiso de los padres con la educación de sus hijos, dándoles información que permita fundar la toma de decisiones. Con esta información, ellos podrán exigir más al colegio y a su hijo o decidir un cambio de establecimiento. Es importante, entonces, mantener esta política de apertura, aun cuando implique un cierto costo el generar nuevos ítems cada año. Esperamos que no sea una política pasajera, fundada sólo en el hecho de que la prueba de 1997 sería reemplazada por una de segunda generación. Sería recomendable que las preguntas de la prueba fueran difundidas al público, ya que esta vez sólo circularon al interior de los establecimientos educacionales. Igualmente, la difusión de los resultados debería ser más eficiente. No basta con comunicar éstos a la dirección de los establecimientos ni con publicarlos en un diario no siempre asequible a la generalidad de la población. Una política correcta sería exigir a los colegios publicar los resultados en un lugar visible o entregar un comunicado escrito a los apoderados. Asimismo, las bases de datos y los criterios técnicos de construcción de la prueba debieran estar disponibles fácilmente para los investigadores. Esto conduciría a una más correcta y completa interpretación de los resultados y a la vez daría la posibilidad de aumentar el cuerpo

de conocimientos sobre el tema. Invertir recursos para poner la información técnica en Internet estimularía la investigación y contribuiría a la transparencia del sistema.

El análisis de los resultados de la prueba de 1997 publicados en 1998 muestra que los resultados son aún menos satisfactorios de lo que se ha dado a entender. Tener en matemáticas un 62,7% y en lenguaje un 65,2% de logro en una prueba que en su mayoría mide conocimientos correspondientes a un 6° básico chileno y a un 5° básico internacional, después de haber cursado ocho años de enseñanza básica, muestra que la eficiencia de nuestro sistema educacional es baja. Una adecuada difusión del significado de estos resultados podría ayudar a comprometer a las personas con el actual proyecto del país de mejorar la calidad de la educación chilena. Por otra parte, los progresos en porcentajes de logro que se han evidenciado a lo largo de los años de aplicación del SIMCE no son fácilmente interpretables, porque los niveles de dificultad de la prueba no son absolutos sino que variables: éstos sufren ajustes de acuerdo a la validación experimental que recibe cada ítem cada año. La política del TIMSS de mantener un número de ítems constantes a lo largo de los años para fines de comparabilidad parece adecuada y es probable que se adopte porque la administración actual está interesada en hacer un seguimiento de los resultados de las reformas a través de las pruebas nacionales.

Los niveles de dificultad de la prueba son bajos, correspondiendo, en promedio, a contenidos y complejidad a dos y tres grados menores que el cursado por los alumnos. Una prueba con un bajo nivel de dificultad puede transformarse efectivamente en un factor limitante de la instrucción, en lugar de estimular aumentos en calidad, pues puede conducir a un conformismo infundado. Nada se saca con demostrar aumentos en los puntajes obtenidos si la prueba en su mayor parte no sobrepasa el nivel de 6° básico. Es mejor enfrentar con realismo el hecho de que nuestros niños, de todos los niveles, están aprendiendo poco y que hay diferencias inmensas con lo que aprenden y saben los niños de países desarrollados. Un cambio de la teoría de construcción de la prueba parece necesario para poder subir el nivel de exigencia de acuerdo a lo que realmente consideramos que es indispensable que los niños de nuestro país aprendan. Si se mantiene la política de seleccionar aquellos ítems que en promedio pueden ser respondidos por un 60% de los alumnos a los cuales se les aplica la prueba experimental, siempre estaremos adecuándonos a lo que los alumnos saben, más que a lo que definimos como necesario. Este período de rediseño de las pruebas es una excelente oportunidad para elevar los niveles de dificultad porque las diferencias de resultados se podrían adjudicar al cambio de

formato, evitándose así que el grueso del público malinterprete la caída de puntajes como un fracaso de la reforma educacional.

Como se desprende del marco teórico, estas pruebas tienen una gran capacidad de marcar rumbos en el desarrollo de los procesos educacionales. Una de las funciones que cumplen estas pruebas en general es ayudar a definir y especificar los objetivos de la enseñanza, y en este sentido el SIMCE cumple un rol aún más importante, dado que nuestros Objetivos Fundamentales y Contenidos Mínimos no explicitan objetivos de logro. Así, los colegios y los autores de textos escolares tendrán que basarse en las señales que dé la prueba SIMCE para esclarecer y determinar el tipo de tareas y el nivel en que deberán plantear su enseñanza. Por lo tanto, es preciso seleccionar con el mayor cuidado el tipo de tareas exigidas, los temas y el nivel de dificultad de la prueba. En matemáticas, por ejemplo, sería importante equilibrar la medición de las distintas destrezas que exige el ramo para enviar una señal clara de lo que es relevante que los profesores trabajen con sus alumnos. El desequilibrio que aparece entre la evaluación de conocimiento de datos, procedimientos rutinarios y de razonamiento, tanto en la prueba de matemáticas de 1997 como en la propuesta para 1998, no está de acuerdo con las teorías actuales de enseñanza de las matemáticas³⁹ ni con la experiencia de los profesores que logran buenos niveles de aprendizaje en matemáticas⁴⁰. Los porcentajes asignados a cada destreza por el TIMSS dan mejor cuenta de la importancia que se le debiera asignar a cada una de ellas. Jerarquiza adecuadamente, dando una mayor importancia al desarrollo de la capacidad de razonamiento sin descuidar el aprendizaje de conceptos y de procedimientos. En lenguaje, como los objetivos del programa de básica están expresados con un alto grado de generalidad, será la prueba la que finalmente establecerá las tareas específicas de comprensión de lectura y redacción que deberán lograr los alumnos. Es, por ello, importante que éstas representen un dominio cabal, completo y bien balanceado de las habilidades de lenguaje que se requieren en la vida real, evitando un énfasis exagerado en determinadas habilidades en detrimento de otras. Nos referimos, por ejemplo, al predominio de la lectura como búsqueda de información en la prueba de 8º y la total falta de habilidades de síntesis o de lectura crítica.

³⁹ John T. Breur, "Mathematics: Making it Meaningful" (1993).

⁴⁰ En el estudio del TIMSS los profesores de los países que logran altos resultados afirman en su mayoría que es imprescindible para el logro en matemáticas el recordar datos, fórmulas, el manejar procedimientos y la comprensión y aplicación de conceptos. Véase informe de resultados de la evaluación del TIMSS en A. E. Beaton *et al.*, *Mathematics Achievement in the Middle School Years: IEA'S Third International Mathematics and Science Study (TIMSS)* (1997).

El escaso número de contenidos que evalúa la prueba de matemáticas de 1997 debiera remediarse incorporando los temas que mundialmente se consideran imprescindibles. En este sentido, es probable que enfrentemos dificultades para incluirlos, por cuanto el nuevo currículo de estudio chileno no se ajusta a estas exigencias; por ejemplo, en octavo básico la mayoría de los países del TIMSS contempla manejo de datos, estadística, probabilidades, álgebra; en cambio, Chile no lo hace. Un problema paralelo enfrenta la asignatura de lenguaje, donde la notable disminución de contenidos de ortografía y gramática en las dos pruebas comentadas hacen temer por la desaparición total de estas materias en nuestra enseñanza, con el consiguiente deterioro de las habilidades expresivas y de la capacidad laboral y académica de los alumnos.

El análisis de la información divulgada sobre las pruebas SIMCE pone en relieve otro problema que aqueja aún a nuestra educación. Se trata de la falta de coordinación entre básica y media. La prueba SIMCE de 8° es especialmente relevante, porque se aplica al finalizar la etapa de enseñanza obligatoria. Desgraciadamente, lo que no se aprendió en la enseñanza básica es, para muchos, imposible de recuperar. Si se consideran los ítems del SIMCE de 8° como las metas para la enseñanza básica, podemos estar seguros que un enorme contingente de niños no estará preparado para enfrentar exitosamente la enseñanza media. No es aventurado pensar que la pronunciada deserción en 1° medio encuentre una explicación en esta discrepancia. Es fácil apreciar que hay una amplísima brecha entre los conocimientos y habilidades que evidencian estos ítems y los que los programas de enseñanza media dan por logrados al iniciarse 1° medio. Basta comparar el nivel de comprensión de lectura evidenciado en esta prueba con el nivel exigido en cualquier texto de física, biología o historia de enseñanza media para apreciar la magnitud de la brecha. Es decir, esta prueba no sólo hace prever dificultades para trabajar en el nivel adecuado en la asignatura de lenguaje, sino también para enfrentar las otras asignaturas por carecer de las habilidades de lenguaje necesarias.

Las nuevas pruebas del SIMCE tienen que restringirse a los nuevos objetivos mínimos y contenidos fundamentales, no pudiendo evaluar los contenidos complementarios que aparecerán en los programas propuestos por el Ministerio, porque pueden no calzar con los que definan los colegios que elaboran programas propios. Por lo tanto, es de suponer que la prueba del SIMCE estará muy por debajo de los estándares que fijarán los establecimientos que elaboran sus propios programas y de los estándares que

determinará el propio Ministerio de Educación a través de sus programas, ya que tendrá que circunscribirse a evaluar el mínimo común. Nuevamente esta señal inducirá a conformarse con bajos niveles de aprendizaje. El problema es difícil de resolver y quizás la única alternativa sea que aparezcan pruebas privadas paralelas que marquen una valla más alta o que decididamente el SIMCE evalúe un espectro más amplio de contenidos y un mayor rango de dificultades que los definidos por el currículo, interpretándose los resultados *en forma diferenciada* para los que logran el currículo mínimo y quienes se exceden de él. Sería recomendable que se mantenga la política inicial de que las pruebas nacionales de evaluación dependan de organismos independientes del Ministerio de Educación para evitar posibles conflictos de interés. La licitación de este proyecto podría abrirse a instituciones extranjeras.

El tema de la evaluación es uno de los pilares del mejoramiento de la calidad de la educación y una herramienta que pueden utilizar los padres y apoderados para hacerse parte de la formación de sus hijos. La tarea de mejorar las pruebas, la de difundir los resultados y la de educar a la comunidad escolar y ciudadana en relación a la naturaleza de esta evaluación ha sido iniciada por el Ministerio de Educación. Sin embargo, queda mucho por hacer para que las pruebas sean un aporte y no una señal que induzca al conformismo con una realidad mediocre. Es evidente que algunas de las falencias descritas en este estudio fueron ya percibidas por los organismos responsables: lo dado a conocer sobre las nuevas pruebas SIMCE para 2° así lo indican. Es de esperar que veamos en la próxima generación de pruebas SIMCE 8° una versión perfeccionada. Esto significa, en lenguaje, preguntas más cuidadosas; mayor coincidencia entre las tareas medidas y los actos comunicativos de la vida real (incluyendo la lectura literaria); un nivel de ortografía y gramática compatible con el término de la enseñanza básica, tareas de comprensión de lectura y redacción de un nivel de dificultad compatibles con las exigencias de la enseñanza media y un mayor énfasis en habilidades cognitivas superiores. En matemáticas, se espera, asimismo, que se eleven los niveles de exigencias para que concuerden con los objetivos fijados por el currículo; que se dé un equilibrio entre la medición de conocimiento básico, destrezas operacionales y razonamiento complejo; y, finalmente, que se incluyan los contenidos que se consideran imprescindibles en la mayoría de los países participantes en el TIMSS. Sólo así tendrá esta evaluación la legitimidad que se necesita para hacer de ella un verdadero instrumento de mejoramiento de la calidad educativa.

BIBLIOGRAFÍA

Artículos y Libros

- Breur, John T. "Mathematics: Making it Meaningful". En John T. Breur, *Schools for Thought. A Science of Learning in the Classroom*. Cambridge, MA.: The MIT Press, 1993.
- Dempster, Frank, N. "Using Test to Promote Classroom Learning". En Ronna F. Dillon, *Handbook on Testing*. Londres: Greenwood Press, 1997.
- Dillon, Ronna F. *Handbook on Testing*. Londres: Greenwood Press, 1997.
- Eyzaguirre, B.; y Fontaine L. (ed.). *El futuro en riesgo, nuestros textos escolares*. Santiago: Centro de Estudios Públicos, 1997.
- "Políticas educacionales comparadas: Consideraciones a partir del tercer estudio internacional de la enseñanza de matemáticas y ciencias (TIMSS)". *Estudios Públicos*, 73 (1999).
- Fontaine, L. "La asignatura de lenguaje vista en programas extranjeros estatales y privados". Centro de Estudios Públicos, Documento de Trabajo N° 285, septiembre, 1998.
- García Huidobro, Juan Eduardo; Cox, Cristián. "La reforma educacional chilena 1990-1998, visión de conjunto". En Juan Eduardo García Huidobro y Cristián Cox (eds.) *La reforma educacional chilena*, Madrid: Editorial Popular, 1999.
- Gardner, H., *La mente no escolarizada: Cómo piensan los niños y cómo deberían enseñar las escuelas*. Buenos Aires: Paidós, 1997.
- Gickling; Thompson (1985) "A Personal View of Curriculum-Based Assessment". En Joyce S. Choate *et al. Assessing and Programing Basic Curricular Skills*. Boston: Allyn & Bacon.
- Hanushek, E. A. "The Economic of Schooling: Production and Efficiency in Public Schools". *Journal of Economic Literature*, 24 (1986).
- y Jorgenson, D. W. *Improving America's Schools: The Role of Incentives*. Washington: National Academy Press, 1996.
- Hirsch, E. D. *The Schools We Need: Why We Don't Have Them*. Doubleday, Nueva York, 1996.
- Koretz, D. *et al.* "The Vermont Portfolio Assesment Program: Findings and Implications". *Educational Measurement Issues and Practice*, 13 (1994).
- "Using Student Assessments for Educational Accountability". En Hanushek E. A. y Jorgenson D. W. (ed.), *Improving America's Schools. The Role of Incentives*. Washington, D.C.: National Academy Press, 1996.
- Mc. Ginn, N. F.; y Borden, A. M. *Framing Questions, Constructing Answers. Linking Research with Education Policy for Developing Countries*. Cambridge, Ma.: Harvard Institute for International Development, 1995.
- Prophet, B.; y Rowell. "Botswana: Curriculum in Action: Classroom Observations in Botswana Junior Secondary Schools, 1987-1988. IEES, Florida, 1988". En N. F. Mc. Ginn y A. M. Borden, *Framing Questions, Constructing Answers. Linking Research with Education Policy for Developing Countries*. Cambridge, Ma.: Harvard Institute for International Development, 1995.
- Ravitch. "Value of Standardized Tests in Indicating How Well Students Are Learning". En Charles W. Daves (ed.), *The Uses and Misuses of Tests*. San Francisco: Jossey-Bas.

- Schmidt, W. H. *Many Visions, Many Aims. Volume 1: A Cross National Investigation of Curricular Intentions in School Mathematics*. Dordrecht: Kluwer Academic Publishers, 1997.
- Smith, M. L. "Put to the Test: The Effects of External Testing on Teachers". *Educational Researcher*, 20 (5), 1991.
- Weinstein, C. S.; y Mignano, A. *Elementary Classroom Management: Lessons from research and practice*. Nueva York: McGraw-Hill, 1993.
- Williams, W.; Blythe, T.; White, N.; Li, J.; Sternberg, R.; y Gardner, H. *Practical Intelligence for School*. Nueva York: Harper Collins, 1996.
- . *Are We Raising Smarter Children Today? School and Home Related Influences on IQ en The Raising Curve*. Editado por Ulric Neisser. Washington, DC.: APA, 1998.

Documentos

- Beaton, A. E.; Mullis, I. V. S.; Martin, M. O.; González, E. J.; Kelly, D. L.; y Smith, T. A. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chesnut Hill, MA: TIMSS International Study Center, Boston College, junio, 1997.
- Department for Education (Inglaterra). *The National Curriculum*. Inglaterra: HMSO, 1995.
- IEA (International Association for the Evaluation of Educational Achievement). "TIMSS Mathematic Items: Released Set for Population 2 (Seventh and Eighth Grades)". Chesnut Hill, MA: TIMSS International Study Center, Boston College, 1996.
- Ministerio de Educación. "Planes y programas de estudio para la educación general básica". *Revista de Educación*, 79 (1980) Santiago.
- . "Planes y programas de estudio de la educación media". *Revista de Educación*, 94 (1980) Santiago.
- . *Manual de orientaciones para la interpretación de resultados: Educación media 1994*. Santiago: Mineduc, 1994.
- . (SIMCE) *Folleto Técnico 1995. Directivos y Docentes*. Santiago: Mineduc, 1995.
- . *Objetivos fundamentales y contenidos mínimos obligatorios de la educación básica chilena*. Santiago: Mineduc, 1996.
- . *Plan y programas de estudio para el primer y segundo año de enseñanza básica*. Santiago: Mineduc, 1996.
- . *Plan y programas de estudio para el tercer y cuarto año de enseñanza básica*. Santiago: Mineduc, 1997.
- . *Objetivos fundamentales y contenidos mínimos obligatorios de la educación media 1998*. Santiago: Mineduc, 1998.
- . *Castellano: Resultados SIMCE octavos años, 1997*. Santiago: Mineduc, julio, 1998.
- . *Sistema de medición de calidad de la educación. Orientaciones para la medición de logros. Asignatura de castellano, segundo año medio*. Santiago: Mineduc, 1998.
- . *Sistema de medición de calidad de la educación. Orientaciones para la medición de logros. Asignatura de matemáticas. Segundo año medio*. Santiago: Mineduc, 1998.
- . *Orientaciones para la medición de logros. Asignatura de matemática: segundo año medio 1998*. Santiago: Ministerio de Educación, 1998.
- . (SIMCE). "Medición y evaluación de la educación nacional: Plan estratégico 1998-2002". Documento no oficial, 1998. □

