

EL SIMCE: SU DESARROLLO Y DESAFÍOS ACTUALES

SIMCE: Current development and challenges

LORENA MECKES G.*

Resumen

En este capítulo se presenta el desarrollo que ha tenido el sistema nacional de medición de logros de aprendizaje los últimos años. Se da cuenta en forma sintética del trabajo y de las principales conclusiones de la *Comisión para el desarrollo y uso del sistema de medición de la calidad de la educación* convocada por el Ministro de Educación que entregó su informe final a inicios de diciembre de 2003. Finalmente se profundiza en la primera de las recomendaciones de la Comisión para el SIMCE: establecer una medición referida a estándares de desempeño.

Abstract

This paper shows the development of the national testing system for students' learning. The committee of development and use of education quality testing system has elaborated a paper that examines the Chilean learning measurement approach. A deep analysis of the committee suggestions is displayed and the use of performance standards in measurement is proposed.

* Coordinadora Nacional del SIMCE.

El SIMCE entre 1990 y 1997

El SIMCE es el sistema nacional de medición de resultados de aprendizaje de los alumnos del sistema escolar. Ese ha sido su foco de atención prioritario desde sus inicios, a pesar de que su denominación (Sistema de Medición de la Calidad de la Educación) sugiere un ámbito de evaluación más amplio. El SIMCE se inició en 1988, precedido por el PER (Programa de Evaluación del Rendimiento Escolar). Inicialmente desarrollado por la Universidad Católica, pasó a ser administrado por el Ministerio de Educación a partir de 1990.

Entre 1990 y 1997 las mediciones censales se concentraron en logros académicos de Matemática y Castellano (la medición de Ciencias Naturales e Historia y Geografía se realizaba a través de una muestra), recogiendo también información sobre desarrollo personal de los alumnos, actitudes hacia el medio ambiente, desarrollo de la creatividad, y sobre la aceptación de la labor educacional por parte de los padres, como también información técnico administrativa de los establecimientos (tasas de aprobación, retiro, reprobación). Los niveles evaluados por el SIMCE son –hasta hoy– 4º, 8º y 2º medio.

La escala de puntajes utilizada en este periodo correspondía al porcentaje de respuestas correctas sobre el total de preguntas presentadas en la prueba, y a partir de 1995 se publicaba, para cada establecimiento, el porcentaje promedio de respuestas correctas de sus alumnos para cada asignatura. El modelo de medición y la construcción de las pruebas sólo permitía comparaciones técnicamente válidas de los resultados de una misma cohorte, por ejemplo entre establecimientos o agrupaciones de establecimientos en una misma medición. Dado que entre los propósitos originales no se contaba la evaluación de tendencias en el tiempo, las comparaciones de resultados de una misma cohorte satisfacían los objetivos iniciales de ofrecer a los apoderados información comparativa que les permitiera hacer opciones para la educación de sus hijos además de ofrecer información a las instancias ministeriales para concentrar recursos y esfuerzos de supervisión. En síntesis, la construcción de las pruebas y el modelo o

teoría clásica de medición utilizado hasta 1997, no permite discriminar si un aumento o disminución en el puntaje de una medición a la siguiente obedece a variaciones en el nivel de dificultad de las pruebas utilizadas en cada caso o a las diferencias de rendimiento de las cohortes que se comparan. A pesar de estas limitaciones, a los propósitos originales del SIMCE se agregó el de evaluar tendencias a lo largo del tiempo tanto a nivel desagregado como a nivel nacional y sus puntajes comenzaron a ser utilizados para establecer comparaciones entre mediciones de años sucesivos.

Durante este periodo el sistema se consolidó, se hizo parte de la tradición del sistema escolar y la publicación de los resultados pasó a ser un hito no sólo en la vida de cada unidad educativa, sino también de la opinión pública a través de la atención que le otorgan los medios de comunicación masiva. Esta importancia ha sido creciente. En el año 1992 se publicaron 40 artículos en el diario El Mercurio, a propósito de los resultados del SIMCE. El año 2000, en cambio, este número ascendió a 137 artículos.

Desarrollos del SIMCE a partir de 1998

Los tres ejes centrales de desarrollo del SIMCE a partir de 1998 han sido los siguientes:

1. Alinear las pruebas nacionales a la reforma curricular en curso y comunicar los nuevos lineamientos de la medición a los docentes, con el objeto de dar mayor impulso a la implementación del nuevo currículum en aula.
2. Incorporar estudios internacionales con el propósito de situar el rendimiento de los estudiantes chilenos en un contexto internacional y retroalimentar el diseño curricular.
3. Modernizar la metodología de medición utilizada, permitiendo comparaciones de resultados y estudio de tendencias en el tiempo.

Las pruebas se orientaron a la evaluación de los objetivos y contenidos del nuevo marco curricular. El nuevo marco curricular hace

énfasis en el desarrollo de habilidades cognitivas de orden superior. El marco curricular para la educación media, por ejemplo, establece entre los objetivos fundamentales transversales, “el desarrollo del pensamiento”, especificando entre las habilidades cognitivas de orden superior, las de investigación, las de comunicación, las de resolución de problemas y las de análisis, interpretación y síntesis de información¹. Por este motivo, uno de los ejes de desarrollo de las pruebas ha consistido en incorporar preguntas orientadas a la medición de algunas de estas habilidades, incluyendo además preguntas abiertas o de desarrollo para todas las áreas curriculares evaluadas. En esta misma dirección y para facilitar la corrección confiable de estas respuestas, se ha desarrollado un software que permite la codificación o corrección en línea, considerando el cálculo de niveles de consistencia entre correctores.

En conjunto con la modificación del tipo de preguntas y del contenido al cual ellas están referidas, se ha puesto énfasis en la publicación de ejemplos de preguntas y criterios de corrección en el caso de las preguntas abiertas. Se ha considerado que la difusión de modelos de preguntas entre los profesores, es una forma de comunicarles cuáles son las expectativas de aprendizaje implicadas en los objetivos del nuevo marco curricular. Es conocida la influencia de los programas de evaluación con altas consecuencias en motivar el cambio educacional. El SIMCE es un sistema de medición de altas consecuencias para las escuelas, en la medida que sus resultados se hacen públicos y pueden tener incidencia en la matrícula y consecuentemente en el financiamiento de las escuelas. En este sentido, se ha concebido el SIMCE como uno de los agentes relevantes en el proceso de reforma, jugando un rol en la dirección y motivación de la implementación curricular.

El segundo eje de desarrollo del SIMCE ha sido la incorporación de estudios internacionales, para poner el rendimiento de los

¹ Ministerio de Educación (1998). *Currículum de la Educación Media: Objetivos Fundamentales y Contenidos Mínimos Obligatorios*.

estudiantes chilenos no sólo en referencia a las demandas del currículum nacional, sino en el contexto del desempeño de los alumnos a nivel internacional. Es así como el SIMCE llevó adelante en Chile el Laboratorio Latinoamericano de la Calidad de la Educación (LLECE), desarrollado por la UNESCO, el cual se concentró en matemáticas y lenguaje en 3° y 4° básico. Esta medición se aplicó en 1997. También se ha participado en dos oportunidades en el Estudio Internacional de tendencias en Matemáticas y Ciencias (TIMSS 1999 y 2003) específicamente en 8° básico. Asimismo, Chile ha participado en el estudio Internacional de Educación Cívica para jóvenes de 8° y 4° medio, en 1999 y 2000 respectivamente. Estos dos último estudios dependen de la IEA. Finalmente, Chile también ha participado en PISA (Programa Internacional de Evaluación de Estudiantes), el cual tiene como propósito evaluar las capacidades básicas de estudiantes de 15 años que son necesarias para desempeñarse en forma efectiva en la sociedad en las áreas de lectura, matemática y ciencia. Específicamente, Chile participó en el ciclo 2001 (PISA +) y está comprometido para la medición de 2006. PISA es conducido a nivel internacional por la OCDE.

El tercer eje central de avance del SIMCE ha consistido en actualizar el modelo de medición utilizado, otorgando mayor solidez a las comparaciones longitudinales de resultados. A partir de 1998 se migró a la Teoría de Respuesta al Ítem (IRT). Este es el modelo de medición utilizado por estudios internacionales como PISA y TIMSS, y es recomendado para mediciones a gran escala como es el SIMCE (Hamilton, L. *et al.*, 2002). Permite hacer comparaciones sólidas de resultados de distintas mediciones y estudiar tendencias. Además, dado que este modelo permite separar el efecto de la dificultad de las pruebas del efecto del nivel de rendimiento de los sujetos evaluados, es posible incorporar preguntas de muy diverso nivel de exigencia y obtener información sobre lo que saben y pueden hacer alumnos de muy distinto desempeño. Por lo mismo, se hace posible describir el desempeño característico de alumnos que se encuentran en puntos distintos de la escala de puntajes. Finalmente, el modelo permite hacer estudios de sesgo de las preguntas.

La posibilidad de describir el desempeño característico de distintos rangos de puntaje, acerca esta medición a una medición referida a criterios. De acuerdo a Linn y Gronlund (2000) una medición referida a criterios es aquella diseñada para proveer una medida del desempeño interpretable en términos del grado de dominio en un área de aprendizaje. Su meta es obtener una descripción del conocimiento y habilidades que demuestra el estudiante. Describir las tareas que son capaces de resolver los alumnos dado su puntaje, es uno de los modos en que es posible interpretar los resultados de una medición referida a criterio, según estos autores. Como veremos más adelante, otra forma de entregar resultados en una medición referida a criterios es comparar el resultado con una definición de estándares de desempeño y decidir si el estudiante alcanza o no un determinado estándar (p. 43).

Finalmente, es importante mencionar una cuarta innovación en la entrega de resultados del SIMCE. A partir de la medición del año 2000, la publicación de los resultados de los establecimientos en la prensa y en la página web incluyen una comparación de los puntajes obtenidos por las escuelas con el promedio de los establecimientos que trabajan con una población de alumnos de similares características socioeconómicas. Ello con el fin de aportar nuevos parámetros de comparación que ayuden a discriminar de algún modo entre variables propias del desempeño del establecimiento y aquellas que son menos controlables por ellos. Asimismo, la publicación de resultados nacionales incluye un análisis considerando esta variable.

La Comisión y su propósito

A fines de mayo de 2003, a solicitud del Ministro de Educación, se conformó la Comisión para el desarrollo y uso del sistema de medición de la calidad de la educación.

El encargo de la Comisión era hacer un diagnóstico sobre el sistema y la información que éste proporciona, además de elaborar re-

comendaciones y propuestas para el desarrollo y uso de la medición tendientes a optimizar su impacto en la calidad de los aprendizajes. Algunas de las preguntas centrales para la Comisión fueron: ¿Cuál debe ser el propósito fundamental de la medición? ¿Qué medir? ¿Son los resultados de aprendizaje en las áreas evaluadas actualmente indicadores suficientes para dar cuenta de la calidad de la educación? ¿De qué modo puede el sistema de medición contribuir de manera efectiva al mejoramiento de la calidad de los aprendizajes? ¿Qué usos de los resultados estimular y qué consecuencias indeseadas evitar? ¿Qué información debe entregarse y cómo para que sea de mayor utilidad para sus distintos usuarios? ¿Debe ser el Ministerio de Educación el que realiza la medición? ¿Qué nos enseña la experiencia internacional sobre la medición de resultados de aprendizaje y su uso?

La Comisión se constituyó buscando combinar solidez técnica, amplitud y pluralidad de visiones sobre el sistema escolar y sobre políticas educativas, la representatividad de docentes, formadores de docentes, sostenedores y la experiencia sobre medición de aprendizajes. A cargo de la Comisión, se nombró como presidenta a la señora Erika Himmel, investigadora y catedrática de la Pontificia Universidad de Chile con una larga trayectoria en el campo de la evaluación educativa, miembro del Consejo Superior de Educación. Fueron miembros de la Comisión la señora Mariana Aylwin Oyarzún (ex Ministra de Educación), Abelardo Castro (Decano de la Facultad de Educación de la Universidad de Concepción), Bárbara Eyzaguirre (Investigadora Centro de Estudios Públicos), Juan Enrique Froemel (ex Coordinador Técnico del LLECE UNESCO), Juan Eduardo García Huidobro (Director de la Escuela de Educación Universidad Alberto Hurtado), Pablo González (Académico del Centro de Economía Aplicada, Facultad de Ingeniería de la Universidad de Chile y Consultor Unicef), Juan Iglesias (Decano de la Facultad de Humanidades y Educación de la Universidad de Atacama), Nicolás Majluf (Académico del Departamento de Ingeniería Industrial y Sistemas, Pontificia Universidad Católica de Chile), Jorge Manzi (Académico Experto en

Medición, Escuela de Psicología, Pontificia Universidad Católica de Chile), Patricia Matte (Presidenta de la Sociedad de Instrucción Primaria), Walter Oliva (Presidente de CONACEP), María de los Ángeles Santander (Investigadora del Instituto Libertad y Desarrollo), Guillermo Scherping (Director a cargo del Departamento de Educación y Perfeccionamiento del Colegio de Profesores) y monseñor Héctor Vargas B. (ex Presidente de la FIDE y actual obispo de Arica). La secretaría técnica estuvo formada por tres miembros del Ministerio de Educación, en donde participaron como miembros permanentes, Cristián Cox (Coordinador Nacional de la Unidad de Currículum y Evaluación) y Lorena Meckes. (Coordinadora Nacional del SIMCE). De acuerdo con los temas abordados, también participó Vivian Heyl del Departamento de Estudios y Estadísticas del Ministerio de Educación.

Se trabajó durante 15 sesiones entre los meses de junio y noviembre de 2003. Para llegar a un diagnóstico acabado del SIMCE se recogieron antecedentes y testimonios sobre el uso de sus resultados por parte de profesores sostenedores, directivos de establecimientos, apoderados, investigadores, medios de comunicación y responsables de políticas en educación. El trabajo de la Comisión contempló la revisión de la experiencia internacional en materia de medición de resultados de aprendizaje, para lo cual se contó con el estudio de cinco casos, realizado por Gilbert Valverde² y Margaret Forster³. Finalmente, se acordaron recomendaciones para incrementar el impacto de la medición en los resultados de aprendizaje.

El diagnóstico de la Comisión

El diagnóstico que la Comisión hizo del SIMCE, lo retrata como un sistema de medición de resultados de aprendizaje que está insta-

² Profesor del Departamento de Administración y políticas educativas de la Universidad de Albany, Nueva York.

³ Directora de Investigación en Evaluación e Información de resultados del Consejo Australiano para la Investigación en Educación.

lado en la tradición del sistema educativo, metodológicamente sólido, que goza de credibilidad y legitimidad, y que ha permitido focalizar progresivamente la atención en los logros de aprendizaje de los alumnos.

Respecto de la utilización de sus resultados, ha predominado su uso para políticas educativas (focalización de recursos, incentivos, información a la demanda, evaluación de impacto de políticas) por sobre su utilización por parte de docentes y padres de familia.

Si bien los resultados del SIMCE son objeto de análisis por parte de los docentes, demandan que el modo como se entrega la información les permita retroalimentar su práctica pedagógica. Concretamente, esperan un parámetro claro para determinar si el rendimiento de sus alumnos es o no suficiente, y permitan describir cuáles son sus logros y debilidades. Por otra parte, la Comisión tomó conocimiento de la brecha de formación que dificulta que los docentes utilicen los resultados de la medición, puesto que el análisis de la información entregada por mediciones nacionales e internacionales es una materia insuficientemente abordada en la formación inicial y en el desarrollo profesional permanente de los profesores.

Respecto de los padres de familia, la Comisión señala que ellos hacen escaso uso de los resultados del SIMCE, probablemente porque los canales de comunicación con ellos no son suficientemente directos o bien porque la información que se les provee no indica en qué medida el logro de los establecimientos es o no satisfactorio.

En relación con usos no deseados de los resultados del SIMCE, la Comisión conoció la preocupación de diversos sectores respecto de prácticas de selección o de exclusión de alumnos por motivos de rendimiento como estrategia para elevar los resultados. También se recogió la preocupación por la tendencia en algunos establecimientos a concentrar sus esfuerzos en la generación que rendirá la prueba, en desmedro de las demás cohortes de alumnos.

Las Recomendaciones de la Comisión

Las principales recomendaciones de la Comisión se concentran en los usos de los resultados del SIMCE que es necesario potenciar. Se concluye que las principales características del SIMCE deben conservarse: su foco en logros de aprendizaje, la medición censal de las escuelas en los mismos grados o niveles evaluados hasta ahora, la publicación de los resultados de cada establecimiento, la confidencialidad de los resultados obtenidos por alumnos individuales, el estudio de información de contexto para la mejor comprensión de los resultados y la realización de estudios internacionales.

Las propuestas de cambio fueron ordenadas según su prioridad, como sigue:

1. En primer lugar, la Comisión asignó la mayor relevancia a que el SIMCE reporte resultados en relación con estándares de desempeño. Es decir, que sus resultados vayan más allá de un puntaje, tengan significado para los usuarios de la información y sean interpretables de acuerdo con la descripción de un nivel de logro que se espera para los estudiantes del país. En términos más específicos, se espera que los resultados del SIMCE indiquen, además del puntaje promedio del de cada establecimiento, la proporción o porcentaje de alumnos que alcanza la expectativa de rendimiento nacional.
2. Comunicación más directa y efectiva de la información a diversos actores del sistema educativo, para involucrarlos responsablemente y para permitir que pidan o den cuenta por los resultados obtenidos. Se propone que, además de la información que hoy se entrega a docentes y directivos, los resultados del establecimiento se hagan llegar directamente a todos los apoderados y no sólo al centro de padres como se hace en la actualidad. Del mismo modo, se espera que se entregue información directamente a sostenedores tanto municipales como privados sobre los resultados de los establecimientos administrados por ellos.

En el caso de los docentes, se propone que el sistema de medición ponga a disposición de ellos la información que les permita enriquecer sus prácticas evaluativas y optimizar el uso que hacen de los resultados de la medición nacional. La expectativa es que el sistema de medición apoye las prácticas evaluativas en lo que se refiere a evaluación diagnóstica y sumativa (no se espera que el sistema nacional de medición sustituya la evaluación de aula que el profesor hace en forma constante). Para este efecto, la Comisión sugirió que se ponga a disposición de los docentes pruebas para otros niveles que los evaluados por el SIMCE externamente, proveyéndoles parámetros de comparación para que puedan contrastar el nivel de rendimiento de sus alumnos con el de los alumnos del país. También se recomendó que, una vez aplicadas las pruebas, el SIMCE libere el mayor número de preguntas posible, con sus resultados para que los docentes puedan hacer un análisis más rico de los mismos.

3. Intensificar la frecuencia de las mediciones para cada nivel y ampliar las áreas curriculares evaluadas. Se recomendó medir más de un nivel cada año, para que más cohortes de alumnos participen de la medición y a la vez se vean beneficiadas por los esfuerzos que hacen los establecimientos respecto de las generaciones que son evaluadas. Esta propuesta permite, además, evaluar cuánto han aprendido los alumnos, de una misma cohorte, entre una medición y la siguiente, lo cual entrega un indicador sobre el aporte del establecimiento en el avance de sus alumnos, sin importar el punto de partida de ellos.

Además de aumentar la frecuencia de la medición de cada grado evaluado, se sugirió incorporar, de un modo más esporádico, otras áreas curriculares como por ejemplo Inglés y Educación Tecnológica para incentivar una implementación más amplia del currículum.

4. La Comisión recomendó que a la hora de reportar resultados, se incluyan indicadores que favorezcan incentivos justos y válidos, en otras palabras, que no estimulen usos indeseados de los resultados. En este sentido, la evaluación del avance de una mis-

ma cohorte entre una medición y la siguiente permite evaluar cuánto ha aportado el establecimiento al aprendizaje de sus alumnos, considerando el punto de partida de ellos. Por otra parte, también se sugirió estudiar y reportar indicadores de selección o discriminación de alumnos por razones de rendimiento, para poder distinguir si el resultado del establecimiento obedece a dichas prácticas. Considerando esto mismo, se recomendó no entregar resultados de alumnos individuales mientras no se estudie el modo de garantizar que esta información no sea utilizada para selección o exclusión de alumnos⁴.

A continuación se describe en mayor profundidad la expectativa de estándares de desempeño, primera prioridad establecida por la Comisión.

Estándares de desempeño y la función formativa que se espera del SIMCE

Una de las expectativas más recurrentes respecto de los resultados entregados por el sistema de medición es que ellos señalen si el rendimiento exhibido por los alumnos alcanza o no un nivel esperado o satisfactorio, y que además les entregue información sobre logros y debilidades de los alumnos en cada área.

Como se ha señalado anteriormente, hasta ahora el SIMCE no ha entregado resultados en relación con una expectativa de rendimiento o desempeño. En ausencia de esta información, los docentes han “construido” este parámetro de comparación utilizando diversos referentes según la escala de puntajes utilizada por el SIMCE históricamente. Es así como en el periodo en que los resultados se reportaban en porcentaje medio de respuestas correctas (desde el inicio del SIMCE hasta 1997), el 100% de respuestas correctas se utilizaba como referente del máximo posible al cual aspirar y, considerando

⁴ Para mayor detalle sobre el diagnóstico y recomendaciones hechas por la Comisión, consultar su informe final (Comisión SIMCE, 2003). *Evaluación de aprendizajes para una educación de calidad*.

prácticas habituales de evaluación de aula, el 60% era significado como un nivel suficiente (equivalente al 4 en la escala de notas en sala de clases). Esta analogía no toma en cuenta que responder el 100% de las preguntas de una prueba no necesariamente implica haber alcanzado un nivel óptimo, ya que depende de las preguntas contenidas en el instrumento. Más adelante, con la adopción de IRT (Teoría de Respuesta al Ítem) y el cambio a una escala de puntajes de promedio 250 puntos⁵, los docentes han tendido a utilizar dicho promedio como sinónimo de un resultado aceptable. En este sentido, los establecimientos de promedio superior a 250 puntos tienden a hacer una evaluación positiva de sus resultados o negativa, si su puntaje promedio es inferior a 250 puntos. De este modo, el promedio nacional se ha transformado, tácitamente, en un estándar de desempeño⁶.

En todos estos mencionados, la demanda implícita por parte de los profesores –y también de los apoderados– es contar con lo que en medición se denomina “punto de corte”, es decir, un punto en la escala de puntajes, que permite distinguir si se ha alcanzado o no el nivel de rendimiento esperado. Es importante señalar, sin embargo, que en ninguno de estos casos ha habido un estudio ni tampoco un consenso que respalde que los puntajes que intuitivamente se han tomado como referentes representen realmente un nivel de rendimiento satisfactorio.

En segundo lugar, los docentes esperan que los resultados de la medición les entreguen información cualitativa sobre el logro de los alumnos, es decir, cuánto saben o de qué son capaces quienes obtienen diferentes puntajes. Las últimas mediciones de 4° básico (SIMCE 1999 y 2002) describieron, a nivel nacional, el desempeño que habían alcanzado alumnos de distintos rangos de puntaje en la prueba,

⁵ El primer año de uso de la escala en cada grado o nivel (año de línea de base), el promedio nacional se fijó en 250 puntos. Sin embargo, a partir de ese punto cada año el puntaje promedio puede ser mayor o menor que 250 puntos, según si los alumnos del país presentan un rendimiento mejor o inferior al del año de base.

⁶ La información sobre las interpretaciones de significado que han hecho los docentes de los puntajes del SIMCE provienen de entrevistas a docentes, realizadas tanto como parte del trabajo del SIMCE como también por la misión de la OCDE durante 2003 (en borrador preliminar de informe).

entregando además la distribución porcentual de los estudiantes del país en dichos niveles. Este modo de entregar resultados constituyó un avance importante en la línea de satisfacer la necesidad de resultados con mayor significado pedagógico (ver recuadro con extractos de la descripción de niveles de desempeño para comprensión de lectura en 4° básico). Esta descripción, sin embargo, no establecía cuál de los tres niveles descritos se consideraba adecuado para el grado evaluado. Por otra parte, al tratarse de una descripción que surge a posteriori, del análisis de patrones de respuesta a las pruebas administradas, está limitada a la caracterización de los contenidos que fueron incluidos en la medición y a las preguntas que se incluyeron en las pruebas.

NIVELES DE DESEMPEÑO EN PRUEBA DE LENGUAJE Y COMUNICACIÓN SIMCE 2002⁷

Nivel Básico:

Los alumnos son capaces de identificar distintos tipos de textos, tales como poemas, cuentos y noticias breves. Pueden identificar información presentada literalmente, sobre todo si se encuentra al comienzo o final de los textos leídos. También son capaces de determinar los sentimientos de los personajes cuando esta información aparece explícitamente en el texto.

Nivel Intermedio:

Además de lo anterior, en este nivel los alumnos son capaces de localizar datos específicos cuando leen, lo que les permite determinar e inferir secuencias de acciones, seleccionar ideas e identificar características de personajes.

Nivel Alto:

Además de lo descrito para los niveles precedentes, los alumnos comprenden el sentido global del texto que leen. Pueden caracterizar la situación comunicativa en que se inserta lo leído, es decir, identificar el emisor, receptor y finalidad del mensaje, de acuerdo con las características del texto.

⁷ Extracto de la descripción de los niveles de desempeño en comprensión de lectura en la prueba de Lenguaje y Comunicación de 4° básico 2002 (Informe de Resultados SIMCE 2002, 4° Básico, p. 25).

El marco curricular, sus objetivos fundamentales y contenidos mínimos obligatorios, constituyen el referente fundamental para la construcción de las pruebas SIMCE. Para Chile, los OFCMO pueden ser definidos como *estándares de contenido*, que definen qué deben saber y ser capaces de hacer los alumnos en determinados niveles o grados (Hansche, 1998). Sin embargo, su formulación está principalmente orientada a entregar lineamientos pedagógicos y no siempre permite establecer cuál es el nivel de dominio de dichos objetivos que se considera apropiado a la hora de evaluar. Por ejemplo, al señalar que los alumnos deben “leer comprensivamente textos breves y funcionales que amplíen el conocimiento de sí mismo y del entorno”⁸, no se describía la extensión y complejidad de dichos textos ni se especificaba el nivel de comprensión que se esperaba que alcanzaran, ni bajo qué criterios dicho objetivo debía considerarse logrado en el curso correspondiente.

Lo anterior no sólo tiene impacto en la medición y en los resultados que ella entrega, sino que en las expectativas y parámetros con que los docentes orientan su práctica pedagógica. Tal como lo muestra un estudio realizado por el equipo de Seguimiento de la Unidad de Currículum y Evaluación del Ministerio de Educación (2001), la interpretación de qué significa “leer comprensivamente” varía según las escuelas, especialmente en NB2, de acuerdo al nivel de exigencia y expectativas de los docentes respecto de sus alumnos. Se observaban variaciones en la extensión de los textos presentados a los alumnos y en el tipo de comprensión o elaboración que se les solicitaba a partir de ellos. Estas diferencias de expectativa y exigencia tienen a su vez impacto en los resultados de los alumnos. Así, por ejemplo, el estudio muestra que los cursos de docentes que pedían a sus alumnos una comprensión más compleja de los textos (por ejemplo hacer inferencias a partir de lo leído) en general pertenecían al grupo de escuelas con mejores puntajes en el SIMCE, a diferencia de los cursos cuyos docentes se limitaban a solicitar sólo identificación de información literal.

⁸ Decreto 240, 1999, marco curricular vigente para la prueba de 4° básico 2002.

Para el ejemplo de objetivo de comprensión de lectura, la modificación del marco curricular de NB1 y NB2, de octubre de 2002, incorpora un mayor nivel de especificación en las expectativas de logro al señalar explícitamente la extensión de los textos tanto literarios como no literarios, que se espera que sean capaces de leer los alumnos, así como el tipo de comprensión (inferencial y literal) que desarrollen. Los estándares de desempeño, sin embargo, darían un paso más y describirían, ejemplificando, qué tipo de inferencias se espera de un alumno en 4° básico, diferenciándola de las inferencias esperables para un alumno de 2° medio.

La expectativa que tienen los docentes respecto de la información que les proporciona el SIMCE puede comprenderse como un requerimiento de que sus resultados no sólo cumplan una función sumativa, sino que cumplan un propósito *formativo*. Si bien el SIMCE es por definición una evaluación sumativa (mide logros de aprendizaje al término de ciclos educativos), se espera que la información sobre los resultados alimente decisiones pedagógicas y de gestión, es decir, que de algún modo cumpla una función formativa. Es por ello que los informes de resultados entregados a los establecimientos entregan información comparativa sobre los puntajes obtenidos por el establecimiento y el promedio nacional, el puntaje máximo y mínimo obtenido por algún establecimiento, y la diferencia con el promedio de establecimientos que trabajan con una población escolar de semejantes características socioeconómicas. Adicionalmente, se publica una muestra de preguntas en que se analizan los contenidos evaluados y se presenta el resultado del establecimiento en cada una de ellas. Asimismo, en este informe, se presenta una guía para la discusión docente.

Sin embargo, la información proporcionada parece no ser suficiente. Para que una evaluación pueda asumir una función formativa, debe ir más allá. Sadler (1989) señala que –cualquiera sea la naturaleza de una evaluación– ella cumplirá sólo una función sumativa (y no formativa) cuando no es usada para ayudar a el o los sujetos evaluados a acortar la brecha entre el nivel logrado y el deseado, o cuan-

do la información está demasiado encriptada o codificada como para que sea posible adoptar acciones apropiadas a partir de ella (por ejemplo, cuando se restringe a una nota o a un puntaje).

Si bien la tipología de retroalimentación y evaluación formativa de Gipps y Tunstall (1997) fue desarrollada para dar cuenta de las prácticas evaluativas de docentes en aula, puede ser útil para comprender la necesidad de otro tipo de información para dar mayor potencia a la contribución del SIMCE al mejoramiento educativo. Las autoras describen cuatro tipos de retroalimentación a partir de la evaluación. Tres de ellos son de interés para una analogía con la retroalimentación que puede entregar el SIMCE. (1) El primer tipo consiste en premiar o castigar a partir del rendimiento mostrado, (2) el segundo es semejante al primero, pero está más dirigido a la tarea y consiste en aprobar o reprobar el trabajo realizado, (3) el tercero también se concentra en la tarea, pero va más allá de la aprobación o reprobación y consiste en describir en forma específica los aspectos logrados y lo que es necesario mejorar.

El primer y segundo de los tipos, cuando se dan puros, son más característicos de las evaluaciones sumativas, mientras que el tercero es más propio de las evaluaciones que cumplen un papel formativo. La investigación sobre aprendizaje y motivación muestra que la retroalimentación entregada a partir de una evaluación es más efectiva a medida que se aproxima al tercero de los tipos mencionados. El sólo refuerzo positivo o negativo (primero de los tipos mencionados) no es suficiente y sólo provee una motivación extrínseca por alcanzar buenos resultados, mientras que la entrega de mayor información promueve más autonomía y motivación intrínseca por la tarea. Siguiendo con la analogía, las consecuencias tanto simbólicas (publicación de los resultados) como concretas (aumento o disminución de la matrícula) asociadas a los resultados del SIMCE son insuficientes. Para que la información que entrega el SIMCE a los docentes incremente su valor formativo, requiere incorporar más elementos del segundo y tercer tipo. Es decir, indicar en qué medida el rendimiento mostrado por los alumnos es o no suficiente (aprobación/reproba-

ción) y describir dichos logros o entregar información sobre los aspectos menos logrados por sus alumnos.

Lo anterior debe comprenderse como complementario al aporte que pueda hacer una evaluación de desempeño docente, puesto que no bastará que ellos comprendan a cabalidad cuáles son los logros que se espera de sus alumnos. Es necesario una retroalimentación sobre las prácticas pedagógicas que facilitan u obstaculizan que los alumnos alcancen o no los estándares esperados.

Los estándares de desempeño describen el nivel de desempeño que se considera suficiente y que se espera que los alumnos demuestren. En otras palabras, los estándares de desempeño comunican qué tan bien se espera que se desempeñen los alumnos en relación con los estándares de contenido (Hambleton, 2001), o en nuestros términos, qué tan bien se espera que se desempeñen en relación con los objetivos y contenidos del marco curricular.

La definición de *estándares de desempeño* para los alumnos y de una medición de logros de aprendizaje alineada a ellos, permitirá responder a las expectativas y necesidades mencionadas: establecer y describir el nivel de logro que se ha considerado como satisfactorio, identificar el puntaje de la prueba que se corresponde con dicho nivel, además de poner en común y comunicar a los docentes, con mayor claridad, cuál es la expectativa nacional de desempeño para el curso evaluado.

Bibliografía

- Comisión SIMCE** (2003). *Evaluación de Aprendizajes para una educación de Calidad*. Informe de la Comisión para el desarrollo y uso del Sistema de Medición de la Calidad de la Educación convocada por el Ministro de Educación.
- Hambleton, R.** (2001). "Setting Performance Standards on Educational Assessment and criteria for evaluating the process". In: Cizek, G. J. (ed.), *Setting Performance Standards, concepts, methods and perspective*, LEA, New Jersey.

- Hamilton, L.; Stecher, B. & Klein, S.** (2002). *Making sense of test based accountability in education*, Washington D.C: Mc Rel.
- Hansche, L.** (1998). *Meeting the requirements of Title I, handbook for the development of performance standards*, prepared for the US Department of Education and the Council of Chief State School Officers, Frost Associates, Maryland.
- Linn, R. L. & Gronlund, N. E.** (2000). "Measurement and assessment in teaching" (8th ed.). Upper Saddle River, NJ: Prentice Hall.
- MINEDUC** (1998). *Currículo de la Educación Media: Objetivos Fundamentales y Contenidos Mínimos Obligatorios*.
- MINEDUC** (2002). *Estudio de la Implementación Curricular del sector de Lenguaje y Comunicación en el primer ciclo de enseñanza básica*. Documento de Trabajo, Componente de Seguimiento, Unidad de Currículo y Evaluación.
- Sadler, D. R.** (1989). "Formative Assessment and the Design of Instructional Systems", *Instructional Science*, 18, 119-144.