

EVALUACIÓN Y ESTÁNDARES: LOGROS Y DESAFÍOS PARA INCREMENTAR EL IMPACTO EN CALIDAD EDUCATIVA

*Assessment and standards: achievements and challenges
to increase the educational quality impact*

LORENA MECKES*

Resumen

Chile ha logrado significativos avances en materia de evaluación educativa través de más de dos décadas de desarrollo del área.

En la actualidad, enfrenta tres desafíos claves para contribuir al mejoramiento de la calidad y equidad de la educación. Primero, asegurar una articulación coherente entre las distintas evaluaciones con que cuenta el sistema. Segundo, lograr que la evaluación de aprendizajes de nivel nacional sea relevante para las prácticas pedagógicas y evaluativas de los establecimientos. El tercer desafío no es para la evaluación. Una política pública efectiva debe combinar la introducción de altas expectativas y exigencia en materia de logros de aprendizaje, a través de estándares y evaluaciones, con estrategias de apoyo y fortalecimiento de las capacidades de los actores del sistema para lograr los resultados esperados.

Palabras clave: medición educacional, evaluación, estándares de desempeño, SIMCE

Abstract

For more than two decades, Chile has been significantly progressing in educational assessment. Currently, it faces three challenges towards improving education quality and equity. First, achieving a coherent articulation among the system's different assessment modes. Second, making the national learning assesment relevant to the schools' teaching and assessing practices. The third challenge however, goes beyond evaluation. An effective public policy should combine high expectations and learning achievement demands –through standards and evaluations– with support strategies as well as the strengthening of skills among those involved in order to achieve expected results.

Key words: students' learning assessment, evaluation, performance standards, SIMCE

* Magíster en Evaluación IOE, Universidad de Londres, Coordinadora Nacional del SIMCE, lorena.meckes@mineduc.cl, lmeckes@gmail.com.

Introducción

Nuestro sistema escolar presenta hoy grandes desafíos: superar los niveles de aprendizaje que alcanzan niños y jóvenes en las distintas etapas de su trayecto escolar y disminuir la disparidad de los resultados según el nivel socioeconómico de sus familias. La evaluación de logros de aprendizaje ha desempeñado un papel muy importante en hacer evidente esta situación tanto para las autoridades educacionales como para el público en general. El papel que la política educacional chilena ha otorgado a la evaluación ha ido más allá de meramente retratar los resultados, ya que le ha asignado un rol en el mejoramiento de ellos a través de asociarles consecuencias e incentivos, al utilizarlos para definir cómo focalizar recursos y recientemente al clarificar las expectativas de logro a través de estándares de desempeño. Este artículo revisa el desarrollo y expansión de la evaluación educacional en los últimos años y los principales desafíos que ella enfrenta para lograr un mayor impacto en la calidad y equidad de los aprendizajes.

1. Evaluación educacional: desarrollo y panorama actual

Chile cuenta –desde inicios de los ochenta– con un sistema nacional de evaluación de logros de aprendizaje de los estudiantes en el sistema escolar (el SIMCE), lo cual convierte a nuestro país en pionero en la materia en Latinoamérica junto a Costa Rica¹. Actualmente se aplican pruebas censales a 4° básico (anualmente) y a 8° y 2° medio (año por medio)², alcanzando una cobertura de 95% de los estudiantes en 4° básico y de 93% en educación media (MINEDUC, 2007b). La aplicación ininterrumpida de pruebas, la entrega y publicación de sus resultados a nivel nacional y de cada escuela han contribuido a la legitimidad y credibilidad con la que hoy cuenta el SIMCE. Al tratarse de una evaluación censal, “toca” a todos los establecimientos educacionales y sus respectivas comunidades escolares, favoreciendo así la instalación de una cultura de evaluación en el país.

Además de la medición de logros relacionados con el currículum nacional, a partir de 1998 el SIMCE ha estado a cargo de los estudios internacionales comparativos en que Chile participa, proporcionando información acerca del rendimiento de los estudiantes chilenos en relación con los conocimientos, habilidades y competencias

¹ El PER (predecesor del SIMCE) comenzó el año 1982 consolidándose como SIMCE a partir de 1988. Su homólogo en Costa Rica se inició en 1986. El grueso de los países latinoamericanos iniciaron sus esfuerzos en evaluación nacional de resultados de aprendizaje durante los noventa (Ferrer, 2006). En la década del sesenta se había aplicado la Prueba Nacional.

² Esta frecuencia se inició en 2006, ya que hasta esa fecha se evaluaba un nivel o grado por año, alternadamente.

que se consideran relevantes a nivel internacional. Las mediciones internacionales en las que Chile participa o ha participado son: TIMSS (1999 y 2003), PISA (2000, 2006 y 2009), Mediciones de Competencias Ciudadanas: CIVED (1999 y 2000) e ICCES (2008), y en los dos estudios comparativos latinoamericanos dirigidos por UNESCO (1997 y 2006).

Un primer logro de esta tradición de evaluaciones nacionales y su expansión a estudios internacionales ha sido el creciente foco de la atención pública y de la política educacional en los resultados y en la equidad de ellos, que ha desplazado la atención prestada a los insumos y al acceso a la educación (Comisión SIMCE MINEDUC, 2003).

El esfuerzo por lograr un alineamiento de la evaluación al currículum nacional como estrategia para comunicar en forma consistente las expectativas sobre el aprendizaje que deben desarrollar los alumnos, también es un avance. Esto se ha traducido institucionalmente a partir de 1998 en la formación de una Unidad de Currículum y Evaluación en el Ministerio de Educación, responsable del desarrollo curricular (del currículum prescrito) y de la medición de logro de dicho currículum por parte de los estudiantes (currículum logrado).

Por último, en el plano de la difusión y uso de la información, el fomento de las prácticas de análisis de los resultados del SIMCE al interior de las escuelas, a través de materiales especialmente elaborados para ello y de una jornada nacional destinada a esta actividad, también representa un avance significativo de los últimos años.

Desde 2003 se han sumado a las evaluaciones de logro de los estudiantes evaluaciones de los demás niveles del sistema educacional: evaluación de desempeño de los docentes y directivos, evaluación del desempeño de las escuelas. Aun cuando la evaluación sistemática de políticas educacionales y de los sostenedores es aún prácticamente inexistente (Banco Mundial, 2007), se puede observar, de acuerdo a la siguiente tabla, que existe en Chile un creciente desarrollo en materia de evaluación educacional en el sistema escolar³.

³ Este artículo se concentra en las evaluaciones en el sistema escolar y no se consideran las evaluaciones de certificación ni de selección como la PSU, ni las de competencias de adultos, tales como el estudio internacional de alfabetización de adultos IALS.

Tabla 1
DIMENSIONES EVALUADAS Y PROGRAMAS DE EVALUACIÓN
EN EL SISTEMA ESCOLAR

Dimensión Evaluada	Programa/estudio	Descripción	Instituciones involucradas en Chile
Logros de aprendizaje de los estudiantes en cada escuela y en el país.	SIMCE Desarrolla la medición nacional y estudios internacionales. Medición externa estandarizada, en relación con currículum y estándares nacionales y en relación con el desempeño de otros países.	Medición nacional censal desde 1982 para 4º, 8º y 10º (segundo año medio)	Unidad de Currículum y Evaluación del Ministerio de Educación.
		Estudios internacionales (muestrales). Sistemáticamente desde 1998: TIMSS, PISA, ICCS, SERCE.	Unidad de Currículum y Evaluación del Ministerio de Educación.
Desempeño de Directivos y Docentes	“Docente más” Sistema de Evaluación del Desempeño Profesional Docente.	Evalúa docentes del sistema municipal. Combina evaluación externa estandarizada (portafolios y análisis de video de clase) con autoevaluación, reporte de superior jerárquico y evaluador par. Iniciado en 2003.	Municipios, Ministerio de Educación y asesoría técnica universitaria (Universidad Católica)
	“Más Directivos”	Programa de evaluación de directivos de establecimientos municipales sobre la base del marco de la buena dirección. Fundamentalmente autoevaluación y establecimiento de compromisos de gestión.	Sostenedores Municipales, establecimientos municipales, Ministerio de Educación.
	TEDS M 2008	Estudio internacional sobre formación docentes de Matemáticas conducido por la IEA.	Unidad de Currículum y Evaluación y Centro de Perfeccionamiento del Ministerio de Educación.
Desempeño escolar	SACGE , Sistema de Aseguramiento de la Calidad de la Gestión escolar.	Autoevaluación voluntaria de las escuelas validada por comisión externa (desde 2003).	Unidad de Gestión y Mejoramiento Educativo del Ministerio de Educación.
	SNED , Sistema Nacional de Evaluación de Desempeño de los establecimientos educacionales subvencionados.	Sistema de evaluación de escuelas para asignación de incentivos salariales colectivos para las escuelas que se basa en un conjunto de indicadores de: efectividad y superación en puntajes SIMCE (65%), iniciativa, condiciones de trabajo, igualdad de oportunidades, integración de profesores y apoderados (desde 1996).	Departamento de Estudios y Desarrollo Ministerio de Educación.

De las evaluaciones mencionadas, la más consolidada y de mayor cobertura es la evaluación de logros de aprendizaje de los estudiantes. El logro más relevante y reciente en el desarrollo de ella ha sido la definición de estándares de desempeño y la entrega de

resultados referidos a ellos. A continuación se profundiza en este avance, que marcará una diferencia en la evaluación de aprendizaje los próximos años.

2. Estándares de desempeño: la transformación reciente más significativa de la evaluación nacional de logros de aprendizaje

La incorporación de *estándares de desempeño* a partir de la entrega de resultados de la medición de 2006 de los cuartos básicos es probablemente la transformación más significativa del SIMCE en los últimos años. La introducción de estándares de desempeño permite complementar la interpretación *relativa* de los resultados, basada fundamentalmente en comparaciones con puntajes previos o con grupos de referencia⁴, con una interpretación de resultados que se contrasta con niveles *absolutos* de logro o competencia claramente descritos e ilustrados, pasando así de un énfasis en puntajes promedio a un énfasis en los aprendizajes que evidencian o no los alumnos y alumnas en la prueba.

Un estándar de desempeño describe el nivel de competencia que debe demostrar un estudiante en un área (por ejemplo, Matemática) y en un grado o curso específico para que su desempeño pueda ser clasificado en una determinada categoría (por ejemplo, “competente”, “avanzado”, etc.). “El estándar describe y define cuán adecuado se considera el desempeño de un estudiante para su nivel o curso” (Forster, 2002). En la medición el estándar se expresa en el puntaje de corte, que es el puntaje mínimo que se debe obtener para determinar que este se ha alcanzado.

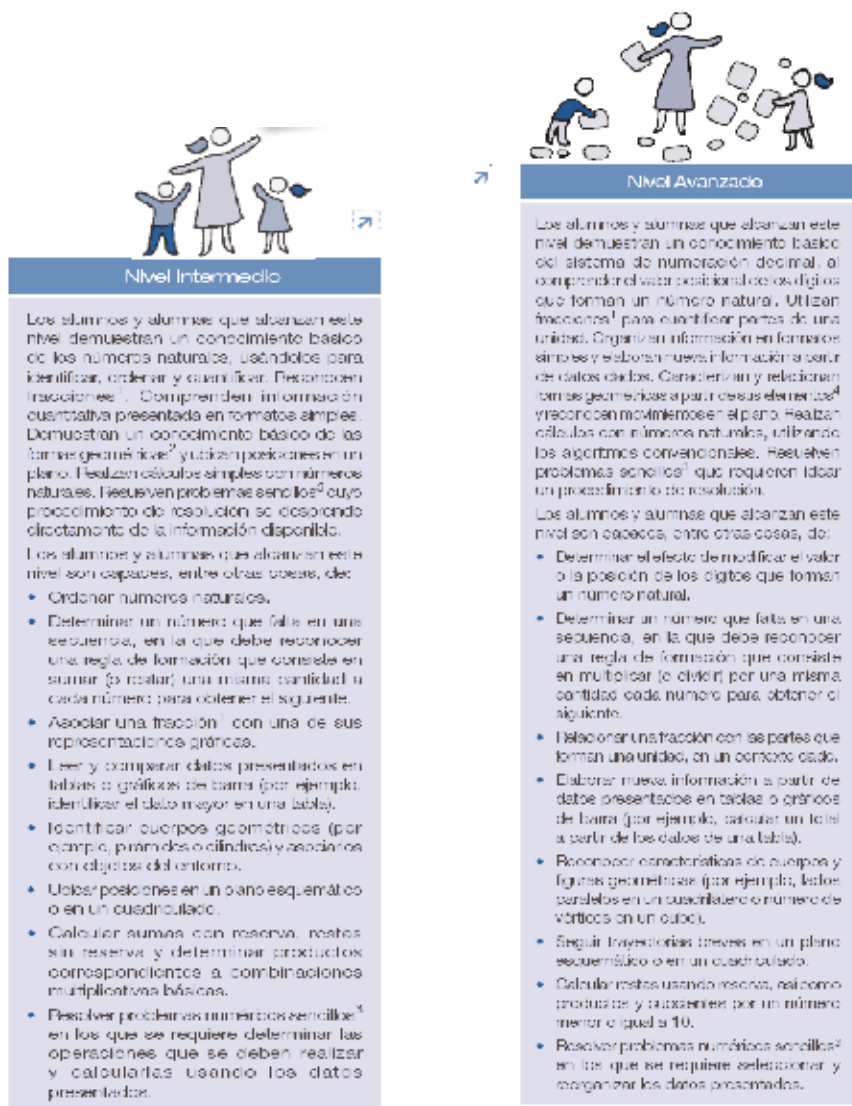
Este puntaje se define a través de una metodología de paneles de jueces (docentes de distintos tipos de escuelas, académicos del área curricular evaluada, curriculistas) Las más conocidas son Bookmark y Angoff y sus modificaciones. En Chile se seleccionó Bookmark y los puntajes de corte para 4° básico fueron definidos a través de un protocolo detallado que fue aprobado por el ETE5⁵.

La siguiente figura muestra los estándares de desempeño (denominado “niveles de logro” por el SIMCE) específicamente para Educación Matemática en 4° básico. Como puede apreciarse, se trata de descripciones detalladas de niveles progresivos de dominio en el área (que dan origen a tres categorías de logro). Para mayor claridad sobre la expectativa que representa cada uno de ellos se ejemplifican ilustrando y comentando el tipo de preguntas que típicamente son capaces de resolver los estudiantes que alcanzan dichos niveles.

⁴ Por ejemplo, el establecimiento educacional comparado con el promedio nacional, con el grupo de escuelas similares, con resultados anteriores.

⁵ Para mayores detalles ver sección Investigadores de www.simce.cl

Figura 1
EDUCACIÓN MATEMÁTICA. NIVELES DE LOGRO O ESTÁNDARES
DE DESEMPEÑO SIMCE PARA 4º BÁSICO



¹ Fracciones: solo manejan algunas fracciones (medios, tercios, cuartos, octavos y décimos).

² Formas geométricas consideran tanto figuras planas como cuerpos geométricos.

³ Problemas sencillos son aquellos de enunciado breve, pocos datos y contexto familiar.

Nivel Intermedio

7. En un consultorio, se entregan mensualmente 2 kilos de leche a cada niño menor de 5 años.
Una familia tiene 3 hijos menores de 5 años.
¿Cuántos kilos de leche recibe esta familia cada mes?
- A. 2 kilos.
B. 3 kilos.
C. 6 kilos.
D. 10 kilos.

Respuesta correcta: C. 6 kilos.

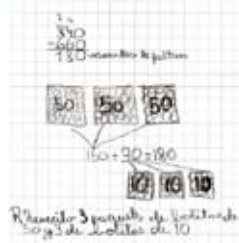
Esta pregunta es un ejemplo del tipo de tareas que realiza un estudiante de **Nivel Intermedio**, ya que requiere resolver un problema numérico sencillo donde es necesario determinar las repeticiones que se deben realizar y calcularlas usando los datos presentados. Es importante destacar que corresponde a una tarea propia de Nivel Intermedio ya que la determinación de qué hacer para resolver el problema (ya sea multiplicar 2×3 o sumar $2+2+2$) se depende de seguir los datos en el mismo orden en que se presentan, para luego traducirlo a un cálculo:

- 2 kilos por cada niño.
Una familia de 3 niños.
¿Cuántos kilos?

Esta pregunta permite ilustrar que los estudiantes que alcanzan el Nivel Intermedio son capaces de resolver problemas en los que el procedimiento de resolución se depende directamente de la información disponible.

Nivel Avanzado

8. Un artesano necesita 840 bolitas de madera para realizar un pedido de collares y pulseras.
El artesano tiene 660 bolitas.
Las bolitas se venden en paquetes de 50 y 10 bolitas.
¿Cuántos paquetes y de cuántas bolitas cada uno, debe comprar el artesano para realizar el pedido sin que le sobren bolitas?
En el siguiente espacio, anota tu respuesta y muestra cómo la obtuviste.



Esta pregunta es un ejemplo del tipo de tareas que realiza un estudiante de **Nivel Avanzado**, ya que requiere resolver un problema numérico sencillo en el que se debe seleccionar y reorganizar los datos presentados.

Esta pregunta permite ilustrar que los estudiantes que alcanzan el Nivel Avanzado son capaces de resolver problemas en los que se requiere usar un procedimiento de resolución.

En este contexto, la selección y organización de la información puede consistir en definir primero cuántas bolitas adicionales necesita el artesano y usar los datos que le permiten llegar a ese resultado, tal como se ve en el ejemplo de respuesta correcta presentada. Esto implica realizar una resta con reserva, lo cual también es característico del Nivel Avanzado. Luego, en una segunda etapa, el estudiante debe pensar en cómo reunir esa cantidad de bolitas usando el dato de las cantidades de bolitas por paquete. El ejemplo de respuesta muestra una forma de hacer eso, cubriendo el máximo de bolitas con paquetes de 50 y luego completar con los de 10.

Nivel Inicial: en esta categoría de desempeño los alumnos y alumnas aún no han consolidado los aprendizajes del Nivel Intermedio, ya que en ocasiones demuestran logros en algunos de los aprendizajes descritos en ese nivel, pero con una menor frecuencia y de manera poco consistente.

Versión completa de los Niveles de Logro en Ministerio de Educación: Niveles de Logro 4° Básico, Lectura y Educación Matemática, SIMCE, mayo 2007 (MINEDUC, 2007a) disponible en www.simce.cl

La definición de estándares de desempeño para el aprendizaje es un rasgo común de los sistemas educacionales exitosos (OCDE, 2004) y la tendencia internacional se ha inclinado a la definición cada vez más precisa de las expectativas de aprendizaje para todos los estudiantes. En Estados Unidos, por ejemplo, en 1993 sólo cinco estados habían definido estándares, mientras que en 2004 todos ellos lo habían hecho. Los estándares son cada vez más precisos y tienden a mostrar una clara progresión entre un grado o nivel y el siguiente (Cohen, 2004).

Tanto el informe de la OCDE (2004) sobre la política educacional chilena como la revisión del SIMCE realizada por la Comisión SIMCE en 2003 habían demandado la necesidad de incorporar estándares a la medición nacional. Adicionalmente, este desarrollo en curso fue respaldado por el Consejo Presidencial en 2006. Las expectativas

sobre los beneficios de incorporar estándares a la evaluación nacional han abarcado un amplio rango de visiones y énfasis. Se ha planteado por un lado la necesidad de otorgar significado a la evaluación y mayor claridad en las expectativas de logro para acercar la medición externa a los docentes y potenciar así el uso educativo de sus resultados. Por otro, se ha enfatizando su posible aporte a generar mayor presión y cuenta pública por resultados como mecanismos privilegiados para el mejoramiento (ver, por ejemplo, Eyzaguirre, 2001 y 2002). Está aún por verse la medida en que Chile logre un balance adecuado entre propósitos de rendición de cuentas y uso de la medición para aprovechamiento pedagógico.

Es claro que incorporar estándares favorece que el foco de la política educacional en resultados adquiera significado, ya que, tal como se puede verificar, los niveles de logro explicitan en forma precisa el tipo de desempeño que debe exhibir un estudiante para que este sea considerado “intermedio” o bien “avanzado” comunicando cuáles son expectativas mínimas y a cuáles se aspira. También es claro el tipo de procesos y conceptos involucrados en cada nivel y el progreso implicado en avanzar de un nivel al siguiente. De este modo, resultados como los que se presentan en la Tabla 2 son cifras que aluden directamente a proporciones de alumnos –en este caso– de cada grupo socioeconómico de 4° básico que, por ejemplo, logran resolver problemas que requieren idear una solución y seleccionar datos (Nivel Avanzado), cuántos sólo serían capaces de resolver problemas muy simples (Nivel Intermedio) y qué proporción aun no son capaces de hacerlo (Nivel Inicial). Así, las brechas de puntaje adquieren significado y se transforman en diferencias en el desarrollo de competencias para distintos grupos de la población.

Tabla 2
PORCENTAJE NACIONAL DE ESTUDIANTES EN CADA NIVEL DE LOGRO (2006)⁶

Nivel de Logro en Matemática	Nivel Socioeconómico					Total
	Bajo	Medio Bajo	Medio	Medio Alto	Alto	
Avanzado	11%	15%	24%	41%	64%	26%
Intermedio	29%	31%	37%	39%	29%	35%
Inicial	60%	54%	38%	20%	7%	39%

(MINEDUC, 2007 b).

El desarrollo de la medición nacional se caracterizará en los próximos años por la incorporación progresiva de las distintas áreas de aprendizaje y de todos los grados

⁶ Dado que los porcentajes están aproximados, los totales pueden no sumar 100%.

evaluados por el SIMCE a esta nueva modalidad de entrega de resultados que describe y define estándares de desempeño para el alumnado.

Queda por discutir si estos estándares de aprendizaje deben traducirse o no en mediciones con consecuencias para los alumnos (desde dar a conocer si cada uno de ellos alcanzó un nivel “intermedio” o “avanzado” hasta definir su promoción o incidir en sus calificaciones, por ejemplo) y de ser así a qué edad esto sería oportuno. Es muy importante hacer notar que la definición de un determinado estándar y su nivel de exigencia nunca es independiente del tipo de uso que se le dará ni tampoco de las consecuencias que tendrá para quienes son evaluados o para quienes son la población objetivo final. Estos elementos de juicio son puestos a disposición de los jueces que establecen los puntajes de corte. En el caso de la descripción y definición del nivel de exigencia de los estándares para el SIMCE de 4° básico, se tuvo en consideración que se trata de una medición que tiene consecuencias para las escuelas (publicación de resultados, asignación de incentivos o decisiones sobre autonomía en el uso de recursos) y para establecer metas de superación a distintos niveles. No se pensó en estándares que determinaran decisiones a nivel de los alumnos (por ejemplo repitencia). Al pensar en alumnos, resulta aun más apropiada la recomendación general que se aplica a las escuelas en sistemas basados en rendición de cuentas: si no se han ofrecido las oportunidades para alcanzar las exigencias, es completamente inadecuado establecer consecuencias para quienes no las logran.

Al estar los Niveles de Logro y el SIMCE referidos al marco curricular, la validación social de ellos se deriva de las instancias de validación y aprobación por las que atravesó dicho currículo. El desafío que queda por delante es lograr un apropiado nivel de compromiso con el logro de estos estándares de desempeño.

La combinación de estándares y evaluación de logros aporta dos de las piezas claves en el mejoramiento de la eficacia de un sistema educacional: propósitos claros y compartidos y monitoreo de la medida en que están siendo alcanzados. Sin embargo, los mecanismos a través de los cuales se espera que estándares y evaluación tengan un impacto en la calidad educativa son variados. En la siguiente sección se revisa el papel que se ha asignado hasta ahora a la evaluación de logros de aprendizaje en la política educacional chilena.

3. El papel otorgado a la evaluación nacional de logros de aprendizaje en las políticas educacionales

El papel que se ha otorgado en Chile a la evaluación nacional de logros de aprendizaje –a través del SIMCE– ha combinado distintas estrategias y supuestos o “teorías de la

acción”⁷ para el mejoramiento. Ellas han ido desde asociar consecuencias y premios a los resultados de la evaluación, hasta propiciar el uso de ellos por parte de los mismos docentes y directivos de las escuelas para reorientar sus propias decisiones. El protagonismo asignado a nivel local y al gobierno central es variable en cada una de estas estrategias, como también lo son los supuestos sobre cuál es la clave fundamental para lograr mayor calidad o equidad en los resultados.

Tabla 3
USOS DE LA EVALUACIÓN DE LOGROS EN LA
POLÍTICA EDUCACIONAL CHILENA

	Rendición de cuentas, incentivos y consecuencias	Apoyo ministerial	Decisiones de las escuelas
Papel de la Evaluación Nacional de Logros de Aprendizaje	–Informar a la demanda en un sistema de financiamiento ligado a matrícula y asistencia. –Proveer información para asignar incentivos.	Identificar establecimientos en los que se focalizarán los programas de apoyo y monitorear los resultados de ellos.	Retroalimentar decisiones pedagógicas y de gestión de las escuelas y sostenedores.
Supuestos de mejoramiento de la calidad o equidad	Los establecimientos orientarán mejor sus esfuerzos para obtener los incentivos o para competir por la matrícula y la subvención.	Se compensan inequidades entre escuelas a través de intervenciones de apoyo.	Prácticas pedagógicas y de gestión escolar orientadas a alcanzar estándares claros e informados por la evidencia de los resultados obtenidos.
Políticas específicas implementadas en Chile.	Publicación de resultados del SIMCE por escuela. Sistema Nacional de Evaluación del Desempeño (SNED) ⁸ .	Políticas focalizadas: Ejemplos: Programa de las 900 Escuelas (desde 1990), Liceo para Todos (2000 a 2006), Escuelas Prioritarias (desde 2002) ⁹ .	–Informes detallados de resultados a cada escuela. –Jornadas nacionales de análisis de resultados SIMCE. –Análisis de resultados SIMCE integrado al SACGE (Sistema de aseguramiento de la calidad de la gestión escolar).

⁷ Esta expresión es tomada de Forte y Hebbler (CCSSO, 2004) que en el contexto de sistemas de responsabilización o rendición de cuentas denominan “teoría de la acción” a la lógica a través de la cual se supone que el sistema funcionará y tendrá los efectos deseados.

⁸ Los resultados en el SIMCE tienen el 65% de ponderación o peso en la asignación de este incentivo salarial a los docentes.

⁹ En el caso de las Escuelas Prioritarias (inicialmente llamadas ‘Escuelas Críticas’), la intervención fue encargada a otros organismos (ONGs e instituciones académicas), pero la contraparte, definición de escuelas objetivos y conducción, ha estado en manos del Ministerio de Educación.

Tal como se esquematiza en la Tabla 3, los resultados del SIMCE han tenido diversos roles que se han ido sobreponiendo en el tiempo en la política educacional chilena. En primer lugar, junto con el diseño del SIMCE en los ochenta se le asignó el rol de proveer información a la demanda en un sistema de financiamiento ligado a asistencia de los alumnos matriculados en cada escuela. El incentivo o consecuencia positiva asociada a buenos resultados es en este caso la atracción de matrícula y recursos. Durante los noventa se incorporó el Sistema Nacional de Evaluación del Desempeño (SNED) que asigna incentivos monetarios colectivos a las escuelas seleccionándolas en gran medida según los resultados del SIMCE obtenidos. El supuesto aquí es que las escuelas orientarán sus esfuerzos de mejor manera para obtener el premio.

La medición nacional también ha tenido un papel relevante en estrategias que otorgan mayor protagonismo a la acción del Estado. A partir de los noventa se la utilizó para decidir sobre focalización de recursos y apoyo a escuelas en el tramo de más bajo rendimiento y –aunque en mucho menor medida– para evaluar estas políticas.

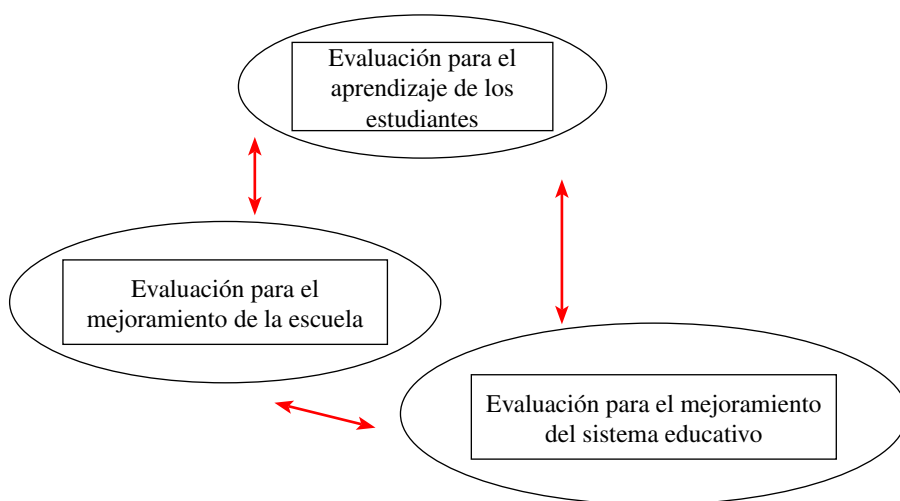
Por último, se ha intencionado que la medición nacional tenga el papel de orientar decisiones pedagógicas y de gestión a nivel local. Con este fin se entregan anualmente informes detallados a cada escuela, además de orientaciones para el aprovechamiento de estos a través de un día destinado al análisis colectivo de los resultados al interior de los establecimientos educacionales.

De los tres tipos de roles asignados a la medición nacional en el mejoramiento de la calidad, en Chile claramente han predominado los dos primeros mencionados, es decir, su asociación a consecuencias e incentivos o bien su uso para la toma de decisiones de los responsables de política¹⁰. A pesar de los avances durante los últimos años, la información entregada por el SIMCE aún es subutilizada por parte de las escuelas y docentes, que son el eslabón clave y determinante en la cadena de calidad. Este tipo de uso apunta además a la construcción de capacidades en los docentes y directivos de las escuelas.

Cualquiera sea el rol asignado al gobierno central o a los responsables en cada escuela y cualquiera sea el supuesto sobre qué moviliza el mejoramiento al interior de ellas (respuesta a los incentivos o decisiones mejor informadas), es necesario que los involucrados tengan claridad de las metas a lograr o estándares, estén identificados y comprometidos con ellas, reciban y analicen la información sobre el logro de las mismas y tengan capacidades instaladas para monitorearlas.

¹⁰ Ver Ministerio de Educación (2003), *Evaluación de Aprendizajes para una Educación de Calidad*.

Lo anterior requiere que la evaluación de los aprendizajes desde el aula hasta la medición nacional “haga sistema” y se potencien recíprocamente (ver el siguiente esquema). Es decir, que las evaluaciones desarrolladas por sostenedores, secretarías ministeriales regionales, directivos de escuela y docentes en el aula sean coherentes entre sí, y transmitan niveles de expectativa semejantes y una misma visión de lo que es relevante aprender. De lo contrario, estudiantes y escuelas recibirán mensajes contradictorios y desorientación en lugar de foco y metas claras. Lograr coherencia es uno de los desafíos más importantes que enfrentará la evaluación de logros de aprendizaje en los próximos años.



OCDE (2005), *Formative Assessment: Improving Learning in Secondary Classrooms*, Paris.

4. Desafíos que enfrentará la evaluación educacional en el futuro cercano

Primer desafío: lograr un estrecho vínculo entre evaluación de aprendizajes a nivel nacional y de aula

Además de contar con estándares y evaluaciones referidas a ellos, los sistemas educacionales exitosos se caracterizan por el alineamiento entre los siguientes elementos: currículo, formación docente, evaluación o monitoreo nacional de resultados de aprendizaje y evaluación de aula para el aprendizaje (NRC, 2003).

La definición de estándares, medición externa de resultados, consecuencias asociadas a ellos, difícilmente tendrán un impacto en los resultados si quienes están a cargo de realizar la tarea –directivos y docentes– no tienen una acabada comprensión

del aprendizaje que se desea desarrollar ni observan permanentemente si sus estudiantes lo están alcanzando.

Existe hoy abundante literatura y experiencias que demuestran cómo la evaluación de aula apropiada e integrada al proceso de enseñanza y aprendizaje da cuenta de progresos notables en los resultados alcanzados por los alumnos. Esta línea de trabajo se considera tan promisorio que la OCDE (2005) le dedicó un número especial de su serie “Lo que funciona en Innovación Educativa”. A este hallazgo se suman numerosas experiencias que pretenden precisamente tender un puente entre la evaluación desarrollada al interior de las unidades educativas y los sistemas estatales o nacionales de monitoreo como es el SIMCE. Uno de los ejemplos de esfuerzos exitosos para incentivar el uso de información de mediciones de gran escala con la información entregada por la evaluación de aula es el “Club de Datos” implementado en el estado de Australia Oriental. Este fue concebido como desarrollo profesional de directivos y docentes favoreciendo no sólo una actitud positiva hacia el programa de medición (WALNA), sino el aprovechamiento de la evaluación para potenciar el aprendizaje de los alumnos (Cook, 2005; Louden & Wildy, 2001).

En Chile encontramos algunas barreras que es necesario superar para que se materialice un uso intensivo e integrado de la evaluación interna y externa, como es el SIMCE, en beneficio del aprendizaje de los alumnos.

En primer lugar, la evaluación de aula es el área más deficitaria de la práctica pedagógica, de acuerdo con las evaluaciones de desempeño docente de los últimos años. Tanto la calidad de los instrumentos de evaluación utilizados por los docentes como el uso de la información entregada por ellos han resultado ser las dimensiones peor evaluadas (Manzi & Flotts, 2007). Típicamente, tanto en Chile como en otros países los programas de formación de docentes incluyen poco de evaluación y menos de interpretación de datos de mediciones externas de gran escala (MINEDUC, 2003).

En segundo lugar, se han hallado importantes discrepancias entre los aprendizajes valorados por las mediciones de gran escala tanto nacionales como internacionales y los aprendizajes que son evaluados por los docentes a través de sus pruebas o evaluaciones de sala de clases. Pellegrino *et al.* (2001) destacan la relevancia de que el aprendizaje que es valorado y evaluado por la medición externa de una determinada área curricular sea un subconjunto (aquel que es abordable a través de una evaluación masiva de papel y lápiz) de los aprendizajes que son valorados y evaluados por los establecimientos en sus propias evaluaciones y retroalimentación a los estudiantes.

El siguiente cuadro contrasta el “peso” que tienen distintas habilidades en la evaluación de Matemáticas en 8° básico en aula en una medición internacional (TIMSS) y en una prueba SIMCE. El dato sobre la relevancia dada a cada tipo de habilidad en aula se obtuvo analizando y clasificando una muestra de 1.048 ejercicios o preguntas de pruebas de sala de clases.

Habilidades	Evaluación de aula	SIMCE 8° 2004	TIMSS 2003
Manejar Conocimientos y Procedimientos.	67%	10%	15%
Usar Conceptos	7%	17%	20%
Resolver problemas de rutina	19%	55%	40%
Razonar	6%	17%	25%

Fuente: Documento de Trabajo, Equipo de Seguimiento a la Implementación Curricular, Unidad de Currículum y Evaluación, Ministerio de Educación.

La discrepancia más llamativa se da entre la gran relevancia otorgada al manejo de procedimientos (algoritmos) en la evaluación de aula, mientras que en mediciones a nivel de sistema predominan la resolución de problemas y el razonamiento. Esto da cuenta de una gran distancia entre las habilidades que son valoradas por la evaluación nacional e internacional y aquellas a las cuales los docentes chilenos están dando mayor importancia en sus propias evaluaciones y, en consecuencia, el mensaje que reciben sus estudiantes sobre lo que es relevante aprender y ser capaz de hacer¹¹.

El Ministerio de Educación ha desarrollado dos iniciativas los últimos años que resultan promisorias para superar las insuficiencias en las prácticas de evaluación de aula y tender un puente entre la evaluación desarrollada internamente por las escuelas y la medición nacional. Se trata específicamente del programa de evaluación de aula de la Unidad de Currículum y Evaluación (ver MINEDUC, 2006) y del desarrollo de un enfoque común para el monitoreo del aprendizaje al interior de las escuelas y la medición nacional a través de los mapas de progreso y niveles de logro del SIMCE, respectivamente.

También se han encontrado barreras en la cultura escolar para el uso de datos en el mejoramiento de las prácticas (Ingram & col., 2004). La cultura docente es una de las principales determinantes de la información que estos consideran válida y significativa para juzgar su propia efectividad. Frecuentemente los docentes desestiman los resultados de los sistemas de evaluación de gran escala por considerar que ellos sólo evalúan un ámbito restringido de aprendizajes, generalmente de tipo instrumental, que estos consideran. No sólo la cultura escolar sino también los padres valoran procesos y resultados que no son visibles en los datos entregados por las mediciones externas. Esto determina que en lugar de considerar las evaluaciones externas e internas de la escuela como fuentes complementarias de información, existe el riesgo de que se experimenten como reñidas y se tienda a utilizar la que es más compatible con el propio sistema de

¹¹ Esta evidencia es contraria a la visión convencional respecto de las mediciones estandarizadas según la cual estas tenderían a evaluar aprendizajes poco relevantes y netamente memorísticos.

valores o que resulte menos disonante con la imagen que el cuerpo docente tiene de su escuela. El desafío aquí es tanto para la medición nacional e internacional que necesita ampliar el ámbito de lo que evalúa incluyendo competencias claves tales como la capacidad para interactuar en contextos socialmente heterogéneos o actuar autónomamente (Schleicher, 2003) como para las prácticas de uso de información al interior de las escuelas, que requieren romper mitos sobre las mediciones de gran escala, pasando de un uso “selectivo” de la información disponible a uno en que se complementan las distintas visiones que se obtienen de los resultados alcanzados.

Finalmente, es necesario considerar que en un sistema de educación tan segmentado desde el punto de vista social como el chileno (OCDE, 2004b) existe un conflicto entre la publicación de los resultados de aprendizaje de cada escuela para usos de rendición de cuentas e incentivos y la utilización de estos resultados por parte de los docentes, especialmente para aquellos que enseñan en contextos de mayor deprivación sociocultural. En estos casos, es más probable el rechazo a la información entregada por un sistema de evaluación nacional si este se percibe como injusto. El desafío aquí es doble; por una parte continuar con la publicación de resultados contextualizada y referida a escuelas de similares características socioeconómicas, incorporando a futuro indicadores de valor agregado y, por otra parte, enfatizar la necesidad de tener altas expectativas para todos.

El desafío de futuro para nuestro sistema educacional y para otros sistemas en el mundo es llegar a ser sistemas que sean “ricos en conocimiento” no solo a nivel de responsables de política, sino también a nivel de los profesionales de la educación (Schleicher, 2006), de modo que tanto las decisiones de política como el juicio profesional de los docentes y directivos sean informados y basados en evidencia.

Segundo desafío: lograr coherencia entre evaluaciones de desempeño de los docentes y escuelas con la evaluación nacional de logros de aprendizaje

En un panorama cada vez más poblado de evaluaciones de los distintos niveles del sistema, existe gran riesgo de disonancia entre ellas. Algunos ejemplos para ilustrar este riesgo y sus posibles consecuencias. En los últimos años han proliferado las iniciativas locales en evaluación de resultados de aprendizaje que se plantean como complementarias al SIMCE. Existen mediciones de logro educativo aplicadas comunalmemente, en algunos casos estableciendo incentivos y consecuencias para docentes y escuelas asociados a sus resultados. Del mismo modo, se desarrollan iniciativas a nivel de Secretarías Regionales y de Departamentos Provinciales del Ministerio que son usadas fundamentalmente para establecer metas de gestión. Estas iniciativas se desarrollan sin estar coordinadas con el SIMCE (CEPP, 2007). Además de las posibles carencias en términos de rigurosidad técnica que presentan estos sistemas de evaluación paralelos, no existe coherencia necesaria

entre el enfoque de uno y otro ni tampoco en sus niveles de exigencia. En consecuencia, los esfuerzos y compromisos de resultados de las escuelas no están necesariamente alineados con los estándares y metas definidas nacionalmente, viéndose las escuelas enfrentadas a incentivos y orientaciones potencialmente contradictorios.

Por otra parte, la evaluación docente (*Docentemás*) establece cuatro categorías de desempeño. Las dos categorías más bajas (“insatisfactorio” y “básico”) concentran un 34% de los evaluados. En contraste, el porcentaje de estudiantes municipales que en el SIMCE de 4º básico muestra un nivel inicial (el más bajo entre las categorías SIMCE) es cercano al 50%. Si bien sería simplista suponer que los resultados de aprendizaje de los alumnos se derivan directamente de la calidad del desempeño de sus docentes, existe una cierta disonancia entre un resultado y el otro. En otras palabras, las “exigencias” de una y otra evaluación no parecen completamente coherentes.

Se ha encontrado que frecuentemente los docentes tienden a disociar su propia práctica de los resultados que obtienen sus estudiantes, especialmente si estos son insuficientes¹², lo cual obstaculiza que esos resultados sean considerados como un “mensaje” que debiera llevarlos a cuestionarse sobre su propio desempeño. Sería interesante indagar sobre las atribuciones causales o explicaciones que se dan los profesores para explicar el rendimiento de sus alumnos, y sobre la potencial contribución de la evaluación de desempeño a reforzar esta posible disociación subjetiva entre la propia práctica y los resultados de sus estudiantes en mediciones como el SIMCE.

Tercer desafío: lograr un balance entre usos para rendición de cuentas y el uso por parte de las escuelas para decisiones pedagógicas y de gestión

Existe una tensión natural entre usar las evaluaciones educacionales para rendición de cuentas e incentivos y para retroalimentar la labor docente (Hamilton *et al.*, 2002). Esta tensión se manifiesta en el tipo de motivación que cada uno de estos usos estimula; más extrínseca en el caso de usos para rendición de cuentas y asociación de consecuencias e incentivos y más intrínseca en el caso de usos para decisiones pedagógicas. El énfasis que se adopte también tiene consecuencias para el diseño mismo de la medición. Frente a recursos y tiempos limitados, mientras mayor sea el énfasis en usos de rendición de cuentas y asociación de consecuencias, mayor deberá ser la prioridad de alcanzar elevados niveles de objetividad (estandarización, precisión y confiabilidad).

El reciente anuncio de una superintendencia de educación a cargo de realizar o de supervisar las evaluaciones de prácticamente todos los niveles del sistema desde los estudiantes hasta el nivel nacional, pasando por sostenedores y escuelas, ofrece

¹² La atribución causal más frecuente del fracaso escolar entre los docentes es el contexto sociocultural de los estudiantes y también las habilidades intelectuales de estos (CIDE, 2003).

la oportunidad para articularlas y lograr que ellas se potencien recíprocamente. Este proyecto institucional permite además independizar la entidad responsable de las políticas e iniciativas de intervención (Ministerio de Educación) de la que está a cargo de la evaluación de dichas políticas, dando mayores garantías de objetividad y despejando posibles conflictos de interés.

Sin embargo, este proyecto de ley apunta a un sistema de aseguramiento de calidad en que la evaluación nacional se inclina más claramente a usos de rendición de cuentas y consecuencias, separando esta función de la función de apoyo.

Al quedar la evaluación educacional albergada en una entidad eminentemente fiscalizadora, con atribuciones directas para recomendar el cierre de escuelas, por ejemplo, lo más probable es que ella adquiera un significado aun más amenazante que hasta ahora para las escuelas y sus docentes y, por lo tanto, una menor disposición a concebir la información entregada como un apoyo a su labor. Es probable que esto gatille una “judicialización” de la medición nacional de logros de aprendizaje, y que se haga necesario priorizar el desarrollo de elevados estándares de precisión y confiabilidad por sobre el desarrollo de estrategias de uso pedagógico de la misma. Adicionalmente, la disociación institucional entre currículum y evaluación nacional de aprendizajes requerirá un esfuerzo mayor que el actual para lograr alineamiento¹³.

Cuarto desafío: comunicar claramente los límites de la precisión y exactitud en un escenario de creciente uso para rendición de cuentas y consecuencias

El creciente uso de los resultados de la medición nacional de aprendizajes para incentivos y cuenta pública del quehacer de escuelas y sostenedores implicará cada vez más exigencias de precisión y estandarización. Esto requiere que la comunicación de resultados de la medición sea cada vez más explícita sobre la adecuada interpretación de ellos y sobre los límites que tiene la precisión. Tal como Linn (2001) puntualizó, uno de los problemas de los sistemas de rendición de cuentas basados en mediciones es la excesiva confianza en la precisión y exactitud de los índices numéricos. Esto implica que es necesario comunicar muy eficazmente que existe un margen de error que se debe considerar. Al respecto, el SIMCE ha venido publicando indicadores de significancia de las diferencias al hacer comparaciones de puntajes entre años o entre escuelas. Sin embargo, estas no siempre son tomadas en cuenta para las decisiones de asignación de incentivos o consecuencias para escuelas o comunas. En el caso de la evaluación referida a estándares, será recomendable indicar la probabilidad que tendría

¹³ La incorporación de modificaciones sustanciales en una medición toma al menos dos años y medio ya que se requiere un proceso de elaboración, experimentación de los instrumentos el año previo a su aplicación definitiva.

un docente o estudiante de quedar clasificado en la misma categoría de desempeño si fuera evaluado más de una vez. Del mismo modo, al usar como indicador la distribución de estudiantes al interior de la escuela en las distintas categorías de desempeño (por ejemplo, el porcentaje en niveles “avanzado”, “intermedio” e “inicial”) también será apropiado comunicar la probabilidad de que esa distribución se mantenga estable si se aplicaran distintas formas de evaluación o una prueba más extensa que ofrezca más oportunidades para que el docente o alumno muestre en qué nivel de desempeño se encuentra. En otras palabras, el margen de “error de clasificación” de la medición. De tomarse en cuenta este margen de error, probablemente se tendrá que en la práctica un número importante de individuos o de escuelas deberían considerarse como teniendo logros equivalentes a pesar de presentar resultados numéricamente distintos.

Sin duda contar con indicadores precisos y sin margen de incertidumbre hace más sencilla y automática la definición de consecuencias e incentivos. Sin embargo, desde el punto de vista de la validez de esas decisiones, es relevante que no estén basadas en una confianza desmedida en una precisión que no es tal.

Quinto desafío: lograr un adecuado balance entre presión y apoyo

El principal desafío no es para la evaluación educacional. La constelación de evaluaciones que exhibe hoy el panorama educacional chileno y el mencionado proyecto de ley sobre la creación de una superintendencia de educación que albergue a la mayoría de ellas advierten sobre un riesgo de desbalance en el énfasis de los esfuerzos de nuestra política educacional. La presencia de sistemas de evaluación e incluso el uso de la información que ellos proveen es condición necesaria pero no suficiente para el mejoramiento de la calidad de la educación.

Los sistemas educacionales varían en la medida que ponen énfasis en exigencia y claridad de metas, por un lado, y estrategias de apoyo, por otro. La combinación de elevadas exigencias y elevado apoyo, siempre que este sea efectivo, deriva en mejoramiento y alto desempeño. En el Reino Unido, por ejemplo, políticas que combinaron estándares y presión ejercida a través de evaluaciones nacionales por un lado con estrategias efectivas de apoyo a los docentes, por otro, resultaron en un mejoramiento sustantivo en las áreas de lectura y matemáticas durante los últimos años. Las evaluaciones nacionales contribuyeron a una mejor comprensión de los objetivos del currículo nacional por parte de los profesores, lo cual fue complementado con la prescripción precisa de las estrategias pedagógicas para mejorar los aprendizajes y con programas intensivos de formación en servicio para que los docentes pudieran implementarlas (Whetton, 2004).

El potencial impacto de la evaluación en un sistema educacional dependerá no sólo de la coherencia y alineamiento entre las evaluaciones de diversa índole y del uso de la información que entregan, sino del apropiado balance de ellas con la construcción

de capacidades y la presencia de oportunidades de calidad en el ámbito de la formación para los docentes.

Las estrategias de formación continua y mejoramiento pedagógico pueden ser más o menos prescriptivas y más o menos centralizadas. Lo relevante es que la política educacional no descansa únicamente en estrategias que acentúan la dimensión de claridad de metas (estándares) y presión (cuenta pública por resultados), sino que combine exigencia con apoyo y oportunidades de formación docente de calidad.

Tampoco el esfuerzo por implementar mediciones y así acumular información crecientemente más rica y sofisticada provocará un mejoramiento por sí solo. Un ejemplo para ilustrar esto. Desde la entrega de resultados del SIMCE de 2004 se proporciona a cada escuela la variación de puntajes que se observa al interior de ella (ver informe de resultados para docentes y directivos de SIMCE 2004). Más recientemente la entrega de resultados del SIMCE en relación con estándares de desempeño informa sobre la proporción de alumnos que al final de un ciclo ha alcanzado diversos niveles de logro en cada establecimiento. Esta información hace evidentes y accesibles para todos los docentes las enormes diferencias de rendimiento entre estudiantes de un mismo nivel o curso. Para tener un orden de magnitud, Ramírez (2007) encontró que ya la prueba de 1999 de Matemática mostraba que el 70% de la variabilidad en los puntajes SIMCE obedece a diferencias al interior de las mismas escuelas: “en Chile un curso típico cubre tres desviaciones estándar de la distribución nacional de puntajes” (p. 14). Riquísima información disponible, presión pública para mejorar la situación y gran desafío para la gestión al interior de las escuelas y para la práctica y formación de los docentes. ¿Cómo enseñar a cursos tan heterogéneos? ¿Cómo un mismo objetivo o contenido puede ser abordado para niños que presentan tan diverso nivel de avance? Y, por lo tanto, ¿cómo lograr que en la próxima evaluación esta distancia se acorte y una proporción menor de alumnos esté en el nivel inicial? Estas interrogantes requieren respuestas muy específicas para la enseñanza de cada disciplina y respuestas diferentes para distintos ciclos de enseñanza. La didáctica de las disciplinas y las prácticas profesionales requerirían capítulos especiales para abordar este desafío con el cual se encontrarán los egresados de las escuelas de pedagogía.

En síntesis, los mayores desafíos para el futuro inmediato son lograr coherencia entre las diversas evaluaciones (de docentes, escuelas y alumnos) y equilibrar su protagonismo con estrategias efectivas de apoyo y construcción de capacidades.

Bibliografía

- Banco Mundial** (2007). El Diseño Institucional de un Sistema Efectivo de Aseguramiento de la Calidad de la Educación en Chile.
- CCSSO** (2004). *A Framework for Examining Validity in State Accountability Systems*, Council of Chief State School Officers, Washington D.C.
- Centro de Estudios en Políticas Públicas** (2007). *Informe del Sistema Educativo en Chile*, Programa Evaluación de Sistemas Educativos, CEPP y Fundación Konrad Adenauer.
- CIDE** (2003). Encuesta a Actores del Sistema Educativo, Santiago, CIDE.
- Cohen, M.** (2004). *Accountability and State Based Reform: lessons from the U.S.* Paper presented at the ECA Education Conference “Linking Policies and Actions to Outcomes”. Bucarest, Rumania.
- Cook, J.** (2005). *Reporting to Systems and Schools*, 2005 Currículum Corporation Conference, Australia. Disponible en http://cmslive.curriculum.edu.au/verve/_resources/Cook_edited.pdf
- Eyzaguirre, B. y Le Foulon** (2001). Estándares Altos para la Educación. *Puntos de Referencia*, 247, pp. 1-10.
- Eyzaguirre, B.** (2002). Los alumnos bajo la lupa: los exámenes con consecuencias individuales. Documento de Trabajo N° 331. Centro de Estudios Públicos.
- Ferrer, G.** (2006). *Sistemas de Evaluación de Aprendizajes en América Latina, Balance y Desafíos*, PREAL.
- Forster, M.** (2002). *Performance Standards and the Measurement of Student Achievement, Options and Challenges*, working paper developed for the Ministry of Education of Chile.
- Hamilton, L.; Stetcher, B. & Klein, S.** (2002). *Making Sense of Test based Accountability in Education*. Washington D.C., Rand Publications.
- Ingram, D.; Seashore, K. & Schroeder, R.** (2004). Accountability Policies and Teacher Decision Making: Barriers to the Use of Data to Improve Practice, in *Teachers College Record*, Vol. 106, N° 6, pp. 1258-1287.
- Linn, R.L.** (2001). En Lewis, A. *Educational Accountability in the 21st Century*, CSE Technical Report, CRESST, CSE, Los Angeles, California.
- Louden, W. & Wildy H.** (2001). *Developing Schools Capacity to Make Performance Judgements*, Final Report, Institute for the Service Professions, Edith Cowan University, Western Australia.
- Manzi, J. & Flotts, P.** (2007). Sistema de Evaluación del Desempeño Profesional Docente, Documento de Trabajo, Centro de Medición MIDE UC.
- Ministerio de Educación** (2007a). Niveles de logro 4° Básico, Lectura y Educación Matemática, SIMCE. Disponible en www.smce.cl.

- Ministerio de Educación** (2007b). Resultados Nacionales SIMCE 2006, 4° y 2° medio, SIMCE, Unidad de Currículum y Evaluación.
- Ministerio de Educación** (2006). *Evaluación para el aprendizaje: enfoque y materiales prácticos para lograr que sus estudiantes aprendan más y mejor*, Unidad de Currículum y Evaluación.
- Ministerio de Educación** (2005). Calidad en todas las escuelas y liceos: sistema de aseguramiento de la calidad de la gestión escolar, Serie Bicentenario.
- Ministerio de Educación** (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.
- National Research Council, NRC** (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment*. Washington, D.C., The National Academy Press.
- OCDE** (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*, Paris.
- OCDE** (2004a). What Makes School Systems Perform: seeing school systems through the prism of PISA.
- OCDE** (2004b). *Chile: Reviews of National Policies for Education*, Centre for Co-operation with non members, París.
- Pellegrino, J.W., Chudowsky, N. Glasser R.** (2001). *Knowing What Students Know: the Science and Design of Educational Assessment*, Board on Testing and Assessment, Center for Education, National Research Council, National Academy Press, Washington DC.
- Ramírez, M.J.** (2007). Diferencias Dentro de la Sala de Clases: distribución del rendimiento en Matemáticas, Estudios Públicos, N° 106.
- Schleicher, A.** (2006). The Economics of Knowledge: Why education is key for Europe's Success. Brussels: The Lisbon Council Policy Brief.
- Schleicher, A.** (2003). Developing a Long Term Strategy for International Assessments, p. 161-179 en Simone D. & Hersh L. (ed.), *Key Competencies for a Successful Life and Well Functioning Society*, Hogrefe & Huber, Germany.
- Whetton, C.** (2004). *Reflections on Fifteen Years of National Assessment: Lessons, Success and Mistakes*. Paper presented at the 30th International Association for Educational Assessment Conference, National Foundation for Educational Research.

FECHA DE RECEPCIÓN: 1 de junio de 2007

FECHA DE ACEPTACIÓN: 28 de junio de 2007