



2008 - Volumen 1, Número 1



<http://www.rinace.net/riee/numeros/vol1-num1.html>

Revista Iberoamericana de Evaluación Educativa

La Revista Iberoamericana de Evaluación Educativa es una publicación oficial de la Red Iberoamericana de Investigación sobre Cambio y Eficacia Escolar.

Dirección:

F. Javier Murillo
Marcela Román

Coordinación Editorial:

Verónica González

Comité Directivo:

Marcela Gajardo (PREAL - Chile)
Sergio Martinic V. (PUC - Chile)
Carlos Pardo (ICFES - Colombia)
Margarita Poggi (IIPE/UNESCO - Argentina)
Francisco Soares (UMG - Brasil)
Héctor Valdés (ICCP - Cuba)

Comité Científico:

Juan Manuel Álvarez Méndez (UCM - España)
Patricia Arregui (GRADE - Perú)
Daniel Bogoya (UPN - Colombia)
Nigel Brooke (UMG - Brasil)
Leonor Cariola (MINEDUC - Chile)
María do Carmo Clímaco (Portugal)
Cristian Cox (PUC - Chile)
Santiago Cueto (Grade - Perú)
Tabaré Fernández (ANEP - Uruguay)
Juan Enrique Froemel (Min. Ed. – Qatar)
Rubén Klein (Fund. Cesgranrio – Brasil)
Luis Lizasoain (UPV - España)
Jorge Manzi (PUC - Chile)
Felipe Martínez Rizo (INEE - México)
Joan Mateo (UB - España)
Liliana Miranda (MINEDU - Perú)
Carlos Muñoz-Izquierdo (UIA – México)
Margarita Peña (ICFES - Colombia)
Dagmar Raczynski (Asesorías para el desarrollo - Chile)
Héctor Rizo (UAO - Colombia)
Guadalupe Ruiz (INEE - México)
Mario Rueda (UNAM - México)
Ernesto Schiefelbein (CIDE – Chile)
Alejandra Schulmeyer (INEP - Brasil)
Javier Tejedor (USAL – España)
Flavia Terigi (UBA/UNGS - Argentina)
Alexandre Ventura (IGE - Portugal)



Editorial

- La Evaluación Educativa como Derecho Humano** 1
F. Javier Murillo y Marcela Román

Artículos/Artigos

- Resultados de Aprendizaje en América Latina a partir de las Evaluaciones Nacionales** 6
F. Javier Murillo y Marcela Román
- Evaluación Educativa. Sobre Sentidos y Prácticas** 36
Margarita Poggi
- Evolución de los Procesos de Evaluación del Sistema Educativo 1950-2008** 45
Ernesto Schiefelbein y Paulina Schiefelbein
- Las Evaluaciones Educativas que América Latina Necesita** 51
Pedro Ravela, Patricia Arregui, Gilbert Valverde, Richard Wolfe, Guillermo Ferrer, Felipe Martínez Rizo, Mariana Aylwin y Laurence Wolff
- Sistemas de Evaluación como Herramientas de Políticas** 64
Leonor Cariola, Gabriela Cares y Rosario Rivero
- Aportes y Desafíos en la Evaluación de Políticas y Programas de Mejoramiento Educativo. Reflexiones desde Nuestra Práctica de Evaluación en Chile** 79
Dagmar Raczynski y Daniel Salinas
- Responsabilização Educacional no Brasil [Educational Accountability in Brazil]** 93
Nigel Brooke
- ¿Qué Estamos Haciendo en Cuba en Evaluación Educativa?** 110
Paul Torres
- Jornada Escolar completa en Chile. Evaluación de Efectos y Conflictos en la Cultura Escolar** 124
Sergio Martinic, David Huepe y Ángela Madrid
- Docentes de Educación Básica y sus Concepciones Acerca de la Evaluación en Matemática** 140
Iraima Moreno y José Ortiz
- Programas de Auditoría Institucional Universitaria. Comparación de la Propuesta Española con el Sistema Británico** 155
Reina Ferrández
- Uso da Teoria de Resposta Ao Item (TRI) para Analisar a Equidade do Processo de Avaliação do Aprendizado Discente** 171
Wagner Bandeira Andriola



EDITORIAL:

LA EVALUACIÓN EDUCATIVA COMO DERECHO HUMANO

F. Javier Murillo y Marcela Román

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/editorial.pdf>



Si recibir educación de calidad es uno de los derechos fundamentales de los seres humanos, ser evaluado y recibir información valorativa sobre la educación que se está recibiendo ha de entenderse también como un derecho. Efectivamente, la evaluación es el único instrumento que tenemos para verificar el cumplimiento de ese derecho superior, internacionalmente reconocido y que supone el acceso igualitario al conocimiento y a las oportunidades disponibles en la sociedad. Garantizar que los niños, jóvenes y adultos estén recibiendo una educación de calidad, requiere de evaluaciones que den cuenta tanto de lo que se aprende, como de las acciones que se desarrollan para su cumplimiento y del contexto en el que ello se desenvuelve. De esta manera los estudiantes, todas las personas en un marco de educación a lo largo de toda la vida, tienen el derecho a recibir información sobre la calidad de los aprendizajes adquiridos. Así como ellos, las familias, la sociedad y los profesionales de la educación, tienen el derecho y el deber de saber si la educación impartida es pertinente, relevante, eficaz, eficiente y equitativa.

Sin embargo, al igual que la educación, no sirve cualquier evaluación. Para que se cumpla ese derecho ésta debe ser de calidad. El incremento en los últimos años de las acciones evaluativas en educación ha supuesto que la cultura de la evaluación, poco a poco, se vaya instalando en los sistemas educativos. Ahora se considera que todos los implicados, todas las instancias y todos los niveles de decisión deben ser evaluados, aunque no siempre de la convicción se haya pasado a la acción. Igualmente es necesario reconocer que en todos los países excelentes prácticas conviven con experiencias deplorables que no sólo no están consiguiendo aportar datos para la mejora de la educación, sino que están contribuyendo a su destrucción.

Pero, ¿qué características debe tener una evaluación de calidad? Buena pregunta. Cuando nos la formulamos, nos percatamos de lo poco que sabemos, de lo mucho que nos queda por recorrer. De hecho, no sabemos exactamente qué significa una evaluación de calidad, ni siquiera si puede haber unos criterios aplicables a toda la gran variedad de evaluaciones que se dan en educación. En todo caso, y a riesgo de ofrecer una imagen sesgada y parcial, es posible defender algunas de las características que debe tener una evaluación de calidad.

En primer lugar, debe ser técnicamente impecable. Independientemente del ámbito, tema o enfoque de evaluación que estamos utilizando, es muy importante que cumpla los requisitos de ser válida, fiable, útil y creíble.

En segundo término, creemos que toda evaluación, independientemente de su tipo, debe ir dirigida a la mejora, a la transformación. Toda evaluación educativa, incluida la sumativa orientada a la toma de decisiones administrativas (por ejemplo, a la contratación de un docente, a la acreditación de una universidad o a la promoción de un estudiante), debe ir orientada a aportar informaciones para transformar la situación evaluada y el contexto en el que se desarrolla.

Para ello, necesariamente deberá estar centrada en aportar informaciones para la reflexión. Dado que el cambio sólo será posible si las personas directamente implicadas toman conciencia de la situación, reflexionan sobre la realidad y toman decisiones pertinentes para su transformación, sólo habrá cambio si se parte de una reflexión. De hecho la evaluación tiene como principal propósito, aportar datos para esa reflexión. De ahí se puede derivar la necesidad de que la evaluación, para que sea de calidad, también ha de ser participativa y no jerárquica.

Se requiere también de una evaluación positiva y no represora. Es probable que muchos de los problemas que tienen los sistemas de evaluación y su rechazo, en muchas ocasiones generalizado, es que son

instrumentos de represión. Sólo en la medida en que las evaluaciones sirvan para destacar los pasos bien dados y sean instrumentos para la comunicación de las altas expectativas y el refuerzo del autoconcepto personal y grupal, podrán a la postre, entregar elementos pertinentes y relevantes para la reflexión y la toma de decisión.

Una evaluación de calidad ha de ser también equitativa y justa. Debe reconocer y asumir la diversidad social y cultural de los estudiantes, para desde allí generar instrumentos y estrategias que permitan evaluar el desarrollo y desempeño de todos y cada uno adecuada y justamente. Una evaluación capaz de dar cuenta de la diversidad y heterogeneidad de los niños, niñas y jóvenes, al mismo tiempo que visibilizar y dimensionar los desafíos que el contexto y sus características, ponen al logro de resultados educativos de calidad.

La última característica que queremos destacar es que sea *adecuadamente comunicada*. La historia nos ilustra cientos de evaluaciones, bien diseñadas e implementadas, pero que después tienen nulos efectos o, incluso, perversos, simplemente por no haber estado adecuadamente comunicadas a la audiencia o audiencias destinatarias.

Si contrastamos mínimamente estos criterios con las evaluaciones que actualmente se desarrollan en Iberoamérica, nos encontramos con la clara necesidad que tienen de mejorar. Efectivamente, esa explosión de evaluaciones que estamos viviendo en la Región, que podría parecer positiva por sí misma, no ha estado acompañada por la suficiente reflexión, análisis e investigación sobre la misma.

Con dolor hemos comprobado que la simple evaluación de la calidad educativa no asegura calidad. Nos hemos pasado mucho tiempo usando la evaluación casi exclusivamente para constatar y ratificar los malos resultados y pobres avances en materia de calidad y equidad educativa de nuestros sistemas. Es hora de validar y usar la evaluación en uno de sus aspectos más relevantes, cual es su potencia para identificar las razones de tales resultados. Una evaluación que identifica, dimensiona y comprende los resultados educativos, como parte de procesos, dinámicas, de juegos de interacción entre actores y sujetos que se actualizan cotidianamente en el deseo de formar y desarrollarse en tanto personas con iguales derechos al conocimiento y a las oportunidades sociales que les aseguren movilidad y un futuro pleno.

La extendida idea de que “sin evaluación no se puede mejorar” no sólo es una falacia, sino que los efectos perversos de una evaluación mal diseñada, implementada o comunicada, pueden ser el elemento clave que impidan el cambio e, incluso, el factor desencadenante de un retroceso en su calidad. Por tanto, mejorar las evaluaciones que se realizan no sólo es un deber técnico, es una obligación moral.

Parece necesario fomentar la investigación, la reflexión y el intercambio de experiencias de buenas prácticas en evaluación educativa en la región, de tal forma que se consiga una mejor evaluación para una mejor educación.

La *Revista Iberoamericana de Evaluación Educativa* nace con ese propósito: contribuir a mejorar la calidad de la Evaluación Educativa que se realiza en Iberoamérica de tal forma que logre incrementar significativamente los niveles de calidad y equidad educativa. Y ello mediante, el fomento de la investigación, el análisis y la reflexión; la generación de redes de estudiosos sobre el tema; y la difusión de estudios, aportaciones teóricas y buenas prácticas sobre el tema.

El ámbito de esta publicación científica de carácter periódico está centrado en evaluación educativa, pero desde una perspectiva amplia: evaluación en el aula, evaluación de docentes universitarios y no

universitarios, evaluación de universidades y de escuelas, evaluación de directivos y de la supervisión, evaluación de programas y reformas educativas, de las administraciones educativas, de sistemas educativos, evaluaciones internacionales Y admite una amplia gama de artículos: evaluaciones empíricas en cualquier ámbito y desde cualquier aproximación metodológica, propuestas de desarrollo metodológico o reflexiones de calidad. También publica reseñas de libros sobre la temática y discusiones sobre los artículos publicados.

La Revista pretende, igualmente, ser un foco de encuentro y debate entre los especialistas en la Región: y ello sólo puede conseguirse mediante una colaboración con las asociaciones que sobre esta temática trabajan en la región, tales como la Red de Seguimiento, Evaluación y Sistematización de América Latina y el Caribe (RELAC), la Sociedad Brasileña de Evaluación Educacional o el Grupo de Trabajo de Evaluación y Estándares de PREAL (GTEE/PREAL), con quienes ya la revista ha trabajado desde antes de su nacimiento y pretende seguir haciéndolo en el futuro.

Como no podía ser de otra forma, la nueva revista nace con una verdadera fiesta de inauguración: una fiesta intelectual en forma de primer número. Así, contamos con una docena de excelentes y variados artículos en los que algunos de los máximos especialistas de la región en evaluación educativa comparten algo de su sabiduría.

Abre el número un extenso artículo de F. Javier Murillo y Marcela Román, donde se aporta una panorámica general de los resultados de aprendizaje de los estudiantes de América Latina y el Caribe a partir del análisis de las evaluaciones nacionales realizadas hasta el momento. Sigue la destacada aportación de Margarita Poggi, directora del IIPE/UNESCO, donde reflexiona sobre la evaluación educativa, tanto de sistemas educativos (o subsistemas), como de instituciones o de programas educativos, aportando conceptos y metodologías que recogen algunas de las discusiones tradicionales sobre el tema. El profesor, ex-ministro, ex-director de la OREALC/UNESCO, Ernesto Schiefelbein, junto con la profesora Paulina Schiefelbein, por su parte, hacen una revisión de cómo ha evolucionado la evaluación del sistema educativo chileno en los casi 60 años de su existencia.

RIEE publica también el interesante documento preparado por el Grupo de Trabajo de Evaluación y Estándares acerca de la evaluación estandarizada de los estudiantes a gran escala, tanto nacionales como internacionales. Esta aportación incumple uno de los requisitos de la revista, al no ser inédito, pero creemos que su calidad hace que sea importante incluirlo en este primer número.

El análisis de los factores asociados al desempeño de los estudiantes dentro de las evaluaciones nacionales como un instrumento para el diseño de políticas públicas, a partir del análisis de los que se hace en Chile es el objeto del artículo de Leonor Cariola, Gabriela Cares y Rosario Rivero, desde el Sistema de Medición de la Calidad de la Educación (SIMCE). También desde ese país, Dagmar Raczynski y Daniel Salinas, reputados investigadores del grupo "Asesorías para el Desarrollo", identifican en su artículo hallazgos claves para el diseño e implementación de la política educativa, y reflexionan sobre algunos desafíos metodológicos que enfrenta la evaluación de políticas y programas de mejoramiento escolar en la actualidad.

El investigador experto en evaluación Nigel Brooke, en su excelente artículo "Responsabilização Educacional no Brasil", publicado en portugués y en inglés, ofrece una respuesta preliminar a la cuestión de si, en América Latina, políticas de *accountability* similares están o no siguiendo la expansión de la evaluación educativa, mediante el estudio del caso brasileño.

Realmente interesante nos resulta también el artículo del destacado investigador del Instituto Central de Ciencia Pedagógicas (ICCP), Paúl Torres, donde nos aporta información sobre el sistema de evaluación de Cuba. Sistema de evaluación que, acorde con la innegable calidad de su sistema educativo, despierta un gran interés entre los expertos.

Los profesores de la Universidad Católica de Chile, Sergio Martinic, David Huepe y Ángela Madrid, por su parte, presentan un estudio en el que se evalúa el peso de la cultura escolar en los efectos que logra la implementación de la Jornada Escolar Completa (JEC), uno de los cuatro ejes de la Reforma Chilena. Así, se sostiene que los efectos de este programa en los aprendizajes escolares, varía de acuerdo a la representación que tienen los docentes de la realidad social del niño y del establecimiento. Mientras más negativa sea la representación, será menor el impacto de la JEC.

Desde Venezuela, los profesores Irima Moreno y José Ortiz expertos en evaluación en Matemática, presentan un artículo donde se analizan las concepciones de docentes de la Tercera Etapa de Educación Básica venezolanos, respecto a la evaluación en matemática y al mismo tiempo, se caracterizan y contrastan con lo establecido en las normativas legales vigentes.

Una visión de la evaluación institucional universitaria la aporta la profesora e investigadora de la Universitat Jaume I, España, Reina Ferrández. En su artículo realiza una comparación entre el sistema de auditoría británico y la propuesta española analizando similitudes y diferencias, ventajas e inconvenientes.

Para finalizar, el profesor e investigador brasileño, especializado en evaluación, Wagner Bandeira nos aporta el artículo de carácter metodológico ineludible: estudia la Teoría de Respuesta al ítem para analizar la equidad del proceso de evaluación de los estudiantes, mediante su aplicación

Creemos que este primer número es un buen ejemplo de lo que pretendemos que sea la revista: de alta calidad, que realice aportaciones de todo tipo (revisiones, reflexiones, evaluaciones empíricas, aportaciones metodológicas), de todos los ámbitos de la evaluación, y con un carácter internacional claramente iberoamericano.

Poner en marcha una nueva publicación científica supone un paso más en la consecución de una irrenunciable utopía: tener una educación de calidad para todos que contribuya a construir una sociedad más justa, equitativa, inclusiva y fraterna. Con la aparición de la Revista Iberoamericana de Evaluación Educativa, la Red Iberoamericana de Investigación sobre Eficacia y Cambio escolar (RINACE) pretende hacer su humilde aportación en el área. Ahora, sólo el tiempo dirá si se ha conseguido.



RESULTADOS DE APRENDIZAJE EN AMÉRICA LATINA A PARTIR DE LAS EVALUACIONES NACIONALES

F. Javier Murillo y Marcela Román

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art1.pdf>



En los últimos años, la mayor parte de los países de América Latina ha desarrollado sistemas nacionales para evaluar la calidad educativa de sus sistemas escolares, instalando procesos permanentes de evaluación del aprendizaje y rendimiento de los estudiantes (Ferrer, 2006). De esta forma se cuenta con una completa información acerca del desempeño de los alumnos de la Región, básicamente en Educación Primaria y en algún caso en Educación Secundaria, a partir de las evaluaciones que los propios países realizan.

En este estudio se analizarán los resultados de aprendizaje de los niños y jóvenes escolarizados de 15 países América Latina a partir de la revisión de los datos obtenidos en evaluaciones nacionales. Está centrado en los resultados de aprendizaje de los estudiantes matriculados en los últimos cursos de Educación Primaria (CINE 1) y de los grados 2º y 3º de la Educación Secundaria Inferior (CINE 2), lo que significa entre 4º y 6º grado de Educación Primaria o Educación Básica y, entre 7º y 9º grado o 1º y 2º de secundaria en el caso de Secundaria Inferior, según la nomenclatura de cada país.

Tiene un doble enfoque: por una parte realiza un análisis de la situación actual, explorando la asociación de algunas variables seleccionadas sobre los resultados de aprendizaje a partir de los datos más recientes disponibles; y, a continuación, un estudio de la evolución desde el año 1999 ó 2000 hasta la actualidad. El documento finaliza con un apartado donde se resumen los hallazgos más importantes y se avanza en la explicación de esos resultados encontrados.

1. DATOS CONSULTADOS

El estudio obtiene la información a partir del análisis de los reportes de evaluación que han elaborado las propias unidades de evaluación de los Ministerios de Educación de los respectivos países para trabajar y para difundir los resultados. De esta forma, los datos e informaciones ofrecidas tienen carácter oficial, pero al mismo tiempo, limitaciones propias de la heterogeneidad de los datos, obtenidos con diferentes metodologías, que abordan variables distintas y medidas en fechas diferentes. Así, se dispone de resultados de evaluaciones de aprendizajes en 4º, 5º y 6º de Primaria; 2º y 3º de Secundaria. En la totalidad de estos grados se ha evaluado rendimiento en Lengua¹ y Matemática y, en algunos grados y países, se ha incorporado además la medición en Ciencias Naturales y Sociales.

Para el análisis de la situación actual, se han utilizados las evaluaciones más recientes disponibles, lo que significa que para Educación Primaria se analizan datos del año 2005 para Brasil, Chile, Colombia, Costa Rica, El Salvador, Guatemala, México, Panamá y Uruguay; del 2004 para Honduras, Paraguay y Perú; del 2003 para Argentina y del año 2002 para Nicaragua. Los últimos datos disponibles para Secundaria Inferior corresponden al año 2005 para Brasil, Colombia, Costa Rica, El Salvador, Guatemala, México y Panamá; del 2004 para Chile y Perú. Venezuela no realiza evaluaciones nacionales del desempeño desde el año 1998, Ecuador desde el 2000 y Bolivia desde 2001; y en ninguno de los tres casos están los datos accesibles en la actualidad. República Dominicana lo hace sólo con fines de ofrecer información individualizada a cada alumno, pero no realiza agregación alguna. No están disponibles los datos de las evaluaciones nacionales de Cuba. En anexo se adjunta el listado de los reportes nacionales de evaluación utilizados en el estudio y que se han obtenido de los sitios oficiales de los Ministerios de Educación, así como de la petición directa a los responsables de todas y cada una de las unidades de medición y evaluación de los Ministerios de Educación de América Latina.

¹ En este documento se utiliza Lengua como término genérico. Cada país ha denominado a esta materia de forma distinta, en algunos casos usando un término sinónimo y en otros casos para destacar que se evalúa un área concreta dentro de esta materia. Así, por ejemplo, Perú las llama Lenguaje; Costa Rica, Español; Brasil, Portugués; Guatemala, Lectura; o México que utiliza Lenguaje o Comprensión Lectora, según las evaluaciones.

Por último, es importante consignar el que los sistemas nacionales asumen distintos principios para la evaluación de los aprendizajes: Criterial o Normativa. Evaluar en relación a una "norma" implica comparar el resultado de cada individuo con los resultados de una población o grupo a los que pertenece (grupo de referencia), fijando para ello una norma, que es la misma para todos. En este caso, la norma del grupo es el estándar. Bajo este principio, no es posible realizar comparaciones entre áreas o disciplinas.

CUADRO 1. DATOS UTILIZADOS PARA EL ANÁLISIS

País	Grado	Grado oficial	Materia	Año	Criterial / Normativa	Nombre
Argentina	6° P 3° S	6° EGB 9° EGB	L, M L, M	2003 (1)	Criterial Criterial	Operativo Nacional de Evaluación
Bolivia	ND: La última evaluación nacional realizada data de 2001, pero no está accesible					
Brasil	4° P 2° S	4° E. Fundamental 8° E. Fundamental	L, M L, M	2005 2005	Normativa Normativa	SAEB
Chile	4° P 2° S	4° Básico 8° Básico	L, M, CNyS L, M, CNyS	2005 2004	Normativa(2) Normativa	SIMCE – MINEDUC
Colombia	5° P 3° S	5° grado 9° grado	L, M, CN y CS L, M, CN y CS	2005 2005	Normativa Normativa	SABER
Costa Rica	4°-5°-6° P 1°-2°-3° S	II ciclo de EGB III ciclo de EGB	L, M, ES, CC L, M, ES, CC	2005 2005	Criterial Criterial	
Cuba	No hay datos accesibles					
Ecuador	La última evaluación es del año 2000 y no está accesible					
El Salvador	6° P 3° S	6° grado 9° grado	L, M L, M	2005 2005	Criterial Criterial	
Guatemala	6° P 3° S	6° de Primaria 3° de Secundaria	L, M L, M	2005 2005	Criterial Criterial	PRONERE 2005 PRONERE 2005
Honduras	6° P	6° grado	L, M	2004	Criterial	
México	6° P 3° S	6° Ed. Primaria 3° Ed Secundaria	L, M L, M	2005 2005	Criterial Criterial	EXCALE
Nicaragua	6° P	6° de primaria	L. M.	2002	Criterial	
Panamá	6° P 3° S	6° básica 9° básica	L, M, CN L, M, CN	2005 2005	Criterial Criterial	Pruebas nacionales de logros académicos
Paraguay	6° P	6° EEB	L, M y ES	2004	Criterial	SNEPE
Perú	6° P 3° S	6° de Ed. Básica 9° de Ed. Básica	L, M L, M	2004 2004	Criterial Criterial	Evaluación Nacional del rendimiento Estudiantil
R. Dominicana	No ofrece análisis, exclusivamente resultados de alumnos individuales					
Uruguay	6° P	6° de Primaria	L, M	2005	Normativa	
Venezuela	Desde 1998 Venezuela no hace evaluación nacional del aprendizaje de los alumnos					

(1) En 2005 Argentina realizó una evaluación, pero en el momento de redactar este informe, los resultados no se habían hecho públicos.

(2) En Chile, hasta el año 2004, la medición realizada en 4° de Primaria era en base a normas. Desde el 2006 en adelante, el SIMCE ha hecho un giro hacia la evaluación de estándares, evaluando y reportando de acuerdo a "Niveles de logro de aprendizajes" de los estudiantes

Fuente: Elaboración propia a partir de los informes de evaluación.

En la evaluación criterial, se miden los aprendizajes y se evalúa a cada estudiante en relación a criterios disciplinarios preestablecidos que reflejan los objetivos curriculares. Los criterios abarcan tanto objetivos relacionados con la comprensión de los conocimientos como con determinadas habilidades y destrezas. En algunos casos, para cada criterio se elaboran descripciones de niveles de logro/desempeño que

constituyen un continuo de calidad, desde un desempeño más básico a uno de excelencia. Se reportan así, resultados en función del porcentaje de estudiantes que se ubica en cada nivel de desempeño y, en algunos casos, estableciendo cuáles deben ser considerados como mínimos exigibles. En otros casos, se opta sólo por ofrecer el porcentaje de respuestas correctas y determinar qué porcentaje de ellas se considera aceptable. El problema que existe con las pruebas criterioles en que el establecimiento del corte siempre es una decisión con un cierto grado de arbitrariedad, por lo que también hay serias limitaciones para hacer comparaciones entre áreas o grados.

Para los años y grados considerados, Brasil, Chile, Colombia y Uruguay, implementaban Evaluaciones basadas en Normas, mientras que la mayoría de los países de la región, utilizaban pruebas criterioles con base curricular (Argentina, Costa Rica, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá y Perú).

Los cuadros 1 y 2 describen los datos utilizados para el análisis (grados, disciplinas evaluadas, según año, criterio y unidad estatal responsable), así como la forma en que los países reportan el rendimiento escolar en Lengua y Matemática.

CUADRO 2. FORMA DE PRESENTAR LOS RESULTADOS DE RENDIMIENTO EN LENGUA Y MATEMÁTICA, EN PRIMARIA (4º, 5º Ó 6º) Y SECUNDARIA INFERIOR (2º Y 3º), EN LOS PAÍSES DE AMÉRICA LATINA, SEGÚN SUS EVALUACIONES NACIONALES

	Rendimiento medido como
Argentina	Porcentaje promedio de respuestas correctas
Brasil	Rendimiento promedio
Chile	Puntuación promedio
Colombia	Puntaje promedio
Costa Rica	Promedio de respuestas correctas (sobre 100) Porcentaje de estudiantes con nota mayor o igual a 65,0 puntos
El Salvador	Puntuaciones promedio (sobre 10) Porcentaje de estudiantes en 3 niveles de desempeño: Básico, Intermedio y Superior
Guatemala	Porcentaje de estudiantes en 4 niveles de logro: No dominio, Dominio cercano, Dominio básico y Dominio Avanzado. No dominio y dominio cercanos conforman el nivel de No logro, Dominio básico y Dominio avanzado conformen el nivel de Logro
Honduras	Puntuaciones promedio (sobre 100) Porcentaje de estudiantes en 4 niveles de logro: Escaso, Bajo, Medio y Alto. Nivel de suficiencia: Medio o Alto
México	Puntuaciones promedio (Media 500, desviación 100) Porcentaje de estudiantes en 4 niveles de logro: Por debajo del básico, Básico, Medio y Avanzado
Nicaragua	Porcentaje de alumnos en 3 niveles de rendimiento académico: Básico, Intermedio y Proficiente
Panamá	Porcentaje de estudiantes en tres niveles de logro: Deficiente, Regular y Bueno.
Paraguay	Porcentaje promedio de respuestas correctas
Perú	Porcentaje de estudiantes en 4 niveles de desempeño: menor que el previo, Previo, Básico y Suficiente. Los estudiantes que se encuentran en los niveles básico o suficiente son los que logran los aprendizajes
Uruguay	Porcentaje de estudiantes en 5 niveles de desempeño: Debajo del Nivel I, Nivel I, II, III y IV

Fuente: Elaboración propia a partir de los informes nacionales de evaluación.

Para hacer el análisis de la evolución de estos resultados desde la Conferencia de Dakar y hasta la actualidad, se ha utilizado el material reportado por los países a partir de sus propios análisis. Ello garantiza la calidad de los datos ofrecidos, pero tiene como limitaciones el que pocos sistemas han realizado este tipo de estudio, y que los años y variables analizadas varían entre uno y otro país. En este caso, se han utilizado resultados de informes de evaluaciones nacionales de Brasil (1999 y 2005), Chile (2002 y 2005), Colombia (2003 y 2005), Costa Rica (1999 y 2005), El Salvador (2003 y 2005), Honduras (1997 y 2004), México (2000 y 2005) y Perú (1998 y 2004).

La comparación temporal, se proponía trabajar con datos de 6° y 9° grados de educación básica (6° grado de Primaria y 3° de Educación Secundaria Inferior). Ello se consiguió con la mayoría de los países, con las excepciones de Brasil y de Chile, que fueron de 4° y de 8°; y de Colombia, que fueron de 5° y 9°. Uruguay y Paraguay sólo estaban accesibles los resultados de 6° grado. En todos los casos se contó, al menos, con datos de Matemática y Lengua.

2. ANÁLISIS DE LOS RESULTADOS DE APRENDIZAJE

A pesar de la existencia de importantes diferencias entre países, es posible sostener que el rendimiento de los estudiantes latinoamericanos es bajo en Lengua como en Matemática y tanto para 6° de primaria como para 3° de secundaria inferior. En aquellos países que han evaluado el rendimiento a través de pruebas criterios, este complicado panorama se ve reflejado en el elevado porcentaje de alumnos que se ubican por debajo del Nivel de desempeño establecido como suficiente o, dicho de otro modo, se expresa en el bajo porcentaje de estudiantes que logran los aprendizajes esperados de acuerdo a su grado y escolaridad previa.

En efecto, los resultados de los países que han definido un nivel de suficiencia mínimo a alcanzar por los estudiantes en las áreas evaluadas, demuestran, inequívocamente, dicha afirmación. Así, en Perú, por ejemplo, sólo el 7,9% de los alumnos de 6° de Primaria y el 6,0% de los que cursan el 3° de secundaria logran los aprendizajes esperados para su nivel en Matemática; en Honduras y en 6° grado, alcanzan el nivel definido como suficiente un 11,1% de los alumnos en Lengua (Español) y un 7,6% en Matemática. En Guatemala, por último el 55% y del 47% de los alumnos que alcanza el nivel de suficiencia predeterminado en Matemática y Lengua respectivamente. Es necesario señalar que los datos de los países no son comparables entre sí, dado que dependen del punto de suficiencia determinado para cada país y en cada evaluación.

La revisión del rendimiento en aquellos países que no establecen niveles de suficiencia directamente, permite conocer los resultados según el porcentaje de preguntas correctas para el colectivo evaluado. Así, Argentina indica que para el conjunto de la nación los alumnos responden entre el 56,4% y el 52,7% de las respuestas correctas, dependiendo del curso y la materia implicada. En Paraguay, el porcentaje medio de respuestas respondido correctamente por los alumnos de 6° de Primaria es de un 51,3% para Lengua (comunicación) y de 45,9% para Matemática.

2.1. Mayor Escolaridad, Menor Aprendizaje: Una paradoja en los resultados

Considerando las limitaciones señaladas de la comparación entre **grados**, la imagen global, respecto del rendimiento en 6° de primaria y 3° de secundaria inferior de los países que utilizan pruebas criterios, muestra que éste es más bajo en secundaria inferior que en los últimos grados de primaria y, que dicha brecha es muy disímil en los resultados entre niveles al interior de los países y entre las áreas evaluadas (cuadro 3).

Así, hay países como México o Panamá con una importante diferencia entre el rendimiento de primaria y secundaria inferior en Lengua y en Matemática. Efectivamente, en México se observa un aumento muy importante en el porcentaje de alumnos de 6° de Primaria a 3° de Secundaria que no alcanzan el nivel básico tanto en Lengua, (del 18,0% a 32,7%), como en Matemática (del 17,4 al 51,1%). Algo análogo acontece en Panamá donde el porcentaje de alumnos que obtienen niveles de regular o bueno pasa de 56,3% en 6° a 28,2% en 3° de Secundaria en Lengua y de 48,3% a 14,7% en Matemática.

En otros países las diferencias son más moderadas. Argentina y El Salvador son ejemplos de este caso. En Argentina la discrepancia en el rendimiento entre 6°P y 3°S es de apenas 1,5 puntos porcentuales en

Lengua y 3 puntos porcentuales en Matemática. En El Salvador las diferencias bordean los 3 puntos porcentuales, siempre a favor de estudiantes de 6° de Primaria (2,9 puntos en Lengua y 3,4 en Matemática).

CUADRO 3. RESULTADOS EN RENDIMIENTO EN LENGUA Y MATEMÁTICA, EN 6° DE PRIMARIA Y 3° DE SECUNDARIA INFERIOR, EN LOS PAÍSES DE AMÉRICA LATINA SEGÚN SUS EVALUACIONES NACIONALES

	Rendimiento medido como	Grado	Lengua	Matemática
Argentina	% de respuestas correctas	6 P 3 S	54,1 52,7	56,4 53,4
Costa Rica	% de estudiantes con nota mayor o igual que 65	4-6 P 1-3 S	77,7 74,5	48,3 22,5
El Salvador	% alumnos en el nivel intermedio o superior	6 P 3 S	73,3 70,4	54,5 51,1
Guatemala	% alumnos alcanzaron el dominio	6 P 3 S	47,9 52,2	55,3 41,7
Honduras	% alumnos en el nivel de suficiencia (medio o alto)	6 P	11,1	7,8
México	% de alumnos en el nivel básico o más	6 P 3 S	82,0 67,3	82,6 48,9
Nicaragua	% de alumnos en el nivel intermedio o proficiente	6 P	30,3	11,9
Panamá	% de alumnos en el nivel regular o bueno	6 P 3 S	56,3 28,8	48,2 14,7
Paraguay	% promedio de respuestas correctas	6 P	51,3	45,9
Perú	% alumnos en el nivel suficiente	6 P 3 S	12,1 15,1	7,9 6,0

Notas: 1. Ningún país aporta información acerca de si las diferencias, tanto entre áreas como entre grados, son estadísticamente significativas

2. Al ser las pruebas de Brasil, Chile, Colombia y Uruguay de carácter normativo, no es posible comparar los resultados por materias o grados

Fuente: Elaboración propia a partir de los informes nacionales de evaluación.

Por último, la comparación del rendimiento por grado, permite constituir un tercer grupo de países en donde no hay un patrón al respecto. Guatemala y Perú se ubican en este grupo.. Por ejemplo, en Perú los resultados muestran una tendencia diferente según el área. Así, el porcentaje de estudiantes que se encuentra en el nivel "suficiente" en Lengua es de 12,1 en 6°, elevándose a un 15,1 en 3° de secundaria; mientras que en Matemática del 7,9% en 6° disminuye al 6,0% en Lengua. Del mismo modo, mientras en Lengua los estudiantes de 6° de Primaria en Guatemala obtienen 4,3 puntos porcentuales menos que los de 3° de Secundaria, logran 13,6 puntos porcentuales más que los de secundaria en Matemática.

Considerando que en la totalidad de los países la Educación Secundaria Inferior es obligatoria y, que al mismo tiempo, las mayores tasas de deserción y abandono escolar se producen en el paso de la Primaria a la Secundaria, cabría esperar un mejor desempeño entre los estudiantes de secundaria puesto que quienes abandonan son principalmente estudiantes con historial de fracaso y bajo rendimiento, permaneciendo de este modo los mejores alumnos en el nivel secundario. Así entonces, sería previsible que el logro aumentara de primaria a secundaria, hecho que como se ha visto no ocurre. Lo anterior lleva a plantearse preguntas que expliquen dicho fenómeno. ¿Son las pruebas pertinentes para dar cuenta de los estudiantes realmente saben?; ¿los conocimientos impartidos en estos grados, son significativos y motivadores para los estudiantes?; ¿la forma de enseñar, el seguimiento y evaluación de los aprendizajes responde a la instalación de procesos de aprendizajes efectivos? ¿hay una búsqueda real y pedagógica de articular la cultura escolar formal con la cultura juvenil y entorno sociocultural?; ¿el dominio teórico de

conocimientos disciplinares de los profesores, dialoga con un dominio de las didácticas necesarias para una enseñanza eficaz?. Este tipo de interrogantes, invisibles en las evaluaciones realizadas no permite disponer de buenas respuestas para mejorar los desempeños de los jóvenes.

2.2. ¿Se avanza igual en Matemática que en Lengua?

Una nueva mirada puede ser hecha mediante el estudio de las diferencias en el rendimiento entre las materias analizadas. Es decir, ¿los estudiantes de América Latina rinden más en Lengua o en Matemática?

El estudio realizado no ofrece ideas conclusivas, encontrándose bastantes diferencias en los resultados reportados por los países.

Así, por una parte, en Argentina, Brasil, México los alumnos parecen obtener mejores resultados en Matemática que en Lengua tanto para Primaria como para Secundaria inferior; en Guatemala sólo ocurre para los alumnos de 6° de Primaria. Hay un segundo grupo de países cuyos resultados son mejores para Lengua que para Matemática. En este grupo se ubican El Salvador, Guatemala (para 3° de Secundaria), Honduras Nicaragua y Panamá. En este grupo destaca Costa Rica por las importantes diferencias en el desempeño entre Lengua y Matemática: mientras que en Lengua el 77,7% de los alumnos obtiene una calificación igual o superior a 65 (puntaje promedio preestablecido), en Matemática, sólo lo alcanza el 48,3%.

En todo caso, parece necesario volver a recordar en este momento la dificultad de hacer extrapolaciones acerca de la superioridad de los resultados entre áreas o entre cursos. Conviene señalar de entrada, que en las evaluaciones nacionales con base normativa no es posible hacer ninguna de esas comparaciones. Por otra parte, tampoco en las evaluaciones criterios tales comparaciones son inmediatas o sencillas, dado que los puntos de corte son arbitrarios y por ende, no necesariamente comparables. De esta forma, las diferencias encontradas en los resultados pueden corresponder simplemente, a criterios más exigentes en una materia o un grado, y no sólo a una disparidad en el nivel de desempeño de los estudiantes.

Es posible postular también, que estos resultados, sean un reflejo de la intencionalidad de la política educativa que busca fortalecer a los estudiantes en las áreas de conocimiento que considera prioritaria, aunque siempre dentro de los aprendizajes más básicos o elementales. Al igual que en el punto anterior, la información disponible no permite argumentar para sostener una u otra hipótesis, mostrando la necesidad de disponer de otro tipo de información que ayude a comprender el por qué la mayoría de los sistemas no aportan al aprendizaje de la misma manera en las distintas disciplinas.

Mucha mayor fiabilidad se obtiene del análisis de los resultados al interior de cada país, a partir de variables como género, nivel sociocultural o zona rural o urbana, dado que los resultados son estrictamente comparables entre esos grupos, tanto en las pruebas normativas como en las criterios. A continuación se abordará el estudio de la asociación de cada una de esas variables con el rendimiento del alumno.

2.3. ¿Aprenden más los niños que las niñas en la escuela?

Desde la mirada de la equidad es importante conocer cuáles son los resultados que obtienen las niñas respecto de los niños. En todos los países analizados, las mujeres obtienen mejores resultados en Lengua, sin embargo, la situación de ellas en Matemática depende del país y grado (cuadro 4).

Las diferencias en los aprendizajes que logran las niñas en Lengua por sobre sus compañeros, se expresa de diferentes maneras en los países:

- En El Salvador, el porcentaje de niñas de 6° de Primaria que está en el nivel de intermedio o superior es igual al 75,1%, porcentaje que disminuye al 71,6% para los varones. En 2° de secundaria el 71,2% de las jóvenes alcanza estos niveles, correspondiendo a un 69,7% en el caso de los chicos.
- En Honduras, en 6° grado, el promedio de las puntuaciones obtenida por las mujeres en Lengua es de 44,9% mientras que el promedio de los varones es de 42,9%.
- En México un 86,6% para las chicas de 6° grado, se ubica en los niveles suficiente o superior frente al 75,6% para los chicos. En el caso de 3° de secundaria, estos niveles lo logran el 71,8% de las niñas y el 62,8% de los varones.
- En Nicaragua, en Lenguaje y para el 6° grado, el 34,7% de las chicas tiene un nivel de rendimiento intermedio o proficiente, mientras que sólo lo logra el 25,4% de los chicos.
- En Perú y en 3° de secundaria, el 16,7% de las chicas alcanzan el nivel suficiente en Lengua, respecto del 13,4% de los varones.

CUADRO 4. RESULTADOS RENDIMIENTO EN LENGUA Y MATEMÁTICA EN ESTUDIANTES VARONES Y MUJERES, EN 6° DE PRIMARIA Y 3° DE SECUNDARIA INFERIOR, EN LOS PAÍSES DE AMÉRICA LATINA SEGÚN SUS EVALUACIONES NACIONALES

	Rendimiento medido como	Grado	Lengua Var-muj	Matemática Var-muj
Chile	Puntuación promedio	4 P	252-259 (ds)	249-246(dns)
El Salvador	% alumnos en el nivel intermedio o superior	6 P 3 S	71,6-75,1 69,7-71,2	56,8-52,3 54,5-47,8
Guatemala	% alumnos alcanzaron el logro	6 P 3 S	47,9-47,2 52,5-47,5	56,6-53,9 47,2-35,7
Honduras	% alumnos en el nivel de suficiencia (medio o alto)	6 P	10-9	13-7
México	% de alumnos en el nivel básico o más	6 P 3 S	75,6-86,6 62,8-71,8	93,2-93,0 50,3-48,3
Nicaragua	% de alumnos en el nivel intermedio o proficiente	6 P	25,4-34,7	12,3-11,5
Perú	% alumnos en el nivel suficiente	6 P 3 S	11,0-13,3 13,4-16,7	9,0-6,8 6,7-5,2

Nota: Argentina, Brasil, Colombia, Costa Rica, Panamá, Paraguay y Uruguay no ofrecen desagregaciones por género.

(ds): Diferencias significativas

(dns): Diferencias no significativas

Fuente: Elaboración propia a partir de los informes nacionales de evaluación.

La excepción está constituida por Guatemala, donde los varones obtienen mejor puntuaciones que las mujeres en los dos cursos estudiados.

Para Matemática, sin embargo, los datos no son tan coherentes. Así, es posible encontrar distintos casos:

- Países donde no hay diferencias estadísticamente significativas para ambos cursos: Chile y Nicaragua

- Países donde las mujeres tienen mejor rendimiento que sus compañeros varones: Perú.
- Países donde los hombres obtienen mejor rendimiento en Matemática: El Salvador, Guatemala, Honduras y México.

Pareciera ser que, de alguna manera el patrón cultural respecto de los géneros, se reproduce en los resultados, adjudicando a las mujeres roles más comunicativos que lógicos. Esta histórica inequidad priva el desarrollo de la expresividad y comunicación en los hombres y de herramientas para comprender y resolver problemas en las mujeres. De esta manera, ambos géneros no logran desarrollar plenamente dimensiones fundamentales para desenvolverse en la vida social, personal y cultural.

Esta compleja situación, demanda mayores y distintos esfuerzos desde la política para revertirla, al mismo tiempo que profundizar en la aparente o mayor equidad que al respecto, han logrado otros países al menos en la enseñanza primaria y/o secundaria.

2.4. Aprendizajes y condición socioeconómica de las familias

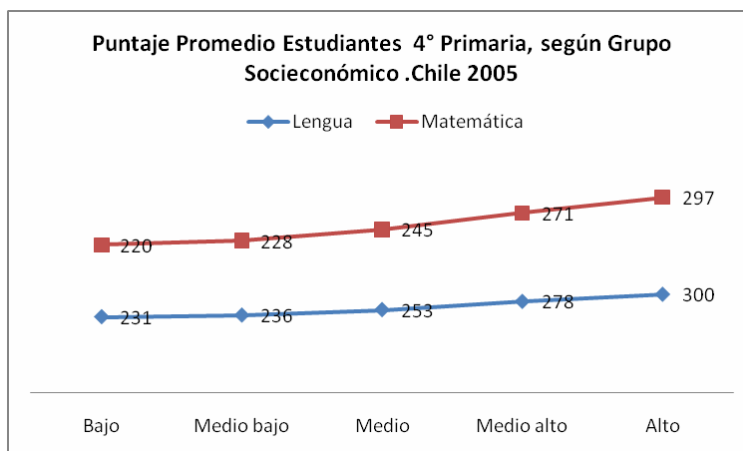
En las evaluaciones nacionales, pocos países de América Latina abordan el estudio de la incidencia de la situación socio-económica de las familias, sobre el rendimiento de los estudiantes, permitiendo así comparar establecimientos que atienden alumnos de similares características socioeconómicas. Lo hacen Chile, Honduras, México, Nicaragua y Uruguay. La diferencia en la conceptualización de dicho factor, así como la forma de su medición, exige ser precavido con la interpretación de los resultados. En el caso de Chile, se clasifican los establecimientos cuyos estudiantes rinden la Prueba SIMCE, de acuerdo con las características socioeconómicas predominantes de sus alumnos. Dicha caracterización se establece promediando los años de estudio del padre y de la madre, el ingreso familiar mensual y el índice de vulnerabilidad escolar del establecimiento². Honduras, por su parte calcula este índice a partir de información aportadas por el propio alumno: la cantidad de bienes y servicios con los que cuenta el hogar, si el niño vive con ambos padres o con algunos de ellos y si saben leer y escribir, y el tipo de material con la que está construido el piso de la vivienda. No hay información detallada al respecto en México.

Uruguay analizó la incidencia del contexto sociocultural de las familias en el rendimiento y Nicaragua del nivel máximo de estudio alcanzados por los padres. Independientemente de lo qué implica esta variable y de cómo ella se mide, todas las aproximaciones informan de la fuerte incidencia del nivel socioeconómico o socio-cultural en el rendimiento de los estudiantes.

Chile, por ejemplo, constata relaciones directas entre nivel socioeconómico de las familias y rendimiento escolar en Primaria y Secundaria, tanto en Lengua como en Matemática. Así, mientras que en Lengua los alumnos del grupo socio-económico bajo han obtenido una puntuación promedio en Lengua de 231 puntos, los estudiantes pertenecientes a grupo socio-económico alto consiguen un promedio de 300 puntos (gráfico 1). En Matemática, los estudiantes pertenecientes al grupo socioeconómico Alto, logran 77 puntos más que los del grupo socioeconómico Bajo.

² La información para las tres primeras variables se recoge en los días cercanos a la prueba, mediante cuestionarios aplicados por el SIMCE, que son respondidos por los apoderados. La última variable corresponde a un índice calculado anualmente por la Junta Nacional de Auxilio Escolar y Becas (JUNAEB) y su valor refleja el porcentaje de los alumnos de un establecimiento que se encuentran en situación de vulnerabilidad escolar.

GRÁFICO 1. PUNTAJE PROMEDIO DE LOS ALUMNOS DE 4º PRIMARIA EN LENGUAJE Y MATEMÁTICA POR GRUPO SOCIO-ECONÓMICO. CHILE 2005



Esta correlación se repite en ambas disciplinas en la Secundaria Inferior, relevando la grave segmentación del Sistema Escolar Chileno. En efecto, las diferencias en los puntajes promedios de los estudiantes de grupos socioeconómicos Alto y Bajo de 2º secundaria inferior, llegan a los 72 puntos en Lengua y a 79 en Matemática (cuadro 5).

CUADRO 5. PUNTAJE PROMEDIO DE LOS ALUMNOS DE 4º PRIMARIA Y 2º SECUNDARIA INFERIOR EN LENGUAJE Y MATEMÁTICA POR GRUPO SOCIO-ECONÓMICO. CHILE, 2004 Y 2005

Grupo socio-económico	4º Primaria 2005		2º Secundaria Inferior 2004	
	Lengua	Matemática	Lengua	Matemática
Bajo	231	220	229	232
Medio bajo	236	228	234	235
Medio	253	245	253	253
Medio alto	278	271	280	282
Alto	300	297	301	311
<i>Total nacional</i>	<i>255</i>	<i>248</i>	<i>251</i>	<i>253</i>

Fuente: SIMCE (2004; 2005).

En Honduras los datos son igualmente concluyentes. Así, en Lengua, el porcentaje promedio de respuestas correctas por parte de los alumnos de 6º grado es de 9,9 puntos porcentuales menos para los niños de familias de condición socio-económica desfavorables, respecto de los estudiantes con mejores condiciones. De manera similar la diferencia en Matemática es de 7,5 puntos (cuadro 6).

CUADRO 6. PORCENTAJE DE RESPUESTAS CORRECTAS DE LOS ALUMNOS DE 6º GRADO EN LENGUA Y MATEMÁTICA POR CONDICIÓN SOCIO-ECONÓMICA. HONDURAS, 2004

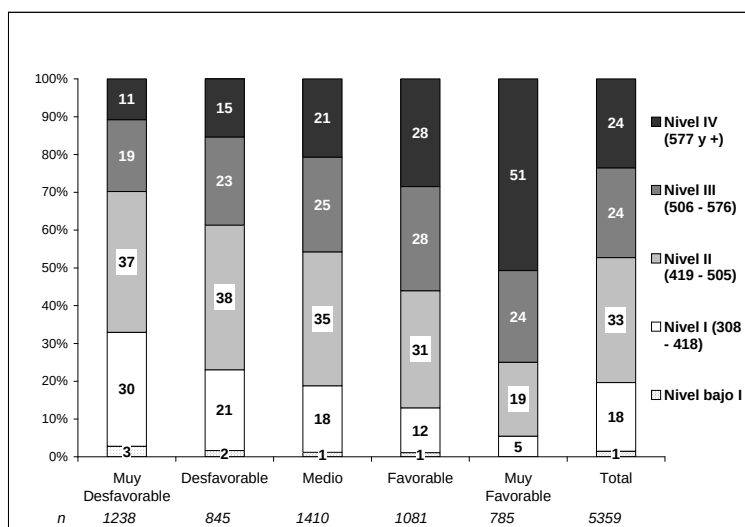
	Lengua	Matemática
Desfavorable	38,8	35,1
Media	42,7	38,6
Favorable	48,7	42,6

Fuente: Rápalo (2005).

En Uruguay, los análisis de la relación entre el nivel sociocultural de las familias y los resultados de los estudiantes, muestran diferencias muy importantes en el logro de los estudiantes según el contexto en el

que se encuentran. Así, por ejemplo, mientras que el 51% de los alumnos de 6° grado de nivel socio-cultural Muy Favorable se encuentran en el nivel de desempeño más alto (IV) en Matemática, sólo lo está el 11% de los estudiantes de contexto socio-económico muy desfavorable (grafico 2). Análogos datos se encuentran en Lengua.

GRÁFICO 2. DISTRIBUCIÓN DE LOS ALUMNOS POR NIVELES DE DESEMPEÑO EN MATEMÁTICA SEGÚN CONTEXTO SOCIO CULTURAL. URUGUAY, 6° GRADO. DATOS DE 2005



Fuente: ANEP (2007)

Nicaragua, por último, analiza las diferencias en el rendimiento de los alumnos según el nivel máximo de estudios finalizado de los padres. Los resultados constatan una estrecha relación entre ambas variables, mostrando mejores resultados escolares a medida que aumenta la escolaridad de los padres. En efecto, menos de un tercio de los alumnos cuyos padres poseen sólo estudios primarios, se ubica en un nivel de rendimiento intermedio o proficiente en Lengua (27,8%). Porcentaje que asciende al 51,1%, cuando los padres son universitarios. En Matemática, la tendencia es la misma: del 10,2% para padres con estudios primarios, al 28,8% para padres con estudios universitarios (cuadro 7).

CUADRO 7. PORCENTAJE DE ESTUDIANTES EN CADA NIVEL DE DESEMPEÑO EN FUNCIÓN DEL NIVEL EDUCATIVO DE LOS PADRES EN 6° GRADO EN LENGUA Y MATEMÁTICA POR NIVEL EDUCATIVO DE LOS PADRES. NICARAGUA, 2002

	Lengua			Matemática		
	Básico	Intermedio	Proficiente	Básico	Intermedio	Proficiente
Educación de Adultos	71,9	24,7	3,4	88,3	10,6	1,1
Primaria	72,1	24,4	3,5	89,8	9,5	0,7
Secundaria	64,8	28,9	6,3	87,9	11,3	0,8
Universidad	49,9	37,3	12,8	81,2	16,6	2,1
Postgrado	47,4	31,6	21,1	70,3	21,6	8,1

Fuente: MECD (2004).

Estos resultados están ratificando lo lejos que aún estamos de hacer realidad la promesa de ampliar las oportunidades educativas de los niños y jóvenes de los sectores más pobres. La calidad de los aprendizajes sigue siendo muy desigual entre quienes provienen de familias de mayores ingresos y quienes forman parte de los segmentos sociales más pobres y vulnerables de nuestras sociedades. Lo

anteriormente se ve gravemente reforzado con ciertas prácticas, en algunos países aceptadas e institucionalizadas, que fomentan aún más la segregación de los sistemas educativos. Entre ellas, los procesos de selección impuestos por los diferentes establecimientos educativos, que cierran la posibilidad de educación a quienes no cumplen con los requisitos previamente definidos por ellos. La mayor parte de las veces, la selección busca quedarse con los mejores estudiantes que resultan ser quienes provienen de segmentos sociales con mayores recursos económicos, sociales y culturales.

La política educativa sigue tensionada por la urgente necesidad de poder constituir espacios y contextos de aprendizaje de alta exigencia académica y socio afectiva para que todos los estudiantes de segmentos socioeconómicos más vulnerables aprendan más y mejor, abriendo así oportunidades reales para su inserción en niveles de estudio medio y superior, de manera de hacer realidad las expectativas que hombres y mujeres cifran en la educación pública de la región.

2.5. Acceso al conocimiento de las minorías étnicas

A pesar del peso e importancia de la población indígena en la región, sólo Guatemala y México informan de la incidencia de la variable grupo étnico, sobre el rendimiento de los alumnos en sus reportes de las evaluaciones nacionales realizadas.

Guatemala, en su evaluación de 2005, analiza las diferencias en el logro del criterio por parte de los estudiantes en función de su origen étnico. A pesar de que el número de estudiantes de algunos grupos étnicos fue muy limitado, los datos muestran grandes diferencias en el logro en ambas materias y en ambos cursos, encontrándose siempre que los alumnos de origen Ladino (mestizos de mayas y europeos que sólo hablan español) obtienen mejores resultados, mientras que los de origen maya obtienen los peores. Sin embargo, en estos datos no se ha controlado por zona geográfica, hábitat o nivel socio-económico, por lo que es necesario tener precaución en su interpretación (cuadro 8).

CUADRO 8. PORCENTAJE DE ESTUDIANTES DE 6º GRADO DE PRIMARIA Y 3º DE SECUNDARIA EN ESCUELAS OFICIALES QUE EL ALCANZAN EL LOGRO EN LECTURA Y MATEMÁTICA SEGÚN ORIGEN ÉTNICO. GUATEMALA, 2005

	6º de Primaria		3º de Secundaria	
	Lengua	Matemática	Lengua	Matemática
Ladino	59,1	61,8	64,2	32,6
Garífuna	53,1	58,9	37,5	25,2
Indígena	42,5	53,1	36,4	32,6
Xinka	39,9	50,7	36,2	35,9
Maya	25,5	42,5	29,6	46,4
Ninguna	54,3	60,1	52,7	40,3

Fuente: CIE/UVG/PRONERE/MINEDUC (2006).

En México la información es menos directa. Allí han separado el rendimiento de los alumnos en función del "estrato educativo" donde estudian. Por estrato educativo se entiende el tipo de escuela, con cinco categorías o estratos: escuelas urbanas públicas, escuelas rurales públicas, escuelas privadas, cursos comunitarios y escuelas indígenas. Estas últimas son escuelas públicas que se encuentran ubicadas en comunidades indígenas y donde se imparte el modelo Educación bilingüe bicultural.

Los resultados muestran que los estudiantes que asisten a estas escuelas tienen un rendimiento significativamente inferior a otras en todas las áreas y grados analizados, incluidas las escuelas rurales públicas. Así, mientras que el 43,2% de los alumnos de 6º de Primaria de escuelas indígenas no alcanza el nivel básico en Lengua (español), esta cifra en las escuelas rurales disminuye a un 23,7%, y al 17,4% para la totalidad de las escuelas (cuadro 9). Datos análogos se dan en Matemática.

En todo caso, tal y como se estudió en la evaluación nacional de México, si se descarta la influencia de variables de contexto tal como el nivel cultural de las familias, las diferencias entre escuelas urbanas públicas y rurales públicas se reduce considerablemente, disminuyendo sustancialmente las diferencias respecto de las escuelas indígenas. Así, es posible sostener que los bajos resultados de los niños que asisten a estas escuelas se deben en gran medida al bajo nivel socio-económico y cultural de las familias de sus estudiantes.

CUADRO 9. PUNTUACIÓN PROMEDIO DE LOS ESTUDIANTES EN LENGUA Y MATEMÁTICA EN FUNCIÓN DEL ESTRATO ESCOLAR. MÉXICO, 2005

	Lengua	Matemática
Escuelas urbanas públicas	541,1	534,2
Escuelas rurales públicas	503,1	494,0
Escuelas indígenas	480,0	468,0
Cursos comunitarios	471,9	470,4
Escuelas privadas	628,4	597,3

Fuente: Backchoff y otros (2006).

La información y resultados comentados, no hacen más que ratificar la gravedad de la situación. Con mayor o menor dramatismo los pueblos indígenas de la región se encuentran en una marcada situación de pobreza que encuentra parte de sus raíces en el ámbito de la educación. Así, a las altas tasas de analfabetismo entre las etnias y comunidades indígenas, se agrega una educación -la mayor parte de las veces- ciega de la diversidad cultural y su riqueza y, al mismo tiempo ignorante de acciones y políticas que permitan desarrollar estrategias pedagógicas, pertinentes y relevantes a dicha diversidad.

En este contexto el sistema educativo y en particular la escuela enfrentan un enorme desafío y reto: generar espacios y condiciones para asumir e incluir "al otro", como legítimo e igual, respetando y valorando los universos simbólicos y por tanto los paradigmas que los sostienen.

Enorme tarea en sociedades que siguen mirando a los grupos étnicos como inferiores y que cada vez son más competitivas, individualistas y segmentadas.

2.6. Ubicación rural o urbana de la escuela

Son bastantes los países que ofrecen datos acerca de la diferencia del rendimiento en función de que la escuela se halle en zona rural y urbana (cuadro 10). Es necesario destacar que es preciso ser muy cautos con estos datos, en primer lugar porque la definición de zona rural y urbana varía de forma considerable entre países, y, más importante, porque detrás de estas cifras se oculta la incidencia del nivel socio-económico y cultural de las familias de los alumnos, habitualmente más bajos en las zonas rurales. Tal como fue anteriormente señalado, México ha estudiado este hecho y en su informe de resultados del año 2006 se afirma que las diferencias en el rendimiento entre escuelas urbanas y rurales desaparecen por completo si se considera las variables del contexto tales como el nivel socio-económico.

En todo caso, los datos apuntan a que los alumnos que asisten a escuelas situadas en zona rural obtienen peores resultados que los de las zonas urbanas. Así lo demuestran los datos de Colombia, donde el puntaje promedio es de 3 puntos menos para los estudiantes que asisten a la zona rural frente a los que lo hacen en escuelas urbanas, en los dos grados y materias analizados. También lo muestran, aunque en menor medida, los resultados de Paraguay, donde los alumnos de zonas urbanas obtienen un 46,5% de

respuestas correctas en Matemática y de 53,0% en Lengua (Comunicación), mientras que esas cifras son de 45,2% y de 49,6% para alumnos que estudian en escuelas situadas en zonas rurales.

CUADRO 10. DIFERENCIAS EN EL RENDIMIENTO ENTRE ALUMNOS ESCOLARIZADOS EN ZONAS URBANA Y RURAL EN LENGUA Y MATEMÁTICA, EN 6° DE PRIMARIA Y 3° DE SECUNDARIA

	Medido como		Lengua		Matemática	
			Urbana	Rural	Urbana	Rural
Argentina	% de respuestas correctas	6 P	51,5-64,0 ¹	51,4	54,1-64,9 ¹	54,0
Colombia	Puntaje promedio	5 P	59,7	57,0	56,8	55,8
		3 S	64,5	62,0	60,7	58,6
El Salvador	% alumnos en el nivel medio o superior	6 P	78,9	67,8	61,4	47,7
		3 S	75,1	63,1	66,1	43,3
Guatemala	% alumnos alcanzaron el dominio	6 P	62,8	41,4	65,7	50,8
Honduras	% alumnos en el nivel de suficiencia	6 P	18,0	5,0	10,0	5,0
Paraguay	% promedio de respuestas correctas	6 P	53,0	49,6	46,5	45,2
Perú	% alumnos en el nivel suficiente	6 P	11,0	2,3	5,8	1,3
		3 S	11,1	3,3	2,4	0,2

(1) Argentina reporta dos informaciones sobre las escuelas urbanas, las urbanas estatales y las urbanas privadas: estatales-privadas.

Fuente: Elaboración propia a partir de los diferentes informes de evaluación.

Interesante es el análisis de Argentina. En sus reportes de la evaluación nacional de 2003, y sólo para 6° grado de Educación General Básica, se dan resultados de los alumnos que estudian en escuelas rurales, de los que estudian en escuelas urbanas estatales, y los que lo hacen en rurales estatales. Así se observa que el rendimiento de los alumnos que estudian en escuelas urbanas públicas es el mismo que en los que lo hacen en las escuelas rurales estatales, las diferencias por tanto, se dan entre escuelas estatales y privadas, no en el hábitat. Ello tanto en Lengua como en Matemática.

Nicaragua, por su parte, ofrece información desagregada no por ámbito de residencia, sino también por gestión, de esta forma ofrece los resultados no solo de escuelas urbanas y rurales, sino también en si es privada subvencionada o no, pública autónoma o no, y rural multigrado o regular, de esta forma ofrece el rendimiento desagregado en nueve categorías. En todo caso, los alumnos de escuelas rurales, tiene peores resultados y, dentro de estos, peores los que estudian en escuelas multigrado.

2.7. Unidad geográfico-administrativa

Una última mirada puede ser hecha estudiando las variaciones que se obtienen en el rendimiento de los alumnos según estudien en escuelas situadas en una unidad geográfico-administrativa menor, tal como Estado, Región o Distrito.

La mayoría de los países realiza este tipo de análisis, y de forma constante aparecen diferencias muy importantes entre regiones; y estas diferencias se mantienen y en el mismo sentido, tanto en los diferentes cursos como en las distintas materias evaluadas. Así lo indican los datos de:

- Argentina, donde las diferencias en el porcentaje de respuestas correctas varía un entre un 8 y 10 puntos porcentuales entre regiones. Por ejemplo, en 9° de EGB en Lengua, el porcentaje se encuentran entre el 46,1% y el 56,0%.
- Brasil, donde la diferencia entre Regiones es muy importante entre 47,5% y 55,7% según la Región.

- Chile; con diferencias de 12 y 10 puntos para Lengua y Matemática, entre las Regiones cuyos estudiantes de 4° de Primaria alcanzan los mejores y más bajos desempeño. La Región Metropolitana obtiene los mejores promedios (significativamente superior al nacional en las tres áreas evaluadas), mientras que las regiones III y IX, obtienen los más bajos. Cabe señalar que las regiones III y IX son de las más pobres del país. Estas importantes diferencias entre regiones, se repite para el logro de los alumnos de 2° de Secundaria inferior (cuadro 11).

CUADRO 11. PUNTUACIÓN PROMEDIO EN LENGUA Y EN MATEMÁTICA DE ESTUDIANTES DE 4° DE PRIMARIA Y 2° DE SECUNDARIA EN CHILE POR REGIÓN

Región	4° de Primaria 2005		2° de Secundaria 2004	
	Lengua	Matemáticas	Lengua	Matemáticas
I Tarapacá	249	245	245	249
II Antofagasta	251	243	249	253
III Atacama	249	239	248	251
IV Coquimbo	254	246	250	250
V Valparaíso	254	247	252	252
VI Libertador	254	245	250	252
VII Maule	253	245	248	252
VIII Bío-Bío	253	245	250	252
IX Araucanía	248	236	244	244
X Los Lagos	254	245	250	249
XI Aysén	258	246	254	257
XII Magallanes	257	252	256	255
XIII Metropolitana (1)	260	253	255	257

(1) En esta región se ubica la Capital de Chile

Fuente: SIMCE 2004-2005.

- Colombia, donde pueden encontrarse importantes diferencias en los resultados de los alumnos por departamentos en todas las materias y grados estudiados. Así, en *Bogotá*, en 6° grado y en Matemática el puntaje promedio es de 70,0; mientras que en *Guaviare* esta puntuación es de 54,8.
- Guatemala, donde la diferencia entre zonas es también muy importante pero varía por cursos y materias. Así, por ejemplo el porcentaje de estudiantes que alcanza el logro en 6° grado en Lectura asciende al 74,8% en *Ciudad Capital*. El departamento con mejores resultados, y de apenas un 25% en *Alta Verapaz*. En Matemática, esas diferencias son menores.
- Honduras, con diferencias muy importantes entre Departamentos; de tal forma que mientras que en *Francisco Morazán* el porcentaje medio de respuestas correctas en Lengua (Español) y en 6° grado es del 48,2%, en *Gracias a Dios* esta cifra es del 30,4%. Análogas diferencias se encuentran en Matemática y en Ciencias Naturales, y con pocas diferencias en el orden de los Departamentos.
- En Perú esas diferencias son incluso mayores. Allí se analizaron las diferencias entre los 25 departamentos, y los datos son reveladores. En 6° y para Matemática, en *Tacna*, el distrito con mejores resultados, el 15,1% de los alumnos alcanza el nivel de suficiencia, mientras que en *Apurímac* sólo lo alcanzan el 0,7% de los estudiantes.
- En los reportes de evaluación de México se ofrecen resultados por entidades federativas (Estados y el Distrito Federal). Como puede observarse en el cuadro 12, estas cifras arrojan diferencias muy importantes según la entidad federativa, y éstos muestran diferencias importantes. Por

ejemplo, en 6° de Primaria y Lengua, mientras que apenas el 6,5% de alumnos que estudian en alguna de las escuelas del Distrito Federal no alcanza el nivel básico, esta cifra asciende al 31,2% de lo que lo hacen en *Chiapas*. Análogos resultados se obtienen en Matemáticas y para 3° de Secundaria.

Se puede afirmar así, que el nivel económico y de desarrollo de la región, distrito o estados, está directamente ligado con los resultados académicos de los alumnos, de tal forma que las zonas con un mayor índice de desarrollo son aquellas cuyos alumnos obtienen mejores resultados, y viceversa. Como muestra de ello es que en todos los países analizados, los alumnos que asisten a escuelas situadas en la ciudad capital del país suelen obtener mejores resultados que los que asisten a escuelas localizadas en cualquier otro lugar.

CUADRO 12. PORCENTAJE DE ESTUDIANTES DE 6° DE PRIMARIA Y 3° DE SECUNDARIA EN MÉXICO POR DEBAJO DEL NIVEL DE LOGRO BÁSICO EN LENGUA Y EN MATEMÁTICA, POR ENTIDAD FEDERATIVA

	6° de Primaria		3° de Secundaria	
	Lengua	Matemáticas	Lengua	Matemáticas
Aguascalientes	15,2	13,5	33,1	45,5
Baja California	12,5	13,1	27,1	49,6
Baja California Sur	11,1	12,5	30,0	50,7
Campeche	21,5	19,8	33,7	51,7
Coahuila	13,5	12,9	33,6	55,4
Colima	17,7	16,7	28,6	49,7
Chiapas	31,2	29,6	51,2	62,8
Chihuahua	15,5	14,4	29,0	47,9
Distrito Federal	6,5	8,7	17,9	39,8
Durango	16,7	14,8	35,0	52,6
Guanajuato	20,4	20,5	33,1	48,1
Guerrero	28,0	27,4	44,4	61,6
Hidalgo	20,6	18,2	34,6	49,6
Jalisco	13,5	13,2	29,5	47,1
México	15,7	15,1	25,4	46,1
Michoacan	27,4	28,2	43,5	61,4
Morelos	13,5	13,4	31,2	50,1
Nayarit	18,6	17,3	37,9	54,5
Nuevo León	14,9	15,4	32,7	52,8
Oaxaca	23,0	20,4	46,1	59,1
Puebla	18,9	18,4	35,7	51,7
Querétaro	11,9	10,5	29,3	46,6
Quintana Roo	16,1	16,4	31,3	52,6
San Luis Potosí	21,5	18,8	35,3	53,1
Sinaloa	12,9	12,2	28,8	45,4
Sonora	14,7	14,2	33,4	56,2
Tabasco	22,8	22,9	40,4	61,6
Tamaulipas	14,5	16,5	30,6	51,5
Tlaxcala	17,0	16,7	31,4	50,2
Veracruz	19,4	17,0	34,0	52,8
Yucatán	19,5	18,2	32,2	53,6
Zacatecas	16,9	14,0	42,5	55,8

Fuente: Backhoff y otros (2006).

3. EVOLUCIÓN EN LOS RESULTADOS DE APRENDIZAJE

El estudio de la evolución del rendimiento de los alumnos ofrece una imagen complementaria a la anterior y del todo necesaria. Dicha mirada, indica si la educación mejora o empeora, a partir del compromiso colectivo que asumieron los gobiernos en la Conferencia de Dakar el año 2000, de actuar y cautelar para que se alcancen los objetivos y metas de la Educación para Todos, con resultados duraderos en la región. Junto con ello, analizar lo que ha pasado con los aprendizajes de los niños y jóvenes durante esta última década, permite dimensionar cambios o estancamiento en sistemas escolares que han visto incrementado sus recursos e insumos a través de los importantes procesos de reformas que, en muchos de los países, se iniciaron a principio de los noventa y que han modificado significativamente no sólo qué se debe enseñar sino cómo hacerlo. Así, es del todo esperable que estas ya largas 'revoluciones' acontecidas en el campo educativo, hagan sentir sus efectos en una de las más importantes indicadores de calidad educativa, como son la pertinencia, relevancia y estabilidad de los resultados escolares, expresados principalmente en el logro cognitivo en las áreas fundamentales del saber: Lengua y Matemática. Aún así, es importante señalar que este tipo de mirada permite responder si los logros escolares son en la actualidad mejor que antes, pero no necesariamente si son todo lo bueno que necesitamos que sean.

Para realizar dicho análisis, en este estudio se han considerado las informaciones de evolución de los desempeños reportadas por los propios países. Por lo anteriormente dicho, se ha intentado analizar la evolución del rendimiento entre los años 1999 ó 2000 y los datos más recientes disponibles. Por último, se ha buscado centrar el análisis en las mismas variables que en la imagen sincrónica antes presentada. Reportan esa información los informes de evaluaciones nacionales de Brasil, Chile, Colombia, Costa Rica, El Salvador, Honduras, México y Perú (cuadro 13).

CUADRO 13. RESUMEN DE LA EVOLUCIÓN DEL RENDIMIENTO DE LOS ESTUDIANTES EN ALGUNOS PAÍSES DE AMÉRICA LATINA

	Medido como	Años		Lengua	Matemática
Brasil	Puntuaciones promedio	1999-2005	4 P 2 S	170,7-172,3 (dns) 232,0-231,9 (dns)	181,0-182,4 (dns) 246,4-239,5 (ds)
Chile	Puntuaciones promedio	2002-2005	4 P	251-255 (dns)	247-248 (dns)
Colombia	Puntuaciones promedio	2003-2005	5 P 3 S	58,2-60,1 60,6-65,3	52,8-57,7 57,2-61,4
Costa Rica	% de estudiantes con nota de examen mayor o igual a 65	1999-2005	4-6 P 1-3 S	91,7-89,4 84,1-71,5	60,3-52,4 23,0-22,5
El Salvador	% alumnos en el nivel medio o superior	2003-2005	6 P 3 S	59-73 62-70	48-55 43-51
Honduras	% de alumnos en nivel de suficiencia	1997-2004	6 P	16-10	1-7
México	Puntuaciones promedio	2000-2005	6 P 3 S	503,9-530,5 (ds) 498,1-505,2 (ds)	503,8-521,6 (ds) 500,1-496,9 (dns)
Perú	Puntuaciones promedio	1998-2004	6 P	308,5-303,2 (dns)	306,3-311,9 (dns)

(ds): Diferencias significativas

(dns): Diferencias no significativas

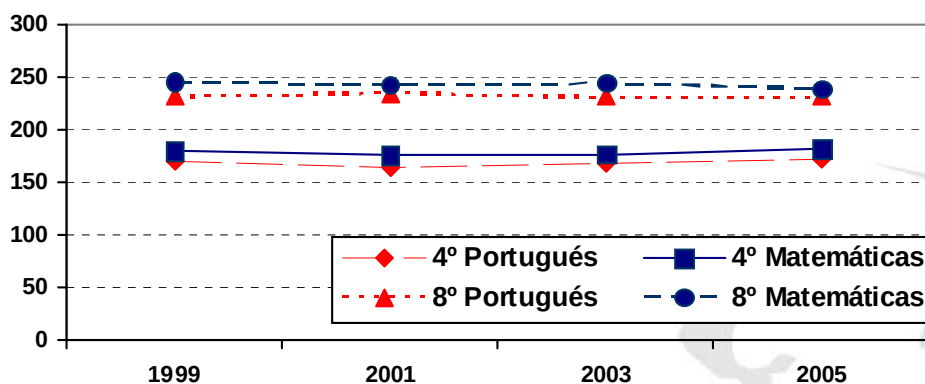
Fuente: Elaboración propia a partir de los informes nacionales de evaluación.

Para tener una imagen fiable de la situación es necesario tener pruebas diseñadas específicamente para ser comparadas en el tiempo, y aunque ese es un objetivo de todos los sistemas nacionales de evaluación, no siempre se consigue. Por ello, la información que sobre este tema puede aportarse es escasa, y ha de tratarse con precaución. Chile, por ejemplo, a pesar que desarrolla evaluaciones nacionales censales desde principios de los noventa, sólo dispone de pruebas comparables a partir de las mediciones del año 1999 para 4º de Primaria y del 2000 para el 2º grado de Secundaria Inferior (8º año de Educación Básica).

De forma general se puede afirmar que el desempeño educativo de la región está estancado, con mínimas variaciones en el tiempo que no parecen seguir una clara tendencia. Este estancamiento puede, sin embargo, esconder un aumento en el rendimiento en estos años moderado por el incremento en el número de estudiantes matriculados. Y, desde esa perspectiva, entenderse como un dato positivo. En efecto, durante los años 90, los sistemas educativos de la mayoría de los países, incrementaron de manera importante su cobertura, expandiendo especialmente la matrícula de primaria, pero también en la educación secundaria.

Como ejemplo sirve los resultados de Brasil, donde los datos de rendimiento de los alumnos no presentan claras evoluciones en estos años. Así, bajan y suben ligeramente, sin poder reflejar claras tendencias ni de mejora ni de retroceso para ninguna de las dos áreas y grados medidos (gráfico 3).

GRÁFICO 3. EVOLUCIÓN DE LOS PROMEDIO DE RENDIMIENTO PARA PORTUGUÉS Y MATEMÁTICA, PARA 4 DE PRIMARIA Y 2º DE SECUNDARIA INFERIOR (4ª Y 8º DE ENSINO FUNDAMENTAL). BRASIL. DE 1999 A 2005.



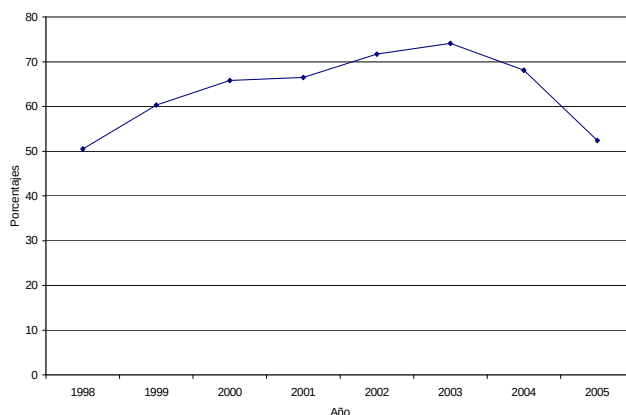
Fuente: Ministério Da Educação/INEP (2007).

En El Salvador se observa una mínima mejoría en los resultados de las evaluaciones de los estudiantes de 6º grado entre 2002 y 2004. Así, mientras que en 2002 el porcentaje promedio de respuestas correctas era de 43,5 para Lengua y de 38,8 para Matemática; dos años después es de 43,9 y 39,1 respectivamente; es decir un mínimo incremento de 0,4 puntos para Lengua y de 0,3 para Matemática, en el reporte no se informa si estas diferencias son estadísticamente significativas.

Perú, por su parte, en el análisis de la evolución de los resultados de la prueba CRECER entre 1998 y 2004, reporta una mínima variación que ha resultado ser estadísticamente no significativa.

El Caso de Costa Rica es especialmente interesante, dado que ha realizado un estudio de la evolución del rendimiento de los estudiantes, tanto en II ciclo como en III ciclo, año a año desde 1998 y hasta el 2005. En el gráfico 3 se han reflejado estos datos y no se observa ninguna tendencia clara. Así, aunque comparando 1999 con 2005 los resultados han disminuido ligeramente (gráfico 4), la tendencia no es claramente decreciente.

GRÁFICO 4. EVOLUCIÓN DEL PORCENTAJE DE ESTUDIANTES CON NOTA DE EXAMEN EN MATEMÁTICA MAYOR O IGUAL A 65,0. COSTA RICA II CICLO, 1998-2005.

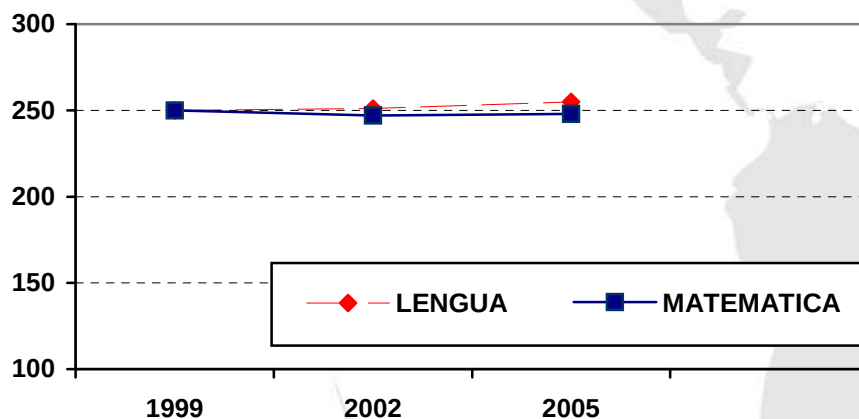


Fuente: Ministerio de Educación Pública de Costa Rica (2005a).

Chile estudia la evolución del rendimiento de los alumnos en Lenguaje y Comunicación, Educación Matemática y Comprensión del Medio entre 1999 y el 2005 en 4° de Primaria y entre el 2000 y 2004 para 2° de Secundaria. Para ambos grados, encuentra que no ha habido variaciones estadísticamente significativas en Lengua y en Matemática, sólo un pequeño avance en Comprensión del Medio. La comparación realizada en 4° de Primaria, muestra una tendencia positiva para Lengua y Comprensión del Medio Social y Natural, y negativa para Matemática (gráfico 5).

Es necesario señalar que la comparación para 2° de secundaria, se realiza entre grupos de alumnos que estudiaron con diferentes currículos durante la Enseñanza Básica. Por tal motivo, es necesario tener cautela al comparar los puntajes obtenidos en ambas evaluaciones. Esto no ocurre en el caso de 4° de Primaria para las 3 mediciones consideradas (1999, 2002 y 2005).

GRÁFICO 5 EVOLUCIÓN DEL RENDIMIENTO EN LENGUA Y MATEMÁTICA, PARA 4° DE PRIMARIA. CHILE. 1999 A 2005



Fuente: Elaboración propia a partir de los Reportes Nacionales SIMCE 1999, 2002, 2005.

Colombia, por último, informa de pequeños avances entre 2003 y 2005 tanto en 5° de como en 9° y en las tres áreas analizadas: Lengua, Matemática y Ciencias Naturales.

3.1. Género

Pocos países reportan datos de la evolución de las diferencias por **género**. Honduras y México sí lo hacen. Los resultados de Honduras, en forma de porcentaje medio de respuestas correctas y para 6° grado, indican que las diferencias se incrementan en Lengua y se mantienen igual para Matemática. Efectivamente mientras que en 2002, las mujeres superaban en 1,1 puntos porcentuales a sus compañeros varones en Español, en el 2004 esa diferencia se ha elevado hasta los 2 puntos. En Matemática, sin embargo, se mantiene la diferencia de 9 décimas superior para los chicos.

México, informa que hay diferencias significativas en el puntaje promedio de los estudiantes en la comparación entre el año 2000 y 2005, pero éstas se producen en ambos géneros y áreas de conocimiento, pero sólo en 6° de primaria. En esos casos parece que la diferencia de rendimiento entre géneros se ha ampliado, de tal forma que cada vez las mujeres sacan mejores resultados que sus compañeros en Lengua (Comprensión Lectora) y peores en Matemática (cuadro 14).

CUADRO 14. EVOLUCIÓN EN EL PUNTAJE PROMEDIO OBTENIDO POR LOS ESTUDIANTES DE 6° DE PRIMARIA Y 3° DE SECUNDARIA GRADO EN LENGUA Y MATEMÁTICA. MÉXICO. AÑOS 2000-2005

		Varones	Mujeres
Lengua	6	494,4 – 521,0	511,5 – 539,6 *
	3	*	500,6 – 507,5
		495,5 – 502,4	
Matemática	6	506,1 - 526,2	501,5 - 517,2 *
	3	*	493,5 - 491,3
		506,8 - 502,6	

(*) Diferencias significativas

Fuente: Backchoff y otros. (2006).

3.2. Evolución según contexto rural-urbano

También Honduras y México han estudiado la evolución en las diferencias de rendimiento con la variable **rural-urbano**. Respecto a Honduras, las diferencias parecen ampliarse, dado que los 4,8 puntos porcentuales diferencia en el porcentaje promedio de respuestas en 6° y Lengua que se daba en 2002, se ha ampliado a 6,3 en 2004. Igual conclusión se obtiene del análisis de Matemática; y en ambos casos por la misma razón: mejoraron los resultados en la zona urbana y bajaron los de la rural.

México reporta la evolución del rendimiento entre el año 2000 y 2005 de las escuelas urbanas públicas y las escuelas rurales públicas en 6° de primaria y tanto para Lengua como para Matemática, y los resultados apuntan a que las diferencias parecen haberse incrementado en estos años, dado que tanto en Lengua como en Matemática, aunque aumenta el rendimiento tanto en las escuelas urbanas públicas como en las rurales públicas, estas últimas lo hacen en menor grado (en Matemática, incluso el incremento es no significativo) (cuadro 15).

CUADRO 15. EVOLUCIÓN DE LA PUNTUACIÓN PROMEDIO DE LOS ESTUDIANTES DE 6° DE PRIMARIA EN LENGUA Y MATEMÁTICA EN FUNCIÓN DEL ESTRATO ESCOLAR. MÉXICO. AÑOS 2002 Y 2005

	Lengua	Matemática
Escuelas urbanas públicas	510,4 - 541,1 *	513,0 - 534,2 *
Escuelas rurales públicas	480,8 - 503,1 *	484,7 - 494,0
Escuelas indígenas	443,7 - 480,0 *	441,4 - 468,0 *
Cursos comunitarios	467,7 - 471,9	466,7 - 470,4
Escuelas privadas	572,7 - 628,4 *	547,3 - 597,3 *

(*) Diferencias significativas

Fuente: Backchoff y otros (2006).

De esta forma, los pocos datos disponibles indican que la diferencia entre rural y urbano no sólo no ha disminuido en estos años de estudio, sino que parece incrementarse.

3.3. Nivel socio-económico

El análisis de la evolución del rendimiento según el diferente nivel socio-económico de las familias no ha sido reportado por casi ningún país. Excepciones son Chile y Honduras.

Las evaluaciones realizadas por el SIMCE Chileno en los años 2002 y 2005 muestran que en Lenguaje y Comunicación los grupos Bajo y Medio Bajo aumentaron significativamente su promedio, mientras que los otros tres grupos se mantuvieron sin variación en esta área. En Matemática no se observan diferencias significativas en ningún grupo socioeconómico. Así, es posible hablar de una disminución parcial de la brecha sólo para Lengua (cuadro 16).

CUADRO 16. EVOLUCIÓN DEL PUNTAJE PROMEDIO DE LOS ALUMNOS DE 4° DE PRIMARIA EN LENGUAJE Y MATEMÁTICA POR GRUPO SOCIO-ECONÓMICO. CHILE. AÑOS 2002 Y 2005

Grupo socio-económico	Lenguaje	Matemática
Bajo	224 - 231 *	218 - 220
Medio bajo	230 - 236 *	228 - 228
Medio	251 - 253	247 - 245
Medio alto	278 - 278	273 - 271
Alto	299 - 300	300 - 297

(*) Diferencias significativas

Fuente: SIMCE (2005).

En Honduras se analizó la evolución desde el 2002 hasta el 2004 en el porcentaje medio de respuestas correctas por condición socio-económica en 6° de Primaria, tanto en Lengua como en Matemática. Los resultados muestran que la situación permanece constante (cuadro 17).

CUADRO 17. EVOLUCIÓN EN EL PORCENTAJE MEDIO DE RESPUESTAS CORRECTAS DADAS POR LOS ESTUDIANTES DE 6° DE PRIMARIA EN LENGUA Y MATEMÁTICA EN FUNCIÓN DE SU CONDICIÓN SOCIO-ECONÓMICA. HONDURAS. AÑOS 2002-2004

	Español	Matemática
Desfavorable	37,6 – 38,8	35,4 – 35,1
Medio	42,6 – 42,7	38,3 – 38,6
Favorable	49,8 – 49,7	42,5 – 42,6

Fuente: Rápalo (2005).

A pesar de la poca información disponible, es posible sostener que la brecha entre los segmentos más pobres y con mayores recursos, se mantiene sin mayores variaciones. Los sistemas educativos no logran así, revertir la inequidad social que los atraviesan.

3.4. Evolución del rendimiento según Unidad Política Administrativa

La evaluación del rendimiento según el área administrativa política ha sido reportada por Brasil, Chile, Honduras y Perú.

- En Brasil se estudiaron las diferencias en el rendimiento promedio de los estudiantes entre 1999 y 2005 entre las regiones. La información obtenida señala que muy pocas de las diferencias encontradas son significativas.
- En Chile, para 4° grado se encuentra que sólo 4 de las 13 regiones aumentaron significativamente su promedio en Lenguaje y Comunicación, entre ellas la Metropolitana; el resto se mantuvo igual. En Matemática ninguna región aumento significativamente su promedio y una de ellas lo disminuyó significativamente (XII Región). Con esos datos, la brecha entre regiones, lejos de aminorarse, parece agrandarse.
- Los datos Hondureños muestran pequeñas variaciones en el rendimiento de los estudiantes de distintas regiones entre 2002 y 2004, pero no se observa tendencia clara en ningún sentido.
- Perú comparó los resultados obtenidos por los alumnos de 6° de primaria en 1998 y en 2004 en Lima y en Provincias, y en ambos casos las diferencias encontradas fueron no significativas, tanto en Lengua (Comunicación Integral) como en Matemática (Lógico Matemática).

En coherencia con lo anterior, puede afirmarse que los resultados de aprendizaje de de los estudiantes apenas han mostrado grandes diferencias entre regiones entre los años estudiados.

4. CONCLUSIONES Y DISCUSIÓN ACERCA DE LAS EXPLICACIONES TENTATIVAS

Esta breve revisión de los resultados obtenidos de las evaluaciones nacionales del rendimiento de los alumnos en los países Latinoamericanos ha aportado algunas ideas clave que es preciso recuperar en este momento.

En primer lugar se verifica que, a pesar de las importantes diferencias encontradas entre países, los resultados de aprendizaje de los estudiantes de Educación Primaria y Educación Secundaria Inferior de América Latina son globalmente poco satisfactorios. Diferentes evaluaciones nacionales con base criterial han mostrado que una gran parte de los estudiantes de América Latina no alcanza el nivel de desempeño mínimo determinado para su grado. Estos resultados son coincidentes con, por ejemplo, lo defendido por la Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe (2001), quien lo señalaba como unos de los puntos críticos para el desarrollo de la región.

Estos resultados son aún más pobres en Secundaria Inferior que al concluir Primaria, y peores en Matemática que en Lenguaje. Y, lo que parece más preocupante, en el análisis de la evolución del rendimiento desde finales de la pasada década de los 90 y hasta la actualidad no se observa mejoría alguna con el tiempo. De esta forma, no sólo se está aún muy lejos de lograr los aprendizajes necesarios para la inclusión y movilidad social de los niños y jóvenes latinoamericanos, sino que los avances son muy escasos y sin tendencias claras. La búsqueda de calidad y equidad educativa, requiere de profundas transformaciones en los sistemas a corto plazo, de lo contrario la situación no cambiará de forma determinante en el futuro. No debemos desconocer sin embargo, que en estos años se ha incrementado significativamente el número de alumnos que acceden a la escuela, por lo que ese estancamiento del rendimiento esconde otros avances y mejoras en los sistemas educativos que deben ser reconocidas y valoradas. Los sistemas escolares han dado pasos relevantes en permitir el acceso a la educación, pero no así, en ofrecer una educación de igual calidad para todos.

Coincidente con lo encontrado en otras regiones del mundo, el análisis también ha mostrado algunas diferencias en el rendimiento en función del grupo de pertenencia: las mujeres rinden más en Lengua, y no se observa una clara tendencia en Matemática; los estudiantes que asisten a escuelas situadas en zona rurales obtienen peores resultados que los que lo hacen en escuelas urbanas; que existen fuertes diferencias en el rendimiento en función de la zona geográfica/unidad administrativa del país donde residan; y, lo que parece clave en explicar muchas de las diferencias anteriores, cuanto más bajo es el nivel socioeconómico o sociocultural de las familias de los estudiantes, peores son sus resultados de aprendizaje. Los pocos datos que se tienen de las evaluaciones nacionales al respecto, indican que esta situación es estable en el tiempo, lo que habla de una porfiada brecha entre todos los grupos y dimensiones analizadas. Sin embargo, siendo la desigualdad una dura característica que acompaña a todos los sistemas educativos analizados, su magnitud, explicaciones y posibilidades de mejora, no es la misma en todos los países.

El reto consiguiente es aportar explicaciones tentativas de la situación y su evolución que permitan proponer medidas que logren realmente impactar en los resultados. Tarea tan necesaria como difícil de alcanzar. En las líneas que siguen se intentará aportar con algunas ideas y acciones, a modo de insumo del necesario debate social sobre esta temática.

Resulta necesario comenzar afirmando que existe consenso en las diferentes investigaciones que han intentado encontrar los factores asociados al logro escolar y, con ello, aportar explicaciones a los bajos resultados en el rendimiento de los alumnos en América Latina, en otorgar un papel preponderante a la situación socio-económica y cultural de las familias de los alumnos y del entorno donde se desarrollan (p.e. Castro, 1984; Vélez, Schiefelbein y Valenzuela, 1994; Murillo, 2003). Efectivamente, el hogar y el contexto social juegan un papel básico en la adquisición de habilidades relacionadas con el Lenguaje, las Matemática y el resto de las áreas escolares. Más concretamente es el nivel cultural de las familias el factor compuesto que más varianza del rendimiento académico de los alumnos es capaz de explicar; y el nivel máximo de estudios alcanzado por la madre del estudiante, la mejor variable simple. Dicho de otra manera, la principal fuente de desigualdad educativa en la región, es la desigualdad económica y socio-cultural de las familias de origen de los estudiantes. Así, los niveles de pobreza y recursos simbólicos de las familias, son variables de gran incidencia en el desempeño de los estudiantes. Con ello se concluye que la mejor estrategia para mejorar el rendimiento de los alumnos es elevar el nivel económico de la población y especialmente su nivel cultural y de estudios. Junto con ello, se reafirma la necesidad de mantener y mejorar las políticas de focalización en educación en la región, de manera de compensar las diferencias de capital cultural de origen de los estudiantes, posibilitando el acceso a procesos y experiencias de calidad a quienes provienen de las familias más pobres.

Tras la influencia del nivel socio-económico y cultural de las familias, la escuela juega un papel fundamental en el desarrollo de los alumnos. Las investigaciones realizadas apuntan a que en América Latina la incidencia de la escuela es mayor que en los países más desarrollados; que aporta más a los aprendizajes en Matemática que en Lenguaje y, que este aporte es mayor para los estudiantes que cursan Primaria que Secundaria. Así, se ha estimado que la magnitud del efecto escolar está entre 25 y el 30%; es decir, estudiar en una escuela u otra hace que el rendimiento de un alumno puede variar hasta un 30%. Más alto para Matemática que para Lengua, dado que la lengua está más sujeta al contexto sociocultural de los alumnos; y más para Primaria que para Secundaria (Fletcher, 1997; Barbosa y Fernandes, 2001; Lastra, 2001; Benavides, 2002; Ferrão, Beltrão y Fernandes, 2003; Fernández y Blanco, 2004; Ferrão, 2006; Murillo, 2007a). Esto sin duda, pone la escuela como centro de las políticas públicas, las acciones han de detenerse en los procesos de enseñanza aprendizaje, en las dinámicas e interacciones sociales y pedagógicas que allí ocurren, así como en la articulación y diálogo que las escuelas han de establecer con las familias y el entorno.

Dentro de las variables escolares, la investigación sobre factores asociados generada en América Latina señala algunas particularidades respecto de la investigación sobre eficacia educativa realizada en los países más desarrollados (Murillo, 2007b). Así, junto a variables tales como el clima escolar y de aula, la calidad del currículo, la metodología docente, el compromiso y trabajo en equipo de los docentes, la implicación de las familias o el trabajo del equipo directivo, entre otras, aparecen algunos factores de especial relevancia para la región (Murillo, 2003).

En efecto, la práctica totalidad de las investigaciones constatan la importancia de las infraestructuras y los recursos didácticos sobre el rendimiento de los alumnos. Así, lo han encontrado, por ejemplo, los trabajos empíricos desarrollados sobre escuelas de América Latina de Concha (1986), Herrera y Lopez (1996), Piñeros (1996), Cano (1997) Barbosa y Fernández (2001), el LLECE (2001), Bellei et al (2003 y Murillo 2007a). En todos ellos se destaca la importancia de que la escuela cuente con unas instalaciones adecuadas tanto en cantidad como en educación y mantenimiento. Igualmente los estudios han encontrado que la disponibilidad y usos didácticos, tanto tradicionales como relativos a las tecnologías de la información y la comunicación marcan importantes diferencias en el logro académicos de los alumnos. Al respecto, son del todo interesante los hallazgos que señalan que tanto o más importante que disponer de recursos tecnológicos, son la frecuencia y el tipo de uso de ellos los que aseguran mejores y más estables aprendizajes (Román, M. *et al.*, 2002). Resulta muy revelador el estudio de Menezes-Filho y Pazello (2004) en Brasil, donde demuestra la importancia de la asignación de los recursos a las escuelas para tener una educación de calidad. En ese sentido parece necesario recordar que uno de los problemas de América Latina es la gran inequidad en la calidad de las infraestructuras escolares, incluso dentro de las escuelas públicas.

Otro factor que ha demostrado tener una fuerte relación con el logro académico de los alumnos es tanto la formación inicial (p.e. Cano, 1997; Barbosa y Fernández, 2001) como la formación permanente de los docentes (p.e. Herrera y Lopez, 1996; Bellei *et al.*, 2003; Razcynski y González, 2005; Murillo, 2007b). Efectivamente, la existencia de docentes que no poseen una titulación adecuada para el desempeño de sus funciones parece tener una clara repercusión en el desempeño de sus alumnos. Igualmente, si el docente de un alumno se preocupa por su formación permanente y asiste a cursos y otras acciones con regularidad, aquél tiene más probabilidades de tener un buen rendimiento. Lo anterior, resulta del todo indispensable para el manejo y apropiación del currículo, así como para la planificación e implementación de procesos de enseñanza aprendizaje pertinentes y relevantes las características y necesidades del grupo curso (Román, 2004, 2006b).

Muy relacionado con ello están las condiciones laborales del profesorado. Aún una buena cantidad de docentes de América Latina no tiene un salario digno que les permite dedicarse con exclusividad a impartir clase en una única escuela y preparar las clases y ocuparse por su desarrollo profesional. La

investigación ha demostrado que los docentes con doble carga lectiva y/o con una inadecuadas condiciones laborales, tiene dificultades para desarrollar su trabajo adecuadamente, lo que necesariamente incide en la calidad de su docencia y el logro de los estudiantes (LLECE, 2001; Murillo, 2007). La sobre carga laboral, la distribución del tiempo profesional en más de una escuela, así como el desempeño en condiciones poco satisfactorias, afecta dos de los factores que han mostrado ser relevantes para los aprendizajes de los estudiantes: la motivación de los docentes y el tiempo disponible para planificar y trabajar en equipo (Murillo, 2007; Román, 2005, 2006a).

Un elemento más, que afecta a diversos niveles, es la gestión del tiempo. La investigación ha demostrado que la cantidad de tiempo de aprendizaje que realmente recibe el alumno es un factor clave de calidad de la educación (Murillo, 2007; Román 2004). En América Latina, sin embargo, ese tiempo se ve muy limitado por problemas que afectan a los diversos niveles del sistema educativo. Así, es muy importante el tiempo de docencia perdido por conflictos laborales de los docentes (UNESCO/OREALC, 2004); pero también por ausentismo de los docentes; y por la deficiente gestión del tiempo en las aulas. En efecto, investigaciones y evaluaciones desarrolladas en diversos países de la región, han mostrado que el tiempo efectivo destinado para el aprendizaje en el aula, es menos de un tercio que el planificado y asignado para cada sesión (Román, M. *et al.*, 2002; Román, 2006a).

Para terminar, reafirmar que los sistemas escolares y la escuela en especial aportan de manera significativa al aprendizaje de los niños y niñas, pero que para que este aporte sea sistémico y permanente requiere de políticas capaces de reconocer y asumir también la inequidad que se genera y se instala en la escuela y la sala de clases. Para desde allí, desarrollar e instalar un conjunto de medidas que aborden esta inequidad, atendiendo diferenciada, pero articuladamente los factores escolares que aparecen asociadas a ella, así como a la calidad de los aprendizajes y logros escolares.

De esta forma, para mejorar el bajo rendimiento de los alumnos de América Latina, sería necesario: dedicar más recursos a la educación, tanto de forma absoluta como relativa, haciendo una asignación de los recursos diferenciada que garantice una educación de calidad para todos; mejorar la calidad de los docentes con especial énfasis en la formación inicial y permanente de los mismo; mejorar las condiciones salariales y laborales de los docentes; disminuir la conflictividad laboral con los docentes, de tal forma que se disminuya el número de días lectivos perdidos por este motivo. Los programas y políticas del campo educativo han de considerar y asumir la complejidad institucional en sus múltiples dimensiones como un todo: gestión escolar, procesos de enseñanza-aprendizaje, vínculos con la comunidad y la familia, resultados y logros de sus estudiantes, gestión de la convivencia y el clima, uso de recursos, infraestructura y equipamiento, entre otros.

Desde esta mirada, emerge la dimensión pedagógica como eje de las políticas educativas, en tanto componente ordenador a partir del cual habrá de considerarse el conjunto de actores y procesos involucrados, así como el contexto socioeconómico y cultural en que se insertan los establecimientos. Aunque parece claro que es necesario desarrollar una estrategia de desarrollo global, porque la escuela no está sola en el aprendizaje de los alumnos. En efecto, la educación por sí sola no puede resolver la desigualdad de los sistemas educativos. Cambios relevantes en este sentido, requieren de concertación, voluntad y esfuerzos permanentes desde otros campos y políticas sociales para proveer a cada niño/a de las condiciones mínimas para que puedan iniciar su proceso escolar y recorrerle con éxito. Si gobiernos y sociedades no asumen y traducen en metas y acciones de política tal compromiso, las brechas de resultados entre estudiantes de distinto origen social y capitales culturales, no sólo se mantendrán, sino que tenderán a profundizarse.

REFERENCIAS BIBLIOGRÁFICAS

- Barbosa, M.E.F. y Fernandes, C. (2001). A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em Matemática dos alunos da 4ª série. En: Franco, C. (Org.). *Promoção, ciclos e avaliação educacional*. Porto Alegre: ArtMed.
- Bellei, C., Muñoz, G., Pérez, L.M. y Raczynski, D. (2003). *¿Quién dijo que no se puede? Escuelas efectivas en sectores de pobreza*. Santiago de Chile: Ministerio de Educación - Unicef.
- Benavides, M. (2002). Para explicar las diferencias en el rendimiento en matemática de cuarto grado en el Perú urbano: análisis de resultados a partir de un modelo básico. En J. Rodríguez y S. Vargas (Eds.), *Análisis de los resultados y metodología de las pruebas CRECER 1998*. Documento de trabajo N° 13. Lima: MECEP/ Ministerio de Educación del Perú.
- Cano, F. (1997). *Factores asociados al logro cognitivo de los estudiantes. Grados 3º, 5º (1993-1994)*. Bogotá: MEN.
- Castro, C. de M., (1984). *Determinantes de la Educación en América Latina: acceso, desempeño y equidad*. Río de Janeiro: FGV/ECIEL.
- Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe (2001). *Quedándonos atrás. Informe del progreso educativo en América Latina*. Santiago de Chile: PREAL.
- Concha, C. (1996). *Escuelas efectivas en Chile: estudio de 32 escuelas exitosas en logros académicos y de alta vulnerabilidad*. Tesis doctoral inédita, Pontificia Universidad Católica de Chile, Santiago de Chile.
- Fernández, T. y Blanco, E. (2004). ¿Cuánto importa la escuela? El caso de México en el contexto de América Latina. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 2(1).
- Ferrão, M.A. (2006). Componentes do efeito-escola no Brasil. En F.J. Murillo (Coord.), *Estudios sobre Eficacia Escolar en Iberoamérica. 15 buenas investigaciones* (pp. 143-168). Bogotá: Convenio Andrés Bello.
- Ferrão, M.E. e Fernandes, C. (2003). O efeito-escola e a mudança - dá para mudar? *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(1).
- Ferrer, G. (2006). *Sistemas de evaluación de aprendizajes en América Latina. Balance y desafíos*. Santiago de Chile: PREAL.
- Fletcher, P. (1997). *À procura do ensino eficaz*. Relatório de pesquisa, PNUD/MEC/SAEB.
- Herrera, M. y Lopez, M. (1996). *La eficacia escolar*. Caracas: CICE/CINTERPLAN.
- Lastra, E.F. (2001). *La efectividad escolar: un estudio de las escuelas primarias públicas en una ciudad mexicana*. Tesis doctoral inédita, Universidad de Stanford, California.
- LLECE - Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (2001). *Primer estudio internacional comparativo sobre Lenguaje, Matemática y factores asociados, para alumnos del tercer y cuarto grado de la educación básica*. Santiago de Chile: UNESCO.
- LLECE - Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (2002). *Estudio cualitativo de las escuelas con resultados destacables en siete países Latinoamericanos*. Santiago de Chile: UNESCO.

- Medina, A. (2002). *Las reformas educativas en acción: eficiencia, equidad y calidad en el sistema educativo de la República Dominicana y América Latina*. Santo Domingo: INDES/BID/Instituto Tecnológico de Santo Domingo.
- Menezes-Filho, N. y Pazello, E. (2004). *Does Money in Schools Matter? Evaluating the Effects of a Funding Reform on Wages and Test Scores in Brazil*. Santiago de Chile: PREAL.
- Murillo, F.J. (Coord.) (2003a). *La investigación sobre eficacia escolar en Iberoamérica. Revisión Internacional sobre el estado del arte*. Bogotá: Convenio Andrés Bello.
- Murillo, F.J. (Coord.) (2005). *Estudios sobre Eficacia Escolar en Iberoamérica. 15 buenas investigaciones*. Bogotá: Convenio Andrés Bello.
- Murillo, F.J. (2007a). School Effectiveness Research in Latin America. En T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement* (pp. 75–92). New York: Springer.
- Murillo, F.J. (2007b). *Investigación Iberoamericana sobre Eficacia Escolar*. Bogotá: Convenio Andrés Bello.
- Raczynski, D. y Muñoz, G. (2005). *Efectividad escolar y cambio educativo en condiciones de pobreza en Chile*. Santiago de Chile: Ministerio de Educación.
- Román, M. et al. (2002). *Estudio sobre dotación y uso de recursos educativos para estudiantes y docentes de las escuelas rurales*. Santiago de Chile: CIDE- Ministerio de Educación.
- Román M. (2004). Enfrentar el cambio en escuelas críticas urbanas: una mirada a la realidad y posibilidades chilenas. *Revista Persona y Sociedad*, 18(3), pp.145-172. Santiago: Universidad Alberto Hurtado.
- Román, M. (2005). *Evaluación Proyecto Enseñar para Aprender. Un modelo de apoyo a escuelas donde todos los alumnos aprenden*. Santiago de Chile: CIDE- Fundación Ford.
- Román M. (2006a). *Evaluación de impacto del Programa Mejoramiento Escolar en cuatro países: Nicaragua, Honduras, El Salvador y Perú*. Santiago de Chile: CIDE-Plan Internacional.
- Román M. (2006b). El desafío de aprender a leer, comprender y razonar en escuelas vulnerables. *Revista Pensamiento Educativo*, Vol.39, pp.69-86. Santiago: Pontificia Universidad Católica de Chile.
- UNESCO/OREALC (2004). *Conflictividad educativa en América Latina: cronología de la actividad sindical docente*. Santiago de Chile: UNESCO/OREALC.
- UNESCO/OREALC (2005). *Condiciones de salud del profesorado*. Santiago de Chile: UNESCO/OREALC.
- Vélez, E., Schiefelbein, E. y Valenzuela, J. (1994). Factores que afectan el rendimiento académico en la educación primaria. Revisión de la literatura de América Latina y el Caribe. *Revista Latinoamericana de Innovaciones Educativas*, 17, pp. 29-53.

Informes nacionales de evaluación utilizados

Argentina

Ministerio de Educación/IDECE (s.a). *Operativo Nacional de Evaluación de la Calidad Educativa: Boletín de Informe de Resultados: TOTAL PAIS. AÑO: 2000, CURSO: 3º año EGB: Resultados en*

Matemática, Lengua, Ciencias Naturales y Ciencias Sociales. Buenos Aires: Ministerio de Educación/IDECE. (<http://www.me.gov.ar/diniece/>)

Ministerio de Educación/IDECE (s.a). *Operativo Nacional de Evaluación de la Calidad Educativa: Boletín de Informe de Resultados: TOTAL PAÍS. AÑO: 2000, CURSO: 6º año EGB: Resultados en Matemática y Lengua.* Buenos Aires: Ministerio de Educación/IDECE. (<http://www.me.gov.ar/diniece/>)

Ministerio de Educación/IDECE [s.a]. *Operativo Nacional de Evaluación: Informe de resultados 2000: Total país. 9º EGB / 2º Medio, 3º Polimodal / 5º-6º Medio.* Buenos Aires: Ministerio de Educación/IDECE (<http://www.me.gov.ar/diniece/>)

Ministerio de Educación/DINIECE. [s.a]. *Operativo Nacional de Evaluación (ONE) 2003: 3º año EGB. 6º año EGB. 9º año EGB. Fin del Nivel Medio o Polimodal.* Buenos Aires: Ministerio de Educación/DINIECE. (<http://www.me.gov.ar/diniece/>)

Brasil

Ministério Da Educação/INEP (2005). *Pesquisa Nacional Qualidade da Educação: A Escola Pública Na Opinião Dos Países: Resumo técnico executivo.* Brasil: Ministério Da Educação.

Ministério Da Educação/INEP (2007). *SAEB – 2005 Primeiros resultados: Médias de desempenho do SAEB/2005 em perspectiva comparada.* Brasil: Ministério Da Educação.

Colombia

Bogoya, D. (2006). Evaluación Educativa en Colombia, En ICFES, Seminario Internacional de Evaluación. Cartagena de Indias: ICFES.

Corpoeducación/Fundación Conona/EXE/PREAL (2006). *Hay avances, pero quedan desafíos: Informe de progreso educativo de Colombia.* Bogotá: PREAL.

ICFES (2006). *Pruebas saber 2005: Marco de Interpretación de Resultados. Estructuras de la Prueba.* Bogotá: ICFES.

Costa Rica

Ministerio de Educación Pública (2005a). *Informe nacional: Resultados II ciclo.* San José: Ministerio de Educación Pública.

Ministerio de Educación Pública (2005b). *Informe nacional: Análisis técnico de ítemes por asignaturas. Adecuación curricular: Tercer ciclo.* San José: Ministerio de Educación Pública.

Chile

SIMCE (2001). *Informe de resultados 2000: 8º básico / Resultados Nacionales.* Santiago de Chile: Ministerio de Educación. (http://www.simce.cl/paginas/res_nacionales_2000.htm)

SIMCE (2005a). *Informe de resultados 2004: 8º básico: Resultados Regionales y Diferencias de Puntaje entre el SIMCE 2004 y 2000.* Santiago de Chile: Ministerio de Educación. (http://www.simce.cl/paginas/res_regionales_2004.htm)

SIMCE (2005b). *Informe de resultados 2004: 8° básico, resultados nacionales: Promedios nacionales y diferencias de puntaje entre el SIMCE 2004 y 2000*. Santiago de Chile: Ministerio de Educación. (http://www.simce.cl/paginas/res_nacionales_2004.htm)

SIMCE (2005c). *Análisis de las diferencias de logro en el aprendizaje escolar entre hombres y mujeres. Análisis y Comunicación de Resultados SIMCE*. Santiago de Chile: Ministerio de Educación.

SIMCE (2005d). *Informe de resultados 2004: 8° educación básica*. Santiago de Chile: Ministerio de Educación.

SIMCE (2006). *Informe de resultados 2005: 4° educación básica*. Santiago de Chile: Ministerio de Educación.

El Salvador

DNME/MINED (2005). *Logros de Aprendizaje de educación básica en El Salvador: Pruebas Censales 2005*. San Salvador: DNME-MINED.

Guatemala

CIE/UVG/PRONERE/MINEDUC (2006). *Informe de Difusión General: Evaluación del rendimiento en lectura y matemática de estudiantes de sexto grado de primaria de escuelas oficiales de Guatemala*. Guatemala: CIE/UVG/PRONERE/MINEDUC.

Honduras

Rápalo, R. (Coord.) (2005). *Informe comparativo nacional del rendimiento académico 2002-2004: Tercero y sexto grado*. Tegucigalpa: Secretaría de Educación/Universidad Pedagógica Nacional Francisco Morazán.

México

Backhoff, E.; Andrade, E.; Monroy, L.; Tanamachi, M.; Bouzas, A.; Sánchez, A.; y Peón, M. (2005). *Estudio comparativo de la educación básica en México 2000 – 2005*. México D.F.: INEE.

Backhoff, E.; Andrade, E.; Sánchez, A.; Peon, M.; y Bouzas, A. (2006). *El Aprendizaje del Español y las Matemática en la Educación Básica en México: Sexto de Primaria y Tercero de Secundaria*. México D.F.: INEE.

INEE (2005). *Proceso de construcción y características básicas: Excale, Exámenes de la Calidad y el Logro Educativos*. INEE, Colección de folletos, N° 8. México D. F.: INEE.

Nicaragua

MECD (2004). *Evaluación del rendimiento académico de los estudiantes de 3ª y 6ª grado de primaria: informe de resultados 2002*. Managua: MECD.

Paraguay

SNEPE. (s.a). *Pruebas Nacionales*. Asunción: SNEPE.

Perú

Espinosa, G. y Torreblanca, A. (2004). *Cómo rinden los estudiantes peruanos en Comunicación y Matemática: Resultados de la Evaluación Nacional 2001*. Lima: Unidad de Medición de la Calidad - Ministerio de Educación.

UMC-MED/GRADE (2000). *Resultados de las Pruebas de Matemática y Lenguaje. ¿Qué aprendimos a partir de la Evaluación CRECER 1998*. Boletín CRECER N° 5/6.: 1-12.

UMC-MED (2005). *Evaluación Nacional del Rendimiento Estudiantil 2004: Informe descriptivo de resultados*. Lima: Ministerio de Educación.

Uruguay

ANEP – PROYECTO MECAEP (1999). *Evaluación Nacional de Aprendizajes en Lenguaje y Matemática: 6° Año Enseñanza Primaria 1999: Primer Informe*. Montevideo: Administración Nacional de Educación Pública.

ANEP – PROYECTO MECAEP (2002). *Evaluación Nacional de Aprendizajes en Lenguaje y Matemática: 6° Año Enseñanza Primaria 2002: Primer Informe Devolución de Resultados de la Muestra Nacional*. Montevideo: Administración Nacional de Educación Pública.

ANEP – PROYECTO MECAEP (2007). *Evaluación Nacional de Aprendizajes en Lenguaje y Matemática: 6° Año Enseñanza Primaria - 2005: Primer Informe*. Montevideo: Administración Nacional de Educación Pública.





EVALUACIÓN EDUCATIVA SOBRE SENTIDOS Y PRÁCTICAS

Margarita Poggi

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art2.pdf>

El presente trabajo se propone aportar algunas reflexiones construidas a partir de distintas experiencias de evaluación en las que he participado, realizadas desde diferentes ámbitos, ya sea que remitan a la evaluación de sistemas educativos (o subsistemas) en organismos de gobierno, como a la evaluación de instituciones o de programas educativos. Dado que siempre he formado parte de equipos (sea como integrante como en su coordinación) estas reflexiones son personales pero, a la vez, resultan de los trabajos colectivos de diseño y desarrollo de estrategias y metodologías de evaluación.

Sostenemos que un modelo de evaluación debe ser construido en y para cada sistema o subsistema educativo o proyecto. Sin embargo, la experiencia acumulada permite ofrecer aportes conceptuales y metodológicos para dicha construcción. Estos aportes recogen algunas de las discusiones tradicionales sobre el tema de la evaluación. Por otro lado, pretenden ofrecer algunas reflexiones que permitan sobrepasar cada experiencia concreta en un intento por sistematizarlas; por ello el artículo concluye con algunas orientaciones generales que pueden ser consideradas en el diseño de propuestas de evaluación.

1. ALGUNAS DEFINICIONES BÁSICAS PARA COMPARTIR

Diferentes conceptos y nociones son las que están involucradas en el proceso de evaluación. Por mencionar sólo algunos podemos enumerar los siguientes: evaluación, referente / referido, evaluación interna / externa, cualitativa / cuantitativa, etc. Si bien no pretendemos reproducir en este apartado definiciones y discusiones que pueden encontrarse en toda la literatura clásica sobre evaluación, nos proponemos dejar establecidas algunas de las nociones clave de modo de compartir sus sentidos.

1.1. Evaluación educativa

Es uno de los términos de los que, tal vez, más acepciones puedan relevarse en la bibliografía especializada en las ciencias sociales y, en particular, en educación. Sin embargo, para aproximarnos al concepto, pueden reconocerse algunas líneas comunes a varios autores¹.

"En primer lugar, evaluar supone efectuar una lectura orientada sobre el objeto que se evalúa, en función de la cual el evaluador se 'pronuncia sobre la realidad'. Dicho de otra manera, no existe una lectura directa de la experiencia. Hay siempre un proceso de interacción entre el evaluador y la realidad a evaluar. En ese proceso, en función de las propiedades de esa realidad, el evaluador construye el referente, es decir, aquello con relación a lo cual se va a efectuar la evaluación, aquello que le permitirá 'pronunciarse sobre la realidad' que evalúa". "Con respecto al producto de la evaluación, también concierne al evaluador construir los resultados de la evaluación, construir los datos, 'lo referido' (con relación al referente)"².

En otros términos, el referente³ puede definirse como el conjunto de normas o de criterios que sirven de grilla de lectura del objeto a evaluar. Constituye una especie de "ideal" o de "deber ser" y hace que la evaluación implique, siempre, una lectura orientada por el marco, por la óptica, por la perspectiva de análisis privilegiada por el evaluador (ya sea éste un sujeto o un equipo). Por el contrario, lo referido está constituido por el conjunto de observables que se recortan de un objeto de evaluación y que se consideran representativos de éste. Cabe aclarar también aquí que el evaluador, durante el proceso de

¹ Las definiciones se retoman de Bertoni, Poggi y Teobaldo, Evaluación. Nuevos significados para una práctica compleja, Buenos Aires, Kapelusz, 1995, particularmente del capítulo primero. Asimismo, se re trabajan las concepciones de diversos autores; entre ellos podemos mencionar los siguientes: Barbier, Jean- Marie, La evaluación en los procesos de formación, Barcelona, Paidós, 1993; Hadji, Ch., L'évaluation des actions éducatives, Paris, PUF, 1992; Hadji, Ch., L'évaluation, règles du jeu. Des intentions aux outils, Paris, ESF, 1992.

² Bertoni, Poggi y Teobaldo, 1995, op. cit., pág. 20 y 23.

³ Recordamos aquí que el término latino referens significa "que reenvía a". Por otra parte, y para no distraer la atención sobre otros aspectos vinculados con el referente, podemos señalar que éste puede ser construido de modo previo al proceso de evaluación o puede ser construido durante dicho proceso. Esto se denomina evaluación apreciativa con o sin referente predeterminado según uno u otro caso.

evaluación, toma decisiones y construye tanto el referente como lo referido. Tomar decisiones significa, entre otras muchas cuestiones, recortar, seleccionar y plantearse desde qué marcos organizar la lectura de la realidad a evaluar.

Este aspecto pone en cuestión tanto la existencia de referentes universales (el evaluador siempre opta por *alguna* perspectiva, *alguna* toma de posición o *algún* recorte teórico en particular), como el hecho de que un programa o proyecto pueda ser evaluado en toda su complejidad (en realidad evaluamos mucho menos de lo que pretendemos, y operamos *como si* hubiéramos atrapado *casi* todo del objeto de la evaluación).

Asimismo, es interesante constatar que la puesta en relación entre el referente y lo referido puede ser abordada en dos sentidos. El primero es el más habitual: del referente a lo referido, de las expectativas a la realidad; el segundo, de lo referido al referente, de lo existente (por lo menos, de aquello que un recorte pone en evidencia) a lo deseado.

Otra cuestión importante a agregar es que la evaluación comprende tanto aspectos políticos como técnicos. Es una actividad política⁴ por naturaleza, ya que debe ser no sólo veraz sino también justa, y por supuesto requiere capacidades técnicas para poder desarrollarse.

A partir de lo afirmado hasta aquí, cabe subrayar entonces las siguientes cuestiones:

- La evaluación desde la perspectiva que sostendremos debe ser entendida como un proceso de construcción de conocimiento sobre el objeto de evaluación.
- Implica, en el mismo sentido, un acto de interrogación sobre aquello que se evalúa; acto de interrogación que se puede formular sobre el objeto de evaluación como sobre las propias estrategias y prácticas del evaluador.
- Siempre supone algún tipo de recorte (tanto desde los referentes que se construyen, como desde los aspectos -lo referido- que se seleccionan del objeto a evaluar y que se consideran representativos de éste).

1. 2. Evaluación interna / externa

Básicamente, los términos aluden, en primera instancia, a la cuestión del origen de una iniciativa de la evaluación. Desde otras perspectivas, puede sumarse también a este primer aspecto el tema del grado de control sobre el proceso y los resultados de la evaluación y el de la participación y grado de implicación de los actores en la realización de la evaluación.

En este sentido, se trata de un binomio de términos que, en algunas ocasiones, son presentados como dilemáticos. Sin embargo, los principios y las modalidades sostenidas en modelos concretos de evaluación pueden permitir articulaciones y potenciaciones mutuas entre una y otra.

Por otra parte es conocido que una y otra modalidad presenta ventajas y desventajas, las cuales aparecen relevadas en la bibliografía especializada. Nos parece importante, sin embargo, señalar en primer lugar algunas fortalezas de la evaluación interna y luego avanzar en su articulación con la externa.

La **evaluación interna o autoevaluación** es el proceso de indagación y estudio de un programa, proyecto o institución, generado, organizado y gestionado por sus propios integrantes.

⁴ Ver al respecto House, E., Evaluación, ética y poder, Madrid, Morata, 1995.

Con respecto a ella, podemos señalar las siguientes ventajas:

- Se incrementa el conocimiento del proyecto o institución que se autoevalúa:
- Se favorece la revisión y adecuación de fines y objetivos de las acciones que se realizan.
- Se clarifican y dimensionan en mayor grado las fortalezas y los problemas.
- Se logra una mejor organización y jerarquización de la información del proyecto.
- Se incrementa la comunicación efectiva entre los actores involucrados en el proyecto.
- Se fortalece la racionalidad de las decisiones.
- Se promueve la capacitación interna y fortalecimiento del desarrollo profesional.
- Se facilita una mayor participación y apropiación de los resultados de la evaluación porque su validez no es cuestionada por los actores implicados en ella.

Si bien es cierto que algunos de los rasgos mencionados pueden ser también adjudicados a la evaluación externa, es decir que no son exclusivos de la autoevaluación, cabe destacar que, en el caso de esta última, se potencia y fortalece la autonomía, el mejoramiento y la profesionalización de los actores de un proyecto y contribuye en gran medida a que se promuevan cambios a partir de los resultados de la evaluación. En otros términos supone un compromiso con el cambio y la innovación.

Sin embargo, también es necesario destacar, además de sus numerosas ventajas, algunas de las debilidades o problemas que presenta la autoevaluación, los cuales se encuentran estrechamente relacionados con ciertas condiciones que deben enmarcar el proceso. Esto es, en la medida en que ciertas condiciones no están presentes, los riesgos se incrementan y la autoevaluación se vuelve más compleja en relación con su factibilidad o más débil desde la perspectiva de su credibilidad. Al respecto, pueden señalarse las siguientes cuestiones:

- Encierra el riesgo de que cada grupo actúe con criterios que sirvan para justificar sólo aquello que está haciendo. Es decir que puede convertirse en una autojustificación.
- No siempre existen condiciones internas en cada proyecto para llevar a cabo el proceso de autoevaluación. Esto es, requiere capacidades metodológicas y técnicas instaladas en los equipos de los proyectos, tiempos y motivación para realizarla y un decidido respaldo institucional.

La instalación y diseminación de una cultura evaluativa democrática y de colaboración en proyectos e instituciones, o su carencia, es un factor importante a la hora de sostener procesos de autoevaluación. No obstante, es necesario aclarar que la instalación de una cultura evaluativa democrática implica un proceso gradual y extenso que sólo puede hacerse efectivo ... *evaluando*. Dicho en otros términos, puede ser considerada como un requisito o condición preexistente para la autoevaluación o como una cuestión que progresivamente se va construyendo en el proceso de evaluación de un proyecto o institución.

Para superar posiciones o planteos dilemáticos, la evaluación externa e interna puede ser pensada de modo que se articulen y complementen entre sí. Más aún, la evaluación externa puede contrarrestar algunos de los riesgos de la autoevaluación, aunque este argumento no es su única justificación.

En realidad, el acceso a la información que provee una evaluación externa sobre un proyecto o institución, especialmente cuando integra aspectos descriptivos, analíticos y reflexivos, encuentra en ellos su mejor fundamentación. Se suma que algunas evaluaciones tienen por propósito central capturar algunas dimensiones de microsistemas, lo que hace inviable (o por lo menos muy dificultoso) el abordaje desde procesos de autoevaluación.

Otra de las ventajas que generalmente se le asigna a la evaluación externa se vincula con el hecho de que es realizada por profesionales con mayor grado de independencia o, lo que es lo mismo, menor grado de implicación que los actores involucrados directamente en los proyectos y en la acción. Mayor grado de independencia significa, centralmente, que sólo tienen menos intereses comprometidos con los resultados de la evaluación; esto no significa mayor grado de objetividad (la que puede lograrse tanto en la autoevaluación como en la evaluación externa mediante ciertos cuidados y recaudos metodológicos).

Cuando las prácticas evaluativas articulan la mirada y el análisis externo, esta combinación permite aumentar la credibilidad y contribuir a que se tome conciencia de los problemas en los que no se ha reparado. Incluso aparece en la bibliografía especializada bajo la figura del "agente provocador" o "amigo crítico", términos que dejan claramente planteada la idea de una interlocución crítica y reflexiva que puede potenciar aún más las ventajas de la autoevaluación.

1.3. Enfoque cualitativo / cuantitativo

Estos términos aluden también a paradigmas y concepciones sobre la evaluación que han tenido un extenso tratamiento en la bibliografía especializada. En muchas ocasiones las concepciones se han presentado de modo fragmentado y polarizado y, en el extremo, como planteos dilemáticos (de modo similar a lo que ha ocurrido con los aspectos tratados en el punto anterior).

No es objetivo de este artículo rastrear el conjunto de las discusiones sobre el tema, que remiten tanto a cuestiones metodológicas como a las concepciones epistemológicas que las sustentan⁵. Sin embargo, conviene recordar que algunas de las discusiones que enfrentan y atraviesan uno y otro enfoque (como por ejemplo, la relación entre paradigma y método; la oposición subjetividad / objetividad como patrimonio de uno u otro; la relación entre validez y confiabilidad en cada uno de los métodos; el énfasis en el proceso o en los resultados; el carácter exploratorio e inductivo en un caso y confirmatorio y deductivo en otro; la posibilidad de establecer generalizaciones; la relación con el contexto de descubrimiento o de verificación; etc.) deben, y más aún en el caso de la investigación evaluativa, ser analizadas a la luz de una nueva reconceptualización del debate que las sustenta.

Al respecto, los autores ya citados afirman que "de hecho, todos los atributos que se asignan a los paradigmas son lógicamente independientes. Del mismo modo que los métodos no se hallan ligados lógicamente a ninguno de los atributos de los paradigmas, los propios atributos no se encuentran lógicamente ligados entre sí... Baste con decir que no existe nada, excepto quizá la tradición, que impida al investigador mezclar y acomodar los atributos de los dos paradigmas para lograr la combinación que resulte más adecuada al problema de la investigación y al medio con que se cuenta". Más adelante, avanzan aún más en sus planteos al afirmar: "Parece entonces que no existe tampoco razón para elegir *entre* métodos cualitativos y cuantitativos. Los evaluadores obrarán sabiamente si emplean cualesquiera métodos que resulten más adecuados a las necesidades de su investigación, sin atender a las afiliaciones tradicionales de los métodos"⁶.

En el mismo sentido que venimos afirmando la importancia de la complementación de estos enfoques, es necesario recordar algunas de las ventajas potenciales sobre el empleo y la articulación de ambas metodologías. Son básicamente tres. En primer lugar, dado que la investigación evaluativa generalmente tiene múltiples propósitos, éstos deben ser atendidos con referencia a una variedad de métodos. Esto supone, en otros términos, atender tanto al proceso como a los resultados y considerar, en muchos casos

⁵ Para una discusión en profundidad sobre este tema puede consultarse Cook, T. D. y Reichardt, Ch. S., "Hacia una superación del enfrentamiento entre los métodos cualitativos y los cuantitativos", Cook, T. D. y Reichardt, Ch. S., *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata, 1986.

⁶ Cook, T. D. y Reichardt, Ch. S., 1986, op. cit., pág. 40 y 41. En este punto se retoman otros planteos de estos autores.

de manera simultánea, los objetivos vinculados con la comprobación, la valoración del impacto y la búsqueda de explicaciones sobre los aspectos evaluados.

En segundo término, la utilización de manera articulada de ambos métodos puede contribuir a una potenciación mutua. Esto implica que uno no puede sustituir al otro, dado que aportan formas diferentes de abordar y develar el objeto de evaluación y que, en consecuencia, requieren su complementariedad. Si bien, cabe aclarar, no se trata de pensar que ésta resulta de una simple sumatoria sino, por el contrario, que implica atender a los problemas que resultan de su articulación.

Por último, dado que no hay método que no presente problemas, dificultades o inconvenientes, la tercera ventaja que ofrece la articulación de ambos enfoques es la de posibilitar la triangulación de métodos⁷ que permitan disminuir los sesgos propios a cada uno de ellos.

Así como consignamos algunas de las ventajas posibles del uso conjunto de ambos enfoques, también es necesario dejar planteados algunos de los inconvenientes que éste presenta. En primer lugar, la disponibilidad de recursos y tiempo que éstos exigen y, en segundo término, la formación metodológica y técnica requerida para sostener la combinación de ambos.

De cualquier manera, más allá de las objeciones que puedan realizarse a la utilización conjunta de estos enfoques, la cuestión principal a considerar es la selección de los métodos más apropiados en función del o de los problemas de evaluación que se hayan formulado en el diseño de este proceso.

1.4. El sentido de la evaluación

Las preguntas por el sentido pueden (y deben) formularse en relación con cualquier actividad, pero se vuelven aún más relevantes en el caso particular de la acción educativa.

Por ello, en el caso particular de la evaluación, también son centrales las preguntas por su sentido, que reenvían a su vez a una interrogación tanto por su concepción y metodología como por sus efectos.

En la evaluación de programas y proyectos, las decisiones metodológicas no están sólo guiadas por una racionalidad técnica. Aún más, "una teoría de la evaluación debe ser tanto una teoría de la interacción política como una teoría del modo de construir el conocimiento"⁸. En esta línea, toda propuesta de evaluación remite, explícita o implícitamente, a posiciones teóricas y no sólo metodológicas, involucra cuestiones vinculadas con las relaciones de poder y produce afirmaciones políticas que tienen efectos políticos (sobre políticas, programas e instituciones concretas).

Las preguntas que orientan una evaluación son relevantes para guiar el proceso y pueden además tener por finalidad el incremento de la autoreflexividad de los equipos responsables de un diseño de evaluación. La indagación sobre el conocimiento práctico puesto en juego en el desarrollo de experiencias evaluativas involucra a los actores implicados, en mayor o menor medida, en las diferentes acciones que supone una evaluación.

En este sentido, la evaluación implica decisiones colectivas y situaciones de negociación; remite a situaciones en donde el quehacer y la acción se vuelven públicos y pueden ser motivo de discusión pública.

Por otra parte, entendemos que siempre la evaluación supone considerar "objetos" singulares; de este aspecto se deriva la necesidad de contextualizar las orientaciones metodológicas en relación con los

⁷ Recordamos que el concepto de triangulación se relaciona, en investigación, con la credibilidad y confiabilidad de la información. Consiste en la combinación de diferentes metodologías que permite analizar los datos desde diferentes ángulos con el propósito de contrastarlos entre sí. La triangulación puede ser metodológica, temporal, interna (en el equipo evaluador), de sujetos o informantes clave, o teórica.

⁸ L. Cronbach, citado en Simons, H., Evaluación democrática de instituciones escolares, Madrid, Morata, 1999, pág. 31.

objetivos priorizados en un diseño y las características propias de lo que se recorta como objeto de la evaluación.

2. ALGUNOS LINEAMIENTOS Y CRITERIOS GENERALES PARA ORIENTAR PROCESOS DE EVALUACIÓN

No obstante el reconocimiento de la singularidad de cada experiencia de evaluación, algunos lineamientos generales pueden resultar pertinentes para orientar cada proceso evaluativo y el modelo que se construya para tal fin. Los lineamientos y las orientaciones metodológicas que aquí se presentan son sólo guías generales para la acción y no pretenden uniformar la evaluación en formatos únicos. Las orientaciones generales pueden permitir enriquecer las propuestas y promover las reflexiones sobre los supuestos (no siempre explícitos) que fundamentan decisiones metodológicas así como anticipar problemas o sistematizar dificultades encontradas en distintas iniciativas de evaluación.

Sin pretender agotar en la enumeración el conjunto de aspectos que pueden contemplarse en un proceso de evaluación, algunos lineamientos y criterios generales que creemos más significativos son:

Asegurar que se presente una visión representativa de aquello que se pretende evaluar, aun cuando la evaluación implica siempre un recorte de las acciones realizadas.

Éste constituye un aspecto central a considerar en el diseño de toda evaluación, porque una de las tentaciones más grandes es la pretensión de querer evaluarlo "todo". En consecuencia, la definición de las preguntas clave que orientan el diseño de evaluación, constituye el hilo que permite marcar el límite pertinente del recorte.

Incorporar, además de la evaluación de resultados, una mirada sobre el contexto del objeto de evaluación y una perspectiva procesual e histórica de éste.

La evaluación supone el abordaje de objetos complejos, ya sea por la diversidad de objetivos y efectos que se pretenden valorar, como por las múltiples dimensiones que siempre conlleva el análisis de un sistema o proyecto educativo. En consecuencia, incluir una aproximación al contexto así como la consideración de una perspectiva procesual e histórica, cuando ésta es factible, permiten una comprensión más integral de los procesos educativos. Ambas cuestiones posibilitan poner en perspectiva la evaluación que se realice de cada sistema o proyecto, analizar tendencias y, de este modo, enmarcar los resultados obtenidos.

Dar cuenta de logros y dificultades en la evaluación de sistemas y proyectos.

Si bien puede ser considerado de "sentido común", esta cuestión apunta a subrayar la relación entre la evaluación y la determinación del valor de un sistema o proyecto (hacia los actores clave comprometidos en el proyecto y hacia otras audiencias o destinatarios). Desde nuestra perspectiva, se articula con la necesidad de ofrecer no sólo una mirada descriptiva, sino fundamentalmente analítica, interpretativa y valorativa. En ese último caso, resaltar los logros (además de señalar las dificultades) permite evitar efectos desmoralizadores en los actores implicados en la evaluación.

Estar al servicio de la producción de conocimiento sobre los procesos y resultados educativos, por un lado, y del aprendizaje de los actores involucrados en éstos, por el otro.

La evaluación debería, desmarcándose de significaciones que la aproximan al control, contribuir a la producción de conocimiento en temáticas clave de la dinámica y procesos educativos. En ese sentido, la evaluación permite no sólo la sistematización de experiencias y su conceptualización, sino que puede contribuir a desarrollar instancias de reflexión sobre el conjunto de factores y dimensiones implicadas en los procesos educativos y sobre las responsabilidades compartidas por parte de los distintos actores

involucrados en ellos. En consecuencia, los procesos de negociación de significado entre dichos actores ocupan un lugar central en el desarrollo de toda propuesta de evaluación.

Asegurar una evaluación válida y confiable.

Una de las cuestiones centrales en este punto se relaciona con el hecho de llevar a cabo un proceso de evaluación que permita construir juicios fundados, elaborados en el marco de un proceso sistemático. Este punto marca la frontera entre una evaluación informal, de aquella que toma en consideración recaudos metodológicos y técnicos para asegurar la construcción de juicios caracterizados por su validez y confiabilidad.

Contemplar los distintos destinatarios, o las diferentes audiencias, de la evaluación: los actores, organismos e instituciones involucrados.

Las preguntas sobre el “para qué” y el “por qué”, cuya formulación es imprescindible en todo proceso de evaluación, remiten en parte a los diferentes destinatarios o audiencias. Más que una referencia a producciones específicas elaboradas para cada audiencia (atendiendo a perfiles diferenciados y a las formas de presentar los resultados obtenidos según el uso de la información) nos interesa llamar la atención sobre la consideración de los diversos intereses y puntos de vista presentes en toda evaluación, que remiten a los diversos sentidos que están en juego en toda evaluación que se realiza.

3. A MODO DE CIERRE

Tanto las discusiones sobre temáticas de la evaluación que en ocasiones se presentan como dilemáticas como los lineamientos presentados no pretenden agotar el conjunto de abordajes que puede realizarse sobre un campo tan amplio como es el de la evaluación educativa. Por ello, nuestra línea de argumentación se orienta a promover la reflexión sobre el sentido de las prácticas evaluativas.

Por otra parte, si la evaluación produce conocimiento sobre una realidad determinada, su propósito no debe circunscribirse a pronunciarse sobre ella, sino que también debe pretender ofrecer indicios para mejorar aquellos aspectos que resultan problemáticos. Para ello, el tema de la construcción de confianza sobre la evaluación (es decir sobre las metodologías que diseña, los actores involucrados, el modo de producir los juicios evaluativos fundados en la evidencia, etc.) resulta clave a la hora de orientar la mejora educativa. Confianza que involucra a distintos actores: aquellos de máxima responsabilidad en la definición de las políticas educativas, los docentes que necesariamente estarán implicados en las estrategias de mejora que se desarrollen, la comunidad educativa en sentido amplio y otros actores sociales preocupados por los temas centrales de una agenda educativa.

Por último, el apoyo político es clave para producir avances o desarrollar nuevas metodologías en materia de evaluación, porque es un área sensible cuyos resultados producen efectos tanto en el sistema educativo como en la sociedad. Por ello, tal como se afirmó anteriormente es fundamental no perder de vista las dimensiones políticas y a la vez técnicas inherentes a toda evaluación educativa.

REFERENCIAS BIBLIOGRÁFICAS

- Barbier, JM. (1993). *La evaluación en los procesos de formación*. Barcelona: Paidós.
- Barlow, M. (1992). *L'évaluation scolaire. Décoder son langage*. Lyon: Chronique Sociale.
- Bertoni, A., Poggi, M. y Teobaldo, M. (1995). *Evaluación. Nuevos significados para una práctica compleja*. Buenos Aires: Kapelusz.

- Carnoy, M, Elmore, R. y Siskin, L. (2003). *The new accountability*. New York and London: Routledge Falmer.
- Cook, T. D. y Reichardt, Ch. S. (1986). *Métodos cualitativos y cuantitativos en investigación evaluativa*. Madrid: Morata.
- De Ketele, J-M. (Ed.). (1992). *L'évaluation: approche descriptive ou prescriptive?*. Bruxelles : De Boeck.
- Elmore, R. (2003). Salvar la brecha entre estándares y resultados. El imperativo para el desarrollo profesional en educación. *Profesorado, Revista de curriculum y formación del profesorado*, 7(1-2).
- Goetz, P. y LeCompte, M. D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.
- Goffman, E. y otros (2000). *Sociologías de la situación*. Madrid: La Piqueta.
- Hadji, Ch. (1997). *L'évaluation démystifiée*. Paris : ESF.
- House, E. (1995). *Evaluación, ética y poder*. Madrid: Morata.
- Martínez Rizo, F. (2003). Una mirada técnico-pedagógica acerca de las evaluaciones de la calidad educativa. En: laies, G. y otros. *Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IIPE-UNESCO Sede Regional Buenos Aires.
- McCormick, R. y James, M. (1996). *Evaluación del curriculum en los centros escolares*. Madrid: Morata.
- McKernan, J. (1999). *Investigación-acción y currículo*. Madrid: Morata.
- Murillo Torrecilla, J. (2003). El movimiento teórico-práctico de mejora de la escuela. Algunas lecciones aprendidas para transformar los centros docentes. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2).
- Poggi, M. (2003). Temas y problemas clave en torno a la evaluación educativa. En C. Ornelas (comp.). *Evaluación educativa: hacia la rendición de cuentas, Memoria del Segundo Encuentro Internacional de Educación*. México: Aula XXI-Santillana.
- Ravela, P. (ed.) y otros (2001). *Los próximos pasos. ¿Cómo avanzar en la Evaluación de Aprendizajes en América Latina?*. Santiago de Chile, PREAL, 2001.
- Simons, H. (1999). *Evaluación democrática de instituciones escolares*. Madrid: Morata.
- Simons, H. (2004). Utilizing Evaluation Evidence to Enhance Professional Practice. *Evaluation*, 10(4).
- Slee, R. y Weiner, G. en colaboración con S. Tomlinson (eds.) (2001). *¿Eficacia para quién? Crítica de los movimientos de las escuelas eficaces y de la mejora escolar*. Madrid: Akal.
- Tiana, A. (1997). Tratamiento y uso de la información en evaluación. Madrid: OEI.
- Tiana, A. (2003). ¿Qué pretendemos evaluar, qué evaluamos y qué conclusiones podemos extraer de la evaluación?. En: laies, G. y otros. *Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa*, Buenos Aires: IIPE-UNESCO.



EVALUACIÓN DE LOS PROCESOS DE EVALUACIÓN DEL SISTEMA EDUCATIVO 1950-2008

Ernesto Schiefelbein y Paulina Schiefelbein

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art3.pdf>



El crecimiento de los sistemas educativos, en las últimas seis décadas, ha estado asociado a profundas revisiones (y significativas mejoras) en la manera de evaluarlos. Tiene especial importancia revisar los cambios producidos en tres dimensiones diferentes (pero complementarias), ya que si bien su impacto final parece ser modesto (McKinsey, 2007:5), permite sugerir nuevas formas de mirar la dinámica instruccional (Ball y Forzani, 2007:530) y, eventualmente, la educacional.

El principal cambio consistió, quizá, en pasar de "contar alumnos en cada nivel" a estimar "niveles de aprendizaje" de contenidos y habilidades o aptitudes. Mientras que hace treinta años cualquier evaluación era considerada como un tabú, en la actualidad la mayoría de países someten a los estudiantes a pruebas desde los primeros años de la escolaridad hasta el final de los estudios universitarios (Schiefelbein y Schiefelbein, 2004). Es importante, además, el haber pasado del "clasificar" a los alumnos (entre los que podían avanzar al siguiente nivel y los que debían repetir), a lograr que cada alumno aprenda al menos el mínimo necesario para avanzar (si tiene la capacidad necesaria, aunque sus conocimientos previos sean limitados). Finalmente, se debe destacar el creciente interés por evaluar no solamente el nivel de "instrucción" (en contenidos y habilidades), sino que también la "formación" (propiamente tal) de la capacidad de tomar decisiones responsables conforme a criterios aceptados.

En este artículo se examina los cambios en cada una de estas tres dimensiones y se comenta la manera en que cada uno de estos cambios ha repercutido en la operación de los sistemas y en los objetivos que privilegian los maestros en el aula. Se examina, además, las brechas que estos cambios han generado entre la formación inicial de los maestros y las demandas sociales de calidad de la educación.

1. PASAR DESDE LAS BRECHAS DE COBERTURA A LAS BRECHAS DE CALIDAD

De los indicadores de analfabetismo y cobertura usados en la primera mitad del siglo 20 se ha pasado, a través de etapas graduales, a usar los puntajes en pruebas de rendimiento académico. Las "tasas de analfabetismo" eran calculadas con las declaraciones que los "jefes de hogar" hacían a los encuestadores que visitaban cada casa cuando se realizaba un Censo de Población (esto solía ocurrir cada diez o más años) y, por lo tanto, era una medida que dependía de lo que cada jefe de hogar consideraba que era ser analfabeto. Se usaban en los estudios internacionales que ha realizado Unesco (desde sus inicios) para comparar el avance de los países en materia de educación y estimular la ampliación de la cobertura.

Chile redujo su tasa de analfabetismo de 25% en 1930, a 16% en 1960, a un 6% en 1992 y a un 4% en la actualidad (MINEDUC, 2005). Si bien Unesco consideraba alfabeto a alguien que había cursado al menos un cuarto grado de primaria, las declaraciones al Censo de población estaban asociadas a que la persona hubiera estado matriculada en algún momento en una escuela.

En los años 60 la mitad de los Ministerios de Educación de América Latina tenía estimaciones de la matrícula en cada grado y, en muchos casos, disponían de datos por edades calculados en los grandes computadoras (denominadas "mainframe" por estar en una gran caja de metal que ocupaba una enorme habitación). Esto permitía calcular las tasas de escolarización (bruta y neta) y, eventualmente, tasas de repetición y de graduación (Schiefelbein, 1974).

Las tasas de repetición permitían suponer, en la década de los 60, que un 20% de los niños de América Latina (que ingresaban a la escuela) tenía problemas de aprendizaje, pero el problema era considerablemente mayor. En efecto, los datos para calcular esas tasas eran proporcionados por los directores de escuela quienes informaban de los alumnos que repetían y de los que abandonaban la escuela, sin saber que muchos de ellos se matriculaban al año siguiente en otra escuela como repitientes. Esto generaba una subestimación de la tasa de repetición y una sobreestimación de la tasa de deserción. Modelos de simulación permitieron calcular, con mayor precisión, que las tasas de repetición del primer

grado estaban cercanas al 50% (Schiefelbein, 1975). Estos indicadores sugerían la necesidad de estimular el aprendizaje de los niños antes de ingresar a la educación primaria (Schiefelbein *et al.*, 2000).

El desarrollo de los lectores ópticos y programas de computación permitió procesar, a fines de los 50, pruebas objetivas administradas a miles de alumnos. Los investigadores y administradores de la educación de doce países desarrollados acordaron en 1958 hacer comparaciones internacionales masivas (Suter, 2001). Estas comparaciones permitieron estimar las brechas en los niveles de instrucción alcanzados en los países. El momento era oportuno porque muchos países de menor desarrollo estaban completando su capacidad de atender a los alumnos de familias con menos recursos de las zonas urbanas. Estos niños tenían problemas para ser promovidos y se acumulaban los alumnos con extra-edad en los primeros grados del sistema.

2. DEL ACREDITAR CONOCIMIENTOS (CLASIFICAR ALUMNOS) AL OPTIMIZAR APRENDIZAJES

Hasta los años 60 era parte de la responsabilidad profesional del maestro el promover un alumno al curso siguiente sólo cuando estaba seguro que podía aprender al nivel esperado en ese curso (y no se sentía responsable por los que no alcanzaban ese nivel mínimo). El maestro evaluaba al alumno a través de Pruebas de Conocimientos o Capacidad (preguntas escritas a las que el alumno daba una respuesta escrita) y observaciones personales. El criterio subjetivo de cada maestro determinaba quienes quedaban clasificados para aprobar el curso y quienes debían repetirlo. En Chile sólo era posible comparar los niveles de aprendizaje al rendir la Prueba de Bachillerato (administrada hasta 1966 por la Universidad de Chile) una vez que se completaba el nivel secundario. Pero sólo un 5 a 10% de cada grupo de edad rendía esa prueba.

Entre 1965 y 2001 la IEA realizó mediciones de lo que aprendían los alumnos en matemática (en 1965, 1982, 1995 y 1999); ciencias naturales (1970, 1986, 1995 y 1999); y lectura (1970, 1991 y 2001) lo que permitió estimar las brechas que existían entre países desarrollados y en desarrollo (Suter, 2001). La identificación de las brechas estimuló el análisis de los factores asociados a los rendimientos y, a continuación, se realizaron estudios para identificar estrategias de enseñanza efectivas (Coleman *et al.*, 1966; Schiefelbein y Simmons, 1979; Velez *et al.*, 1992; Schiefelbein, 1994). Este tipo de estudios produce, gradualmente, una separación entre los que ejecutan programas educativos y los que los evalúan (McKinsey, 2007:36).

Los estudios sobre escuelas y maestros efectivos de los años 70 llevaron a medir el logro de conocimientos y habilidades específicas (por ejemplo, capacidad de identificar la idea principal de un texto de 100 palabras) mediante pruebas referidas a criterios específicos (Bond, 1996). La selección de ítems de comprensión de lectura ha permitido constatar que en América Latina la mitad de los niños de tercer o cuarto grado no logra entender los mensajes de textos breves y que esta dificultad estaría asociada a insuficiente fluidez de lectura y a un vocabulario relativamente reducido. En la medida que se han detectado estos problemas, las pruebas nacionales se tienden a concentrar en esclarecer las características del problema y dejan de medir otros aspectos de la lecto-escritura como ortografía, gramática o morfosintaxis.

La evaluación referida a norma (clasificar a los alumnos de acuerdo al puntaje en una prueba) vuelve a tomar importancia cuando se constata que en los establecimientos que atienden a niños de nivel socioeconómico bajo sólo se estudia parte del currículo. Esto hace que esos alumnos tengan bajos puntajes en las pruebas de ingreso a la universidad. Sin embargo, los mejores alumnos de cada curso (el 10% con más alto rendimiento) tienden a lograr excelentes rendimientos (en el siguiente nivel) a pesar de sus bajos puntajes en las pruebas de conocimiento (Gil, 2005). Se recupera —mediante esta manera de

seleccionar a los alumnos— a jóvenes de talento excepcional, que por falta de una instrucción adecuada (en el nivel anterior) quedarían marginados de continuar estudiando en el nivel superior.

Pero los resultados de las pruebas no permiten avances significativos en los aprendizajes, ya que los usuarios enfrentan numerosos obstáculos. Algunos ejemplos permiten ilustrar los tipos de obstáculos: (i) distinguir entre acertar respuestas y saber las respuestas; (ii) diferenciar las probables “causas” de un cambio en una variable independiente, de las asociaciones o correlaciones espurias o (iii) estar alertas frente a procesos de “desorientación” en los medios de prensa, por ejemplo, distinguir entre los aportes positivos y negativos de las Tecnologías de la Información y la Comunicación (TIC). Muchas veces los resultados solo permiten suponer soluciones alternativas (Schiefelbein, 2005a, 2004a).

Los que toman decisiones (que no son expertos en el tema) no reciben el mensaje que menos de la mitad de los alumnos de cuarto grado entienden lo que leen y muchos quedan con la sensación que “unos dos tercios” comprende lo que lee. Esto se genera porque los resultados (para el nivel primario) se suelen difundir en términos de porcentajes “brutos” de respuestas correctas. Si en un grupo de 100 alumnos hay 52 que responden correctamente (tasa “neta” de 52% que refleja los que “saben” la respuesta), los 48 restantes (en promedio, cuando hay 4 alternativas) acertarán doce veces, por lo que la tasa bruta llegará al 64% de aciertos. En el SIMCE 2006 el 61,4% acertó con la idea principal del texto de dificultad “intermedia”, pero sólo el 48,7% “sabía” cual era la idea principal (SIMCE, 2007b). Este resultado dista mucho de “un buen rendimiento” donde cada uno de los niños tiene éxito (McKinsey, 2007,37).

También la difusión de resultados suele “desinformar o alejar a la opinión pública de los verdaderos problemas” (Brunner, 2007). Esto ocurre, por ejemplo, cuando los titulares de los medios de prensa dicen que “la educación privada logra 60 puntos más que la educación municipal lo que refleja su mejor calidad”. Lo que no mencionan los medios de prensa es que cuando los resultados se controlan por el nivel socioeconómico de la familia no hay diferencias en los puntajes (el nivel alto logra unos 300 puntos, el medio obtiene 250 y el bajo entre 220 y 230). La conclusión es inversa (a lo afirmado en la prensa): no hay diferencia de calidad entre los establecimientos (privados y públicos) y, en cambio, los profesores no están preparados para compensar las diferencias en los niveles iniciales de educación (0 a 6 años).

Los mensajes sobre las TIC suelen desorientar al lector desprevenido. Se señala, por ejemplo, que los nuevos audífonos o los mensajes de texto y fotografía de los teléfonos celulares han elevado la cantidad de “copia” o fraude en las pruebas. El mensaje es que el problema estaría generado por la tecnología. Sin embargo, el verdadero problema está generado por el énfasis en que los alumnos memoricen información en vez de que aprendan a utilizarla. En este caso la tecnología ayuda a encontrar la verdadera solución: que las preguntas impliquen un mejor nivel de aprendizaje (por ejemplo, aplicación, comparación, inferencia o juicio crítico, según la taxonomía de Bloom) que no es posible copiar de un libro o un mensaje de texto. En realidad la mayor parte de las pruebas se podrían hacer con derecho a consultar los libros y apuntes (pruebas con libro abierto).

Finalmente, el uso de la información que proporcionan las evaluaciones, para mejorar la educación, implica un uso profesional del conocimiento acumulado hasta ahora, tomando en cuenta sus diversos niveles de confiabilidad (Ravela, 2005; Chapman y Mahlck, 1993). Por ejemplo, si el problema fundamental es el bajo nivel de comprensión de lectura, será necesario examinar las revisiones de las investigaciones pertinentes (Show *et al.*, 1998; Abadzi, 2006; Oliveira, 2006; Bloom, 1976). En muchos casos usar la información que proporcionan las evaluaciones implicará reanalizar los datos con teoría o modelos más adecuados (por ejemplo, puntajes de un grupo de ítems de comprensión de lectura o nivel de vocabulario) o realizar estudios complementarios de los alumnos (por ejemplo medición de aptitudes y personalidad) o de otros aspectos, por ejemplo, examinar características de los profesores o de sus formadores como maestros (Schiefelbein *et al.*, 2007; Schiefelbein, 2004b; Schiefelbein, 1994). En muchos casos se trata de entender el contexto en que ocurren los procesos y las interacciones entre los

diferentes elementos mediante una o varias visitas de un grupo de especialistas a una escuela específica cada cierto número de meses o años, que puede ser menor para las escuelas con mejores indicadores de rendimiento. Esto ha permitido reducir a la mitad el tiempo que se dedica a evaluar en Inglaterra (McKinsey, 2007:37).

Este considerable aumento de los instrumentos disponibles para evaluar y de información acumulada sobre procesos de enseñanza y aprendizaje no debe hacernos olvidar que: (i) evaluamos en función de ciertos objetivos que se consideran importantes y (ii) no siempre es posible medir con precisión los objetivos importantes. Es por esto que se evalúa, fundamentalmente, la instrucción, y pone menos atención a la "educación propiamente tal" (porque no podemos medir con facilidad actitudes y procesos de decisión).

Por ahora todavía es útil la evaluación de niveles de instrucción (tal como antes fue la declaración en los censos o la cobertura), pero no hay duda que se necesita más precisión para medir procesos de formación de personas (Schiefelbein, 2003a). Es por esto que se debe revisar qué es lo que realmente queremos obtener cuando evaluamos. ¿Nos interesa saber más, mejorar el mundo (ciencia o técnica) o desarrollarnos como personas?

3. DEL EVALUAR CONOCIMIENTOS (INSTRUIR) A FORMAR PERSONAS (EDUCAR)

A mediados de los 50 se pensaba que el problema de la educación consistía en construir más escuelas y contratar más profesores (y por eso se evaluaba el sistema con los indicadores comentados en el punto 1). Se esperaba que, eventualmente, los niños que estudiaran en esas escuelas se graduaban con niveles similares a los de los alumnos que se graduaban en los 50. Los Bancos internacionales de fomento prestaron fondos y los países ampliaron su cobertura. Pero la ampliación implicó el ingreso de alumnos cuyas familias quedaban por debajo del nivel socioeconómico promedio del país (hasta los años 50 se atendía sólo a la mitad de los niños en edad escolar de primaria cuyas familias quedaban por encima del nivel socioeconómico promedio del país). Los recién ingresados tenían un menor vocabulario, asistían, en promedio, menos días a clases que los de familias con más recursos —la necesidad de cuidar a hermanos o la casa, el caminar mayores distancias (y a veces con ropa inadecuada), la desnutrición y menor protección de la salud o la desorganización familiar están asociadas a mayores inasistencias a clases— y, en general, sus aprendizajes eran menores que los del grupo anterior.

En las última dos décadas, entonces, el problema ha sido como lograr ciertos niveles mínimos de instrucción, principalmente la capacidad de leer y entender los mensajes fundamentales de un texto y realizar ciertos cálculos aritméticos básicos. La etapa siguiente podría ser la capacidad de usar la información (más que la capacidad de recordar datos o aplicar algoritmos que se mide en la actualidad) y en ese caso se usarían pruebas para medir aptitudes (por ejemplo, pruebas con libro abierto, quizá con el estímulo de la tecnología que facilitad el "copiar" información a distancia mediante audifonos o celulares). Pero esto está muy lejos de un esfuerzo de cambiar lo que pasa en las mentes y corazones de los niños (McKinsey, 2007:5).

En la medida que la educación (y la evaluación) se ha reducido, en gran medida, a una mera instrucción, es necesario recuperar *"un justo equilibrio entre la libertad y la disciplina. . . [en el que] . . . la educación lograda es una formación para el uso correcto de la libertad. Hay que aceptar el peligro de la libertad, pero con la corrección de las ideas y decisiones equivocadas, sin apoyar en los errores o fingir que no los hemos visto o, aún peor, que los compartimos, como si fueran las nuevas fronteras del progreso humano"* (Benedicto XVI, 2008). Esto implica empezar a medir lo que se logra en "formación" (porque de otra manera sólo se pone atención a la "instrucción" por ser lo que se mide).

¿Qué indicadores usar para saber cuál es el nivel actual o el cambio en el uso de la libertad? ¿Queremos medir con precisión o saber la magnitud del problema? ¿Cuál es el costo de usar los indicadores de formación? ¿Es posible grabar clases, discusiones o dramatizaciones --en las cuales los participantes representan teatralmente, el papel del ocupante de un determinado rol en la clase-- y luego re-analizar la actuación de los participantes (un psicólogo puede colaborar en analizar el desempeño de cada participante)? ¿Sería posible usar pruebas psicométricas para evaluar aptitudes o rasgos innatos de los participantes a fin de tratarlos y transformarlos en habilidades o capacidades personales para tomar decisiones razonadas, conforme a la conciencia moral? ¿Hay elementos de selectividad en los instrumentos (que generen nuevos problemas de inequidad)? En todo caso es posible visitar las clases y obtener indicadores de los procesos de enseñanza relacionados con formación (Schiefelbein, 2005b).

Por ahora, cuando la mitad de los niños de cuarto grado de Chile no logra comprender textos simples, las alternativas de evaluación son más simples y se analizan en la última sección de este artículo.

4. CÓMO USAR HOY EN CHILE LOS RESULTADOS DE LAS EVALUACIONES: ¿CASTIGAR O AYUDAR?

Desde fines de los 60 se ha medido el nivel de aprendizaje de los alumnos para elevarlo, pero no ha sido evidente la forma de hacerlo y, de hecho, no ha aumentado. La prueba de 8° grado (1967-1971) se limitó a mostrar a los profesores los tipos de habilidades que debían estimular en sus alumnos. La PER, creada a principios de los '80, buscó determinar el nivel del servicio educativo que se ofrecía y "monitorearlo" con ayuda de los padres, que tratarían de seleccionar las mejores escuelas para sus hijos. El SIMCE midió, a partir de 1988, los rendimientos de las escuelas y trató de identificar factores que pudieran explicar las diferencias y evaluar el impacto de los programas de los municipios y del MINEDUC.

Sin embargo, no hay avances y conviene repensar este poderoso instrumento, que es el SIMCE, si se quiere mejorar los aprendizajes. Un rol alternativo del SIMCE sería, según algunos expertos, ayudar a los profesores a enseñar mejor y revisar cuáles son los temas importantes del currículo, más que el "clasificar" rendimientos. Este segundo rol plantea diferencias nítidas con respecto al primero: ¿Seleccionar o desarrollar? ¿Poner nota o medir logro de meta? ¿Conocer el ranking o el aprendizaje de objetivos específicos (criterio)? ¿Castigar o ayudar? "Son dos funciones diferentes (los especialistas hablan de Evaluación Sumativa y Formativa, respectivamente). Si la primera no elevó hasta ahora los rendimientos, convendría probar la segunda (Schiefelbein, 2003b).

Esto implica fuertes cambios en el SIMCE. En efecto, para "seleccionar" o "castigar" basta calcular un puntaje total que "discrimine entre buenos y malos" y difundirlo públicamente. En este rol es fundamental no divulgar los ítems utilizados para que mantengan su poder de discriminación (no ser conocidos por los profesores y evitar que preparen a sus alumnos para contestarlos mecánicamente). Basta calcular el puntaje de cada escuela para que los padres seleccionen entre escuelas o preparar datos e indicadores globales para que los funcionarios del MINEDUC modifiquen las estrategias.

En la segunda función, en cambio, el usuario de la información del SIMCE es el profesor en su sala de clases. En efecto, para "desarrollar" o "ayudar a mejorar" se necesita entregar información detallada a cada profesor sobre los aspectos en que cada uno de sus alumnos logró los niveles adecuados y aquellas habilidades o conocimientos que todavía no posee o no domina suficientemente. El profesor debe saber lo que contestó cada alumno en la prueba y revisar con cada uno los errores que cometió, hasta que el alumno internalice las deficiencias y reforme adecuadamente sus procesos de pensamiento. Sin información detallada el profesor no puede identificar los aspectos de su enseñanza que debe cambiar o cómo ayudar a cada estudiante... *"En resumen, una redefinición clara del objetivo de la prueba nacional de medición de la calidad de la educación debe orientar la reflexión del grupo de expertos sobre los cambios que debe tener el SIMCE..."*



LAS EVALUACIONES EDUCATIVAS QUE AMÉRICA LATINA NECESITA

*Pedro Ravela, Patricia Arregui, Gilbert Valverde, Richard Wolfe,
Guillermo Ferrer, Felipe Martínez Rizo, Mariana Aylwin y Laurence Wolff¹*

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art4.pdf>

Referencia original:

Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez, F., Aylwin, M. y Wolff, L. (2008). Las Evaluaciones Educativas que América Latina Necesita. Documento de Trabajo N. 40. Santiago de Chile: PREAL.

¹ Los autores de este documento son los miembros del Comité Gestor del Grupo de Trabajo sobre Estándares y Evaluación del PREAL.

1. ¿POR QUÉ SON IMPORTANTES LAS EVALUACIONES NACIONALES DE LOGROS EDUCATIVOS?

La realización de evaluaciones estandarizadas como forma de conocer mejor la dinámica de procesos y resultados en los sistemas educativos es cada vez más frecuente a nivel regional y mundial, en países de muy diversas culturas y orientaciones ideológicas de gobierno. Prueba de ello es la creciente participación de los países en las evaluaciones internacionales como PISA, TIMSS y PIRLS, y regionales como SERCE (en América Latina) y SACMEQ (en África), así como el desarrollo de diferentes tipos de sistemas nacionales y subnacionales de evaluación.

En algunos casos, la evaluación está motivada por una preocupación por la formación ciudadana y la consolidación de una sociedad democrática; en otros, por la productividad de la fuerza de trabajo y la competitividad de la economía nacional, por las oportunidades para el desarrollo integral de las personas y sus posibilidades de participación en la sociedad del conocimiento, o por la equidad y la visión de la educación como uno de los caminos para superar la pobreza. Muchos sistemas de evaluación parten de una combinación de los intereses anteriores. En casi todos los casos se asume que la evaluación puede servir:

- como base para adoptar políticas educativas mejor fundamentadas;
- para mejorar la gestión de los sistemas educativos;
- como instrumento para la colaboración y el aprendizaje continuo al interior de los mismos.

A continuación se destacan los aportes principales de este tipo de evaluaciones.

La evaluación estandarizada ayuda a visualizar los resultados educativos del conjunto de estudiantes

La educación es una actividad “opaca”, en el sentido de que sus resultados no son directa ni inmediatamente observables. Un buen docente puede apreciar si sus estudiantes están aprendiendo o no y cómo, pero no todos los docentes tienen los mismos criterios de valoración, los cuales están fuertemente vinculados a su experiencia profesional: su formación general y específica, su conocimiento de la disciplina que enseña, su capacidad para percibir los procesos y dificultades de los estudiantes, su familiarización con diversos tipos de alumnos, etc. Dada la heterogeneidad del cuerpo docente de un país, no es posible obtener una visión de la situación del conjunto por la mera agregación de los puntos de vista individuales. La evaluación estandarizada apunta a lograrlo.

Las evaluaciones entregan información sobre el real acceso al conocimiento y a las capacidades que alcanzan los alumnos, más allá de la cantidad de años de estudio que tengan

En décadas pasadas, la equivalencia entre permanencia en el sistema educativo y acceso al conocimiento y a las capacidades simbólicas se daba por descontada y los indicadores para valorar a los sistemas educativos estaban relacionados con el acceso (matrícula, cobertura, retención, etc.). En esos tiempos, las mayorías más pobres y con menor capital cultural no accedían al sistema educativo o apenas cursaban unos pocos años de educación básica. Con la progresiva universalización del acceso al sistema educativo, hoy muchos estudiantes procedentes de los sectores sociales más desfavorecidos llegan en desventaja en términos de manejo del lenguaje oral (oficial) y escrito, con lo cual la equivalencia entre años de estudio y acceso al conocimiento y a las capacidades simbólicas ya no es tal. Por otra parte, el crecimiento del acceso a la profesión docente no fue acompañado por garantías de calidad en su formación, lo cual también cuestiona la mencionada equivalencia. Así, más años en el sistema educativo no necesariamente significa que todos los niños y jóvenes estén incorporando los conocimientos, actitudes y capacidades necesarias para la vida personal y social, los que son cada vez más sofisticados y complejos. Las evaluaciones buscan dar luces sobre lo que está ocurriendo al respecto.

Las evaluaciones estandarizadas ayudan a hacer visibles un conjunto de aspectos centrales de la labor educativa

Entre otras cosas, las evaluaciones estandarizadas aportan información sobre:

en qué medida los alumnos están aprendiendo lo que se espera de ellos al finalizar ciertos grados o niveles;

- cuál es el grado de equidad o inequidad en el logro de dichos aprendizajes;
- cómo evolucionan a lo largo de los años, tanto los niveles de logro como la equidad en el acceso al conocimiento por parte de los diversos grupos sociales;
- en qué medida y cómo las desigualdades sociales y culturales inciden sobre las oportunidades de aprendizajes de los estudiantes;
- cuál es la diversidad de prácticas educativas existentes en escuelas y maestros y cómo las mismas se relacionan con los aprendizajes de los estudiantes en diversos contextos sociales;
- cómo influyen las condiciones de la enseñanza (situación de los maestros, recursos disponibles, tiempos de estudio, etc.) en los progresos de los alumnos;

qué efecto tienen en los logros educativos las inversiones en programas educativos, los cambios en la estructura del sistema, los cambios curriculares, los programas de formación, la adquisición de material educativo, etc.

Un sistema de evaluación de aprendizajes y/o logros educativos puede aportar información importante diversos actores sociales

En la medida en que el sistema produzca y comunique adecuadamente información sobre los aspectos antes indicados, puede ser un instrumento clave de mejora, enriqueciendo la comprensión de la situación educativa y la toma de decisiones en diversos ámbitos:

- *Las autoridades y los encargados de formular políticas educativas* pueden comprender mejor los problemas de la enseñanza y del aprendizaje; hacerse cargo de las carencias en que se desarrolla la labor docente y desarrollar políticas pertinentes para apoyar el trabajo de las escuelas. Las evaluaciones también les permiten valorar, sobre una base de evidencia empírica sólida, el impacto de las políticas y programas que han impulsado y los probables efectos de las que se proponen impulsar.
- *Los directivos y docentes* pueden, a partir de una mirada externa sobre los logros educativos en el conjunto del sistema, comprender mejor lo que están logrando y lo que no están logrando sus propios estudiantes, cómo están aprendiendo y qué dificultades tienen. Pueden aprender de las experiencias de otros docentes y escuelas que trabajan con estudiantes de características tanto similares como diferentes a los propios. Pueden tomar decisiones más apropiadas acerca de qué aspectos del currículo enfatizar y enriquecer sus propios modos de evaluar el aprendizaje de sus estudiantes.
- *Los supervisores y encargados de la formación de docentes* pueden apoyarse en la información sobre los logros y dificultades educativas en el conjunto del sistema para estudiar en profundidad las debilidades en el enfoque o en la práctica de la enseñanza que probablemente generan algunas de las insuficiencias constatadas en los aprendizajes. Desde esa nueva perspectiva, pueden mejorar su labor tanto de orientación a los docentes como de formación de los mismos. En particular, los supervisores pueden beneficiarse de contar, como instrumento de apoyo para

su labor, con un mapa de las escuelas que las caracterice tanto en términos de la composición sociocultural de su alumnado como de sus logros educativos.

- *Los padres y madres de estudiantes*, adecuadamente informados, pueden comprender mejor qué se espera que sus hijos aprendan, qué es lo que están logrando y qué pueden hacer para colaborar con la escuela y con el aprendizaje de sus hijos.
- *La ciudadanía* en general estará mejor informada sobre lo que acontece al interior del sistema educativo y, por tanto, estará más atenta a los temas y problemas de la educación. Estará, además, en mejores condiciones para exigir, tanto a los poderes públicos como a los profesionales de la docencia, la mejora continua de la educación que se brinda a los niños, niñas y jóvenes, y el uso responsable de los recursos que se destinan a la educación.

El desarrollo de un sistema nacional de evaluación estandarizada puede obligar a realizar una discusión informada sobre qué aspectos del currículo formal son exigibles a todos los estudiantes y a definir con claridad qué es lo que todos los estudiantes deberían haber aprendido al final de cada ciclo educativo.

La mayoría de los currículos en la región están constituidos por largas listas de objetivos y temas, todos deseables pero no todos realizables. La elaboración de pruebas nacionales obliga a definir qué es lo que debe ser considerado como fundamental y, por tanto, lo que todos los estudiantes deben saber y ser capaces de hacer. Se pueden usar distintos términos para designar a estas definiciones: estándares, competencias fundamentales, indicadores de logro, niveles de desempeño, metas de aprendizaje, criterios de suficiencia, entre otros.

PARA NO CONFUNDIR...

Al desarrollar estos sistemas nacionales de evaluación, es necesario explicitar tanto el rol que estos cumplen así como las funciones que no cumplen y cuáles son sus limitaciones, para evitar así los riesgos de mala interpretación respecto a lo que son y la información que recogen y difunden.

- **Una evaluación estandarizada aporta información fundamental e indispensable sobre la "calidad educativa", aunque no es un indicador completo de la misma.** No todos los objetivos valiosos de la educación están incluidos en este tipo de evaluaciones. Hay una gran cantidad de saberes, actitudes, valores y aprendizajes relevantes que no pueden – por la dificultad de medirlos de esta manera – o no deben – porque son propios de cada entorno local y, por tanto, no son exigibles a todos los estudiantes del país – formar parte de una evaluación estandarizada.
- **La evaluación estandarizada de aprendizajes y/o logros educativos es un componente esencial de un sistema integral de evaluación educativa, pero no es el único tipo de evaluación relevante.** También tienen importancia la evaluación realizada en el aula por los docentes, la evaluación del desempeño docente, la evaluación de los centros educativos, la evaluación de las políticas educativas, la evaluación del uso de los recursos, la evaluación de la relevancia del currículo, etc.
- **La evaluación es condición necesaria, pero no suficiente, para mejorar la educación.** Si bien existe alguna evidencia de que la mera existencia y difusión de información tiene algún impacto sobre ciertos actores, es necesario recordar que la evaluación es apenas uno de varios elementos clave de la política educativa que debe procurarse estén adecuadamente alineados: la formación docente inicial y en servicio, la carrera y condiciones para el trabajo docente, la gestión escolar, la supervisión, los diseños curriculares, los libros y materiales educativos, una inversión de recursos proporcional a las necesidades de las diversas poblaciones y una acción decidida por parte de los responsables educativos para resolver los problemas detectados, entre otros.
- **La evaluación estandarizada externa solo tendrá efectos positivos sobre la educación si es concebida, percibida y empleada como un mecanismo de responsabilización pública de todos los actores vinculados al quehacer educativo.** Existe siempre el riesgo de que la política educativa se concentre en la implementación de evaluaciones, pero que luego no se tomen acciones concretas para enfrentar y resolver los problemas que las mismas ponen de manifiesto. Muchas veces las autoridades se limitan a informar de los resultados y transferir toda la responsabilidad por la solución de los problemas detectados a las escuelas y familias, como si se tratase de un asunto privado entre ambas. Otras veces, toda la responsabilidad se atribuye a los docentes, quienes, por su parte, tienden a transferir toda la responsabilidad a los padres o al contexto. Lo importante es evitar ingresar en una dinámica de culpabilización y, por el contrario, intentar construir una lógica de responsabilidad compartida en torno a la educación.

2. LA SITUACIÓN DE LA EVALUACIÓN ESTANDARIZADA EN AMÉRICA LATINA

Los sistemas nacionales de evaluación estandarizada se desarrollaron con fuerza en toda la región durante los años 90. Algunos países han mantenido sus sistemas en funcionamiento con continuidad, aun cuando se produjeran cambios de enfoque o variaciones en la institucionalidad de los mismos. Otros países han tenido importantes discontinuidades y han tenido que comenzar prácticamente desde cero en repetidas oportunidades o tendrán que hacerlo en el futuro próximo. Además de ello, actualmente:

- Dieciséis países están participando en el Segundo Estudio Regional en 3° y 6° de primaria que lleva adelante la OREALC/UNESCO: Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay.
- Seis de estos países participan en PISA 2006 y otros tres se incorporarían al ciclo PISA 2009.
- Algunos países de la región han participado además (o lo están haciendo) en los estudios de Matemática y Ciencias (TIMSS), Lectura (PIRLS) y Educación Cívica que lleva adelante la *International Association for the Evaluation of Educational Achievement* (IEA).
- Durante los últimos 12 años ha estado activa en la región la red de los sistemas de evaluación organizada en torno al Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de OREALC/UNESCO, en la que participan actualmente casi todos los países latinoamericanos. Esta red se reúne dos veces al año y es un espacio de intercambio de experiencias y formación en el ámbito de la evaluación.
- Más recientemente se ha constituido el Grupo Iberoamericano de PISA, integrado por Argentina, Brasil, Chile, Colombia, España, México, Portugal y Uruguay.

Las evaluaciones a gran escala en la región no solo son más frecuentes sino también están mejorando

Los principales cambios y mejoras que se observan en los últimos años incluyen:

- *Mayor transparencia en la difusión de los resultados.* Parece haberse superado la etapa en que las autoridades en varios países interferían en la divulgación de los resultados de las evaluaciones cuando estos no les parecían favorables.
- *Tendencia creciente a pasar de pruebas normativas –que tienen como propósito principal ordenar comparativamente a los estudiantes– a pruebas de criterios, que se enfocan en qué es lo que los estudiantes saben y son capaces de hacer.* Asimismo, crecientemente se incluye en las pruebas de criterios una definición de cuál es el resultado que todos los estudiantes deberían alcanzar para que su desempeño pueda ser considerado satisfactorio.
- *Mejoría de las capacidades técnico-metodológicas para la construcción de pruebas y para el procesamiento de los datos.* Existen esfuerzos por desarrollar pruebas que evalúen un rango de conocimientos y capacidades más amplio y que incluyan preguntas de respuesta construida. Se han ido incorporando también metodologías más sofisticadas para el tratamiento de datos, como la Teoría de Respuesta al Ítem y el Análisis Multinivel.
- *Creciente atención a la difusión y uso de los resultados.* Se ha ido comprendiendo que no basta con implementar una evaluación y publicar un informe, sino que es necesario desarrollar una estrategia de divulgación y un conjunto de reportes apropiados a cada una de las audiencias a las que se quiere llegar. Las Unidades de Evaluación tienen una mayor conciencia acerca de los usos apropiados e inapropiados para cada tipo de evaluación, aunque esto muchas veces no está suficientemente claro entre quienes toman decisiones de política.

- *Mayor preocupación por la investigación de los factores que inciden sobre los aprendizajes.* Si bien falta mucho por hacer en términos metodológicos e interpretativos para producir buenos trabajos de investigación, en la mayoría de los países hay conciencia de que es preciso avanzar en la investigación y la construcción de hipótesis sobre cómo las dinámicas de la gestión de los sistemas educativos, los procesos escolares, las prácticas de enseñanza y las decisiones de política educativa influyen en las oportunidades de aprendizaje de los niños y niñas.
- *Creciente participación de los países evaluaciones internacionales.* Esto ha tenido consecuencias positivas en las Unidades de Evaluación y en la calidad de su trabajo: ha contribuido a la construcción y acumulación de capacidad técnica en evaluación, ha facilitado el intercambio y el desarrollo de un lenguaje común entre países y ha ayudado a mejorar la calidad de diversos procesos técnicos (elaboración de pruebas, incorporación de preguntas de respuesta abierta, muestreo, control de las aplicaciones, análisis de datos, modos de difusión, entre otros).

Las principales debilidades presentes en los sistemas de evaluación estandarizada en la región se observan en la estrategia de difusión, en el uso de los resultados y en la calidad técnica de las evaluaciones

A pesar de las mejoras previamente mencionadas, persiste un conjunto importante de debilidades que es preciso abordar:

- Es frecuente que desde el nivel político se demande la implementación de grandes operativos de evaluación sin propósitos claramente definidos y en tiempos y con recursos absolutamente insuficientes para hacerlo técnicamente bien.
- Aún no se tiene claro que no cualquier evaluación sirve para cualquier propósito y que es necesario diseñar cuidadosamente el sistema de evaluación para el largo plazo, en función de propósitos y usos claramente establecidos. Esto es determinante para las decisiones relativas a la utilización de censos o muestras, a los grados y disciplinas que serán evaluados, la periodicidad de las evaluaciones, el tipo de pruebas y escalas de reporte, entre otras.
- Falta una mayor discusión pública sobre qué debe ser evaluado y qué aspectos del currículo deben haber sido enseñados a todos los estudiantes al concluir determinados niveles educativos. Como resultado de ello, las metas y/o estándares de aprendizaje no son claros. Mientras no exista claridad al respecto, tampoco existirá claridad ni para la enseñanza ni para la evaluación. Esta tarea requiere de un amplio debate social y de un trabajo técnico coordinado entre las unidades de currículo y de evaluación.
- Para mejorar la investigación sobre los factores que explican los resultados, falta mejorar mucho los instrumentos complementarios que se aplican junto con las pruebas y concebir diseños de investigación más ambiciosos: estudios longitudinales, estudios de "valor agregado" y estudios cualitativos, entre otros.
- Los Ministerios de Educación han tenido, por lo general, una limitada capacidad para concebir, formular e implementar políticas que den respuesta a los problemas identificados en las evaluaciones. Falta para ello mayor articulación entre las Unidades de Evaluación y otros actores educativos relevantes, dentro y fuera de los Ministerios de Educación, y mayor articulación entre la evaluación, el desarrollo curricular, la formación inicial y el desarrollo profesional de los docentes.
- Las acciones para la divulgación y uso de los resultados de las evaluaciones aún son insuficientes, en particular para lograr que los docentes los comprendan y utilicen y para que los resultados se incorporen a la cultura escolar.

- En la mayoría de los países perdura la práctica de divulgar resultados de tipos de escuela, escuelas individuales, estados o provincias, sin una adecuada consideración de los contextos socioculturales en que operan las instituciones y subsistemas educativos. Esto conduce a interpretaciones y conclusiones erróneas sobre la eficacia educativa de dichas instituciones o subsistemas. Para ello es necesario realizar evaluaciones de "valor agregado", que impliquen dos mediciones sobre una misma población en distintos momentos. Sobre este aspecto la experiencia en la región es mínima.
- En muchos países persisten importantes debilidades técnicas en el diseño de las evaluaciones: excesiva simplicidad de la mayor parte de las preguntas y dificultad para elaborar preguntas que permitan evaluar capacidades cognitivas complejas; focalización de las evaluaciones en los logros educativos correspondientes a un grado específico, lo que impide saber qué han aprendido quienes aún no logran lo esperado para dicho grado; deficiencias en la conformación de las muestras y en la forma de estimar y reportar los errores de medición; y debilidades para establecer mediciones que sean comparables en el tiempo.
- No hay suficientes profesionales calificados para diseñar y conducir este tipo de evaluaciones, lo que se agrava por la falta de continuidad de los equipos técnicos de los países, muchas veces por razones de índole político-partidaria. Esto dificulta la acumulación de conocimiento y experiencia en la región y ha dado lugar a que muchos países tengan que comenzar desde "cero" con sus sistemas de evaluación, algunos años después de haber tenido uno en funcionamiento y haberlo desmantelado.

EL RIESGO DE LAS MALAS EVALUACIONES

Sólo una evaluación técnicamente buena y cuyos resultados sean conocidos y utilizados de manera apropiada, puede tener algún impacto en la mejora de los aprendizajes. Como es obvio, una evaluación cuyos resultados son poco conocidos y utilizados supone un ejercicio estéril y un desperdicio de recursos.

Más graves aún son aquellas situaciones en que evaluaciones técnicamente malas son ampliamente difundidas o en que las evaluaciones son empleadas para fines diferentes de aquellos para los cuales fueron concebidas y diseñadas, lo cual puede tener efectos dañinos para el sistema educativo. El razonamiento según el cual lo que importa es evaluar y difundir resultados como mecanismo de rendición de cuentas, sin importar la calidad de la evaluación, es una falacia que puede hacer retroceder los sistemas de evaluación e imposibilitar una discusión seria de las ventajas y desventajas de la rendición de cuentas o responsabilización por los resultados.

3. LOS PROPÓSITOS Y LOS USOS DE LOS RESULTADOS DE LAS EVALUACIONES ESTANDARIZADAS

Sea en la implementación de un sistema de evaluación o en su reforma, es necesario considerar algunas opciones básicas de acuerdo a lo que se espera de él.

El primer paso es definir cuál será el propósito del sistema de evaluación y para qué se usarán sus resultados

Es preciso considerar diversas opciones, que no son necesariamente excluyentes. Un sistema puede combinar varias de ellas, pero es importante tener conciencia de que cada opción requiere de un determinado diseño y tiene exigencias técnicas y de costos diferentes. Se debe analizar si el sistema nacional de evaluación será:

- *de carácter diagnóstico* (para alumnos, escuelas o sistema educativo), sin consecuencias directas para los estudiantes, cuya finalidad principal es enriquecer con información de calidad las percepciones, decisiones y acciones de diversos actores del sistema –autoridades y cuerpos

técnicos, supervisores, directivos, docentes, estudiantes y sus familias— para mejorar la enseñanza y el aprendizaje.

- *de certificación de los logros educativos de los estudiantes*, cuyo propósito central es establecer quiénes han alcanzado los conocimientos y desempeños necesarios para aprobar un curso o nivel y, en consecuencia, aprueba o reprueba.

Es necesario también establecer un adecuado balance entre la implementación de pruebas nacionales y la participación en pruebas internacionales

Las pruebas nacionales pueden ofrecer una mirada más apropiada acerca de qué aprenden los estudiantes de aquello que se les está enseñando, mientras que las pruebas internacionales pueden servir para ubicar la situación del país en relación a lo que saben y son capaces de hacer los estudiantes en otras sociedades y enriquecer el debate sobre el currículo nacional y el enfoque de la enseñanza. Los países deberían analizar cuidadosamente en qué pruebas internacionales participar, teniendo en cuenta lo que cada una pretende evaluar y su pertinencia en relación a los objetivos nacionales. Parece recomendable participar periódicamente en al menos una evaluación regional o internacional, en el marco de la política de prioridades que cada país defina.

Evaluaciones para la certificación de los aprendizajes de los estudiantes

Un sistema de evaluación para certificar los aprendizajes de los estudiantes a través de exámenes nacionales de alta calidad técnica tiene ventajas asociadas a la transparencia y responsabilidad por los resultados

Hoy es común que dos estudiantes que han aprobado un mismo nivel educativo en dos regiones distintas de un país, posean en realidad niveles de conocimiento muy diferentes. Los sistemas de evaluación para la certificación dan transparencia al valor de los certificados educativos ante la sociedad. Además, dichos sistemas hacen responsables tanto a los docentes como a los propios estudiantes por alcanzar los conocimientos y capacidades que evalúa el examen, lo cual tiene impactos positivos sobre el aprendizaje. En todo caso, la evaluación externa con propósitos de certificación es más apropiada para los ciclos superiores del sistema educativo, sobre todo hacia el final de la educación media.

Un sistema de evaluación para la certificación implica algunos requerimientos que son costosos

Por una parte, las pruebas deben ser de carácter censal y cada estudiante debe tener más de una oportunidad para rendirla, lo cual requiere implementar varios operativos de evaluación cada año. Por otra parte, las pruebas deben tener una amplia cobertura curricular, lo cual requiere que sean extensas, en lo posible con preguntas de respuesta construida (que tienen un importante costo de codificación) y deben abarcar varias asignaturas o disciplinas.

Un sistema de evaluación para la certificación puede generar tensiones importantes que es necesario anticipar

Si las pruebas son exigentes, ello puede conducir a niveles de reprobación muy altos, que afectarían principalmente a los sectores sociales más vulnerables y podrían agudizar problemas como la deserción y el desempleo juvenil, lo cual las haría insostenibles desde el punto de vista social y político. Por ello, es preciso acompañar este tipo de evaluaciones (en realidad todo tipo de evaluaciones, pero estas en particular) con acciones pedagógicas de carácter complementario y compensatorio. Proponer estándares altos implica asumir responsabilidad sistémica por proporcionar todas las oportunidades para aprender: textos, materiales, instalaciones y preparación de maestros, entre otros. Existen diversos caminos para enfrentar este problema:

- Establecer que la evaluación estandarizada constituya solo una parte del proceso de certificación de aprendizajes de los estudiantes (por ejemplo, el 40% de su calificación final). El resto de la nota correspondería a las evaluaciones que realizan los docentes a cargo de los estudiantes. Si bien en este caso se mantiene el problema respecto a la heterogeneidad de los criterios de evaluación empleados por los docentes, es una forma de combinar la evaluación externa y la interna, e incluso de aportar a los docentes una mirada externa que les permita reflexionar sobre sus propios criterios.
- Establecer un periodo de transición hacia estándares únicos y universalmente aplicables. Durante dicho período, el foco principal se centra en la mejoría o progreso de los estudiantes de cada escuela respecto a su situación anterior, y no únicamente en el cumplimiento del estándar absoluto de certificación.

La peor respuesta que podría darse a este problema sería establecer un porcentaje fijo de reprobación, por ejemplo, no debe reprobado más del 10% de los estudiantes. Ello implica utilizar pruebas más fáciles y, por lo tanto, enviar a estudiantes, familias y educadores una señal equivocada respecto a las expectativas de aprendizaje.

Evaluaciones de carácter diagnóstico y formativo

Un sistema de evaluación de tipo formativo y sin consecuencias directas para los estudiantes tiene ventajas en términos de costos, posibilidades de diseño y de establecimiento de estándares altos, y aporte a una cultura de evaluación

Los costos de aplicación de estas pruebas pueden ser menores, dado que pueden hacerse en base a muestras y aplicarse solo en algunos grados claves y cada cierto número de años. Por otra parte, las pruebas pueden diseñarse con carácter matricial, en que no todos los estudiantes responden a las mismas preguntas sino a bloques de las mismas, lo cual permite trabajar con una cantidad muy grande de preguntas y obtener un análisis más detallado de los distintos aspectos del currículo. Estos sistemas permiten definir estándares o niveles de expectativa exigentes sin que ello produzca un fracaso generalizado. Asimismo, ayudan a construir una cultura de la evaluación y a acumular capacidad técnica, de modo que, cuando se analice la posibilidad de establecer un sistema con consecuencias, existan las condiciones para hacerlo apropiadamente. Una de las alternativas a considerar entre los usos de carácter formativo es la de “liberar” una prueba completa para que sea aplicada en forma autónoma por parte de los maestros, con el fin de ayudarles a identificar dificultades de alumnos individuales y de enriquecer su repertorio de instrumentos de evaluación.

El principal problema de estas evaluaciones es que pueden carecer de impacto si no se toman algunas acciones complementarias

Las evaluaciones de carácter diagnóstico y formativo pueden resultar intrascendentes si no van acompañadas de una estrategia precisa y de una inversión importante para asegurar la divulgación de los resultados y su uso en acciones educativas posteriores, puesto que su efectividad se juega en que los diferentes actores reciban, comprendan y utilicen los resultados. Por esto hay que tener presente que:

- *Para que los resultados de este tipo de evaluaciones tengan impacto sobre las políticas educativas, se debe invertir tiempo en el análisis y discusión de los mismos por parte de diversas áreas del Ministerio de Educación y otros actores relevantes, en la comprensión de los problemas y deficiencias que los resultados ponen de manifiesto y en la concepción de acciones e inversiones apropiadas para hacerles frente. Las autoridades deben estar dispuestas a someter sus políticas y decisiones al escrutinio de la ciudadanía, para lo cual es necesario también invertir en la comunicación apropiada y permanente de los resultados a la opinión pública.*

- *Para que los resultados tengan impacto sobre las prácticas de enseñanza*, es imprescindible asimismo invertir tiempo en el análisis e interpretación de sus implicancias didácticas: si los estudiantes no son capaces de resolver cierto tipo de situaciones, ¿qué es lo que se está haciendo de manera insuficiente o inapropiada en las aulas y qué es lo que se debería hacer? Este tipo de análisis debe ser realizado tanto por especialistas en la didáctica de las áreas evaluadas como por los docentes, creando para este último efecto espacios permanentes y sistemáticos de formación en servicio y de trabajo colectivo al interior de las escuelas. Es muy importante que los maestros puedan analizar la mayor cantidad posible de ítems para poder identificar cuáles revelan un bloqueo importante para el desarrollo de nuevos conceptos o capacidades. Sin embargo, siempre será necesario conservar el carácter confidencial de una parte de los ítems para poder realizar mediciones comparables en el tiempo.
- *Para que los resultados tengan impacto sobre la motivación de los estudiantes y sobre la actitud de las familias hacia el aprendizaje escolar*, es necesario proveerles de información apropiada y comprensible acerca de lo que se considera imprescindible haber aprendido en cada grado o nivel educativo y de las acciones que pueden contribuir al logro de dichos aprendizajes.

EVALUAR CON UNA PERSPECTIVA AMPLIADA

Es muy importante que las pruebas no se limiten a evaluar los conocimientos y competencias correspondientes a un único grado (aquel en que se aplica la evaluación), sino que incluyan una perspectiva más amplia de niveles de desempeño —desde más básicos a más complejos— a lo largo de varios grados, de modo de identificar qué han aprendido los estudiantes en grados anteriores y qué necesitan ahora. De este modo, los docentes podrán constatar y subsanar las deficiencias en los aprendizajes de cursos previos, que impiden a los estudiantes continuar avanzando y los resultados serán útiles no sólo para los docentes del grado evaluado, sino también para los de grados anteriores. El énfasis no estará puesto en comunicar a los docentes que los estudiantes “aprueban” o “desaprueban”, sino en comunicar que están en diferentes puntos de un continuo de aprendizaje, en el que todos pueden y necesitan progresar.

Las pruebas de tipo diagnóstico pueden tener carácter muestral o censal, dependiendo de la estrategia de cambio educativo

El carácter censal o muestral tiene diversas implicancias, aunque también es posible combinar una aplicación por muestras controlada con una distribución censal de las pruebas para su aplicación en forma autónoma por parte de las escuelas, con fines formativos y de análisis de resultados e identificación de estudiantes que necesitan apoyos complementarios.

- Las *pruebas a base de muestras* sirven para tener un diagnóstico global del sistema. Requieren de un cuidadoso diseño de la muestra, de modo de obtener información representativa para los niveles de desagregación en los que se desee actuar y tomar decisiones (regional, provincial, municipal; urbano y rural; escuelas indígenas, etc.). El impacto de los resultados de estas pruebas depende principalmente de las medidas de política educativa que se tomen a nivel central y de una estrategia de difusión apropiada que llegue a todas las escuelas.
- Las *pruebas censales* sirven para obtener información de cada una de las escuelas e incluso de los alumnos. El impacto de sus resultados depende de la devolución de la información a cada comunidad educativa, con un enfoque y un formato adecuados para promover una mayor participación y compromiso a nivel local. La información también puede resultar muy útil para dirigir mejor las políticas hacia distritos o escuelas con mayores problemas, ya que se puede contar con un “mapa” de resultados de todas las escuelas, zonas, provincias, tipos de escuela, etc.

Uso de las evaluaciones para establecer incentivos

Además de las opciones examinadas, existen políticas de evaluación estandarizada dirigidas a establecer incentivos económicos en función de los resultados o a propiciar un mercado competitivo entre las escuelas

Tres son las principales modalidades en este tipo de políticas:

- Utilizar los resultados para construir rankings de escuelas y hacerlos públicos, como forma de fomentar la responsabilidad de las escuelas por sus resultados; entregar a las familias información para la toma de decisiones sobre la escuela a la cual enviar a sus hijos; promover la competencia entre los centros educativos por obtener mejores resultados.
- Utilizar los resultados para entregar incentivos económicos a las escuelas que obtienen mejores resultados o mejoras en relación a sus resultados en evaluaciones anteriores.
- Utilizar los resultados como indicador de la calidad del trabajo de cada docente y como criterio para otorgar incentivos económicos.

Estos enfoques de uso de la evaluación a veces se constituyen, con o sin intención deliberada, en un mecanismo por el cual el Estado renuncia a su responsabilidad por los resultados del sistema educativo

En ocasiones, el Estado limita su accionar a la realización de evaluaciones, la entrega de resultados y el establecimiento de incentivos en función de estos resultados, transfiriendo la responsabilidad por los resultados a la relación entre escuelas y familias, como si se tratase de un asunto entre actores privados, en lugar de ocuparse de crear las condiciones para que la enseñanza resulte efectiva proveyendo los recursos necesarios, constituyendo un elenco docente adecuadamente formado y estableciendo mecanismos idóneos para la evaluación y la orientación de la labor de las escuelas. Este modo de operar no considera la complejidad de la labor educativa, en especial en medios socialmente desfavorecidos, y la necesidad de invertir en la creación de capacidades como herramienta principal para la mejora de la enseñanza y el aprendizaje.

3.1. Criterios básicos para la divulgación y uso de los resultados

La responsabilidad por los resultados debe ser compartida por diversos actores

Debería evitarse todo uso de los resultados con el fin deliberado o implícito de culpabilizar o responsabilizar de manera exclusiva a ciertos actores. La responsabilidad debe ser compartida entre las autoridades nacionales y regionales, los cuerpos docentes, los estudiantes y sus familias.

La responsabilidad por los resultados requiere de una adecuada combinación entre las exigencias y los apoyos a las escuelas y a los docentes

Por un lado, es necesario que escuelas y docentes asuman su responsabilidad respecto al objetivo de que todos los estudiantes aprendan lo que se espera de ellos. Simultáneamente, es deber de las autoridades establecer las políticas de apoyo necesarias para que escuelas y docentes puedan desarrollar su labor de manera apropiada. Un esquema de exigencias altas sobre escuelas y docentes sin los correspondientes apoyos solo puede generar malestar y desánimo. Un esquema de apoyos sin contrapartes y exigencias puede generar autocomplacencia.

Es inapropiado utilizar los resultados de las pruebas estandarizadas como indicador principal de la calidad del trabajo del docente o de la escuela

Esto es importante, en particular si no se ha controlado el efecto de otros factores internos y externos al sistema educativo y si no se tiene en cuenta que el aprendizaje depende también de la motivación y esfuerzo personal del estudiante. Además, dado que los resultados de pruebas estandarizadas no son el

único indicador de la calidad de la educación, es preciso considerarlos en conjunto con otros aspectos relevantes como la retención, la pertinencia y relevancia de lo que se enseña, la formación de valores y hábitos o la formación ciudadana. Debe evitarse la identificación del término “calidad” con resultados de pruebas estandarizadas.

Para facilitar la comprensión y uso de los resultados, no basta con ofrecer resultados numéricos

Es fundamental ilustrar a los distintos actores acerca del tipo de tareas que los estudiantes debieron resolver en las pruebas. Este criterio debe combinarse adecuadamente con la necesidad de mantener el carácter confidencial de un conjunto de preguntas, que puedan emplearse en futuras evaluaciones para establecer mediciones comparables en el tiempo. Si bien no se pretende que los docentes utilicen pruebas estandarizadas para evaluar a sus estudiantes, el conocimiento y la comprensión de las mismas pueden contribuir a mejorar sus modos de evaluar y a desarrollar una cultura de la evaluación.

Al reportar las diferencias de resultados en función de la composición sociocultural del alumnado de las escuelas, se debe cuidar de no crear un sistema de expectativas de logro diferenciado por grupos sociales

En todo caso, las diferencias deberían producirse en las políticas para crear condiciones apropiadas para enseñar a los grupos más desfavorecidos. Se debería descartar también cualquier tipo de uso de los resultados que, directa o indirectamente, propicie la selección de estudiantes por parte de las escuelas a los efectos de mejorar sus resultados.

3.2. Advertencias sobre la comparación de resultados entre escuelas

Aunque los resultados de pruebas estandarizadas no dan cuenta en forma exhaustiva de la calidad educativa de una escuela, sí aportan información relevante de los niveles de desempeño alcanzados.

La información sobre resultados en dos pruebas estandarizadas –en general, lenguaje y matemática– no constituye por sí misma una evaluación de la “calidad educativa” de las escuelas ni debería presentarse al público como tal. La “calidad” de una escuela incluye otros aspectos relevantes y valorados por los docentes, estudiantes y familias, tales como el desarrollo emocional, las relaciones interpersonales, la formación cívica y en valores, entre otros. Aun así, la información comparativa acerca de los niveles de desempeño alcanzados por los estudiantes en una diversidad de escuelas puede ser relevante para los equipos docentes de las escuelas, en la medida en que enriquece la percepción de la propia labor, permitiéndoles ubicar los logros de sus estudiantes en el contexto de los alcanzados por los estudiantes en otras escuelas.

Para que la comparación entre los niveles de desempeño de los estudiantes sea apropiada, se deben considerar sus condiciones sociales de origen

Los resultados académicos de las escuelas deberían poder compararse con los de escuelas de similar composición social, dado que los retos y dificultades que implica enseñar a estudiantes de origen desfavorecido (o de lengua materna indígena) son muy diferentes que en el caso de estudiantes de familias con educación secundaria completa y/o terciaria. También hay que tener en cuenta las tasas de deserción de las escuelas y sus políticas de selección de estudiantes, dado que una escuela puede mejorar sus resultados a través de la exclusión de estudiantes con dificultades.

Las comparaciones entre escuelas deben considerar la diferencia entre medir el “logro educativo” y medir el “aprendizaje”

En rigor, si el propósito es entregar a las escuelas, familias y/o autoridades, información acerca de la capacidad de enseñar de los docentes y las escuelas, es necesario medir el progreso de los estudiantes a lo largo de un período lectivo (“aprendizaje”) y no únicamente el resultado final (“logro”). Las diferencias entre estos dos términos se expresan en que:

- El *aprendizaje* puede ser definido como el cambio ocurrido en los conocimientos y capacidades de cada estudiante a lo largo del año lectivo. Su medición requiere la aplicación de dos pruebas, una al inicio y otra al final. De este modo es posible constatar el progreso realizado por cada alumno.
- El “logro educativo” en cambio, se mide con una única prueba y refleja la acumulación de conocimientos y capacidades a lo largo de toda la vida del estudiante, lo cual incluye el capital cultural familiar y sus experiencias en otras escuelas o con otros docentes.

Debe tenerse siempre presente que el aprendizaje y el logro educativo de los estudiantes dependen tanto de lo que hacen las escuelas y docentes, como del esfuerzo de los propios estudiantes, del apoyo de las familias a la tarea escolar, del contexto comunitario y cultural y de las políticas educativas. Los problemas educativos no pueden ser resueltos apelando exclusiva ni principalmente a mecanismos de mercado.

Los resultados expresados en términos de rankings de escuelas deben ser mirados con precaución

La mayoría de rankings de escuelas transmiten una falsa imagen de ordenamiento. Una escuela puede ocupar el primer lugar y otra el número 40 y, sin embargo, las diferencias entre sus promedios pueden no ser estadísticamente significativos. Esto significa que no puede decirse si un promedio es realmente superior al otro, porque las diferencias entran dentro de los márgenes de error de la medición. Pero aun cuando las diferencias en los promedios sean estadísticamente significativas, pueden ser irrelevantes en términos de las proporciones de alumnos que logran los niveles de desempeño esperados.

3.4. Itinerario para la toma de decisiones

Considerando las alternativas existentes para una política de evaluación, antes de poner en marcha un sistema de evaluación (o de modificar uno existente) es indispensable seguir un itinerario de decisiones para definir las características del sistema que incluya al menos las siguientes preguntas básicas:

- ¿Para qué se quiere evaluar? ¿Quiénes usarán los resultados y con qué propósitos? ¿Qué se espera saber que no se sepa ya?
- ¿Cuáles serán las unidades de análisis para el reporte de resultados: estudiantes individuales, grupo de clase/maestros, escuelas, tipos de escuela, entidades subnacionales, sistema educativo?
- De acuerdo a los propósitos definidos, ¿qué es más apropiado evaluar –aprendizaje o logro educativo– al egreso de determinados grados o ciclos?
- ¿Qué tipo de consecuencias tendrán los resultados y para quiénes?
- De acuerdo a los propósitos definidos, ¿es necesario trabajar en forma censal o son suficientes evaluaciones en base a muestras?
- ¿Qué grados y qué disciplinas es importante evaluar?
- ¿Cada cuánto tiempo es necesario y adecuado realizar las evaluaciones?

Las respuestas a estas preguntas deberían plasmarse en un plan de evaluación a corto, mediano y largo plazo, claro y explícito. En la elaboración de dicho plan es muy importante considerar los costos en dinero implicados en cada opción y los recursos humanos necesarios para su adecuada implementación. En particular, se debe balancear la inversión a realizar en la recogida de información con la inversión en la difusión y uso de los resultados. Muchos Ministerios de Educación destinan importantes sumas de dinero a recoger una gran cantidad de información todos los años, que luego casi no se analiza, ni se difunde ni se utiliza. Es absolutamente inconducente implementar un primer operativo de evaluación si no se ha elaborado un plan de trabajo de largo plazo.

"MANDATOS" ÚTILES AL MOMENTO DE TOMAR DECISIONES POLÍTICAS RESPECTO AL SISTEMA DE EVALUACIÓN

- *"No harás daño"*. Una de las primeras normas del juramento hipocrático señala que el médico se abstendrá de todo aquello que pueda generar daño o perjuicio a sus pacientes. Del mismo modo, a la hora de pensar en un sistema de evaluación es importante considerar los riesgos de efectos no deseados y perjudiciales que la estrategia de evaluación diseñada podría tener sobre el propio sistema educativo que se desea mejorar.
- *"El consumo excesivo de este producto puede resultar perjudicial para la salud"*. La evaluación no puede estar por encima de la educación. El exceso de evaluaciones puede producir perjuicios para la salud del sistema educativo, en especial si solo hay evaluaciones pero no hay políticas para dar respuesta a los problemas que estas ponen de manifiesto.
- *"El consumo exclusivo de este producto también puede resultar perjudicial para la salud"*. La evaluación no produce mejoras por sí misma. Debe estar alineada con otro conjunto de elementos para que aquellas se produzcan. Se debe evitar la ingenuidad de pensar que evaluando mucho – y poco más que eso – se va a mejorar la educación.
- *"No evaluarás en vano"*. Se debe evitar recoger todos los años información que nunca será analizada ni utilizada. Para que un sistema de evaluación tenga impacto, debe tener una periodicidad que garantice que los datos puedan ser analizados, discutidos, comprendidos y utilizados. Incorporar la nueva información y traducirla en decisiones y acciones lleva tiempo. Los cambios en el sistema educativo requieren más tiempo aun.
- *"Vísteme despacio Sancho, que estoy apurado"*. Es muy importante que quienes toman las decisiones en los Ministerios de Educación desestimen la creencia ingenua en la existencia de caminos fáciles y rápidos. No es posible montar un programa de evaluación serio en tres meses. No existe coyuntura ni "ventana de oportunidad" que lo justifique: las consecuencias de la improvisación se harán sentir tarde o temprano. La evaluación requiere reflexión cuidadosa acerca de sus fines y usos; discusión pública sobre qué es lo fundamental a evaluar; diálogo e involucramiento de diferentes actores; conformación de equipos técnicos competentes en diversos aspectos; e información previa para que los actores se involucren con la evaluación.

4. LOS DESAFÍOS DE CALIDAD TÉCNICA DE LAS EVALUACIONES

Una vez definida la política de evaluación, es necesario implementarla de acuerdo con diversos estándares apropiados de calidad técnica. Esto implica abordar los siguientes desafíos:

Elaborar un referente o marco conceptual que defina con precisión los conocimientos y desempeños que son considerados apropiados al finalizar el grado o ciclo educativo que será evaluado

Si bien esta construcción exige antes que nada un debate y decisiones de carácter político, las definiciones tomadas en dicho terreno deben luego ser adecuadamente traducidas en especificaciones y estándares de carácter técnico. Asimismo, el debate político-educativo debe estar alimentado por información sobre los avances conceptuales recientes en cuanto a la enseñanza y a los desempeños de los estudiantes en las disciplinas consideradas.

Incluir en las pruebas actividades con diversos grados de complejidad

Se debe contemplar actividades cuya solución requiera de competencias cognitivas suficientemente complejas, apropiadas a los desafíos de la sociedad del conocimiento, así como también actividades sencillas que permitan informar sobre el punto en que se encuentran los estudiantes menos avanzados. Las actividades deben poseer una serie de *propiedades psicométricas* que es necesario garantizar y cuyo control requiere pruebas piloto y análisis cuidadoso. También es importante, en la medida de lo posible,

ampliar el uso de preguntas de respuesta construida, con el doble propósito de evaluar capacidades más complejas y de mejorar la articulación entre la evaluación externa y la cultura docente y escolar.

Conformar las pruebas a partir de la apropiada integración de las actividades en bloques y cuadernillos

Este es un desafío técnico importante y complejo, que requiere de conocimientos especializados y experiencia. Especial atención debe otorgarse a la decisión de utilizar la Teoría Clásica o la Teoría de la Respuesta al Ítem para la construcción y el análisis de las pruebas, lo cual requiere de programas de procesamiento modernos, analistas bien entrenados, análisis rigurosos y asesoramiento de alto nivel.

Definir los "puntos de corte" que establecen los límites entre niveles de desempeño en una prueba

Vinculado con esto, se debe definir la metodología a seguir para establecer cuál de los niveles debe ser considerado como aceptable para un estudiante al finalizar el grado o ciclo educativo evaluado. Un rendimiento aceptable no puede ser establecido automáticamente como equivalente al 51% o más del máximo posible en una prueba.

Diseñar muestras apropiadas a los propósitos de la evaluación

Esto tiene por objeto evitar operaciones más grandes y costosas de las estrictamente necesarias y que, al mismo tiempo, tengan un nivel de precisión adecuado. En relación a este último aspecto debe incluirse, además, la estimación y reporte de los márgenes de error de las mediciones.

Definir cómo se equiparán las evaluaciones

La equiparación de las evaluaciones se refiere a la metodología mediante la cual los resultados de pruebas aplicadas en distintos años se hacen comparables. Este es uno de los desafíos técnicos más importantes que los sistemas de evaluación deben afrontar si uno de sus propósitos es informar sobre los avances o retrocesos en los logros educativos a lo largo del tiempo. Este proceso es indispensable para poder afirmar que los eventuales cambios que se constaten responden a cambios en la realidad educativa y no meramente a cambios en el instrumento de medición. Requiere tomar recaudos tanto desde el punto de vista estadístico, como desde el punto de vista de la continuidad de los conocimientos y competencias evaluadas, y la estabilidad en la longitud y la estructura de la prueba.

Construir datos longitudinales que permitan apreciar la evolución en el tiempo de los aprendizajes de un mismo conjunto de estudiantes

Este tipo de datos, que exigen más de una medición en el tiempo, es el que evalúa el "aprendizaje" entendido como cambio y el que mejor permite establecer cuáles son los procesos propiamente escolares que tienen influencia sobre el mismo. Este enfoque es el que permitiría obtener información más relevante para la política educativa y para la investigación.

Establecer mecanismos de control del cumplimiento de las condiciones estandarizadas que deben regir durante la aplicación de las pruebas

El cumplimiento de las condiciones de aplicación de las pruebas involucra aspectos técnicos complejos y muchas veces descuidados. Sin embargo, son fundamentales para que la información sea confiable y comparable. Esto apunta a temas como:

- La calidad de la capacitación de los aplicadores.
- el establecimiento de controles de calidad durante la aplicación de las pruebas.
- La motivación a los estudiantes para realizar las pruebas.
- La logística de la distribución.
- El retorno seguro de los materiales.

Combinar las evaluaciones con estudios cualitativos

Es importante combinar las evaluaciones nacionales estandarizadas, que ofrecen una mirada sobre el conjunto del sistema, con la realización de estudios cualitativos que permitan profundizar en los procesos escolares y los procesos que ocurren en las aulas. Esta combinación de enfoques es el mejor modo de aportar información rica y compleja para pensar en las políticas educativas y en las prácticas de enseñanza.

LA NECESARIA TRANSPARENCIA

Sobre todos los procesos técnicos debe existir información transparente y accesible. Es necesario dar un especial énfasis a la documentación de los procedimientos técnicos seguidos en:

- La construcción de los instrumentos.
- La estimación de la precisión de las mediciones (y, por consiguiente, su margen de error).
- El diseño de las muestras y la cobertura alcanzada.
- La aplicación y control de calidad de la misma.
- La definición de niveles de desempeño y puntos de corte.
- La equiparación y comparabilidad de los resultados con mediciones anteriores.

5. SOBRE LA CONSTITUCIÓN DE LAS UNIDADES DE EVALUACIÓN

Para llevar adelante una política de evaluación de logros educativos o aprendizajes apropiada, es necesario contar con Unidades de Evaluación con las capacidades y los recursos adecuados que implica este complejo emprendimiento.

La puesta en marcha de un sistema de evaluación serio requiere de un tiempo mínimo de entre dos y tres años

- Este periodo es necesario para llevar adelante los siguientes procesos fundamentales:
- Discusión, definición y difusión pública de los propósitos del sistema de evaluación, del tipo de consecuencias, de los usos esperados y de qué debe ser evaluado.
- Diseño del plan de evaluación a largo plazo.
- Conformación de equipos técnicos con la diversidad de capacidades requeridas (elaboración de pruebas; elaboración de cuestionarios; conocimiento de las disciplinas a evaluar y de su didáctica; currículo y estándares; muestreo; logística de la aplicación y control de calidad de la misma; captura y limpieza de datos; procesamiento y análisis; construcción de escalas; contextualización sociocultural de los resultados y análisis de factores asociados; interpretación de los datos en términos de políticas educativas, de didáctica y de prácticas de enseñanza).
- Elaboración de instrumentos, pilotaje y supervisión externa de los mismos.

Las Unidades de Evaluación necesitan tener una cierta estabilidad en el tiempo de sus elencos técnicos

El desarrollo de un sistema de evaluación requiere de una planificación a 10 ó 15 años. Si los equipos técnicos cambian con frecuencia, se desperdicia el conocimiento y la experiencia acumulada en un área compleja y se desacreditan los procesos de evaluación ante la sociedad y los educadores.

Las Unidades de Evaluación necesitan tener independencia para reportar los resultados de su trabajo

Al igual que en el caso de las Unidades de Estadísticas sociales y económicas, las unidades que realizan las acciones de evaluación y la divulgación de los correspondientes resultados no deben depender de los

tiempos e intereses político-partidarios. Se ha discutido mucho respecto a si las Unidades de Evaluación deberían estar dentro o fuera de los Ministerios de Educación. El argumento principal para esta última opción es justamente la necesaria independencia y transparencia. Sin embargo, la experiencia en la región muestra casos de países con unidades estables e independientes dentro de los Ministerios de Educación y otros con unidades inestables que no han logrado consolidar su trabajo aun estando en una institución externa. La respuesta a esta disyuntiva es que no importa tanto el lugar institucional como la cultura de continuidad y transparencia que se cree en torno a la evaluación. Esto se logra cuando existe un mandato claro y una institucionalidad sólida en relación a la evaluación, lo cual implica algún tipo de estatuto jurídico para el sistema de evaluación. Establecer esto por Ley es uno de los caminos a considerar, dado que exige un acuerdo amplio, en lo posible supra-partidario, que permita establecer un plan de evaluación de la educación de largo plazo. En este terreno, al igual que en el económico, se requiere de cierta estabilidad en las reglas de juego. Si la política de evaluación cambia permanentemente, se genera desconfianza y se pierde credibilidad.

Una institucionalidad sólida requiere de órganos de gobierno y de asesoría técnica independientes y plurales, y de un presupuesto apropiado y plazas de trabajo que garanticen la operación de la unidad con la calidad técnica requerida

La independencia de la Unidad de Evaluación no debe dar lugar a su desvinculación de la política educativa. Por el contrario, la evaluación debe responder a un proyecto político-educativo con amplio respaldo y debe mantener una estrecha relación con otras áreas claves de la política educativa como la formación docente, el desarrollo curricular, el planeamiento y formulación de proyectos, la evaluación de programas y la investigación.

Para que las evaluaciones estandarizadas tengan impacto sobre las políticas y las prácticas educativas, es necesario generar nuevas interfases y modalidades de trabajo que aseguren el alineamiento de los distintos actores y ámbitos de decisión

Las Unidades de Evaluación deben asumir que su tarea va más allá de producir datos. Si su misión es producir información para otros actores, se requiere de personas preparadas y con tiempo para establecer líneas de diálogo con instancias y actores responsables de:

- *Análisis de política educativa*, con el fin de mejorar tanto los diseños de las evaluaciones como los planes de procesamiento y los reportes de resultados, teniendo en cuenta preguntas relevantes para la política educativa (por ejemplo, la muestra puede ser diseñada de modo que permita evaluar el impacto de políticas o programas específicos en ciertos grupos de escuelas).
- *Análisis didáctico*, con el fin de interpretar los resultados y los problemas de aprendizaje de los estudiantes desde la perspectiva de la enseñanza y la didáctica de la disciplina evaluada, elaborar reportes con significado didáctico para los docentes, diseñar programas de formación en servicio a partir de los resultados, pensar la articulación de las evaluaciones estandarizadas con la evaluación en el aula.
- *Comunicación*, con el fin de elaborar una diversidad de reportes de resultados, con diversos formatos y lenguajes, que sean apropiados y comprensibles para diversas audiencias.

Evaluar bien requiere inversión

Más vale no hacer ninguna evaluación que hacer una evaluación mala o deficiente. Esto debe ser tenido en cuenta al tomar las decisiones acerca del Plan de Evaluación. Asimismo, más vale un sistema de evaluación modesto pero con costos asequibles y sostenibles en el tiempo, que una gran evaluación, amplia y sofisticada, que sólo podrá realizarse una vez y que luego no podrá replicarse.

La inversión en evaluación debe ser valorada en función del uso que se hace de sus resultados, más que a partir de otros indicadores como el costo por alumno

Los costos de las evaluaciones son relativamente bajos cuando se los compara con los presupuestos nacionales y con las inversiones alternativas. Pero cualquier inversión en evaluación, alta o baja, es inútil si no se hace uso de los resultados.

6. DIEZ RECOMENDACIONES ACERCA DE LAS EVALUACIONES QUE LA REGIÓN NECESITA

La evaluación debe concebirse como un elemento articulado en un conjunto más amplio de acciones y políticas educativas. La evaluación por sí misma no produce mejoras. Deben existir articulaciones estables entre el área de evaluación y las áreas de desarrollo curricular, formación docente, investigación, diseño de políticas, comunicación y prensa, entre otros.

La evaluación debe contemplar un proceso de reflexión colectiva sobre el estado de la educación y los caminos para mejorarla. El primer paso para ello es la consulta y el debate públicos acerca de qué deben aprender los estudiantes y acerca de los propósitos y consecuencias de la evaluación. Del mismo modo, es fundamental la discusión pública constructiva a partir de los resultados, con el fin de encarar las insuficiencias e inequidades en el acceso al conocimiento por parte de los y las estudiantes. Para ello es necesario invertir en comunicación y difusión tantos o más recursos que en la propia evaluación, antes, durante y después de la misma.

La evaluación debe estar al servicio del desarrollo de un sentido de responsabilidad compartida por la educación como bien público. Debe promover el compromiso con la educación de todos los actores, cada uno según su lugar y ámbito de acción. Por este motivo, debe evitarse utilizar la evaluación para culpabilizar a actores específicos por los problemas detectados.

Los sistemas de evaluación de la región necesitan ampliar progresivamente el abanico de fines educativos que son objeto de evaluación. Es importante en este sentido incluir a la formación ciudadana, otras disciplinas además de Lenguaje y Matemática, así como un espectro más amplio de competencias y capacidades.

Los sistemas de evaluación de la región deberían progresivamente diseñar evaluaciones de progreso de los estudiantes, dado que son las que pueden aportar mayor información acerca del impacto de las políticas educativas, de las acciones de las escuelas y de las prácticas de enseñanza, en los aprendizajes de los y las estudiantes.

Un sistema de evaluación es un proyecto de largo plazo, por lo que requiere de un compromiso del Estado y de una planificación cuidadosa del diseño del sistema. Ello exige tomar decisiones acerca de los fines, las consecuencias, los aspectos a evaluar, las áreas y los grados, y la periodicidad de las evaluaciones, entre otras cosas. Todo ello exige tiempo, por lo que no es conveniente pretender la implementación de sistemas de evaluación en plazos reducidos.

Un buen sistema de evaluación requiere inversión, principalmente en la conformación de equipos humanos calificados, así como recursos económicos suficientes para una adecuada implementación de todos los procesos implicados.

El sistema de evaluación debe sustentarse en una actitud firme de transparencia en relación a los resultados y de rendición de cuentas a la sociedad.

Los Ministerios de Educación deben asumir un compromiso serio y consistente con los resultados de la evaluación, lo cual implica promover el diálogo acerca de los problemas detectados y los modos de

enfrentarlos, diseñar líneas de acción apropiadas para resolver los problemas e invertir los recursos necesarios para llevarlas adelante.

Los sistemas de evaluación deben ser objeto de evaluación periódica, con el fin de analizar la calidad técnica de la información que producen y su relevancia para diversos actores educativos y sociales.





SISTEMAS DE EVALUACIÓN COMO HERRAMIENTAS DE POLÍTICAS

M. Leonor Cariola, Gabriela Cares y Rosario Rivero

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art5.pdf>

* Ponencia Presentada en el 1er Congreso Iberoamericano de Eficacia Escolar y Factores Asociados, 12 -14 de Diciembre de 2007, Santiago de Chile.

1. INTRODUCCIÓN

La evaluación cobra creciente importancia en todos los ámbitos de la vida pública. Al mismo tiempo, desde hace años se viene discutiendo por qué el diseño de políticas, especialmente en el campo de la educación, no presta mayor atención a los resultados de investigación. Con distintas denominaciones, se hace referencia a esta necesidad: Política basada en evidencia, lecciones aprendidas, transferencia de políticas, entre otras (Evans, 2004).

El diseño de políticas educacionales tiene complejidades adicionales a otros campos de la política por dos motivos a lo menos. Por una parte, la relativa ausencia de recomendaciones claras y unívocas para emprender acciones, ha hecho posible que muchas veces los políticos desdeñen los resultados de la investigación en ciencias sociales, tachándolas de teóricas o poco prácticas. Por otra, la relación entre investigación y diseño de políticas en el caso de la educación está mediada por una opinión pública que en cualquier caso, se considera a sí misma ilustrada y conocedora de la realidad. Esto obedece a que todos tenemos experiencias directas con la educación, sea como alumnos o como apoderados, y adicionalmente se reconoce que la educación tiene una gran relevancia en la vida personal y para la sociedad en su conjunto.

Concretamente, nos referiremos a la experiencia del Sistema de Medición de la Calidad de la Educación (SIMCE) en Chile. El SIMCE entrega información sobre logros de aprendizaje de los estudiantes y sobre factores que permitan contextualizar y/o explicar sus diferencias. El tipo de información que recogen y entregan los sistemas de evaluación, se relaciona con el enfoque político y el uso que se le quiera dar desde el sistema educacional en cada momento histórico.

Lo que pretendemos mostrar aquí, es cómo la evidencia producida por la evaluación estandarizada de los aprendizajes, incide en la política de una manera tal vez lenta e indirecta, pero decisiva.

En la siguiente sección describiremos brevemente el Sistema de Evaluación en Chile y el tipo de información que recoge, para contextualizar una tercera parte, donde haremos referencia explícita al análisis de factores asociados y su incidencia en la política educacional, así como los condicionantes que aparentemente gatillan el éxito de esta relación. En una última sección, se describen los proyectos en curso tendientes a perfeccionar la evidencia proporcionada por el SIMCE para incidir en el diseño de política.

2. EL SISTEMA DE EVALUACIÓN DE APRENDIZAJES EN CHILE

2.1. Historia y características de la participación en evaluaciones estandarizadas

La evaluación en Chile tiene una larga data. Comienza en los años 60, con la participación en el estudio internacional comparativo de la IEA, denominado "Six Studies" en referencia al número de áreas del conocimiento que cubrió. En los 70, se intentó usar pruebas estandarizadas al final de la Educación Básica como certificación de los aprendizajes alcanzados en este nivel, pero esta iniciativa duró un corto período.

Entre los años 1982 y 1985, se aplicó el Programa de Evaluación del Rendimiento (PER), a cargo de la Universidad Católica de Chile. Mediante este sistema, se buscaba que los establecimientos y autoridades, al enfrentarse a los datos de la evaluación, reaccionaran modificando o reforzando las prácticas educativas. Finalmente, en 1988 se inaugura el Sistema de Medición de la Calidad de la Educación (SIMCE) que por un tiempo estuvo a cargo de la Universidad Católica y cuyos objetivos serán expuestos en esta sección.

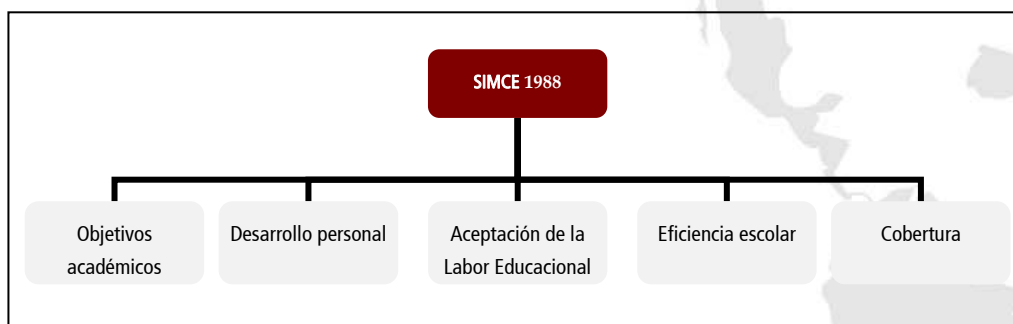
TABLA 1. SÍNTESIS HISTÓRICA DE EVALUACIONES ESTANDARIZADAS DE APRENDIZAJE

Años	Sistema de Medición	Institución
1967 - 1971	Prueba Nacional	Ministerio de Educación
1982 - 1985	PER	Universidad Católica – Subsecretaría de Educación
1988 - 1990	SIMCE	Universidad Católica– Subsecretaría de Educación
1991- a la fecha	SIMCE	Ministerio de Educación

El SIMCE surge en el Gobierno Militar como un elemento complementario a la privatización y municipalización de la educación, consonante con la economía de mercado que se imponía en ese momento. Bajo los supuestos que los padres buscarían la mejor educación para sus hijos, que tendrían libertad para cambiarlos de establecimiento educacional y que serían ampliamente informados sobre los resultados en la prueba SIMCE, se esperaba que el mercado, provisto de información para todos los actores, regulara por sí mismo la calidad de la educación. La expectativa era que los establecimientos, para no perder financiamiento, movilizarían acciones de promoción de calidad: Por ejemplo, al no querer perder matrículas y sus respectivas subvenciones, y, en la búsqueda de aumentar sus ingresos, harían lo posible por mejorar la calidad de los aprendizajes y retener a los estudiantes.

Contradictoriamente, en los inicios del SIMCE, no se entregó información pública de los resultados de escuelas individuales, que es la información requerida por el mercado para operar. Más que describir y explicar la diversidad de rendimientos entre las escuelas, el interés era hacer un diagnóstico de los resultados lo más acabado posible para entregar información a la comunidad escolar. Los resultados de la escuela, se le entregaban a ella en privado, aunque se suponía que la escuela debía compartirlas con el resto de la comunidad. En los primeros reportes, se entregaba muchísima información respecto a distintos aspectos y a logros de los estudiantes, tanto académicos como de su desarrollo personal¹. A modo de ejemplo, se puede citar el primer informe de resultados desarrollado por el SIMCE en 1988, donde se dedica un volumen completo a los indicadores de “desarrollo personal” y “aceptación de la labor educativa” (SIMCE, 1988).

FIGURA 1. INDICADORES DE LA CALIDAD DE LA EDUCACIÓN MEDIDOS POR EL SIMCE EN 1988



En la Figura 1, se muestran los indicadores de calidad de la educación evaluados en 1988², donde los objetivos académicos representan los contenidos propios del Currículum de la época; Desarrollo personal y Aceptación de la labor educativa corresponden a factores afectivos que pueden incidir sobre el

¹ Nos referimos a logros no académicos en el sentido que los describe PISA, donde se señala que: los niños llegan a la escuela con curiosidad y ganas de aprender. Mantener estas inquietudes, desarrollar actitudes propicias para aprender y ser capaces de manejar su propio aprendizaje, tanto para aprender en la escuela como a lo largo de la vida, se puede considerar como un logro. (OECD, 2003a:120).

² Denominados en la época “estimadores de la calidad de la educación”.

rendimiento académico o ser considerados como logros en un sentido amplio; eficiencia escolar a la medición anual de tasas de aprobación, reprobación, retiro y retención de estudiantes en cada establecimiento; y cobertura alude al nivel de cobertura curricular declarada por los profesores.

En 1990, un día antes de asumir el gobierno democrático, se promulgó en Chile la Ley Orgánica Constitucional (LOCE) que obligaba a los gobiernos posteriores a evaluar el aprendizaje de los estudiantes y a hacer públicos los resultados de las escuelas (no de los alumnos). Junto con esta exigencia, se aseguraba el financiamiento del sistema nacional de evaluación de aprendizajes, que pasó a formar parte del Ministerio de Educación.

El Ministerio, una vez en democracia, reconoce la importancia de la evaluación, pero también comprende que el mercado por sí solo no mejoraría la calidad de la educación, por lo que al enfoque informativo, se incorporan iniciativas de apoyo a las escuelas. El Estado refuerza su rol subsidiario, con programas de mejoramiento universal de las escuelas y focalizando apoyos en aquellas con menores rendimientos. Así, por ejemplo, surge, el Programa P-900 para ayudar a las escuelas con más bajos rendimientos (Meckes y Carrasco, 2006).

En los informes SIMCE de los primeros años en democracia, se mantienen con fuerza los mensajes sobre indicadores del área afectiva, expresados de manera reiterada en los reportes de resultados. En la Tabla 2, se observa que durante sus primeros siete años, el SIMCE reportó mediciones sobre dos variables afectivas que pueden influir el aprendizaje de los estudiantes: Desarrollo personal, evaluado mediante cuestionarios a alumnos, y Aceptación de la Labor Educativa, para lo cual se aplicaron cuestionarios a alumnos, apoderados y docentes. Por otra parte, la Eficiencia Escolar era medida a través de un instrumento aplicado a todos los establecimientos al final de cada año académico, por el Ministerio de Educación.

TABLA 2. INDICADORES DE CALIDAD DEL RENDIMIENTO REPORTADOS POR EL SIMCE 1988 - 1997

Año	Niveles evaluados	Indicadores de calidad reportados
1988	4° básico, censal	Indicadores de calidad medidos recurrentemente: - Desarrollo personal - Aceptación de la Labor Educativa - Eficiencia Escolar
1989	8° básico, censal	
1990	4° básico, censal	
1991	8° básico, censal	
1992	4° básico, censal	
1993	8° básico, censal 2° medio, muestral	Indicadores de calidad incorporados en algunas mediciones durante los 90: - Desarrollo de la creatividad (1995, 1996) - Estrategias de aprendizaje (1993, 1994, 1995) - Actitudes hacia el ambiente (1993, 1994, 1995, 1996)
1994	4° básico, censal 2° medio, censal	
1995	8° básico, censal	
1996	4° básico, censal	
1997	8° básico, censal	Se aplica un cuestionario específico para medir autoestima en alumnos (antes contenida en "desarrollo personal"). Deja de medirse "Aceptación de la labor educativa". Los cuestionarios a docentes se especifican de acuerdo a su área de enseñanza (lo cual se mantendrá hasta la fecha). Se siguen aplicando cuestionarios a padres y apoderados.

Para recoger información sobre las variables afectivas que influyen en el rendimiento académico, se emplearon instrumentos que, al no tener continuidad ni un marco de referencia común entre años, fueron perdiéndose en el tiempo. Si bien había un marcado interés en variables afectivas y de percepción de alumnos, padres y docentes (desarrollo personal y aceptación de la labor educativa), no se mantuvo una estrategia común clara, en la medición de éstas.

Desde principios de los 90, se impulsa en Chile una importante reforma al currículum, apostando principalmente en ello, el progreso del sistema educacional. Por su parte, los instrumentos de evaluación

aplicados por el SIMCE no experimentaron cambios de gran envergadura hasta una vez que se inició la implementación del nuevo currículo.

Al inicio de los noventa, la primera señal de apoyo a los establecimientos es utilizar los resultados del SIMCE, para la detección de escuelas con requerimientos de apoyos específicos, realizando una labor de discriminación positiva.

En el año 1992, los reportes del SIMCE a cada escuela introdujeron una asociación entre los puntajes y la dependencia administrativa de los establecimientos. En 1993, los resultados comienzan a presentarse asociados al nivel socioeconómico de los establecimientos.

En 1995, por primera vez se dan a conocer en un inserto de prensa, los resultados por escuela, donde además se enviaba un mensaje público sobre la importancia de focalizar los recursos en los establecimientos con menor rendimiento. En 1996, los resultados del SIMCE comenzaron a estar disponibles en el sitio Web del Ministerio de Educación. La entrega pública de los resultados, se asume como una ganancia en transparencia y permite responsabilizar y controlar a las escuelas. Por lo mismo, se quiere también dar más apoyo y empieza un movimiento en búsqueda de retroalimentar a cada establecimiento con los logros de sus estudiantes, a fin de que pueda mejorar este proceso de enseñanza aprendizaje.

En el transcurso entre la creación del SIMCE y el segundo quinquenio de los 90, se percibe un doble movimiento. Por una parte, la medición se concentra en los resultados de aprendizaje, abandonando la medición de variables socio-afectivas. Por otra, se mueve hacia la entrega de información más precisa y retroalimentación a los establecimientos. De esta manera se complementa la información dirigida al funcionamiento del mercado, con apoyos para mejorar la calidad educativa desde "adentro" de la escuela. En este sentido, la revisión de los reportes de resultados generados por SIMCE a través de los años, permite observar un formato y redacción, cada vez más amigables y enfocados a sus públicos objetivos.

A partir de 1997, pero ya más concretamente en 1998, cambia el tipo de información recogida por los cuestionarios del SIMCE, los cuales son aplicados cada año a padres y apoderados y a docentes. En pocas ocasiones, también se aplicaron a directores y estudiantes. En esta nueva etapa, desde 1998 hasta la fecha, se ha dado especial relevancia a la recolección de información de variables familiares, especialmente nivel socio-económico de la familia, que permite hacer comparaciones más justas entre escuelas.

En la Tabla 3, se muestran los factores asociados que actualmente, mide y reporta el SIMCE.

En la actualidad, el SIMCE tiene como misión contribuir al mejoramiento de la calidad y equidad de la Educación en Chile, para lo cual se encarga de proveer información acerca de los resultados de aprendizaje a nivel nacional y de establecimientos educacionales. Los actuales propósitos del SIMCE son (Informe MINEDUC, 2003):

- Monitorear la evolución de los resultados del sistema escolar y orientar decisiones de política educativa nacional.
- Orientar y retroalimentar acciones de docentes y directivos.
- Promover el compromiso responsable de escuelas, sostenedores y apoderados.

TABLA 3. FACTORES ASOCIADOS Y FUENTES DE INFORMACIÓN PARA LA ENTREGA DE RESULTADOS DEL SIMCE

Reporte de factores asociados	Fuentes
Dependencia	- Base de Datos RECH (Registro de Estudiantes de Chile) - Base de datos de subvenciones - Directorio de establecimientos
Género del alumno (hombre / mujer)	
Género del establecimiento (mixto, hombres o mujeres)	
Ruralidad	
Modalidad de enseñanza (Científico-humanista o Técnico-profesional)	
Tamaño del establecimiento	
Financiamiento compartido	
Niveles impartidos por el establecimiento	
Multigrado	
Tamaño de los cursos	
Cantidad de establecimientos a cargo de un mismo sostenedor	
Región	
DEPROV (Departamentos provinciales de Educación)	
Comuna	- Cuestionario de Padres y Apoderados - Cuestionario de Profesores - Índice de vulnerabilidad construido por JUNAEB.
Grupo Socioeconómico	
Índice de vulnerabilidad del establecimiento	
Escolaridad del padre	
Escolaridad de la madre	
Expectativas de los profesores	
Cobertura Curricular	
Preparación para enseñar	
Prácticas pedagógicas	
Expectativas de los padres	

Para cumplir estos objetivos, el SIMCE ha trabajado en aumentar y mejorar la precisión de la información que distribuye, orientándola además, a un mayor número de públicos. Los resultados del SIMCE se informan anualmente a través de la prensa, con los datos de cada establecimiento del país; en el Informe de Resultados a Profesores y Directivos se entrega información individualizada de cada escuela; y en el Informe de Resultados a Padres y Apoderados, se presentan los datos de la escuela en una versión simplificada. Además, se elabora un Informe Nacional que entrega los datos del país y por regiones.

Por otra parte, se ha hecho un gran esfuerzo en la construcción y continuo mejoramiento de la página Web, donde se incluyen todos los reportes de resultados generados por el SIMCE, informes de estudios internacionales, bases de datos y una herramienta de análisis de resultados. En general, cada tipo de información está especialmente diseñado para un perfil de usuario (educador, investigador o apoderado).

Desde fines de los 90, Chile se incorporó de manera sistemática en evaluaciones de aprendizaje internacionales, las que presentan un nuevo panorama de posibilidades metodológicas, influyendo en el tipo de cuestionarios aplicados por el SIMCE a nivel nacional y en el mejoramiento de la metodología de elaboración y análisis de las pruebas cognitivas.

Las evaluaciones internacionales permiten tener información sobre los resultados de aprendizaje de los estudiantes chilenos con relación a otros países y de algunos factores posiblemente asociados a los resultados, contrastados con referentes internacionales.

TABLA 4. ESTUDIOS INTERNACIONALES EN LOS QUE CHILE HA PARTICIPADO DESDE LA DÉCADA DE LOS '90

Estudio	Año aplicación	Foco principal	Ejemplos de factores de contexto evaluados
LLECE SERCE	1997 2006	Lenguaje y Matemática	- Características del sistema educacional de cada país (por ejemplo, contenidos curriculares; financiamiento de las escuelas; duración jornada escolar). - Características de los establecimientos (por ejemplo, dependencia administrativa; ubicación urbano/rural; gestión organizacional; recursos humanos y materiales con los que cuenta; proyectos extracurriculares). - Características de los estudiantes y sus familias (por ejemplo, nivel de educación de los padres; nivel socioeconómico; bienes materiales y culturales en el hogar).
TIMSS	1998 2002	Matemática y Ciencias Naturales	
CÍVED ICCS	1999 2008	Conocimientos cívicos y habilidades de interpretación, actitudes cívicas	
PISA	2001 (plus) 2006	Alfabetización en Lectura, Matemática y Ciencias Naturales	

En los párrafos anteriores, se puede observar cómo el SIMCE ha profundizado en su rol de informar sobre los aprendizajes de los estudiantes en Chile, a la vez que se ha ido acompañando de manera creciente, por políticas educacionales de apoyo al rendimiento de las escuelas.

Actualmente, el SIMCE busca cada vez más apoyar a las escuelas con información que sea relevante para ellas y que les permita diseñar acciones para mejorar el aprendizaje de sus estudiantes. En este sentido, se ha ampliado y diversificado la comunicación de los resultados del SIMCE, a la vez que la recolección y análisis de la información se ha hecho en forma cada vez más rigurosa, tanto de las pruebas de evaluación de aprendizaje como de los cuestionarios de contexto. Si bien durante este último período los cuestionarios se han ido desarrollando en base a requerimientos contingentes, cabe destacar que, actualmente, el SIMCE impulsa un proyecto con el objetivo de mejorar estos instrumentos, elaborando un marco de referencia posible de proyectarse en el mediano y largo plazo, con un plan de análisis que seguirá una determinada racionalidad.

Como ya se ha señalado, la forma en que se analicen y entreguen los resultados de la evaluación, incide fuertemente en las consecuencias que esta tenga tanto para los actores, como para la política educacional en general.

Habiendo descrito la evolución de la información que se recoge y analiza en el SIMCE, en el punto siguiente analizaremos el rol que han jugado los factores asociados con esta evolución.

2.2. Recolección y análisis de los factores asociados

Entre los factores asociados que se miden y analizan con relación a los resultados en pruebas estandarizadas, hay factores internos y externos a las escuelas. Los factores internos son aquellos que dependen del manejo propio de la escuela, tales como rotación de profesores, capacitaciones docentes, manejo curricular, gestión y organización, entre otros. Por otra parte, los factores externos a las escuelas, se caracterizan porque esta no los puede alterar por sí misma. En este grupo, tiene gran injerencia el nivel socio económico, cuyo análisis en Chile, demuestra históricamente una asociación importante con el rendimiento académico de los estudiantes. También aquí se encuentran variables como género, área urbano – rural, entre otras. Los factores externos permiten hacer comparaciones de rendimiento más justas, ya que a partir de ellos se puede categorizar escuelas semejantes según una o más variables que no dependen de la gestión del establecimiento.

Para recoger información sobre factores asociados, hemos visto que el SIMCE administra cuestionarios autoaplicados a apoderados y docentes, los cuales se distribuyen en conjunto con los instrumentos de evaluación de aprendizajes. También las evaluaciones internacionales contemplan este tipo de cuestionarios, dirigidos a alumnos, apoderados y directores de escuela. Estas herramientas recogen la información censal, en el caso de las pruebas nacionales, o muestral en estudios específicos y evaluaciones internacionales.

Al interior del SIMCE, los equipos técnicos se encargan de sistematizar y analizar la información recopilada en los cuestionarios, la cual posteriormente, es puesta a disposición de investigadores para que realicen los análisis que estimen pertinentes.

El análisis de factores asociados consiste en identificar formas de elevar el aprendizaje, a través de la alteración de dichos factores. Estos análisis, en general, no pretenden determinar causalidad en el sentido más estricto, pero sí se espera aislar el impacto de distintas variables y buscar aquellas que son posibles de modificar a través de políticas o medidas tomadas a distintos niveles de decisión³. Esto exige una tremenda cautela para no tomar medidas y decisiones en base a información sesgada.

Con la publicación de los resultados obtenidos el año 2000, se empezó a entregar información por “grupos comparables” según nivel socio económico. Con esto se buscaba hacer una entrega más justa de resultados, ya que así se podría comparar el rendimiento académico de establecimientos con características internas y de entorno, semejantes. Además, se buscaba posicionar el tema en la opinión pública, donde hasta ese entonces, se asignaba fuerte relevancia al efecto de la dependencia administrativa sobre los resultados académicos, empañando la importancia del nivel socio económico de los estudiantes (Meckes y Carrasco, 2006).

Los “grupos comparables” se forman categorizando a los establecimientos en función de las siguientes variables:

- Nivel educacional de la madre
- Nivel educacional del padre
- Nivel de ingreso de las familias
- Índice de vulnerabilidad escolar

Con estos datos, se establecen cinco grupos socioeconómicos, que permiten comparar establecimientos que entre sí tienen una composición socioeconómica de su alumnado similar.

Los factores asociados permiten observar los contextos en los cuales se enmarcan los resultados de aprendizaje, así como establecer variables que acompañan al mejor o peor rendimiento de los estudiantes. En general, estos análisis se cuidan de caer en determinismos difíciles de establecer en ciencias sociales y, consecuentemente, se habla de “factores asociados” en oposición a causales o determinantes. Sin embargo, se sabe que el interés es conocer qué se puede hacer para incrementar el aprendizaje de los estudiantes, qué es lo que funciona. Algunos autores advierten este riesgo, enfatizando la necesidad de utilizar métodos experimentales y otros consideran la necesidad de refrendar las conclusiones con distintos métodos, antes de considerar que las conclusiones de las investigaciones son evidencia para el diseño de políticas (Cook y Gorard, 2007).

En la medida que los sistemas de evaluación quieren mejorar los resultados de aprendizaje y fomentar la responsabilidad de la comunidad escolar en estos logros, intentarán entregar información que sea útil

³ Para un análisis detallado de la importancia de determinar causalidad para la política educacional y sobre metodologías y sus riesgos, ver Cook y Gorard (2007).

tanto para diseñar políticas a nivel nacional o local y/o para orientar las prácticas de docentes y directores.

3. CONSTRUCCIÓN DE EVIDENCIA PARA POLÍTICA

A partir de los datos obtenidos por el SIMCE, se espera promover la generación de políticas educacionales específicas, coherentes con la evidencia, que tiendan a solucionar, mejorar y/o promover las situaciones observadas, tanto a nivel de las escuelas como del sistema educacional en su conjunto.

La evidencia puesta a disposición de la comunidad escolar, investigadores y políticos, debe ser robusta, clara y accesible. La validez y confiabilidad de los datos están a la base del éxito de cualquier intento por intervenir la Educación de manera informada. La certeza sobre lo que se está comunicando, constituye un compromiso ético y es parte de la cultura laboral con que se trabaja en el SIMCE, como debería ser en cualquier sistema de medición. Además, los sistemas de medición deben hacerse cargo de informar cómo se cumplen los requisitos técnicos de la medición, con fines de documentación y de difusión a los especialistas que, seguramente, querrán examinar la información con todo detalle. La credibilidad y legitimidad del sistema es uno de los bienes más preciados que hay que resguardar y desarrollar.

Por otra parte, es necesario que la información presentada, haga sentido en el ámbito de la opinión pública, que tenga "face validity". Es problemático ir contra el sentido común de la gente o intentar posicionar en la agenda pública, un tema cuyos componentes sean demasiado técnicos y complejos. Aunque sea redundante, la evidencia debe ser "evidente" para muchos.

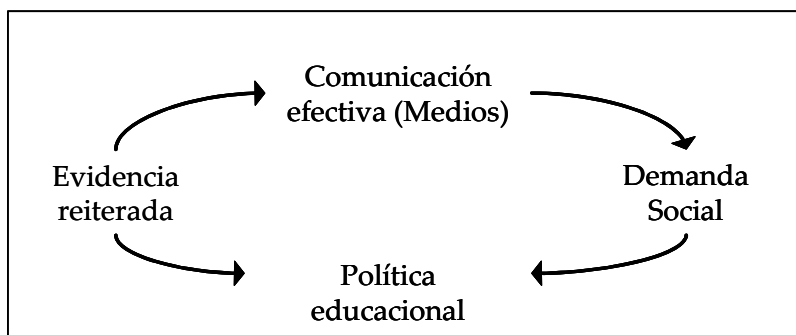
Considerando la importancia que juega la opinión pública como mecanismo de coacción hacia el mundo político y el interés del Sistema por impulsar políticas basadas en la evidencia de evaluaciones estandarizadas, es necesario un esfuerzo por simplificar los mensajes comunicacionales, de modo que estos puedan ser absorbidos por un máximo de personas: internas y externas al campo de la educación, pero también entre los que están a favor y en contra del gobierno. Como se señalaba al principio de este artículo, cada persona y sociedad tiende a tener una postura crítica frente al sistema educacional. En la medida que esta crítica sea más informada, se convierte en un mecanismo de presión más valioso y efectivo para la generación de nuevas y mejores prácticas educacionales.

En la Figura 2, se muestran dos posibles mecanismos de generación de políticas educacionales a partir de la evidencia. Hacia la derecha, creemos que se refleja la demanda y conquista de iniciativas políticas de gran envergadura que, la gran mayoría de las veces, son mediadas por la opinión pública. Por otra parte, también existe una forma de generación de políticas con influencia más directa desde la evidencia, lo que podría ejemplificarse en el impacto del SIMCE sobre programas de otras unidades del Ministerio de Educación. En general, es difícil que una iniciativa de política educacional no sea mediatizada, antes de su puesta en marcha.

Se puede constatar que, a pesar de las constantes críticas al sistema de medición por sus posibles efectos no deseados, finalmente, a la larga, la información entregada va permeando a la opinión pública y llega a constituir evidencia, contribuyendo así al diseño de políticas más coherentes con la realidad.

La teoría de redes sociales, habla de "hacer puentes" (brokerage) mediante intermediarios que vinculen actores dentro de una misma o varias estructuras. En el caso que revisamos, significaría vincular el conocimiento que puede surgir a partir de la evidencia, con la toma de decisiones políticas.

FIGURA 2. MECANISMOS DE GENERACIÓN DE POLÍTICAS A PARTIR DE LA EVIDENCIA



La OCDE ha demostrado interés por el estudio y promoción del *brokerage* entre investigadores y tomadores de decisiones políticas en distintos ámbitos. En un estudio enfocado específicamente al ámbito de la Educación (OCDE, 2003), se señala que la política educacional basada en evidencia enfrenta principalmente, las siguientes dificultades:

- Bajos niveles de inversión en investigación educacional
- En general, bajos niveles de capacidad para investigación, especialmente cuantitativa.
- Débil vinculación entre investigación, política e innovación.

El diseño de política basado en evidencia (evidenced informed policy) es el uso conciente y explícito de la evidencia al tomar decisiones o al elegir entre políticas alternativas (OCDE, 2007). Davies (2004), contrasta esta forma de hacer política con el diseño de política basado en la opinión, que usa evidencia selectivamente, o en puntos de vista individuales. La política basada en evidencia requiere una búsqueda sistemática de información, de una evaluación crítica de los estudios identificados y una comprensión equilibrada de las fortalezas y debilidades de la evidencia entregada por la investigación.

En este artículo no se pretende insinuar que toda la información recogida y/o difundida por el SIMCE constituya evidencia para política, de acuerdo al señalado enfoque. Ni siquiera que haya habido una acción sistemática e intencionada en esa dirección. Para constituir evidencia para el diseño de políticas, la información, además de robusta y clara, debe estar disponible de manera fácil y oportuna. Creemos que esto último sí se ha cumplido crecientemente en el caso de la información entregada por SIMCE, pero aún no se ha hecho el procesamiento requerido de juntar y complementar resultados de otras investigaciones.

En un afán por aumentar su llegada al público, el SIMCE ha avanzado en la ampliación de la información entregada y en la generación de bases de datos válidas, que puedan ser útiles para los investigadores. Aún cuando el SIMCE forma parte del Ministerio de Educación, permanece el desafío de promover una mejor gestión de la información generada por esta unidad, de modo de vincular más directamente los resultados de la evaluación y la toma de decisiones políticas.

Habiendo dicho lo anterior, queremos señalar que, la información entregada por este sistema de evaluación puede, debe y, en ocasiones, ha entregado evidencias ampliamente consideradas en el diseño de algunas políticas.

Lo que postulamos y queremos ejemplificar en las siguientes secciones, es que al no haber intermediario oficial, como los *brokerage* que menciona Burns y Schuller (2007) o los que existen en Gran Bretaña (Davies, 2004), la opinión pública y los medios de comunicación pueden ser un canal útil para articular el mundo de la investigación y de los datos (incluido el SIMCE) con el mundo político.

Para lograr posicionar un tema en la agenda pública, creemos que hay aspectos del sistema de evaluación que inciden. La duración en el tiempo, la credibilidad, y la rigurosidad del sistema de evaluación (Meckes y Carrasco, 2006) son características claves que posibilitan la recepción de la información evaluativa tanto por parte de la opinión pública como del mundo político.

Consideraremos tres tipos de variables que han tenido distinto peso y desarrollo en la historia del SIMCE y cuya influencia en la política nacional ha variado. Estas son: nivel socioeconómico, prácticas de enseñanza y género.

3.1. Grupos de escuelas según nivel socioeconómico

Como ya se ha señalado, el nivel socioeconómico es el factor asociado que se ha trabajado por más tiempo y con insistencia.

A partir de los resultados del año 2000, los resultados por escuelas han sido entregados públicamente contextualizados según el nivel socioeconómico, de tal modo que es posible comparar los rendimientos de las escuelas, teniendo en cuenta estas características.

Los datos SIMCE han entregado evidencia de grandes inequidades en la calidad de la educación y de las diferencias en los aprendizajes de los estudiantes, con lo cual se ha ayudado a crear consenso sobre la importancia del contexto socioeconómico en los aprendizajes. Indiscutiblemente la equidad educativa está posicionada como un hito ineludible de la agenda pública y el SIMCE ha contribuido a ello.

Con los resultados de PISA 2000 (MINEDUC/UCE, 2007), se hipotetiza que algunos establecimientos pueden lograr un rendimiento superior al que se esperaría según su condición socioeconómica.

Luego, en el Informe Nacional SIMCE 2006 se comprueba con información censal la hipótesis anterior y además se entrega evidencia de las características diferenciales de los establecimientos con rendimiento sobre y bajo lo esperado dentro de su grupo socioeconómico. Se demuestra así que a partir de sus decisiones y gestión interna, las escuelas pueden impulsar el aprendizaje de sus estudiantes, sin estar determinadas por su nivel socioeconómico.

Recientemente, los resultados de PISA 2006 entregan evidencia adicional sobre la inequidad, lo cual permitirá seguir profundizando este tema en el tiempo.

El efecto de esta evidencia en las políticas educacionales se puede apreciar en el diseño e implementación de políticas educativas que promueven la equidad a nivel nacional.

Posterior al *"Programa de las 900 escuelas"*, que también utilizó información SIMCE, se implementó el programa de Escuelas Prioritarias, Liceo para Todos y el Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales Subvencionados (SNED), entre otros. El SNED entrega un premio al 35% de los establecimientos con mejores desempeños entre aquellos que son comparables, considerando tipo de enseñanza, región, niveles impartidos y nivel socioeconómico.

El proyecto de ley sobre Subvención Escolar Preferencial, la cual es aún discutida en el Senado, busca que las escuelas que atiendan alumnos con mayores carencias socioeconómicas, reciban un bono adicional a la subvención común. Con esto, al igual que con la exigencia que todo establecimiento subvencionado incluya en su matrícula al menos un 15% de niños vulnerables, se estimula a los establecimientos a fomentar la integración social.

El gran movimiento social denominado de los "Pingüinos", desarrollado en 2006, sirve para graficar cómo el conjunto de evidencias de inequidad, proveniente del SIMCE y de otras fuentes, se cristalizó en acciones que obligaron al país en general a buscar nuevas propuestas y acuerdos en educación. Así el país ha sido testigo de la discusión sobre la Ley de Subvención Escolar Preferencial y la reforma a la Ley

General de Educación, donde la equidad ha estado al centro del debate con medidas como la no selección de alumnos y la supresión del lucro.

3.2. Factores asociados relativos al proceso de enseñanza y aprendizaje

Estos factores serían de los más potentes, pero tienen complejidades inherentes que dificultan las certezas de las conclusiones a las que se puede llegar. Por una parte, las fuentes de información con que cuenta el SIMCE son cuestionarios autoaplicados y la información es, en cierta medida, de carácter subjetivo ya que las personas responden de acuerdo a sus propias percepciones. Por otra parte, aunque se tenga información censal y un gran número de casos, es difícil controlar todas las variables intervinientes en un proceso de tal complejidad que involucra relaciones humanas y asuntos cognitivos en contextos que pueden ser muy diversos.

Es aquí donde mejor se ejemplifica la necesidad de recopilar y procesar evidencia proporcionada por distintas investigaciones y fuentes, así como de ratificarla con diversas metodologías. Frente a estos mayores requerimientos, debemos reconocer que no se ha trabajado con una intensidad equivalente.

Sin embargo, en un esfuerzo por apoyar la labor docente, los informes de resultados en los últimos años han señalado la asociación entre el rendimiento de las escuelas y los siguientes factores internos a la escuela:

- Preparación para enseñar (cuán preparados se sienten los profesores)
- Cobertura curricular (objetivos curriculares cumplidos)
- Expectativas de profesores y padres (nivel educacional que esperan para sus hijos)
- Usos de resultados SIMCE (si se analizan en la escuela)
- Algunas características de las prácticas pedagógicas

Con este tipo de análisis y la evidencia que ellos arrojan, se busca informar a las escuelas en la forma más clara y explícita posible, para afectar tanto sus políticas, como las prácticas docentes.

A través de estudios específicos, el SIMCE analiza el impacto de la información que llega a las escuelas, lo que permite determinar en qué medida se logran transmitir mensajes comprensibles y útiles, a la vez que entrega pistas sobre cómo se podría mejorar la comunicación para incidir más, tanto en la gestión de las escuelas, como en sus prácticas pedagógicas.

3.3. Género

Este factor ha sido el menos analizado y ejemplifica una preocupación incipiente, surgida de análisis parciales, en distintas mediciones y según distintas fuentes.

En un análisis realizado por el SIMCE⁴ sobre la asociación entre el rendimiento de los estudiantes y el género, se resumen los siguientes resultados:

- En la prueba del año 2004 de 8° Básico, las mujeres superan a los hombres en Lenguaje y Comunicación y los hombres alcanzan mayores niveles de aprendizaje que las mujeres en Educación Matemática, Estudio y Comprensión de la Naturaleza y Estudio y Comprensión de la Sociedad.

⁴ SIMCE, 2006: Análisis de las diferencias de logro en el aprendizaje escolar entre hombres y mujeres.

- La diferencia en Lenguaje y Comunicación es significativa en todos los grupos socioeconómicos, y tiende a aumentar entre los estudiantes de sectores más acomodados.
- Así como en el SIMCE, en PISA 2001 las mujeres también tuvieron un mayor logro que los hombres en Lenguaje.
- En TIMSS 2002, los hombres alcanzan un mayor nivel de aprendizaje que las mujeres, en Matemática y Ciencias.
- En la prueba PISA 2001, donde los alumnos debieron contestar preguntas de situaciones de la vida real e integraron conocimientos y habilidades de Ciencias y Matemáticas con habilidades de interpretación de información y reflexión, mujeres y hombres tienen un similar desempeño.
- Entre la prueba SIMCE 2000 de 8° Básico y la prueba SIMCE 2004 del mismo nivel, la brecha del puntaje entre mujeres y hombres se mantuvo constante en Lenguaje y Sociedad.
- En este mismo período, la brecha entre mujeres y hombres aumentó en Matemática y en Naturaleza. Este aumento se debió a que solo los hombres mejoraron su rendimiento entre el año 2000 y el 2004.
- Al comparar las diferencias entre mujeres y hombres en las pruebas SIMCE 2002 para 4° Básico, SIMCE 2004 para 8° Básico y SIMCE 2003 para 2° Medio, tanto en Lenguaje como en Matemática, se observa que la brecha entre hombres y mujeres aumenta en el transcurso de la enseñanza básica: hay mayores diferencias en el rendimiento 8° Básico son mayores que en 4° Básico.
- En 2° Medio, la distancia entre mujeres y hombres es menor que en 8° Básico, pero sigue siendo importante: a las mujeres les va mejor en Lenguaje, a los hombres les va mejor en Matemática.
- Las mujeres tienen actitudes más positivas hacia la lectura que los hombres: disfrutan más de ella, de comentar libros, e ir a bibliotecas; por su parte, los hombres declaran en mayor medida que leen por obligación y que leer es una pérdida de tiempo.
- Los hombres perciben más apoyo y mayor presión de sus padres para rendir bien en el colegio: declaran obtener más ayuda para estudiar y hacer las tareas y se sienten más exigidos para sacarse buenas notas que lo que declaran las mujeres.

Actualmente, los resultados de PISA 2006 han impactado aún más, ya que las diferencias por género en Chile difieren a los resultados de otros países. En la mayoría de los países el rendimiento en Ciencias o bien no se diferencia por género o la diferencia es a favor de las mujeres. Chile está entre los pocos países con una diferencia importante e inferior en el rendimiento en Ciencias de las mujeres. Además, la diferencia a favor de los hombres en el aprendizaje de Matemáticas, en Chile, es mayor que en todos los otros países evaluados en PISA. Finalmente, en Lectura la diferencia es a favor de las mujeres y la brecha se ha acortado en relación a PISA 2001.

Para llegar a una evidencia que permita fundamentar el diseño de políticas, habrá que desarrollar análisis más profundos para especificar dónde y en qué casos se genera esta inequidad y si es posible establecer a qué factores se asocia y determinar si hay o no discriminación por género.

Con estos ejemplos se ha querido mostrar que a medida que más se informe y se vuelva sobre los temas, los medios de comunicación y la opinión pública más se apropien de las ideas, las que luego se transforman en demanda social por políticas educacionales que favorezcan los rendimientos.

Entre los avances y progresos que están desarrollándose en el SIMCE, en la siguiente sección, señalaremos aquellos que nos ayudarán a comunicar y mostrar mejor la evidencia recogida por el SIMCE.

4. PROYECTOS EN CURSO

4.1. Valor Agregado

Actualmente, se encuentra en desarrollo el proyecto de valor agregado, el cual pretende ser un gran paso en la línea de caracterizar el trabajo de los establecimientos, manteniendo bajo control las variables externas.

La forma actual de entregar resultados de aprendizaje (puntajes promedios y niveles de logro), permite mostrar una "fotografía" del rendimiento de los estudiantes en un momento determinado del tiempo. Sin embargo, el rendimiento de los estudiantes es influenciado por las "diferencias de entrada", o mejor dicho por el bagaje cultural que traen los estudiantes desde su familia. Por lo tanto, las diferencias de puntaje entre establecimientos no sólo responden a diferencias en la calidad de la enseñanza impartida y resulta más justo responsabilizar a las escuelas por lo que efectivamente aportan.

En este contexto, nace y cobra relevancia el proyecto de "valor agregado", mediante el cual se busca realizar comparaciones más justas y válidas de los resultados de las escuelas, evaluando el *progreso* de los estudiantes y la *agregación de valor* por parte de sus establecimientos.

Desarrollar medidas de este tipo, pone el foco en la equidad, al reconocer las diferencias de los estudiantes al inicio de la vida escolar. Así se destacan y valorizan las escuelas que logran superar las posibles desventajas de sus estudiantes y no ser reproductoras de las condiciones con que llegan los estudiantes.

Con el proyecto sobre valor agregado se espera avanzar en los siguientes aspectos:

- Identificar escuelas que entregan más y menos valor agregado, lo que permitirá comparaciones más justas.
- Determinar los factores asociados a las variaciones en el valor agregado que entregan las escuelas. De esta manera, se podrán fomentar las características de los establecimientos que son relevantes para que los estudiantes tengan un alto progreso, independiente de sus condiciones de entrada.

El proyecto de valor agregado plantea al SIMCE el desafío de comunicar de manera sencilla, una metodología compleja, tanto a la comunidad escolar, como a las autoridades y la opinión pública.

Herramienta de análisis fue inaugurada recientemente en la Página Web www.simce.cl Esta herramienta permite obtener fácilmente distintas estadísticas asociadas al puntaje, con el objetivo de fomentar el uso de los datos provistos por el SIMCE e incrementar las fuentes de evidencia para política⁵.

5. CONCLUSIONES

En este artículo hemos mostrado cómo la información desarrollada por el SIMCE ha constituido insumo para las políticas educacionales en el país.

⁵ Muy posiblemente, la simplicidad de estos análisis no constituirán fuente de evidencia, pero en la medida que involucre a más investigadores y educadores, se irá desarrollando un círculo virtuoso.

Si bien este ha sido un proceso lento y probablemente no planificado con anterioridad, el solo hecho que exista información y se haga disponible hace que los medios y la opinión pública tengan un conocimiento más sofisticado de las restricciones y posibilidades de la política educacional.

En un contexto de gran visibilidad de los resultados de aprendizaje y de centralidad de la educación, la formación e información a la opinión pública puede constituirse en un mecanismo poderoso para promover políticas educacionales, como ha quedado ejemplificado en Chile con el "movimiento de los pingüinos".

El Ministerio debe enfrentar el desafío de entregar más y mejor información. La tarea pendiente para la sociedad en su conjunto, es desarrollar un mecanismo de procesamiento científico de la evidencia donde esta pueda ser contrapuesta, analizada, evaluada y complementada a fin de avanzar con mayor rapidez y certeza hacia una mejor educación.

REFERENCIAS BIBLIOGRÁFICAS

- Burns, T. y Schuller, T. (2007). The evidence agenda. En *Evidence in Education. Linking research and policy*. Paris: CERI/OECD.
- Cook, T. y Gorard, S. (2007). What counts and what should counts as evidence. En *Evidence in Education. Linking research and policy*. Paris: CERI/OECD.
- Davies, P. (2004). Is evidence-based government possible? *4th Annual Campbell Collaboration Colloquium*. Washington D.C.
- Evans, M. (2004). At the interface between theory and practice – policy transfer and lesson drawing. En *Learning from comparative public policy: a practical guide*. New York: Routledge.
- Meckes, L. y Carrasco, R. (2006). SIMCE: Lessons from the Chilean Experience in National Assessment Systems of Learning Outcomes. *World Bank Conference on Latin American Lessons in Promoting Education for All*, Cartagena de Indias, Colombia.
- MINEDUC (2003). *Evaluación de Aprendizajes para una Educación de Calidad. Comisión para el Desarrollo y Uso del Sistema de Medición de la Calidad de la Educación*. Santiago de Chile: MINEDUC.
- MINEDUC/UCE (2007). *PISA 2006: Rendimientos de los estudiantes de 15 años en Ciencias, Lectura y Matemática. Resumen ejecutivo*. Santiago de Chile: Ministerio de Educación, Unidad de Currículum y Evaluación.
- OECD. (2003). *Literacy Skills for the World of Tomorrow. Further results from PISA 2000*. Paris: OECD/PISA.
- OECD/CERI (2003). *New challenges for educational research*. Paris: OECD.
- OECD/CERI. (2007). *Evidence informed policy. Linking research and policy*. Executive Summary. Paris: OECD.
- SIMCE/Ministerio de Educación (1988). *Manual de interpretación de resultados y orientaciones pedagógicas*. Santiago: MINEDUC.
- SIMCE/Ministerio de Educación (1990). *Manual de orientaciones pedagógicas*. Santiago: MINEDUC.
- SIMCE/Ministerio de Educación (1991). *Manual de interpretación de resultados*. Santiago: MINEDUC.



**APORTES Y DESAFÍOS EN LA EVALUACIÓN DE POLÍTICAS
Y PROGRAMAS DE MEJORAMIENTO EDUCATIVO.
REFLEXIONES DESDE NUESTRA PRÁCTICA DE
EVALUACIÓN EN CHILE**

Dagmar Raczynski y Daniel Salinas

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art6.pdf>



1. FORTALECIMIENTO DE LA INVESTIGACIÓN Y LA EVALUACIÓN DE POLÍTICAS Y PROGRAMAS

En el año 1993 José Sulbrandt planteaba un severo diagnóstico del estado de la evaluación de políticas sociales en América Latina, al afirmar que “la mayoría de las políticas y programas sociales en la región no reciben ninguna evaluación y en los casos en que ésta se ha aplicado se ha llevado a efecto de una manera que no satisface plenamente los requisitos de una investigación profesional” (Sulbrandt, 1994:310). Además, afirmaba que muchas veces los informes de evaluación realizaban “un deliberado ocultamiento de todos aquellos puntos que puedan servir para formular algún juicio técnico desfavorable al programa”, y que “salvo casos muy excepcionales, no ha existido relación alguna entre las evaluaciones realizadas y los procesos de gestión de los programas respectivos” (Sulbrandt, 1994:311).

Quince años después, podemos decir que hoy día en América Latina y en Chile estamos bastante mejor que lo que denota esa descripción. En estos años se ha avanzado en la institucionalización y profesionalización de la evaluación de políticas públicas en distintas áreas, incluida la de educación. Las evaluaciones de programas y proyectos se han multiplicado, las hay de distinto tipo: procesos, resultados, impacto, costo y costo-efectividad, satisfacción. Junto con ello se dispone de más y mejores diagnósticos de las realidades concretas y de segmentos sociales afectados por una problemática. Como consecuencia, las decisiones de política se apoyan y fundamentan en mayor medida que en el pasado en evidencia empírica. No obstante, permanece la tendencia de buscar hallazgos e información que sean coherentes con la postura de quién toma una decisión y la solución que prefiere y a la cual ha optado a priori, dejando de lado evidencia que cuestiona la solución.

Los factores que han facilitado o empujado la mayor disponibilidad, uso y análisis de información empírica, son varios y se listan e ilustran, a modo de ejemplo, en la Tabla 1. De un lado, está la importancia que la evaluación ha adquirido en los organismos internacionales y los esfuerzos que éstos han puesto en generar indicadores comparativos entre países. Este trabajo se ha visto facilitado por la disponibilidad de instrumentos y técnicas de medición (en particular en el ámbito de aprendizaje) y el menor costo relativo que recientemente tiene la recolección y el procesamiento de información. Por último, a nivel nacional, destaca la voluntad política de invertir esfuerzos y recursos en medir resultados y evaluar los programas por parte del poder ejecutivo, legislativo y el mundo privado.

La formación de un acumulado de información, estudios e investigaciones sobre variados temas de la educación, incluyendo evaluaciones de programas de mejoramiento educativo ha llevado a reconocer las importantes insuficiencias existentes en la implementación de la política y reforma educacional chilena. Adicionalmente, el reconocimiento de esas insuficiencias y límites plantea nuevos desafíos temáticos y metodológicos a la propia evaluación educativa, que debe perfeccionarse para contribuir a entender y superar los límites de la reforma. Es en torno a estos nuevos desafíos que se estructura este artículo, basado en nuestra experiencia en realizar evaluaciones y estudios sobre programas de mejoramiento educativo y sus efectos en los establecimientos educacionales¹.

La sección 2 del artículo define el contexto de reforma educativa y los desafíos temáticos que han ido surgiendo, entre ellos la importancia de profundizar las evaluaciones de procesos, que es el tema de la sección 3. La sección 4 y 5 plantean algunos de los desafíos metodológicos que se enfrentan al identificar y medir los procesos escolares claves que inciden sobre los aprendizajes de los estudiantes, delimitando

¹ Los trabajos que subyacen a esta reflexión corresponden a una evaluación del programas de las 900 escuelas de escasos recursos (Asesorías para el Desarrollo y Santiago Consultores (2001); Sistematización y evaluación del programa de escuelas críticas aplicado en la región metropolitana de Santiago (PNUD-Asesorías para el Desarrollo, 2004); un trabajo en curso de monitoreo, seguimiento y evaluación de la Estrategia de Apoyo a Establecimientos Prioritarios (PNUD-Asesorías para el Desarrollo, 2007-08), a lo que se suman un estudio sobre supervisión educacional en Chile. Experiencias públicas y privadas (Raczynski y Muñoz, 2005); y una investigación sobre Efectividad Escolar y Cambio Educativo en Condiciones de Pobreza en Chile (Raczynski y Muñoz, 2005); y otro sobre escuelas efectivas (Bellei y otros, 2004).

las dimensiones claves de observación. Las secciones 6, 7 y 8 se detienen en tres temas de política educativa: el difícil balance entre un diseño estructurado y al mismo tiempo flexible para lograr un “empalme” fructífero de lo que trae el programa con la dinámica de la unidad educativa; la importancia de evaluar políticas y no sólo programas; y la urgencia de acercar los hallazgos de las evaluaciones y estudios a los tomadores de decisiones de política.

TABLA 1. FACTORES QUE HAN CONTRIBUIDO A FORTALECER LA EVALUACIÓN EDUCATIVA EN CHILE

Factores	Ilustraciones-ejemplos de Chile
1. Organismos Internacionales	
Préstamos de organismos internacionales que apoyan el diseño e implementación de programas específicos, requiriendo componentes de evaluación en el diseño del programa	- Banco Mundial apoyó técnica y financieramente los programas Mejoramiento de la Calidad y Equidad de la Educación (MECE) y ha requerido de evaluaciones intermedias y de impacto. En el contexto de estos programas se realizó un exhaustivo diagnóstico de la situación de la enseñanza media. - Banco Mundial y el BID están asesorando al Programa Chile Crece Contigo (desarrollo integral del niño desde el embarazo de la madre hasta los 5 años de edad) y diseñando su evaluación de impacto. Este programa es parte del Sistema Nacional de Protección Social y combina acciones en salud y estimulación temprana, apoyo al cuidado que el niño recibe en el hogar, más cobertura y calidad instancias de educación pre-escolar y estímulo a factores proyectores en el espacio comunitario.
Organizaciones internacionales han promovido mediciones comparativas de la situación de la educación entre países y algunas han elaborado Informes sobre la situación y política educacional. elaboran informes en temas específicos en países particulares	- Chile ha participado en las pruebas PISA, TIMSS, Alfabetización de adultos, Sites, Llece, el Estudio internacional de Educación Cívica. Variados organismos (OECD, BID, Banco Mundial, UNESCO-OREAL, PREAL han elaborado informes que comparan la situación educacional de distintos subconjuntos de países. La OECD publicó en 2004 una Revisión de las Políticas Nacionales de Educación de Chile. ón en donde le tamde la educación Al mismo tiempo organismos.
2. Avances en enfoques y tecnologías de evaluación	
A nivel internacional	- Especialización creciente en el tema de evaluación, disponibilidad de instrumentos de medición y análisis, desarrollo de tecnologías sofisticadas y eficientes de manejo de gran cantidad de datos. - Seminarios de intercambio y reflexión sobre los desafíos de la evaluación
A nivel internacional	- Incorporación del tema de la evaluación en programas de formación postsecundaria en las mallas curriculares en ciencias sociales, ingeniería, ciencias de la salud y de la educación, entre otras; diplomados que profundizan en el tema. - Seminarios y Talleres sobre el tema.
3. Demanda Social y decisión política del país de invertir en evaluación	
Instalación en el gobierno de oficinas de evaluación	-La DIPRES ha desarrollado un Sistema de Control de Gestión y Presupuesto por Resultados que tiene dos líneas: i) la evaluación de programas gubernamentales y ii) la evaluación de instituciones (evaluación comprensiva del gasto). La primera se inicia en 1997 y contempla dos modalidades de evaluación: una de procesos y resultados y otra de impacto. Varios programas de educación han sido evaluados dentro de este esquema.
Demandas por “evaluaciones” en temas específicos desde los ministerios sectoriales y servicios públicos, y también, crecientemente desde Fundaciones y Organismos no gubernamentales que deben o quieren demostrar la pertinencia y efectividad de su labor.	-El Ministerio de Educación ha encargado a instituciones externas evaluaciones de procesos, resultados e impacto de varios de sus programas (Programa de las 900 escuelas de escasos recursos, programa de escuelas críticas, estrategia de apoyo a escuelas y liceos prioritarios). -Algunos pocos municipios han contratado evaluaciones diagnóstico de la situación de los establecimientos educativos a su cargo como paso previo a la toma de decisiones en torno a medidas específicas a tomar para mejorar la calidad de la educación que imparten. - El Hogar de Cristo (Organismo de Iglesia comprometido con la opción por los pobres) ha encargado evaluaciones de algunos de sus programas, en el área infanto-juvenil y otras. - ONG y Fundaciones crecientemente están demandando apoyo en temas de seguimiento y evaluación de sus programas

Fuente: elaboración de los autores.

2. LÍMITES DE LA REFORMA EDUCACIONAL CHILENA: AUSENCIA DE UN ENFOQUE DE ACTORES Y DE UNA MIRADA A LOS PROCESOS Y PRÁCTICAS QUE IMPERAN EN LA UNIDAD EDUCATIVA

Los cuatro Gobiernos de la Concertación de Partidos por la Democracia, que se suceden a partir de 1990, han realizado esfuerzos contundentes por cumplir con su promesa de mejorar la calidad y equidad en la distribución social de los aprendizajes². En pos de esa meta hubo incrementos graduales significativos en el gasto público en educación; se formuló e implementó un nuevo currículo para los distintos niveles de enseñanza; se amplió la oferta de cursos de perfeccionamiento docente; se extendió la jornada escolar; se invirtió fuertemente en infraestructura y equipamiento (textos para niños y profesores, bibliotecas de aula, recursos didácticos, computadores e internet, laboratorios, bibliotecas, etc.); mejoraron los salarios de los profesores; se incrementaron los beneficios asistenciales para los estudiantes en situación de vulnerabilidad social (alimentación, salud, útiles escolares, becas de retención y para minorías); se pusieron en marcha varios mecanismos de incentivo al buen desempeño de las escuelas y liceos y para profesores individuales y un proceso de evaluación de desempeño docente; se diseñaron e implementaron diversos programas de mejoramiento educativo que representan acciones en las unidades educativas dirigidas a mejorar el trabajo de los profesores en el aula, la gestión escolar y la convivencia.

Pese a estos y otros esfuerzos, los resultados de aprendizaje a nivel agregado no mejoraron y la brecha de aprendizaje por nivel socioeconómico de los estudiantes no ha disminuido significativamente. Este resultado no significa que la reforma educacional de los años 90 y hasta el presente no tuvo efectos ya que ha habido avances significativos en acceso, cobertura y permanencia de los estudiantes en el sistema; mejoras significativas en las condiciones básicas y materiales en las cuales ocurre la enseñanza-aprendizaje, incluyendo mejoras en las condiciones laborales de los docentes. Lo que flaquea son los avances en resultados de aprendizaje y la distribución social de ellos.

¿Por qué, a pesar de todos los esfuerzos e inversiones realizadas, el sistema educativo no ha logrado constituirse en una fuente de desarrollo y de distribución equitativa de las oportunidades? Las evaluaciones y estudios apuntan a dos grupos de factores que han sido gravitantes. En primer lugar están los factores que dependen casi exclusivamente de la escuela³: la gestión directiva y técnico-pedagógica, la interacción profesor-alumno y el trabajo de los docentes en el aula, la estructuración del tiempo escolar, el clima organizacional y compromiso de la comunidad escolar con la escuela. Por otra parte, hay factores exógenos a la unidad educativa, que no dependen de ella y que con frecuencia se denominan “variables estructurales”. Encontramos acá elementos como la formación inicial docente, la estructura administrativa del sistema educativo, el marco regulatorio que lo rige, la modalidad de financiamiento. Los estudios verifican la presencia de factores de ambos tipos y que modificaciones en los factores estructurales por sí mismo no conducen al resultado esperado de mejores aprendizajes. Para que ello ocurra es indispensable poner en el centro de la política a la escuela, esto es, enfocar la escuela como organización o sistema y trabajar con el conjunto de los actores involucrados en ella, no aisladamente uno del otro, sino que en su mutua relación. Es la dinámica entre los actores y las prácticas escolares imperantes la que debe comprenderse, para a partir de ellas identificar los nudos centrales que dificultan que la escuela mejore y a partir de ahí identificar claves que permitan que los actores en ella inicien una senda que los lleve a procesos de mejora continua. Encarar este desafío no es tarea sencilla. Requiere de un enfoque y modalidad de trabajo, insuficientemente desarrollado hasta hoy, que

² Son muchos los trabajos que analizan la reforma educacional chilena. Entre ellos destacan la colección de artículos en Cox (2003) y García-Huidobro (1999). Una visión sintética y más actualizada puede verse en Raczynski y Muñoz (2007).

³ En el texto se utilizan indistintamente los términos escuela, colegio, establecimiento educacional, aludiendo a unidades educativas de la enseñanza básica y media.

identifique, conozca y comprenda la lógica de acción de y entre los actores, y construya desde ellos y con ellos caminos de cambio.

3. NECESIDAD DE PROFUNDIZAR EL ESTUDIO Y EVALUACIÓN DE PROCESOS ESCOLARES EN LAS UNIDADES EDUCATIVAS

Distintos observadores de la educación (desde especialistas hasta actores políticos y la opinión pública) tienden muchas veces a reducir todo análisis del fenómeno educacional a los resultados de aprendizaje y sus tendencias agregadas de variación (caída, estancamiento, mejora a nivel país). En Chile, esto se da porque la información más estandarizada y comparable con la que se cuenta sobre el sistema educativo completo es la que ofrece el SIMCE, la prueba que mide aprendizajes cognitivos en tres niveles educativos (al cuarto, octavo y décimo año de enseñanza) y en algunos sub-sectores del currículum obligatorio. Desde la perspectiva de la evaluación educativa, esto se ha traducido en la preponderancia que a la hora de evaluar las políticas educativas (estrategias, intervenciones, programas de mejoramiento, etc.) es otorgada a la medición del impacto en los resultados de aprendizaje según la prueba SIMCE. El desarrollo de esta prueba ha sido aporte fundamental a la educación del país.

No obstante, desde el punto de vista de los *policy-makers* de la educación, es decir, de las autoridades políticas y técnicas encargadas de elaborar e implementar las políticas educativas, la información de *impacto* tiene una utilidad limitada. Permite saber qué efectividad han tenido una iniciativa, pero dice muy poco respecto de los factores que explican los buenos o malos resultados obtenidos: no ofrece una mirada en profundidad al cómo funcionan las cosas en la práctica cotidiana de las unidades educativas, cómo es que los actores interpretan y se apropian de los apoyos externos, etc. Esto es lo que permiten las evaluaciones de *procesos*. Poco a poco las autoridades han ido comprendiendo que, para tomar decisiones, necesitan de esta información, que la mayor parte de las veces es cualitativa, se fundamenta en estudios de casos, entrevistas abiertas, grupos focales y de discusión, observación *in situ*, por nombrar las técnicas principales. Importantes reflexiones teóricas concuerdan en esta conclusión. En *Development as Freedom*, Amartya Sen aborda el tema de la evaluación y llama a no perder de vista el proceso de instalación de capacidades y competencias en los actores beneficiarios, de modo tal que éstos puedan actuar con autonomía y de un modo sustentable en el tiempo en la definición de sus objetivos, estrategias y acciones (Sen, 1999:54-86)⁴. Cada vez más estudios en el área de la educación plantean evidencia en la misma dirección (Fullan, 2005, 2006; Elmore, 2004). La evaluación de procesos implica observar de qué manera cada una de las partes que componen la organización escolar –y en especial directivos, docentes, alumnos y apoderados– se nutren de la mejora institucional a través de innovaciones en sus propias prácticas de trabajo, ya sea al nivel de la gestión escolar y pedagógica como de la enseñanza directa en el aula o la interacción escuela-familia. El estudio de estos tópicos obligan a adentrarse en la “caja negra” de las escuelas.

⁴ Sen ha formulado adecuadamente los límites de lo que llama el “consecuencialismo”, esto es una perspectiva que sólo juzga las decisiones o proyectos por los resultados que generan. Decir que “sólo los resultados importan”, argumenta Sen, implica valorar una acción únicamente por sus consecuencias, excluyendo la consideración de que existen ciertos principios o derechos que son importantes o valorables con independencia de sus resultados (al margen de que muchas veces el resguardo de estos derechos sea *además* funcional a los resultados de calidad y equidad deseados en el caso de la política educativa). De la misma manera, Sen toma distancia de una postura que se centra enteramente en los principios y pierde de vista la importancia del logro de resultados. En ambos casos, la evaluación de políticas públicas estaría restringiendo su base de información posible (“informational base”) sin considerar o deliberadamente excluyendo a todo un conjunto adicional de hechos significativos. Sen llama a integrar ambas perspectivas y a no perder de vista el proceso de instalación de capacidades y competencias en los actores beneficiarios, de modo tal que éstos puedan actuar con autonomía y de un modo sustentable en el tiempo en la definición de sus objetivos, estrategias y acciones (Sen, 1999:54-86).

4. LOS ACTORES ESCOLARES SON LOS MEJORES INFORMANTES, PERO NO SIEMPRE LO QUE DICEN ES CONFIABLE: ES NECESARIO COMPARAR Y CONTRASTAR LAS MIRADAS INTERNAS CON UNA EXTERNA

Reconocer la necesidad de abrir esa “caja negra” que es el aula y también la escuela, con sus procesos institucionales claves, desde el punto de vista de la evaluación implica el tremendo desafío de producir una reconstrucción de la realidad de la escuela que equilibre adecuadamente las *miradas internas* de los actores escolares (sostenedores, directivos, profesores, estudiantes, apoderados) y la *mirada externa* de los propios investigadores (categorías de análisis, información previa, etc.). El desafío metodológico es grande y las dificultades se manifiestan en varios frentes:

Las visiones de los distintos actores de una comunidad escolar no sólo son distintas, sino que a veces no son consistentes entre sí. Los distintos estamentos y personas que la integran tienen versiones propias, distintas y muchas veces contradictorias sobre los procesos claves y las prácticas pedagógicas en ella. Por ejemplo, el déficit de aprendizaje de un determinado curso, que para un profesor se debe al peso de los factores de vulnerabilidad social-familiar, para un apoderado puede deberse a la falta de expectativas que el docente ha puesto en las capacidades de los alumnos, para un directivo al exceso burocrático de programas ministeriales ineficaces que entorpecen la gestión técnica de la escuela, o para un jefe técnico pedagógico a las debilidades en las competencias de los profesores. El discurso de cada estamento tiende a acomodarse *ad-hoc* a las características de su rol en la escuela, y el investigador queda en la difícil situación de conciliar esas interpretaciones heterogéneas y derivar de ellas conclusiones sobre los procesos institucionales en una escuela, dónde están los nudos y porque se producen e identificar potenciales claves de mejora.

En el contexto chileno –donde a nivel de los actores escolares, especialmente en el sector público, existe una sensación dominante de estancamiento en materia de calidad y equidad– los actores tienden a tener un discurso defensivo que atribuye la responsabilidad de los resultados siempre a otros actores o factores. La reticencia de los profesores frente a las evaluaciones externas sobre su desempeño es sintomática en este sentido. Muchas veces las entrevistas a directivos o sobre todo a docentes se convierten en pequeñas catarsis en las que éstos descargan sus frustraciones cotidianas, a través de discursos caracterizados por la victimización frente a las condiciones adversas. Más allá de lo efectivamente duras que son las condiciones de la docencia en contextos de vulnerabilidad, desde el punto de vista metodológico se enfrenta la necesidad de ir más allá de ese discurso aprendido y cerrado, que no informa sobre el núcleo de lo que el investigador busca aprehender, esto es, la realidad concreta del establecimiento escolar.

La escuela está cruzada por una serie de relaciones de poder e interés entre sus miembros que hacen que muchas veces los entrevistados opten por responder según el “deber ser” o por la respuesta más “conveniente” según su posición. Los cuestionarios y encuestas muchas veces se prestan más a este tipo de distorsión que las entrevistas co-presenciales. No es descartable la hipótesis de que esta tendencia de los directivos y docentes a responder según la “deseabilidad social” esté relacionada con la falta de expectativas en la posibilidad de cualquier clase de cambio radical en el sistema, producto del condicionamiento que experimentan cotidianamente frente a una realidad educativa injusta “contra la que nada se puede hacer”. Adaptan deseos y expectativas –y por lo tanto sus respuestas– a lo que les parece factible. Las preguntas de satisfacción (tales como ¿qué nota le pondría usted a tal o cual proyecto? ¿cómo calificaría la asesoría brindada a la escuela?, etc.) son, por lo general, engañosas: “cualquier aporte está bien” en medio de una realidad tan difícil. Muchas veces los actores suben sus evaluaciones calculando que una nota negativa podría desembocar en la interrupción del apoyo y “siempre es mejor algo que nada”.

Por razones de este tipo, los actores no son fuentes de información transparentes de su propia práctica, pero al mismo tiempo son una de las principales fuentes de información de la que dispone la evaluación

educativa. Al mismo tiempo, se sabe que la autoevaluación es un momento indispensable y clave en los procesos de mejora. A la hora de evaluar procesos, la mirada externa del investigador consiste precisamente en poner en relación los discursos de los distintos actores para ver más allá de lo que cada uno de ellos ofrece explícitamente, y triangular esa realidad particular con el saber previo que ha producido la literatura y que ofrece las categorías a partir de las cuales se han definido las variables de observación y con su propia observación de la realidad.

En nuestros evaluaciones hemos aprendido que las preguntas más informativas son las que remiten a las prácticas efectivas; que hay que preguntar y simultáneamente observar distintas situaciones, por ejemplo, conversaciones espontáneas en las salas de profesores, las entradas y salidas del recinto escolar, la presencia del director, el uso de los espacios y equipamiento, las reuniones con padres, de profesores, del equipo directivo, las salas, la conformación grupos, etc. Los profesores informan con más honestidad sobre las prácticas del director que lo que este señala sobre sus propias prácticas, del mismo modo que los alumnos dicen más sobre las prácticas docentes que éstos. Lo que queremos decir es que en determinados temas las respuestas de algunos informan más que las de otros que, por su posición o cargo, responden siguiendo sus percepciones del “deber ser” y deseabilidad social.

5. CONCRETIZANDO EL DESAFÍO: EVALUACIÓN COMPRENSIVA DE ESTABLECIMIENTOS EDUCACIONALES

En nuestros trabajos hemos desarrollado una metodología de “estudios de casos” que representan una evaluación comprensiva de una muestra de establecimientos escolares, centrada en procesos institucionales claves y la identificación de actitudes, expectativas y conductas de los actores involucrados. A través de un trabajo eminentemente cualitativo reconstruimos la realidad y dinámica del establecimiento, cubriendo dimensiones como las que lista la Tabla 2.

Nuestros estudios en Chile en escuelas en sectores de pobreza nos han llevado a identificar una jerarquía entre las características relevantes de los establecimientos que obstaculizan la efectividad escolar y procesos de mejora. Identificamos un piso de condiciones sin las cuales resulta casi imposible avanzar, que denominamos “normalización” y que refiere al funcionamiento básico de la organización escolar. Abarca dimensiones como existencia y cumplimiento de horarios y normas de funcionamiento, división del trabajo y roles, disciplina, convivencia y respeto de la persona, aseo, mantención e higiene, etc. Supone la existencia de insumos mínimos: infraestructura, equipamiento básico, personal directivo y docente, etc.

Concomitantemente, identificamos lo que denominamos “columna vertebral de la efectividad”, y que refiere a procesos escolares y actitudes, expectativas y conductas de directivos y docentes que verifican una gestión escolar centrada en lo pedagógico: preocupación permanente por el aprendizaje de los alumnos, expectativas altas respecto al potencial de aprendizaje de los mismos⁵, un trabajo activo (no

⁵ Los profesores y directivos creen en el potencial de sus alumnos y transmiten sus expectativas a ellos, haciéndolas públicas, motivándolos a superarse, a aprender. La motivación de los alumnos es una de las principales preocupaciones del profesor en clase. No se escuchan opiniones que indican que el principal problema de su enseñanza es la falta de motivación de los niños. Más bien sostienen que la motivación es una variable que depende del profesor y de la escuela y ellos son responsables de activarla. Las expectativas altas están presente no solo en la relación profesor – alumno sino que también en los directivos y líderes pedagógicos respecto al potencial profesional de los docentes. Existe en las escuelas efectivas una visión de la escuela como una “que puede más” y que se esfuerza por superarse. Directivos y profesores proyectan una visión positiva del futuro de la escuela.

rutinario) de los profesores en el aula, trabajo que es evaluado y cuenta con el respaldo y apoyo regular del equipo directivo (hay observación en aula de pares y del equipo directivo). Estas características van asociadas con un fuerte liderazgo pedagógico que se expresa en apoyo y altas exigencias del equipo directivo hacia los profesores y de éstos hacia los alumnos; y se expresan en un sistemático trabajo colaborativo de reflexión y análisis pedagógico entre docentes que gira en torno a lo que se hace y logra en el aula; y, más en general, en una planificación del trabajo que se traduce en metas concretas y priorizadas, que se monitorean regularmente en cuanto a su aporte a una buena enseñanza y aprendizaje; directivos que tienen un diagnóstico claro sobre las fortalezas y debilidades del cuerpo docente y de los alumnos; trabajan con los docentes de modo de aprovechar sus fortalezas y suplir sus debilidades; los docentes, a su vez, conocen muy bien a sus alumnos (cada uno) y la escuela lleva adelante mediciones para ver en qué fallan y cómo suplir esto; los directivos negocian “apoyos” que puedan favorecerla y los que le llegan por iniciativa de otros, como por ejemplo, el MINEDUC, una empresa o una universidad no desordenan ni dispersan el trabajo y foco pedagógico, la carta de navegación, que ella se ha propuesto.

TABLA 2. DIMENSIONES MÍNIMAS PARA UNA EVALUACIÓN COMPRENSIVA

El establecimiento y su entorno: características estructurales, difícilmente modificables desde dentro de la escuela o liceo, pero que no pueden ser ignoradas porque es el medio (comunidad) en el cual el colegio opera y del cual recibe porcentajes importantes de sus alumnos. Incluye características del entorno local, características del alumnado y rasgos relativamente estables del colegio como lo son antigüedad, dependencia administrativa, niveles y modalidades de enseñanza que imparte.

Resultados de aprendizaje en los últimos años, por nivel, curso y asignatura, promedio y dispersión.

Otros resultados: tendencias en matrícula, retención / abandono de alumnos, presencia de señales de discriminación y exclusión de alumnos.

Insumos y recursos: recursos con que cuenta el colegio, materiales y humanos. Incluye la adecuación de la infraestructura y equipamiento básico, la disponibilidad y características de los recursos humanos directivos, docentes y no docentes, y la disponibilidad de recursos didácticos y de aprendizaje.

Procesos o condiciones institucionales internas, esta dimensión es central e incluye varias subdimensiones. Todas ellas dicen relación con la forma en que el colegio se organiza para la enseñanza-aprendizaje de sus alumnos y las dinámicas de trabajo que se generan a partir de esa organización. Las subdimensiones más importantes son:

Capacidad de liderazgo del director y del equipo directivo que lo acompaña.

Existencia y características del proyecto educativo y el lugar que ocupa la enseñanza aprendizaje en éste, y gravitación que tiene sobre las decisiones que se tomen, planificación y evaluación global.

Forma en que se organizan e implementan las tareas pedagógicas en el colegio: presencia o no de trabajo sobre la base de objetivos compartidos, características de la planificación, apoyo concreto del equipo directivo a los profesores de aula, trabajo en equipo entre éstos, reflexión pedagógica colectiva, seguimiento y evaluación del trabajo de aula por parte del equipo directivo.

Organización curricular, preparación de la enseñanza, acción docente en el aula y prácticas en uso para evaluar la implementación curricular.

Disciplina, convivencia escolar y apoyo al desarrollo psicosocial, afectivo y físico de los estudiantes.

Clima organizacional, gestión de recursos humanos y oportunidades de desarrollo profesional del personal.

Imágenes de directivos y docentes sobre los alumnos y su familia (señales de estigmatización y mal trato), expectativas de los docentes respecto al potencial de aprendizaje de los alumnos.

Acciones del colegio para incorporar a los apoderados y la familia a la tarea educativa.

Relaciones institucionales externas. Las escuelas y liceos no operan de modo aislado, sino que están en interacción permanente con otros agentes, y las características de estas interacciones inciden sobre la situación de los recursos e insumos, los procesos internos y los resultados. De especial relevancia son las relaciones siguientes:

Con el sostenedor a cargo del colegio y responsable final de los resultados que obtiene (en qué y como se involucra, qué tipo de apoyo presta, etc.).

Relaciones con los niveles provincial y regional del MINEDUC; específicamente, cuando corresponde la relación con él o los encargados de la supervisión formativa, sea ésta directa o delegada en una institución experta; y emitir un juicio sobre este trabajo, estado en que se encuentra y avances que se han tenido.

Redes de apoyo con que cuenta y en que participa el establecimiento con universidades, empresas, ONGs, Fundaciones y con otras escuelas y liceos.

Fuente: Elaboración de los autores.

Otra dimensión constitutiva de la efectividad escolar es la capacidad para adecuarse a los cambios que enfrenta desde el exterior e internamente. En estas escuelas los problemas, tanto internos como del entorno, se reconocen y solucionan. Directivos y docentes se preocupan por identificar y reflexionan sobre los cambios que la afectan, considerándolos al momento de tomar decisiones. En este plano, la información sobre alumnos y profesores y el entorno se constituyen en insumos claves para construir un diagnóstico sobre la escuela, sus desafíos, sus fortalezas y debilidades, y definir cursos de acción concretos para avanzar y funcionar de modo mas fluido y de esta forma apuntar a mejorar los resultados de aprendizaje. La escuela efectiva busca y procesa información sobre su entorno.

A la normalización y la gestión pedagógica centrada en el aprendizaje se suman otras características que facilitan la efectividad, pero cuya existencia por si mismo no la aseguran. Estas son dos:

a) El clima, compromiso e identidad con la escuela. Para que una escuela funcione se requiere una cuota de confianza entre los integrantes de la unidad educativa, una identidad y compromiso y un clima que permitan un fluido trabajo pedagógico colaborativo, donde se revisa lo que cada uno y todos en conjunto hacen, se critica, se enmienda el rumbo, se plantean mejoras. Pero lo que define la efectividad no es la confianza, identidad, clima y compromiso sino en el cómo éstos se ponen al servicio del proceso de enseñanza – aprendizaje, y no de otros fines o propósitos. De hecho, un exceso de buen clima o un cuidado exagerado de éste puede llevar a desvirtuar un trabajo pedagógico exigente en la escuela.

b) Las características de la relación escuela – familia. Las escuelas efectivas no esquivan el tema de los padres y apoderados. Cada una lo asume y desarrolla acciones concretas las que, sin embargo, son muy disímiles entre sí. Hay escuelas que sólo exigen de los padres un mínimo de responsabilidad para con sus hijos y basan su alianza en un compromiso básico de que la familia intentará no ir en contra de la educación que la escuela le está entregando a los niños. En el otro extremo, es posible encontrar escuelas que han incorporado a los padres de familia en la sala de clases, a través de sistemas de tutoría. Lo central es que, aún con estrategias muy distintas, todas las escuelas entienden que el tema de la familia y su participación en la educación de los hijos no pueden dejarse de lado. No detectamos, sin embargo, una relación directa en el sentido que una mayor cercanía escuela – familia se encuentre asociada necesariamente a mejores resultados educativos.

En términos de evaluación desde nuestro punto de vista realizar evaluaciones comprensivas de los establecimientos educacionales en dos o más puntos en el tiempo, por ejemplo, cada 3 años, es una alternativa real para identificar avances y retrocesos en los procesos escolares. Si estas mediciones se combinan con resultados de aprendizaje, y mejor, estimaciones de valor agregado, se contaría con una fuente de información potente para evaluar establecimientos educacionales e identificar oportunidades u obstáculos para sus procesos de mejora. No obstante, previamente habría que “codificar” y definir una métrica para cualificar los procesos. A modo de ilustración se incluye en el Anexo un primer intento en esa dirección.

6. DESAFÍOS PARA LA POLÍTICA EDUCATIVA

6.1 El difícil balance entre estructuración y flexibilidad. Los límites de intervenciones rígidas y altamente estandarizadas

Los programas ministeriales, casi sin excepción, son intervenciones estandarizadas en el sentido que su diseño predefine las acciones a desarrollar en la escuela, los tiempos y secuencia de las mismas, la modalidad de trabajo, etc. Este molde deja escaso espacio para acoger la situación específica de cada escuela, dificultando una adecuado “empalme” entre lo que trae el programa y las necesidades,

fortalezas y problemas de cada escuela. La literatura sin excepción revela que la mejora en aprendizaje de los estudiantes sólo tiene éxito si se construye desde dentro de la escuela. No cabe duda que las escuelas requieren de apoyo externo para mejorar. Sin embargo, este apoyo no logra resultados sostenibles sino se entrelaza con liderazgos y voluntad interna de asumir el desafío del cambio, darle vida y liderarlo. Esto es, si el apoyo externo se transforma en una mejora depende de elementos en el plano subjetivo: compromiso, expectativas, mística, voluntad de cambio y creencia de que es posible. La escuela debe “endogenizar” los apoyos que recibe, hacerlos suyos y transformarlos en su proyecto.

Lo anterior sugiere la necesidad de diferenciar en el momento de diseñar un programa entre un diseño de primer y segundo orden. El diseño de primer orden corresponde a las decisiones macrosociales y políticas (propósitos y objetivos que se buscan, orientaciones estratégicas y principios que se espera orienten las acciones, por ejemplo integralidad o participación) mientras que el de segundo orden corresponde al diseño que elaboran los “operadores locales” del programa en el marco del diseño general. El diseño de segundo orden corresponde así a los responsables de implementar la política, categoría que en general es vista como “ejecutor” sin ingerencia en el diseño. Rescatar su papel y responsabilidad en el diseño y prepararlos para esta tarea es un reto importante para la política.

La evidencia de nuestros estudios verifica que diseños rígidos y altamente estandarizados dificulta que las escuelas le den un sentido propio al cambio y se apropien de él, negociando para adaptar el programa a su realidad y así aprovecharlo mejor. En este sentido, las políticas de mejoramiento educativo deben cuidar que el cambio educativo que buscan sea vivenciado y comprendido desde la escuela y demandado por ésta. Desde los programas se requiere de una modalidad de trabajo abierta, menos estandarizada, con mayores opciones para la escuela, más a la medida de ésta. Entre otros, esto requiere la llegada e invitación a participar en un programa o acción ministerial considere una conversación con el equipo directivo (y sostenedor) sobre su proyecto educativo, las iniciativas en curso y los aportes que en ese contexto haría el programa, y las responsabilidades que ellos asumen en la implementación del programa. Como complemento a esta conversación, se debería contar con una evaluación comprensiva del establecimiento: la forma en que se está desempeñando, sus fortalezas y debilidades, sus planes y capacidad de mejora, poniendo atención en los procesos y condiciones institucionales y diseñando un plan de mejora con la escuela.

El punto anterior plantea tremendos desafíos al diseño y gestión de la política educativa. En general en nuestros países se detecta una multiplicación de programas y medidas que llegan y/o afectan a la escuela en distintos tiempo, con reglas distintas, con actividades parcialmente superpuestas. A modo de ejemplo, en Chile hay programas de desarrollo profesional docente y al mismo tiempo los programas de mejoramiento e innovación pedagógica implementan acciones de capacitación y perfeccionamiento; está el programa de distribución de textos escolares y diversos programas a su vez distribuyen textos y material didáctico; varias iniciativas incluyen apoyos en gestión y solicitan al establecimiento la formulación de planes de acción concretos, los que se superponen; todos los programas solicitan información a los establecimientos parte de la cual se repite, pero la solicita en momentos del tiempo y con formatos distintos. Esta modalidad y forma de trabajo del Ministerio ha ido en desmedro de una visión integral de la unidad educativa y sus necesidades concretas y particulares. La bajada de la política por programas debe privilegiar un enfoque que conjugue la diversidad de las ofertas ministeriales con las necesidades y posibilidades del establecimiento, sus directivos, profesores y alumnos, para un mejor empalme entre ambos. Se plantea así nuevamente un rol de primera importancia para los responsables de implementar la política.

6.2 Evaluar políticas y no sólo programas

En nuestros países existe una tendencia de confundir políticas con programas y de evaluar los programas más que la política. Los avances en el tema de evaluación se han expresado en la evaluación de medidas programáticas y rara vez en la evaluación de la política o del conjunto de programas y medidas que son parte de una política. Evaluar la política, entre otras cosas, implica preguntar por la coherencia de los instrumentos programáticos y otros, tanto a nivel de su concepción como en cuanto a su implementación (instrumentos, tiempos, sinergias, competencia). En esta perspectiva, las preguntas de evaluación refieren no a programas individuales sino a cómo el conjunto de los programas “aterrizan” y “empalman” con la situación y dinámica de la unidad educativa. Evaluar políticas así mismo implica hacer meta-análisis de las evaluaciones de programas específicos poniendo atención en sus fortalezas y debilidades, amenazas y oportunidades comunes, y las que son específicas a programas particulares.

6.3 Fortalecer el vínculo entre evaluación y sus resultados y el policy-making

Uno de los principales argumentos a favor de las evaluaciones de procesos es la utilidad práctica que éstas tienen para los *policy-makers* y tomadores de decisiones, pues les permite comprender en detalle dónde están los problemas de implementación, qué situaciones no fueron anticipadas en el diseño y, en general, hacer sentido sobre lo que está pasando con las políticas educativas a un nivel micro.

Sin embargo, la relación entre evaluación y toma de decisiones ha sido siempre un territorio incierto. Veíamos en un comienzo que José Sulbrandt observaba una completa desconexión entre los resultados de las evaluaciones con los procesos de gestión y toma de decisiones de los programas y políticas educacionales. ¿Cuánto se ha avanzado en este sentido? Hoy en día no existe en Chile un estudio que se haya propuesto observar qué función cumple o qué influencia tiene la evaluación educacional en las decisiones de gobierno y del Ministerio de Educación. Aún así, es razonable pensar que ya no cabe hablar de una “completa desconexión”, y en el proceso de estrechar ese vínculo los propios investigadores han comenzado a reconocer que no corresponde simplemente sentarse a esperar a que los escuchen sino más bien tomar ellos mismos ciertas iniciativas.

El principal desafío de los estudios de evaluación, es sin dudas, el de la rigurosidad metodológica y la capacidad de ofrecer información y conclusiones válidas. No obstante, el “evaluador” a nuestro entender debiera asumir también otras responsabilidades. Entre ellas, la más importante es realizar todos los esfuerzos posibles para que los hallazgos de los estudios efectivamente permeen las estructuras existentes de toma de decisiones entre los gestores. Cuatro elementos son importantes: a) dedicar tiempo a comprender en detalle la demanda del que contrata la evaluación, el por qué y para qué de la demanda, qué le preocupa respecto al programa o la política, y cuáles son las hipótesis que maneja sobre sus resultados y procesos; b) asimilar el marco de referencia y lenguaje de los responsables del programa / política, y plantear los hallazgos y recomendaciones en ese marco; c) asegurar un trabajo que implique contactos periódicos con la contraparte en los cuales se ocurran acciones de retroalimentación, esto es, se plantean hipótesis sobre resultados preliminares, fortalezas y debilidades, dificultades en la evaluación, etc., para asegurar que la información producida sea adecuadamente procesada e internalizada por los mismos que implementan los programas; d) preparar e insistir en la importancia, una vez terminada la evaluación, de difundir los resultados de modo de que otros actores sociales legítimamente interesados en los programas evaluados (incluyendo a los beneficiarios y destinatarios de los mismos) conozcan las evaluaciones y puedan participar y aportar al análisis crítico de las políticas.

REFERENCIAS BIBLIOGRÁFICAS

- Asesorías para el Desarrollo y Santiago Consultores (2000). *Evaluación del Programa de Mejoramiento de la Calidad de las Escuelas Básicas de Sectores Pobres, P-900*. Santiago: MINEDUC.
- Bellei, C., L. M. Pérez, D. Raczynski y G. Muñoz (2004). *¿Quién dijo que no se puede? Escuelas Efectivas en Sectores de Pobreza*. Santiago: UNICEF.
- Cox, C. (editor) (2003). *Políticas educacionales en el cambio de siglo. La reforma del sistema escolar de Chile*. Santiago: Editorial Universitaria.
- Elmore, R. (2004). *School Reform from the inside out*. Cambridge: Harvard Education Press.
- Fullan, M. (2005). *Leadership & Sustainability – System Thinkers in Action*. California: Corwin Press, Thousand Oaks.
- Fullan, M. (2001). *Leading in a culture of change*. Jossey-Bass, San Francisco, California, USA
- García-Huidobro, J.E. (editor) (1999). *La Reforma Educacional Chilena*. Madrid: Editorial Popular.
- PNUD-Asesorías para el Desarrollo (2004). *Sistematización y evaluación del Plan de asistencia técnica a escuelas críticas de la Región Metropolitana*. Informe Final para el Ministerio de Educación.
- PNUD-Asesorías para el Desarrollo (2008). Monitoreo, seguimiento y evaluación de la Estrategia de Apoyo a Establecimientos Prioritarios. estudio en curso.
- Raczynski, D. y G. Muñoz (2005). *Supervisión educacional en Chile. Experiencias públicas y privadas*. Santiago: Ministerio de Educación.
- Raczynski, D. y G. Muñoz (2005). *Efectividad Escolar y Cambio Educativo en Condiciones de Pobreza en Chile*. Santiago: Ministerio de Educación.
- Raczynski, D. y G. Muñoz (2007). Chilean educational reform: the intricate balance between a macro and micro policy. W.T. Pink and G. W. Noblit (Eds.). *International Handbook of Urban Education*. Springer: 2007.
- Sen, A. (1999). *Development as Freedom*. New York: Anchor Books.
- Sulbrandt, J. (1994). La evaluación de los programas sociales: una perspectiva crítica de los modelos usuales. En: Kliksberg, B. *Pobreza, un tema impostergable*. México: Fondo de Cultura Económica.

ANEXO

Propuesta de operacionalización de las variables que contribuyen a la efectividad escolar

La tabla a continuación lista las variables, define indicadores y entrega una definición operacional de las mismas. La definición operacional se traduce en una escala de puntuación que fluctúa entre 1 y 5, donde 1 representa el valor más bajo, la ausencia total de señales que apuntan a la presencia de la variable y 5 una presencia decidida y fuerte de la variable en cada uno de sus indicadores. Las fuentes de información son variadas, entrevistas, documentos (libros de clase, actas, proyecto educativo, reglamentos, etc.), encuestas, registros estadísticos, observación directa.

Dimensiones y variables de efectividad escolar		
Variables	Indicadores	Operacionalización y puntuación
Dimensión A. NORMALIZACIÓN		
Normalización	Existencia y cumplimiento de normas básicas de funcionamiento en la escuela Existencia de disciplina, convivencia, respeto entre las personas Puntualidad y aprovechamiento adecuado del tiempo Asistencia alta y pocos permisos del personal Asistencia regular y a la hora de los alumnos a clases Aseo, mantención e higiene de las instalaciones de la escuela Presentación personal de directivos, docentes, paradocentes y alumnos Claridad en roles y funciones de los equipos de trabajo y las personas	La normalización es deficiente cuando no hay normas y ni reglas que la rigen y/o cuando los problemas de asistencia, disciplina, puntualidad, aseo etc. son generalizados (dejan de ser puntuales) Valor 5. Hay normalización en todas los indicadores Valor 4. Hay normalización en la mayoría de los indicadores Valor 3. Hay normalización en la mitad de los indicadores Valor 2. Hay normalización en unos pocos indicadores Valor 1. Ninguno de los indicadores muestra normalización
Dimensión B. COLUMNA VERTEBRAL DE LA EFECTIVIDAD		
1. Gestión pedagógica	Proyecto educativo real compartido por la comunidad escolar cuyo eje es la enseñanza-aprendizaje de todos los niños y que efectivamente orienta la gestión escolar Gestión pedagógica fuerte que (i) entrega soporte al trabajo pedagógico en el aula y el trabajo colaborativo entre docentes; (ii) trabaja con metas realista evaluables y que se evalúan regularmente Observación de aula por parte del equipo directivo y presencia de actividades de perfeccionamiento en servicios (en el establecimiento) Planificación de las clases, articulación entre niveles y asignaturas Equipo directivo conoce fortalezas y debilidades de los profesores y les asigna cursos, asignaturas y tareas aprovechando sus fortalezas Presencia de directrices institucionales para diagnosticar y apoyar a los alumnos mas lentos más allá del tradicional "reforzamiento" Tareas exigentes y desafiantes para los alumnos (no se baja exigencia para evitar repitencia) Motivar a los alumnos está por sobre la disciplina como preocupación principal de los docentes en la	La gestión pedagógica se aleja de la que es propia de las escuelas efectivas cuando no hay liderazgo pedagógico directivo, el proyecto educativo es formal, no compartido, con objetivos vagos (no operacionalizados) o no es considerado al momento de tomar decisiones; cuando el equipo directivo no da soporte y no evalúa el trabajo en aula y no fomenta la reflexión pedagógica en la escuela; cuando hay ausencia de planificación y articulación entre niveles y asignaturas; cuando el horario y tareas de cada docentes se definen solo con criterio de disponibilidad horaria; cuando no se realizan esfuerzos por diagnosticar a los alumnos y atender a los más lentos o atrasados; cuando se bajan las exigencias a los alumnos; y problemas de no disciplina en la sala de clase son la principal preocupación del profesor Valor 5. El colegio muestra efectividad en la mayoría de los indicadores Valor 4. El colegio muestra efectividad en la

	sala de clase. Escuela maneja información y hace diagnósticos de sus alumnos y del entorno Escuela busca solución oportuna a problemas del entorno que la afectan	mitad de los indicadores Valor 3. El colegio muestra efectividad en menos de la mitad de los indicadores Valor 2. El colegio ha iniciado preocupación por mejorar resultados de aprendizaje, con acciones concretas asociadas, pero la presencia de estas es muy incipiente. Valor 1. Ninguna de los indicadores apunta a efectividad y no hay una preocupación en torno a ella.
2. Expectativas altas de directivos y profesores para la escuela y el aprendizaje de los alumnos	Docentes y directivos creen que pueden mejorar el colegio y que ellos son los principales responsables. Docentes creen en el potencial de aprendizaje de los alumnos, los incitan a superarse, su preocupación es motivarlos para que así sea Directivos y líderes pedagógicos tienen expectativas respecto al potencial de sus docentes y los apoyan para actualizar este potencial. Directivos y docentes en ningún momento culpan a la familia del alumno y al entorno por las déficit de aprendizaje de los alumnos	Valor 5. Presencia segura de todos los indicadores (situaciones) Valor 4. Presencia de cuatro de los situaciones Valor 3. Presencia de dos o tres de la situaciones Valor 2. Presencia incipiente de algunas de las situaciones Valor 1. No hay evidencia de ninguna de las situaciones. Directivos y docentes no creen en el potencial de aprendizaje los alumnos y culpabilizan a su familia y el entorno de los problemas de aprendizaje.
Dimensión C. FACTORES FACILITADORES / OBSTACULIZADORES		
1. Capital simbólico	Existe un buen clima interno reconocido públicamente Ausencia de conflictos entre directivos y docentes Ausencia de conflictos entre los docentes Ausencia de conflictos de los profesores con los alumnos y los apoderados . El buen clima y actividades sociales no inhibe la crítica del trabajo, el ser exigente y riguroso, reconocer las debilidades propias y las de otros. Existe una imagen/identidad positiva de la escuela	Valor 5: Presencia de todos las situaciones (no pueden faltar las dos últimas) Valor 4. Presencia de solo las dos últimas situaciones Valor 3. Presencia solo de la penúltima situación (el buen clima, la actividad social y sociabilidad no interfieren con un trabajo riguroso y exigente) Valor 2. Conflictos notorios (generalizados) entre docentes y de éstos con los apoderados Valor 1. Conflictos generalizados que marcan una escisión entre los directivos y los docentes
2. Relación Familia-escuela	La escuela tiene una postura explícita de vinculación con los padres que va más allá de lo tradicional (CEPA, aporte en recursos, asistir a reuniones) La escuela esta abierta a la comunidad (presta infraestructura y/o equipamiento) La escuela entrega servicios a los padres y la comunidad (talleres, capacitaciones) La escuela incorpora a los padres a la labor escolar (padres colaboran con los profesores en tareas educativas) La escuela firma cartas de compromiso con los padres y se preocupa de su cumplimiento	Valor 5. Presencia de al menos 4 de las situaciones, no pudiendo faltar la primera ni la cuarta Valor 4: Presencia de al menos 3 de las situaciones, no pudiendo faltar la primera ni la cuarta Valor 3. Presencia sólo de la primera y cuarta Valor 2. Presencia solo de la quinta situación Valor 1. Solo existe el CEPA y hace lo tradicional. Escuela no tiene una postura explícita frente a su relación con los padres.



RESPONSABILIZAÇÃO EDUCACIONAL NO BRASIL

Nigel Brooke

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art7port.pdf>



1. INTRODUÇÃO

O crescente uso de sistemas externos de avaliação para medir o desempenho dos alunos, somado à confiabilidade cada vez maior dos instrumentos utilizados, facilitou o desenvolvimento de políticas de responsabilização educacional que tornam as equipes escolares responsáveis pelos resultados de suas escolas (Linn, 2000; West e Peterson, 2003). Em alguns países, notadamente nos EUA, esses resultados podem determinar consequências administrativas tanto positivas como negativas, logo aumentando o impacto das avaliações (Hess, 2003). Essas políticas não encontram amplo apoio entre os professores por uma série de razões. Dentre as quais se destacam os riscos educacionais reais associados a currículos orientados por testes, a aparente ameaça à autonomia dos professores e dúvidas quanto à validade das conclusões a respeito do desempenho das escolas, conclusões estas provenientes de sistemas de mensuração de larga escala que são pouco sensíveis às características específicas das escolas ou distritos escolares. De todo modo, há evidências sugerindo que medidas de responsabilização podem promover ganhos no desempenho dos alunos (Carnoy, Loeb e Smith, 2001; Jacob, 2003; Carnoy e Loeb, 2004).

Mediante o estudo do caso brasileiro, o propósito desse artigo é oferecer uma resposta preliminar à questão de se políticas similares de responsabilização estão seguindo a expansão da avaliação educacional na América Latina. Ao longo das duas últimas décadas a maioria dos países latino-americanos adotou alguma forma de avaliação dos seus sistemas nacionais de educação, além de ter participado de avaliações internacionais, como TIMSS, PISA e PERLS (Ferrer, 2006). O Brasil é um caso especialmente interessante, existindo avaliações nacionais em todos os níveis educacionais, incluindo o ensino superior.

Duas questões mais específicas são se o processo de implementação dessa política está sujeito à mesma resistência por parte dos docentes e funcionários administrativos, e se há alguma evidência em relação aos efeitos benéficos da política de responsabilização sobre o desempenho dos alunos no contexto de um país em desenvolvimento. Define-se aqui política de responsabilização como uma tentativa de melhorar os resultados das escolas mediante a criação de consequências para a escola ou para professores individuais, sejam elas materiais ou simbólicas, de acordo com o desempenho dos alunos medido por procedimentos avaliativos estaduais ou municipais.

Sistemas estaduais de avaliação educacional existem em treze dos vinte e seis estados brasileiros, e o mais velho desses, em Minas Gerais, remonta a 1991. Também, há um sistema de avaliação nacional por amostragem conhecido como SAEB¹, que em seu rigor e sofisticação técnica está entre os mais completos de seu tipo. Originalmente lançado na década de 1980, o SAEB coleta dados de desempenho acadêmico e dados socioeconômicos de alunos de três diferentes séries escolares, bem como dados das escolas através de questionários de professores e diretores, sendo uma amostra nacional de escolas públicas e privadas. Além disso, o governo federal foi pioneiro em lançar uma prova nacional não-compulsória para alunos concluintes do ensino médio. Hoje ela é feita por quase três milhões de estudantes em seu último ano de escolaridade obrigatória. O governo também lançou um sistema sem paralelo de testes nacionais para a avaliação dos cursos de ensino superior, realizados com alunos no primeiro e no último ano dos cursos de graduação. Nesses dois últimos casos, o interesse do governo era criar mecanismos para o monitoramento externo da qualidade da educação, e, através da publicação dos resultados, estimular a demanda por melhorias (Castro, 1999).

Há sofisticação nos métodos de construção dos testes, nas técnicas para garantir a comparabilidade ao longo do tempo, na logística de se aplicar provas simultaneamente de norte a sul e na produção de

¹ O Sistema Nacional de Avaliação da Educação Básica, modelado com base no NAEP, assumiu sua forma atual em 1995.

relatórios ao nível das escolas e dos indivíduos (Ferrer, 2006). Em 2005 a primeira avaliação nacional censitária de Português e Matemática no ensino fundamental, chamada *Prova Brasil*, foi feita na 4ª e na 8ª série de todas as escolas urbanas, tanto públicas como particulares. Em suma, o Brasil acumulou domínio técnico considerável ao longo de duas décadas de experiências em avaliações educacionais de larga escala, e está entre os líderes nesse campo na América Latina.

O que as avaliações da educação básica (séries 1-12) no Brasil mostram, é que os resultados escolares médios são extremamente fracos. Como informado pelo Ministério da Educação, os resultados do SAEB de 2001 indicaram que 59% dos alunos da 4ª série conseguia ler apenas as frases mais simples ou não conseguia ler nada, e que apenas 4,8% nessa série obteve desempenho considerado adequado para a idade (MEC, 2003). Os resultados das provas do SAEB aplicadas em 2005 mostram uma melhora fracionária em relação aos anos anteriores no desempenho dos alunos de 4ª série, mas os resultados da 8ª série e da 3ª série do Ensino Médio continuaram a cair tanto em Português quanto em Matemática, e agora estão entre 3,8% e 11,2% menores do que em 1995 (MEC, 2007).

A impressão que o governo tem de que a qualidade da educação básica está falhando em alcançar até mesmo os padrões mais baixos é corroborada pelo desempenho medíocre do Brasil em avaliações internacionais de larga escala que envolvem países de outras regiões. O Brasil participa dessas provas desde o segundo *International Assessment of Educational Progress* (Avaliação Internacional do Progresso da Educação) que foi aplicado em vinte países em 1990-91 para estudar o desempenho em Matemática e Ciências de alunos de nove e treze anos (Lapointe, Askew e Mead, 1992a, 1992b). No caso do Brasil esse estudo se restringiu a alunos de treze anos nas cidades de São Paulo e Fortaleza, e, a despeito das dúvidas quanto à representatividade da amostra, mostrou que o país estava apenas marginalmente à frente de Moçambique, um país predominantemente rural com uma taxa de analfabetismo adulto de 80% na época.

O Brasil também esteve entre os trinta e dois países que participaram do PISA em 2000, obtendo a nota média mais baixa na escala combinada de competência em leitura e na escala de conhecimento matemático, ficando atrás do outro participante latino-americano, o México. Para corrigir as diferenças entre países no que tange ao número médio de séries escolares que os alunos da amostra tinham frequentado – o resultado da matrícula mais tardia e da taxa de repetência maior nos países mais pobres – os responsáveis pelo relatório do Brasil desenvolveram comparações entre estudantes com o mesmo número de séries escolares (INEP, 2001). Mesmo com esses controles, a desvantagem do Brasil era clara. Por exemplo, a proporção de estudantes brasileiros com nove ou mais séries escolares que obteve nível três ou maior na escala de competência em leitura foi de apenas 25%, comparado a 76% para a Coréia do Sul, 59% para a Espanha e 30% para o México. Para um observador "o que o PISA nos mostra é que de todas as mil coisas que uma escola [brasileira] faz ou tenta fazer ela se esquece do mais importante, que é ensinar o domínio da língua" (Castro, 2001). Os resultados do PISA 2003 não foram mais confortantes, tendo entre 22,5% e 28,2% dos estudantes brasileiros atingido nível de proficiência 2 ou maior nas quatro escalas de Matemática, em comparação a de 75,2% a 79,3% para os países da OECD (OECD, 2004).

Dentro da estrutura legal da federação brasileira a responsabilidade dessa situação cabe aos governos estaduais e municipais. A atual divisão de atribuições, definida na Constituição de 1988 e ratificada pelo Lei de Diretrizes e Bases da Educação de 1996, dá maior responsabilidade pelo ensino público pré-escolar e fundamental para os governos municipais e pelo ensino médio para os governos estaduais, e enfatiza a responsabilidade dos governos federais pelo ensino público superior. A tabela 1 mostra a atual distribuição da matrícula dos estudantes entre as diferentes esferas do governo de acordo com suas responsabilidades definidas na Constituição, ao mesmo tempo demonstrando a importância do ensino privado, especialmente no ensino superior. Enquanto a maioria dos estudantes de pré-escola e ensino fundamental está matriculada em escolas municipais, mais de doze milhões deles ainda estão em escolas

estaduais. Ao mesmo tempo, há aproximadamente 186.000 estudantes matriculados em escolas secundárias municipais ao invés de estaduais (MEC, 2006).

TABELA 1. MATRÍCULAS POR NÍVEL DE EDUCAÇÃO E TIPO DE ESTABELECIMENTO. BRASIL. 2006

Nível	Federal	Estadual	Municipal	Particular	Total
Pré-escolar	2.471	242.979	4.820.236	1.950.409	7.016.095
Elementary	25.031	11.825.112	17.964.543	3.467.977	33.282.663
Secondary	67.650	7.584.391	186.045	1.068.734	8.906.820
Adult	1.203	3.226.780	2.180.391	207.917	5.616.291
Higher* (undergraduate)	579.587	477.349	135.253	1.753.184	4.453.156
Total	675.942	23.356.611	25.286.468	8.448.221	59.275.025

Data for 2005

Sources: MEC/INEP, Censo Escolar – 2006; MEC/INEP, Sinópsse da Educação Superior – 2005.

Dada a situação geral de resultados fracos no ensino fundamental e médio junto com um grau relativamente alto de experiência na área de avaliação educacional, seria razoável esperar a implementação de políticas de responsabilização escolar baseadas em medidas de desempenho de alunos naqueles níveis de governo em que existe responsabilidade pelo fornecimento da educação básica. No Brasil isso significaria o crescimento de políticas de responsabilização dentro dos sistemas estaduais e municipais de educação, em vez de ser um componente da política educacional do governo federal, e uma diversidade de políticas para refletir as diferentes condições das secretarias locais de educação.

Essa expectativa deriva da história recente da responsabilização no Reino Unido e nos EUA. Nesses países a crença de que os padrões educacionais estavam sob ameaça trouxe a educação para o topo da agenda política e levou os governos a demandarem maior responsabilidade de suas escolas (Jones, 1989; West e Peterson, 2003; Gillard, 2004). Essa expectativa também reconhece a pronta transmissão de idéias e políticas educacionais através das fronteiras nacionais como resultado da influência das agências internacionais e da crescente circulação de informações sobre políticas públicas.

No entanto, o que nós observamos é que há poucos exemplos de políticas formais de responsabilização no Brasil, e das que apareceram, algumas já foram abandonadas e todas as outras passaram por grandes modificações desde suas primeiras aparições.

Não há nenhuma coleta centralizada de dados sobre as políticas educacionais dos estados ou municípios no Brasil. Não se vê isto como papel do Ministério ou do Conselho de Secretarios Estaduais de Educação (o equivalente do *Council of Chief State School Officers*) e nem da União de Dirigentes Municipais de Educação. Esse é um exemplo da considerável autonomia dos governos estaduais e municipais, fundada em princípios federativos relacionados à não-interferência em assuntos pertinentes a outras esferas do governo, o que até recentemente dificultava a coordenação da política pública educacional (Brooke, 1992).

A informação disponível indica que políticas de responsabilização já existiram ou ainda estão em uso nos sistemas estaduais de Rio de Janeiro, São Paulo, Ceará e Paraná, e também nos sistemas municipais de Sobral e Belo Horizonte. Entre elas, as do Rio de Janeiro, São Paulo, Ceará e Sobral podem ser consideradas sistemas *high stakes* (com muito em jogo, oferecendo bônus salariais ou prêmios monetários, enquanto que os casos do Paraná e do município de Belo Horizonte são sistemas *low stakes* (com pouco em jogo), de boletins, envolvendo a publicação e disseminação de indicadores de desempenho, sem conseqüências materiais explícitas para a equipe da escola.

O estudo se propõe a salientar os detalhes proeminentes de cada sistema de responsabilização para determinar se a responsabilização encontrou resistência e, em caso positivo, explicar o que impediu tanto

a implementação como a subsequente manutenção das políticas de responsabilização no Brasil. Qualquer evidência do possível impacto das políticas existentes também será apresentada.

2. MÉTODOS

Os dados dos casos incluídos nesse estudo foram retirados de três fontes. Em primeiro lugar, de correspondência de caráter pessoal com os responsáveis pela gestão das políticas estaduais e municipais; em segundo lugar, de informações apresentadas em seminários ou disponíveis na internet; e, por fim, das raras publicações produzidas seja para as escolas participantes de programas de responsabilização ou por funcionários governamentais como registros de suas atividades. Os dados coletados através de correspondências foram obtidos por uma espécie de amostragem tipo bola de neve, na qual me baseei em referências de contatos iniciais para gerar novos informantes. Ao resumir e analisar essa informação, faço uso de minha própria experiência e observações como membro da Faculdade de Educação da Universidade Federal de Minas Gerais e, por nove anos, como responsável pelo Programa de Educação da Fundação Ford no Rio de Janeiro. Não obstante a busca por informação secundária ter sido razoavelmente rigorosa, essa investigação não pretende ser exaustiva. A breve história de cada programa de responsabilização, baseada em suas características mais proeminentes, é qualitativa e projetada para ser muito mais uma descrição preliminar do que um resumo estruturado dos elementos comuns.

3. PROGRAMAS DE RESPONSABILIZAÇÃO *HIGH-STAKES*

3.1. Estado do Rio de Janeiro

O programa Nova Escola é a política de responsabilização de maior duração no Brasil. Foi criada em janeiro de 2000, juntamente com o Sistema de Avaliação Permanente das Escolas Públicas do Estado do Rio de Janeiro, para atingir tanto os professores quanto os administradores escolares. A longa lista de mudanças desde a criação do programa dá uma idéia da dificuldade em se estabelecer um consenso em relação ao propósito e os métodos para comparar o desempenho das escolas.

O programa foi projetado conforme um modelo longitudinal para medir os ganhos de um determinado coorte, começando com os alunos da terceira série do ensino fundamental, de modo a conseguir uma medida mais precisa das contribuições da escola para o aprendizado dos alunos. No entanto, a ausência de qualquer avaliação externa no terceiro ano do programa, 2002, devido a uma mudança de governo, ocasionou uma mudança de planos. Novos instrumentos, alinhados com a escala nacional do SAEB (para permitir comparações com outros estados), foram aplicados a 180.000 alunos da 4ª e da 8ª séries do ensino fundamental e da terceira série do ensino médio no final de 2003. Medidas de gestão e eficiência (baseadas em estatísticas de fluxo de alunos) das escolas foram incluídas, e foram criados cinco grupos de comparação contendo escolas com alunos em situações socioeconômicas similares.

Em 2004, o governo do estado entregou o contrato para o desenvolvimento e gestão do Nova Escola para uma organização de testes educacionais diferentes, que reformulou os critérios para se concentrar na administração da escola, no desempenho dos alunos e no fluxo de alunos. Os grupos de comparação também foram eliminados, mas se manteve o modelo SAEB de testar a 4ª, 8ª e 11ª séries, bem como a concessão de dois tipos de premiação, um por desempenho absoluto e outro por nível de progresso.

Os prêmios são dados às escolas baseado na acumulação de pontos. O maior número de pontos que uma escola pode acumular é 25, dos quais 10 se devem ao fluxo de alunos (90% ou mais dos alunos ainda matriculado no final do ano escolar), 5 à gestão da escola (com critérios relacionados a transparência,

relações com a comunidade, assiduidade da equipe e gestão de matrículas), e 10 (ou 40%) ao desempenho dos alunos no sistema estadual de avaliação. No caso do desempenho dos alunos, a escola alcança o número máximo de pontos se 80% de seus alunos nas séries testadas atingir o nível de desempenho considerado satisfatório na escala do SAEB.

O prêmio dado à equipe da escola tem variado de acordo com o desempenho. Em 2004, o nível máximo de uma escala de cinco níveis equivalia a R\$500,00 para professores de tempo integral e R\$250,00 para pessoal auxiliar. O dinheiro da premiação, dois terços do qual foi distribuído pela nota geral do desempenho da escola e um terço pelo “progresso” da nota da escola entre os anos escolares, foi dividido entre um total de 58.008 professores e 19.443 funcionários auxiliares.

Um aspecto importante do programa é que ele é o único a ter sido avaliado por meio de uma comparação entre os resultados das escolas secundárias do Rio de Janeiro e aqueles de dois outros estados no sudeste brasileiro de 2000 a 2005 (Rodrigues, 2007). Os resultados em questão são a média das notas do SAEB para os três estados (coletadas a cada dois anos entre 1995 e 2005) e as médias da relação aprovação/reprovação e da taxa de evasão, coletadas todos os anos como parte do conjunto de dados do censo escolar. Apesar da falta de controles para outras possíveis diferenças de políticas educacionais ou para mudanças na situação econômica relativa dos três estados, o estudo mostra ganhos maiores no desempenho em Português no Rio de Janeiro em relação aos outros estados, invertendo uma posição de inferioridade vigente durante o período anterior à criação do programa Nova Escola (Rodrigues, 2007). Apesar de apresentar pioras ao invés de melhorias, as estatísticas de fluxo também mostram que a reprovação média dos alunos no final do ano cresce mais devagar no Rio de Janeiro do que nos outros estados. No entanto, outros resultados tornam essas descobertas menos conclusivas. Os resultados do desempenho em Matemática mostram tendência inversa à de Português, e o Rio de Janeiro perde terreno comparativamente aos outros estados. Como a Matemática é uma matéria mais “escolar” do que é o Português (Brandsmo e Knuver, 1989), o resultado esperado era o oposto. Também se mostra que o nível de evasão aumentou em comparação com os outros estados, levantando a possibilidade de as taxas de reprovação, menores do que o esperado, serem fruto de uma auto-seleção por parte dos alunos. Apesar do grande esforço em sujeitar o programa a uma avaliação objetiva, a ambigüidade dos resultados indica que outros estudos serão necessários.

Não obstante as mudanças recentes no programa, incluindo a redução do peso do critério de gestão e a implementação de critérios mais específicos em relação ao fluxo de alunos, a resistência ao programa é ainda alta, abundando críticas técnicas, políticas e doutrinárias. Uma dessas críticas diz que o programa reflete uma preocupação neoliberal com as necessidades dos empregadores: “O programa Nova Escola é um empobrecimento da educação. Ele é mais do que isso, é uma tentativa de satisfazer as demandas de um mercado de trabalho produtivo” (Gama, 2004). Até 2006 a continuidade do programa foi possível devido ao suporte dado na assembleia estadual por representantes dos setores populares, evangélicos e de baixa renda, que não tinham relação com os sindicatos de professores ou outros grupos sociais organizados. O governador recém-eleito ainda tem que declarar seu apoio ao programa, e há incertezas quanto à sua sobrevivência no novo ambiente político. Em agosto de 2005 o sindicato dos professores foi responsável por cinco ações legais movidas contra o governador da época, três das quais relacionadas ao programa Nova Escola e hostis a ele, principalmente devido à interferência do mesmo na carreira dos professores concursados e em suas escalas salariais. Se o novo governador enxergar a necessidade de obter apoio político dos representantes dos professores na assembleia estadual, o programa dificilmente sobreviverá em sua forma atual.

3.2. Estado de São Paulo

Entre os anos de 2001 e 2003 diretores e professores escolares no estado de São Paulo receberam um complemento único aos salários anunciado no fim do ano escolar, um pouco antes do Natal. Esse bônus

foi calculado de acordo com a assiduidade média da equipe, o desempenho global da escola medido durante o ano escolar anterior através do sistema estadual de avaliação, o SARESP², e outros indicadores de qualidade, principalmente o nível de retenção de alunos pelas escolas. Essa política de responsabilização, envolvendo cálculos para determinar pagamentos individuais, foi uma expressão da opinião declarada da Secretaria Estadual sobre a importância de tornar público o resultado das avaliações anuais:

“Os resultados do desempenho das escolas deverão ser amplamente divulgados, de forma que tanto a equipe escolar como a comunidade usuária seja capaz de identificar a posição da sua escola no conjunto das escolas de sua Delegacia, de seu bairro e de seu município. Isso possibilitará à escola a busca de forma diversificadas de atuação, com o objetivo de implementar a melhoria dos resultados escolares. Por outro lado, permitirá também à população acesso às informações, de modo que possa fiscalizar, participar e cobrar a qualidade do serviço que lhe deve ser prestado.” (São Paulo, 1995, p. 310).

Em 2002 o governo do estado pagou um total de R\$370 milhões³ para 170.974 professores e 18.744 diretores e supervisores, um aumento de 58% em relação ao ano anterior. O bônus mais alto para professores foi de R\$4.500 e para diretores de R\$5.500, e o mais baixo de R\$1.000 e R\$1.500, respectivamente, de acordo com qual dos seis níveis de desempenho era atingido pela escola. Em 2003, 42.000 funcionários escolares auxiliares também foram incluídos no programa, e receberam um bônus fixo de R\$500 que não dependia de critérios de desempenho. Em 2004, o montante gasto subiu para R\$514 milhões, pagos a 194.348 professores que receberam entre R\$1.220 e R\$6.000, a diretores e outros profissionais escolares que receberam entre R\$1.200 e R\$7.000, e à equipe de auxiliares que recebeu a mesma quantia estabelecida no ano anterior. No entanto, nesse ano os critérios foram modificados para, em primeiro lugar, incluir um mínimo de 200 dias de trabalho efetivo e o desenvolvimento de projetos envolvendo a comunidade da escola, e, em segundo lugar, para excluir o uso dos resultados de desempenho do SARESP. Em 2006, o SARESP foi suspenso e os resultados do desempenho dos alunos novamente retirados do cálculo do bônus. Isso se deveu, em grande medida, à aplicação no ano anterior do novo *Prova Brasil*, a avaliação do governo federal dos alunos de 4ª e 8ª série de todas as escolas urbanas de ensino fundamental (APEOESP, 2006). A decisão simultânea de realizar a prova do SARESP de dois em dois anos, ao invés de anualmente, torna incerto se, no futuro, se voltará a usar dados do desempenho de alunos no programa de bônus para determinar o valor dos pagamentos suplementares a professores.

A oposição aos bônus anuais tem sido forte. A mais vociferante, do Sindicato dos Professores do Ensino Oficial do Estado de São Paulo, expressou uma rejeição à “transferência abusiva da lógica de mercado” para a educação (Cavaliere, 2003, p.30) e à competição entre as escolas e os diretores das escolas, que seria fomentada por essa política (CDEP, 2006). Outras críticas tiveram mais a ver com problemas de trapanças, havendo meios de os professores ajudarem os alunos a preencher as folhas de respostas ou alterarem as respostas dos alunos, uma vez que são os professores que aplicam e corrigem as provas. Esses casos são “bem conhecidos” entre professores de escolas estaduais (Dissidência, 2006), bem como o são os problemas burocráticos associados ao cálculo e pagamento individual dos professores. Outro

² O Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo, criado pela Secretaria de Educação, em 1996, diferencia-se dos outros sistemas estaduais na medida em que seus testes são elaborados, aplicados e corrigidos com a ajuda dos professores. As séries testadas variaram de ano a ano. Em seus primeiros anos, de 1996 a 1998, os alunos da amostra foram testados de acordo com o currículo do ano anterior, usando um modelo longitudinal. Em edições mais recentes, o sistema aplicou testes de Português e Matemática simultaneamente para 90% dos alunos em todas as séries do ensino fundamental e em aproximadamente 80% dos alunos nas séries do ensino médio que voluntariamente fizessem o teste. Em 2004, o SARESP se tornou universal e testou aproximadamente 4,5 milhões de alunos, dos quais 4,1 milhões eram das 5.422 escolas estaduais. O restante dos alunos vinha das menos numerosas escolas municipais e particulares que escolheram participar.

³ O Real (R\$) vale aproximadamente meio dólar (US\$0.48) à atual taxa de câmbio.

argumento, aparentemente bem justificado, diz respeito ao uso de resultados de desempenho médio dos alunos sem qualquer tentativa de se controlar por diferenças socioeconômicas entre as escolas ou pelo impacto de diferenças no nível anterior de aprendizagem.

A despeito das intenções iniciais, o programa de bônus parece ter se tornado mais uma maneira de justificar um benefício natalino do que a construção de uma conexão clara entre níveis de desempenho escolar e pagamentos suplementares às equipes. De fato, a Secretaria foi acusada de falta de transparência com relação ao método de cálculo do bônus (Dissidência, 2006), o que também contradiz a idéia de que era sua intenção transformar o desempenho em um componente claro do pagamento dos professores. Portanto, sem nenhuma declaração nesse sentido ou quaisquer explicações a respeito dos resultados terem sido satisfatórios ou não, o programa de responsabilização que se iniciou em 2001 parece ter terminado. O uso contínuo de taxas mínimas de comparecimento em cursos de treinamento para os professores, tal como incluído entre os critérios de 2005, é um reflexo de problemas perenes de absenteísmo dos professores, e não a expressão de uma política destinada a atribuir responsabilidade por resultados escolares aos professores ou outros membros da equipe da escola (Fórum Estadual em Defesa da Escola Pública, 2003).

3.3. Estado do Ceará

A relação entre o sistema estadual de avaliação da educação, SPAECE⁴, e o Prêmio Escola do Novo Milênio foi legalmente estabelecida em 2001. As cem melhores escolas ganhavam um prêmio com base nos resultados médios de Matemática e Português dos estudantes incluídos na amostra do estado. Para as cinquenta melhores escolas, em 2002, o prêmio valia R\$800 para professores de tempo integral e R\$300 para a equipe administrativa. Para as cinquenta escolas seguintes os prêmios equivaliam à metade do valor apontado acima. Também havia prêmios para os melhores estudantes com base no seu desempenho nos testes. O propósito explícito desses prêmios era promover o reconhecimento das conquistas obtidas pelas escolas públicas e medidas pelo SPAECE, melhorar o ambiente escolar e influenciar seus resultados através de um clima de qualidade, e elevar o padrão da educação pública (Lei No 12.203, 21 de fevereiro, 2002).

Os resultados da avaliação do SPAECE de 2003 dos estudantes da 8ª série do ensino fundamental e da terceira série do ensino médio causaram preocupação considerável na medida em que mostraram uma queda acentuada no desempenho médio. Em Português, os resultados da 8ª série estavam 7,5% mais baixos se comparados a 2002, e, conquanto os resultados de Matemática de 2003 estivessem iguais aos de 2002, os das duas disciplinas estavam 9,4% mais baixos se comparados a 2001. Quanto ao nível do desempenho em Português, a proporção de alunos que a Secretaria considerou estar no nível crítico ou pior chegou a 60,8%, com uma queda correspondente na proporção de alunos que atingiram os níveis intermediários ou adequados. Em Matemática, 97,7% dos alunos foram classificados como estando no nível crítico ou abaixo, e apenas 0,2% chegou ao nível de desempenho considerado adequado para a série (Lima *et al.*, 2005). Apesar de o governo não ter oferecido nenhuma explicação para esses resultados e nem ter discutido os possíveis impactos de seu programa de responsabilização, é evidente que o Prêmio Escola do Novo Milênio não conseguiu combater a tendência geral de queda no desempenho dos alunos. Se isso se deve ao valor dos prêmios ou outros aspectos do programa, ou se o sistema de prêmios foi simplesmente irrelevante, são questões que não ficaram claras, mas o fato de a política de responsabilização ter sobrevivido indica que o governo acreditava que ela ainda tinha um papel a cumprir.

⁴ O Sistema Permanente de Avaliação da Educação no Ceará, criado em 1992, avalia o desempenho em Português e Matemática de uma amostra de alunos da 4ª e 8ª séries do ensino fundamental e da 3ª série do ensino médio. Em 2001, o SPAECE adotou testes computadorizados.

Como parte de uma revisão das políticas vigentes, em 2004 e 2005 o programa Prêmio Escola do Novo Milênio foi revisado através de leis que revogaram procedimentos anteriores, criaram um programa de certificação das escolas e, para as escolas certificadas, criaram o Prêmio Escola Destaque do Ano, com procedimentos novos para sua outorga. Esses procedimentos estipulavam que os prêmios seriam entregues às cinquenta melhores escolas com base em uma variedade de indicadores, incluindo taxas de aprovação e evasão e o desempenho dos alunos em Matemática e Português como avaliado pelo SPAECE. Com esses novos procedimentos um total de 41 prêmios são agora outorgados a escolas com base na melhoria de um ano para o outro, e 9 prêmios com base em níveis absolutos do desempenho das escolas (Ceará, 2006).

A política reformulada tem como objetivo declarado “valorizar a gestão escolar comprometida com o sucesso dos alunos; promover o reconhecimento público das escolas estaduais pela qualidade de ensino; desenvolver uma cultura de acompanhamento e avaliação da escola, contribuindo para a melhoria do ensino e da aprendizagem” (Ceará, 2006, p.3). Três categorias (Fundamental, Fundamental e Médio, e Médio) foram criadas para permitir às escolas competirem com instituições similares quanto ao nível de educação oferecido, e dentro de cada categoria o número de indicadores de qualidade varia de 10 a 22, com os indicadores do desempenho dos alunos no SPAECE representando 60% do total. Outros indicadores incluem taxas médias de aprovação e de desistência para cada série. Cada categoria de escola é apta para concorrer a 3 dos prêmios de desempenho global e a um número de prêmios de melhoramento escolar que varia de 8 a 18, dependendo do número de escolas na categoria.

Com essas mudanças alguns dos problemas anteriores associados à comparação de escolas com diferentes níveis de educação foram amenizados. A outorga de prêmios por melhoramento escolar de um ano para outro também reduz outros riscos associados a abordagens transversais da responsabilização baseadas em uma única medida de desempenho médio, dado que a composição do corpo de estudantes provavelmente não mudará de modo significativo entre uma medida e outra. No entanto, o grande problema de comparar escolas com alunos de diferentes origens socioeconômicas ainda não foi resolvido. Embora esse modelo de responsabilização de “mudança de status” reduza o impacto do nível anterior de aprendizado (Hanushek e Raymond, 2003), ele não oferece controles pelos diferentes ritmos de aprendizagem de diferentes grupos sociais, ou pelas mudanças de um ano para outro nos outros fatores associados à aprendizagem dos alunos.

Não houve nenhuma avaliação externa das políticas de melhoramento escolar do governo do Ceará. Todavia, a permanência da idéia de concessão de prêmios, embora em versão reformulada, sugere que esse tipo de responsabilização, na qual as escolas não são obrigadas a participar, provoca menos resistência do que programas que abrangem o sistema todo e em que todas as escolas precisam ser comparadas, como no programa do estado de São Paulo, mesmo que não haja nenhuma consequência negativa para as escolas com as avaliações mais baixas.

3.4. O Município de Sobral (Ceará)

Durante o último ano da gestão de 1997-2000 do governo municipal, o nível de alfabetização das crianças na escola em Sobral foi medido duas vezes (INEP, 2005). O primeiro teste, realizado por um especialista da Universidade Federal do Ceará, mostrou que metade das crianças no final da segunda série ainda não conseguia ler. Na segunda avaliação, feita em todos os alunos da 1ª à 4ª séries pela Fundação Ayrton Senna, Sobral apareceu em penúltimo entre os mais de 400 municípios assessorados por essa Fundação em seu programa Acelera Brasil. Com base nesses resultados, o governo que entrou em seguida estabeleceu como seu principal objetivo o melhoramento do ensino da leitura e da escrita, e criou novas políticas baseadas na reorganização do sistema de apoio pedagógico da secretaria de educação e na maior autonomia das escolas. A primeira envolveu a definição de metas, o contínuo fornecimento de materiais e treinamento, a contratação de avaliações externas e a disseminação de uma

nova cultura de monitoramento do progresso dos alunos. Com sua nova autonomia, as escolas poderiam selecionar os melhores professores para as aulas de alfabetização, lidar individualmente com os problemas de cada aluno e constantemente reavaliar os resultados e mobilizar as famílias dos alunos para a tarefa de atingir os objetivos da escola.

Um bônus mensal para professores de alfabetização no valor de R\$100 (aproximadamente US\$47) também foi criado, em 2001, como parte da nova política de melhorar os resultados da leitura e da escrita. Os professores tinham que atingir 75% de alfabetização na primeira série básica (alunos de seis anos) ou 90% de alfabetização na série seguinte (alunos de sete anos) ou 100% na 2ª à 4ª séries. A escola com 90% de alfabetização entre as crianças da 1ª à 6ª séries, desde que atingisse 75% na 1ª série básica e 70% na escola como um todo, também podia concorrer ao Prêmio Escola Alfabetizadora, criado em 2001.

Houve diversos problemas com esse sistema de incentivos, tais como a questão do que fazer com professores que atendessem aos critérios em um ano, mas não o fizessem no ano seguinte. Em resposta a isso foi aprovada uma nova lei municipal permitindo reduções de 25% e 50% no nível do bônus. Um segundo problema, criado pela pressão maior para mostrar notas altas, foi que as escolas começaram a impedir que alunos fizessem os testes, dizendo que eles haviam faltado ou que haviam sido transferidos (INEP, 2005). Isso foi atacado via a criação de uma nova lei que requeria que as escolas provassem que os estudantes haviam sido transferidos e demandava um mínimo de 98% de comparecimento no dia do teste externo.

O uso de avaliações externas, desenvolvidas por especialistas da Universidade Federal do Ceará e da Fundação Carlos Chagas, instituição respeitada na área de organização de testes, permitiu que se produzissem bons resultados globais com a política de alfabetização de Sobral. De acordo com uma fonte (INEP, 2005), entre 2001 e 2004 o nível de alfabetização na primeira série básica (alunos de seis anos) aumentou de 33,7% para 89,1%, e na primeira série regular, para alunos de sete anos, de 49,1% para 92,2%.

Como componentes de uma política mais ampla contendo diversos outros investimentos, as contribuições do bônus para professores de alfabetização e do Prêmio Escola Alfabetizadora são de difícil mensuração. Entretanto, como nem um relatório recente sobre a experiência de Sobral (INEP, 2006) nem o sítio na internet da Secretaria Municipal menciona o bônus aos professores, parece provável que, tendo criado amplo e disseminado descontentamento, essa parte da política de responsabilização foi agora abandonada. A continuação do Prêmio Escola Alfabetizadora indica que as autoridades consideram que ele tenha mais efeitos positivos do que negativos e que ele é uma contribuição válida para a política de alfabetização como um todo.

4. PROGRAMAS LOW-STAKES, DE BOLETINS

4.1. Estado do Paraná

Um programa pioneiro de boletins, chamado Boletim da Escola, foi criado em 2001 pela Secretaria Estadual de Educação como parte de uma estratégia de aumentar a responsabilização das escolas e dar poder aos pais para demandar resultados (Ayres, 2003). Um propósito menos divulgado foi o de fortalecer as associações de pais e contrabalancear o poder do sindicato de professores que na época

estava em desavença com o governo do estado⁵. A teoria por trás dessa estratégia parece ser a de que as escolas se sentiriam mais pressionadas em demonstrar bons resultados se os pais estivessem mais bem informados a respeito da sua escola em comparação com as outras.

Distribuído a todos os pais e membros da equipe da escola, o boletim era organizado em um informativo de quatro páginas que provia três blocos de informação. O primeiro bloco, com resultados do programa de avaliação estadual de 2000, continha a média das notas dos alunos de 4ª e 8ª série em Português e Matemática da escola, do município e do estado, classificados de acordo com quatro níveis de desempenho. O significado desses níveis de desempenho era explicado em um suplemento que descrevia as habilidades em Matemática e Português associadas a cada nível. Em 2002 o Boletim da Escola adicionou um indicador que informava se a escola tinha atingido ou não o nível de desempenho “esperado” para ela levando em consideração a condição socioeconômica média dos alunos de cada série. Os diferentes níveis foram estabelecidos para o estado todo com base nos dados socioeconômicos dos alunos retirados de um questionário aplicado a todos na mesma hora da avaliação. Esse resultado foi então incluído no Boletim, e chamado a medida do “valor agregado” pela escola.

O segundo bloco de informações continha dados do censo escolar oficial, incluindo taxas de aprovação, reprovação e evasão, além de comparações municipais e estaduais. O terceiro bloco apresentava afirmações subjetivas em relação à escola, retiradas dos questionários aplicados a alunos, diretores e pais (Ayres, 2003). O questionário não-obrigatório para pais, também introduzido em 2001, representou a primeira tentativa de captar as opiniões e percepções dos pais quanto a elementos-chave do ambiente escolar, incluindo o envolvimento dos mesmos. Nas palavras do então secretário da educação, a intenção era “promover um processo através do qual os pais aumentariam a sua percepção de seu papel de educadores, de cidadãos e de consumidores de serviços educacionais – para seus filhos” (Saliba, 2003). Em 2001, 53.000 pais concordaram em visitar a escola de seus filhos para responder ao questionário, número este que aumentou em 50% quando a pesquisa foi repetida no ano seguinte.

Também se pediu aos pais para darem uma nota global para a escola, mas quando os diretores começaram a temer que essa nota poderia ter algum efeito sobre sua avaliação de desempenho, passaram a influenciar a resposta dos pais. Então, tal nota não mais podia ser considerada uma expressão verdadeira da opinião dos pais (Ayer, 2003). Com a mudança da administração estadual em 2003 e a nomeação de um Secretário da Educação com fortes laços com os sindicatos dos professores, o sistema de boletins foi completamente abandonado. As dificuldades em se informar os pais quanto à interpretação e uso do boletim, a falta de comunicação real entre pais e professores e a falta de confiança geral dos professores e diretores com relação ao propósito da política provavelmente levaram ao seu abandono precoce, apesar de nenhum estudo ter sido feito. A despeito de consideráveis esforços em treinar os pais e em se estabelecer um processo multiplicador no qual os pais já treinados treinariam outros, o fracasso do Boletim da Escola do Paraná mostra que, embora a associação de pais tenha sido um elemento-chave no desenvolvimento da estratégia de responsabilização, ela foi impotente para defender essa estratégia frente à oposição do sindicato de professores.

4.2. O Município de Belo Horizonte (Minas Gerais)

Criado em 2006 e desenvolvido para ser usado por administradores e professores escolares, o Informativo da Escola Municipal de Belo Horizonte é o exemplo mais recente do sistema de responsabilização educacional *low-stakes* sobre o qual se tem informação. Em sua primeira versão, o Informativo de quatro

⁵ O Sindicato dos Trabalhadores da Educação no Paraná liderou greves contra as políticas “neoliberais” do governo em 2000 e 2001, e alguns de seus membros fizeram greve de fome em 1998 e 2000.

páginas foi distribuído apenas às escolas, e não aos pais, e só se deu treinamento prévio aos funcionários de enlace que agem como intermediários entre as escolas e a Secretaria Municipal de Educação.

O primeiro Informativo foi a tentativa inicial da secretaria de dar às escolas uma visão geral de seus próprios resultados bem como uma oportunidade de se comparar com outras escolas de características similares. Isso correspondeu ao propósito explícito do Informativo de prover as escolas com os dados necessários para que elas se auto-avaliassem, e assim promovessem uma revisão dos seus planos e objetivos, estabelecendo-se parâmetros para a comparação ao longo do tempo (SMED, 2006). Incluíram-se como indicadores gerais, de modo a contextualizar as escolas, o tamanho e nível de complexidade delas, a raça/etnia e condição socioeconômica dos estudantes, e também o número de questões administrativas/financeiras pendentes como um indicador de eficiência administrativa. As informações dos professores incluíam o nível e tipo de educação recebida, e, pela primeira vez, o número médio de dias por turma e por ano que foram perdidos pela escola devido a ausências de professores por motivos de saúde. A inclusão desse indicador, que ia de 3,3 dias/turma a surpreendentes 98,5 dias/turma, é mais um indício do problema do absenteísmo de professores do setor público.

Os indicadores pedagógicos incluíam resultados de avaliações externas estaduais (SIMAVE) e nacionais (*Prova Brasil*), mostrando a percentagem de alunos em cada um de quatro níveis de desempenho e fazendo-se comparações somente com escolas que tivessem condições socioeconômicas similares. Outros indicadores mostram os níveis de equidade socioeconômica e racial atingidos pela escola, calculados com base nos dados sobre o desempenho dos alunos para a cidade de Belo Horizonte como um todo. Esses indicadores foram considerados especialmente importantes, dado o objetivo de inclusão social da política educacional "Escola Plural" que a cidade desenvolve e que é reconhecida como a bandeira do governo do Partido dos Trabalhadores em Belo Horizonte. A diferença entre os níveis de desempenho para alunos de diferentes grupos socioeconômicos e de diferentes grupos raciais dentro da escola é comparada às diferenças na população estudantil da cidade como um todo. Informa-se às escolas se elas produziram um nível de igualdade que reflete a média da cidade ou se as diferenças entre seus alunos mostraram uma maior ou menor igualdade do que a da população como um todo. Ao colocar indicadores de equidade no boletim junto a indicadores tradicionais de desempenho de alunos, o município transformou a equidade em um componente de sua política de responsabilização, demonstrando a importância atribuída a estratégias de ensino social e racialmente inclusivas (SMED, 2006).

É muito cedo para avaliar o impacto do programa, mas apesar de um ambiente altamente politizado não há sinais de nenhuma reação contrária. Isso é relativamente surpreendente, dado que o sindicato local dos professores do setor público sempre foi um crítico das avaliações externas, e a cidade de Belo Horizonte foi palco de uma reforma educacional radical do Partido dos Trabalhadores para melhorar a inclusão social que eliminou as avaliações dos alunos de final de ano. Relatórios de inspetores escolares sugerem que tem havido interesse e engajamento no Informativo da Escola ao invés de indiferença, o que pode ser atribuído à ausência de quaisquer consequências explícitas para as escolas com desempenho abaixo da média e ao esforço metódico da secretaria em explicar a utilidade da informação para o planejamento da escola (SMED, 2006).

5. DISCUSSÃO/CONCLUSÕES

A primeira observação a se fazer é que, a despeito do uso disseminado da avaliação externa de desempenho de alunos no Brasil, há muito poucos programas de responsabilização que usam esses dados de maneira a responsabilizar as escolas ou os professores por seus resultados. Para os programas de responsabilização existentes não há dados empíricos sobre suas possíveis consequências, exceto no caso do Rio de Janeiro. O estudo desses programas relativamente aos seus impactos nos resultados das

escolas não deve ser tarefa fácil, dado que a maioria tem sofrido mudanças contínuas desde sua implementação e que três dos seis exemplos foram eliminados após dois a quatro anos de existência.

Quanto às características dos diferentes programas, vale notar que todas as quatro políticas de responsabilização *high-stakes* estabeleceram alguma forma de incentivo salarial relacionada aos resultados da escola e dos professores. Não há exemplos de programas que usam outras formas de incentivo, tais como recursos adicionais ou vantagens de carreira para membros da equipe escolar. Também não há exemplos de programas que usam desincentivos, tais como o fechamento da escola, avaliações de desempenho dos professores mais baixas ou métodos mais públicos para a disseminação dos resultados da escola. Todos os programas existentes escolheram dar recompensas pecuniárias a algumas escolas (ou professores, no caso de Sobral) e deixar outras escolas sem essas recompensas, sob a hipótese de que a equipe da escola faria um esforço adicional para garantir um prêmio ou bônus mais alto. A queda nos níveis de desempenho no Ceará certamente traz dúvidas quanto à eficácia até mesmo desse tipo de programa de responsabilização, mas à luz da queda nos resultados de outros estados também (Rodrigues, 2007), é possível que o Ceará tivesse resultados ainda piores sem o Prêmio Escola do Novo Milênio.

Dos quatro programas descritos de responsabilização *high-stakes*, dois assumem a forma de prêmios e um terceiro a forma de um bônus anual ao salário, distribuído perto do Natal. Isso parece sugerir que, apesar de os governos municipais ou estaduais se interessarem em estabelecer uma conexão entre o resultado dos alunos e o pagamento final dos professores, a maneira de criar uma conexão mais direta para todas as escolas, como no caso do Rio de Janeiro, ou não se apresentou ou não foi politicamente viável. Como os prêmios são uma parte crescente da cena educacional, e representam uma maneira de se recompensar desempenhos superiores sem o fomento de discórdia provocado por outras formas de cálculo de suplementos diferenciais de pagamento, é fácil ver como eles se tornaram um aspecto central da política de responsabilização.

Outra característica é que apenas as versões mais recentes dos programas do Ceará e do Rio de Janeiro incorporaram maneiras de se determinar mudanças no desempenho dos alunos e da escola ao longo do tempo. Mesmo sem serem medidas legítimas do "valor agregado" pela escola, uma vez que nenhum dos dois programas emprega um modelo longitudinal para acompanhar os mesmos alunos, o grau do progresso de uma escola ao longo do tempo é claramente uma indicação melhor de sua eficácia e uma base melhor para o cálculo de suplementos salariais do que médias de desempenho de um único ano escolar. Isso significa que os outros programas, bem como as versões antigas dos programas do Ceará e do Rio de Janeiro, estão todos sujeitos à críticas quanto à simplicidade com que tratam a medida crucial do desempenho acadêmico dos alunos. A coleta de dados sobre outros aspectos do desempenho da escola também tem sido bastante rudimentar. Tem havido, portanto, dificuldades em legitimar e manter procedimentos para a coleta de dados, bem como em estabelecer uma conexão entre indicadores das avaliações e suas conseqüências.

Uma razão para o avanço vagaroso e aparentemente hesitante da responsabilização é a oposição dos sindicatos de professores. Essa oposição tem sido tanto reativa, por exemplo, nos estados do Rio de Janeiro, Paraná e São Paulo, como preventiva. No caso do estado de Minas Gerais, o uso de informação proveniente da avaliação de alunos para o propósito de responsabilização é expressamente excluído pelo Artigo 126 da legislação de 2000 que criou o Sistema Mineiro de Avaliação da Educação Pública (SIMAVE)⁶. Esse artigo proíbe o uso de resultados do sistema estadual de avaliação para o propósito de "classificar as escolas ou outros componentes do Sistema Mineiro de Educação com a intenção de alterar o processo educacional". Nessa rejeição explícita do uso dos resultados de testes para o propósito da

⁶ Resolução 14, Feb. 3, 2000

responsabilização parece que os legisladores estaduais foram influenciados pelo poder político do sindicato de professores e de outros críticos da avaliação externa.

Enquanto os professores do setor público e seus sindicatos fazem parte da base de poder dos governos, é pouco provável que as políticas de responsabilização baseadas no desempenho dos alunos sejam implementadas, devido a suspeitas constantes com relação à avaliação externa e devido à resistência ativa à avaliação do desempenho da escola. Mesmo os programas *low-stakes* podem ser combatidos por medo de consequências negativas. Quando os empregados do setor público, especialmente os professores, não representam fonte significativa de apoio ao governo, eles possam ser considerados até mesmo como opositores dos esforços do governo em melhorar a educação dos segmentos mais pobres da comunidade, como no caso do Rio de Janeiro, e a oposição desses funcionários ser ignorada. A sobrevivência da política Nova Escola no Rio não é uma medida de sua popularidade e nem de sua eficácia, mas sim de uma decisão do governo de procurar apoio político para suas políticas fora do setor educacional.

O precário desenho inicial das políticas estaduais e municipais de responsabilização, freqüentemente postas em práticas sem estudo e preparação adequados, parece indicar outro problema comum: a falta de assistência técnica e/ou financiamento a partir das autoridades educacionais federais. Como já foi dito, isso é consequência da divisão de responsabilidades entre as esferas do governo que perfazem a federação brasileira, o que inibe o envolvimento do governo federal no específico das políticas educacionais estaduais e municipais. Além de garantir a sua anuência com a Lei de Diretrizes e Bases aprovada em 1996, o Ministro da Educação no Brasil não exerce nenhum controle executivo ou legislativo sobre as secretarias de educação estaduais e municipais ou seus professores. A influência financeira que ele pode exercer também é limitada, sendo responsável por uma parte pequena do financiamento do ensino fundamental através da transferência de recursos federais de acordo com critérios de equalização fiscal. Sob a distribuição de deveres e responsabilidades atualmente aceita e o alto nível de autonomia local em termos políticos e financeiros, o Ministério tem buscado expandir suas próprias políticas de avaliação da educação sem tentar promover o uso da avaliação por outras esferas do governo. Sob essas regras nenhuma assistência técnica na criação de programas de responsabilização no âmbito estadual ou municipal foi pedida ou concedida.

Outra chave para se entender a incipiência da política de responsabilização é o nível de consciência em relação à qualidade da educação daqueles cujos filhos estão matriculados em escolas públicas do ensino fundamental ou médio. Não obstante a progressiva descentralização da administração escolar desde o retorno à democracia no princípio da década de 1980 e os consequentes esforços para envolver os pais em questões escolares, a fuga da classe média para o setor de educação privada (Chauí, 1999) significou menores níveis de mobilização e menor pressão por mudanças entre aqueles pais cujos filhos continuam a freqüentar a escola pública. Com a falta de provisão regular de informação na mídia para além da publicação dos resultados da avaliação nacional, refletindo a ausência de qualquer movimento político de maior vulto a favor de mudanças educacionais, os pais que mais têm mais a ganhar com programas de responsabilização que se concentram nos resultados dos alunos são justamente aqueles com menos condições de demandar tais políticas.

A despeito da ausência de um debate acadêmico e de apenas expressões limitadas de apoio popular às políticas de responsabilização existentes, tem havido sinais de interesse do Instituto Nacional de Estudos e Pesquisas Educacionais pertencente ao Ministério da Educação e Cultura (INEP/MEC). Esses sinais ainda estão longe de se tornar diretrizes para os governos estaduais e municipais, mas uma vez que o volume de recursos a ser transferidos para esses níveis de governo deve crescer em razão de legislação recém aprovada, é possível que o Ministério assuma um papel mais substancial na indução de políticas locais de responsabilização. Até agora o Ministério tem cumprido um papel completamente subsidiário na determinação de padrões educacionais, distribuindo livros-texto e outros recursos materiais e

fomentando a adesão aos parâmetros curriculares nacionais. No entanto, se estiver interessado em assumir maior liderança ele poderia usar seus recursos de acordo com os resultados de desempenho dos alunos disponibilizados pela recém-criada *Prova Brasil*, de sua autoria, estabelecendo critérios para comparar o desempenho não só de escolas, mas também das secretarias de educação locais.

Embora o boletim do Paraná não tenha sobrevivido por muito tempo, a implementação bem-sucedida de um instrumento similar na cidade de Belo Horizonte, um ambiente potencialmente hostil, sugere que esse pode ser o caminho a se trilhar no Brasil. Ainda não há evidência objetiva de mudanças, mas é mais fácil imaginar a aceitação e uso de informação *low-stakes* para promover discussões sobre o desempenho no âmbito da escola do que exercícios externos de avaliação que logo perdem sua capacidade de diagnóstico em virtude da importância ofuscante dos benefícios a que são associados.

Conquanto muitas das críticas referentes às deficiências técnicas dos sistemas de responsabilização existentes parecem justificadas, é cedo para descartar as políticas por inapropriadas ou ineficazes. No Brasil, fontes legítimas de pressão pelo melhoramento de resultados escolares, sejam essas fontes os pais, os alunos ou os próprios professores, são quase inexistentes. Sem essas fontes primárias de pressão para o melhoramento do padrão educacional há boas razões para os governos estaduais e municipais continuarem a experimentar com o uso de políticas de responsabilização que oferecem a possibilidade de melhores resultados. Em um cenário dramático de desempenho escolar abaixo de qualquer padrão, com efeitos desproporcionais sobre as chances de vida da população pobre, os governos locais devem reavaliar suas próprias atividades além de assumir um papel de liderança no melhoramento de suas escolas.

REFERENCIAS BIBLIOGRÁFICAS

- APEOESP (2006). Sindicato dos Professores do Ensino Oficial do Estado de São Paulo. Fax nº 48, 22/08/2006 At: http://www.apeesp.org.br/teste/Fax/48_06.htm. Access 1/04/07
- Ayres, S. (2003). Boletim da Escola: Diagnóstico Multidimensional das Escolas do Paraná. *Ensaio*, 38(11), jan-março. p. 33-44.
- Brandsma, H. and Knuver, J. (1989). Effects of School Classroom Characteristics on Pupil Progress in Language and Arithmetic. *International Journal of Educational Research*, 13.
- Brooke, N. (1992). The Equalization of Resources for Primary Education in Brazil, *International Journal of Educational Development*, 12(1) pp. 37-49.
- Carnoy, M. e Loeb, S. (2004). *A Responsabilidade Externa tem Efeito nos Indicadores Educacionais dos Alunos? Uma análise entre os estados dos E.U.A.* Documento 29. Santiago de Chile: PREAL.
- Carnoy, M., Loeb, S. y Smith, T. (2001). Do Higher State Test Scores in Texas Make for Better High School Outcomes? *Consortium for Policy Research in Education, University of Pennsylvania* (CPRE). Research Report Series RR-047.
- Castro, M.H.G. (1999). *Educação para o século XXI: o desafio da qualidade e da equidade*. Brasília : Instituto Nacional de Estudos e Pesquisas Educacionais.
- Cavaliere, A.M. (2003). Entre o Pioneirismo e o Impasse: A reforma paulista de 1920. *Educação e Pesquisa*, 29(1), pp.27-44.
- CDEP (2006). *Comitê em Defesa da Escola Pública*. <http://www.adusp.org.br/PEE/PEE.rtf> Accessed: 12/12/2006.

- Ferrer, G. (2006). *Educational Assessment Systems in Latin America: Current practice and future challenges*. Washington: PREAL.
- Fórum Estadual em Defesa da Escola Pública (2003). Plano Estadual de Educação: Proposta da Sociedade Paulista. *Fórum Estadual em Defesa da Escola Pública*, São Paulo, 14 de outubro, 2003. At: <http://www.adunicamp.org.br/noticias/educacao/PEE.pdf> Accessed: 01/04/2007.
- Gama, Z.J. (s.d.). *Nova Escola: Erros e acertos do programa: Folha Dirigida*, 18/03/ 2004 At: http://www2.uerj.br/~clipping/marco04/d18/dirigida_nova_escola.htm Accessed: 8/10/2005.
- Gillard, D. (2004). *Education in England: a brief history*. At <http://www.dg.dial.pipex.com/history/index.shtml> Accessed 12/01/07.
- Hanushek, Eric A. y Raymond, Margaret E. (2003). Lessons about the Design of State Accountability Systems. In Peterson, Paul E. and West, Martin R. (editors), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington, Brookings Institution Press.
- Hess, Frederick M. (2003). Refining or Retreating? High-Stakes Accountability in the States. En: Peterson, Paul E. and West, Martin R. (editors), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington, Brookings Institution Press, 2003.
- INEP (2001). *PISA 2000. Relatório Nacional*. Brasília: INEP.
- INEP (2005). *Vencendo o desafio da aprendizagem nas séries iniciais: a experiência de Sobral/CE*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
- INEP (2006). Prêmio Inovação em Gestão Educacional 2006, A Política de Alfabetização como estratégia para elevação do desempenho escolar nas séries iniciais do Ensino Fundamental, Sobral / CE. Laboratório de Experiências Inovadoras em Gestão Educacional, Relatório de Avaliação em Loco, September 2006.
- Jacob, Brian A. (2003). A Closer Look at Achievement Gains Under High-Stakes Testing in Chicago. In Peterson, Paul E. and West, Martin R. (editors), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington, Brookings Institution Press, 2003.
- Jones, Ken. (1989). *Right Turn: The conservative revolution in education*. London, Hutchinson Radius.
- Lapointe, Archie E., Askew, Janice M. and Mead, Nancy A. (1992a). *Learning Science*. Report No. 22-CAEP-02, Educational Testing Service, Princeton, NJ, Feb.
- Lapointe, Archie E., Askew, Janice M. and Mead, Nancy A. (1992b). *Learning Mathematics*. Report No. 22-CAEP-01, Educational Testing Service, Princeton, NJ, Feb.
- Lawton, Dennis and Chitty, Clyde. (1988). *The National Curriculum*. Bedford Way Papers 33. Institute of Education, University of London.
- Lima, Alessio C., Coelho, Sylvia M. and Soares, Lindomar da S., (2005). *Sistema Permanente de Avaliação da Educação Básica do Ceará: uma leitura dos resultados*. In Secretaria da Educação Básica do Ceará, *Gestão para o sucesso escolar*. Fortaleza, Edições SEDUC, 2005.
- Linn, R. (2000). Assessment and Accountability. *Education Researcher*. 25: 4-16
- MEC (2003). *Qualidade da Educação: uma nova leitura do desempenho dos estudantes da 4a série do Ensino Fundamental*. Ministério da Educação/INEP, Brasília. Abril 2003.
- MEC (2006). Tabelas do Censo Escolar 2006. Ministério da Educação/INEP, Brasília. At:

http://www.inep.gov.br/imprensa/noticias/censo/escolar/news07_02.htm

MEC (2007). *SAEB – 2005 Primeiros Resultados: Médias de desempenho do SAEB 2005 em perspectiva comparada*.

OECD (2004). *Programme for International Student Assessment. Learning for Tomorrow's World. First Results from PISA 2003*. OECD, Paris, 2004.

Rodrigues, J.R.S. (2007). Responsabilização e Resultados Escolares no Rio de Janeiro. Master's Degree Thesis. Pontifícia Universidade Católica, Rio de Janeiro.

Saliba, Alcyone V. (2003). Have You Heard About The Paraná (Brazil) School Report Cards? Mimeo.

Secretaria da Educação de São Paulo (1995). Coordenadoria de Estudos e Normas Pedagógicas. En: Rama, L. M. da S. *et al.* (org.). *Legislação de ensino de 1º e 2º graus: compilação*. São Paulo: Secretaria da Educação de São Paulo.

SMED (2006). Prefeitura Municipal de Belo Horizonte/Secretaria Municipal de Educação *Monitoramento da Inclusão: Manual do Informativo de Contexto e Desempenho para as Escolas Municipais de Belo Horizonte*. Belo Horizonte: SMED.

West, Martin R. y Peterson, Paul E. (2003). The Politics and Practice of Accountability, (2003). In Peterson, Paul E. and West, Martin R. (editors), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington: Brookings Institution Press.





¿QUÉ ESTAMOS HACIENDO EN CUBA EN EVALUACIÓN EDUCATIVA?

Paul Torres

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art8.pdf>



Cuando en fecha reciente iniciamos, en el Grupo de Investigación de Evaluación de la Calidad de la Educación del Instituto Central de Ciencias Pedagógicas (ICCP), un estudio del estado del arte de los Sistemas Evaluativos Nacionales, con vistas a desarrollar una de las principales tareas del Proyecto de Investigación Construcción de un Sistema de Evaluación de la Calidad de la Educación (SECE), nos llamó poderosamente la atención la limitada cobertura informativa que, sobre las acciones que se han venido realizando en Cuba en ese campo, se aprecia en la literatura especializada regional (Tiana, 2002) (Ferrer-Arregui, 2003) (Jaies, G. *et al.*, 2003) (Arregui, 2006) (Ferrer, 2006) (Murillo *et al.*, 2006).

La sospecha se hizo aún más evidente cuando, hace unos pocos días, un dúo de jóvenes pero competentes investigadoras foráneas visitaron nuestras instalaciones en búsqueda de información para un estudio comparado que sobre cinco países, en materia de logística y organización metodológica de Operativos Nacionales de Evaluación de la Calidad de la Educación, viene realizando su universidad. Llegaron acompañadas de un amplio soporte documental; sin embargo, el dossier cubano estaba compuesto por apenas unos pocos trabajos (Valdés, 2000) (Milán *et al.*, 2004) (Torres *et al.*, 2004) (Valdés, 2006) (Proenza-Leyva, 2006).

Fue así que experimentamos la necesidad de poner a disposición de especialistas e investigadores interesados información básica y actualizada sobre los pasos que se han venido dando en el país en la dirección de la Evaluación de la Calidad de la Educación y las proyecciones futuras más inmediatas en ese campo; algunas de las cuales pueden considerarse experiencias innovadoras; pero sobre todo, acciones ajustadas a las condiciones específicas del país.

Su cabal comprensión debiera estar precedido de un estudio del origen y evolución del Sistema Educativo Nacional, el que le ha permitido a este pequeño país del Tercer Mundo pasar, en pocos años, de un considerable nivel de analfabetismo a elevados índices educativos y culturales, avalados por resultados académicos destacados en el Primer Estudio comparativo y explicativo del Laboratorio Latinoamericano de la Evaluación de la Calidad de la Educación (LLECE, 2001). Pero esa empresa pudiera restar, lamentablemente, mucho espacio al propósito fundamental de este trabajo. En su defecto, el lector interesado pudiera conocer los rasgos fundamentales de ese desarrollo en (Gómez-Alonso, 2007:135-172) y en (CAB, 2003:257-292), donde aparecen descripciones de orden jurídico, organizativo y metodológico suficientemente documentadas.

La presentación de las ideas en torno al propósito fundamental declarado se hará en tres direcciones temporal y espacialmente diferenciadas. En primer lugar, se ofrecerá información acerca de la ejecución de los doce Operativos Nacionales de evaluación del rendimiento de los escolares cubanos que se han realizado ya en el país. Posteriormente, se explicarán los pasos que, en el orden proyectivo, se vienen dando para concertar las acciones evaluativas alrededor de un sistema evaluativo nacional, sustentado por una prolongada actividad investigativa teórico-práctica que ha venido movilizándolo en los últimos años a centenares de evaluadores educativos en la nación. Finalmente, se describirán los pasos que se vienen dando para crear y fortalecer una cultura nacional de la evaluación de la calidad de la educación.

1. EL PUNTO DE PARTIDA: LA EVALUACIÓN DEL RENDIMIENTO DE LOS ESTUDIANTES

Cuba, tras numerosas y valiosas investigaciones pedagógicas durante cuatro décadas de educación revolucionaria (CAB, 2003:257-292), se ha insertado sólidamente en los últimos doce años en el esfuerzo regional por la evaluación científica de la calidad de los servicios educativos prestados y del impacto de las políticas educativas introducidas.

Como prácticamente todos los Sistemas de Evaluación Educativa creados en la región, el primer foco de atención lo constituyó la evaluación del rendimiento de los estudiantes, especialmente de carácter

cognitivo, aunque se han realizado sistemáticamente mediciones de rendimientos *no cognitivos* y de *factores asociados al aprendizaje*.

En la tabla 1 se resumen las características fundamentales de los doce *Operativos Nacionales de Evaluación de la Calidad de la Educación* realizados hasta el momento en el país, los primeros de los cuales se organizaron a la luz de la participación cubana en el *Primer Estudio Internacional Comparativo sobre lenguaje, matemática y factores asociados*, para alumnos de tercer y cuarto grado de la educación básica, realizado por el LLECE en 1997.

Como puede apreciarse, las evaluaciones nacionales de *rendimiento* de los estudiantes han logrado sistematicidad en su ejecución, con la realización de eventos anuales. Las educaciones implicadas han ido en crecimiento; de una participación inicial de sólo la *educación primaria*, se han ido incorporando gradualmente las educaciones: *media básica, media superior, técnico-profesional, de adultos y de formación de docentes*, para el próximo año se prevé la inclusión además de la *educación especial*.

Las *áreas curriculares* evaluadas también han ido en crecimiento, con una estabilidad en torno a cuatro *asignaturas priorizadas*: Matemática, Español, Historia y Ciencias Naturales. En relación al *rendimiento no cognitivo* de los estudiantes y a los *factores asociados al aprendizaje*, como se aprecia, no se ha tenido la misma estabilidad pero se trabaja seriamente en ello.

Este gigantesco y sostenido esfuerzo ha sido posible gracias a la voluntad política de costear las evaluaciones proyectadas y de articular un aparato técnico-administrativo que, desde el Ministerio de Educación (MINED) hasta los territorios más apartados, se encargue de organizar y ejecutar los Operativos Nacionales y dar seguimiento a sus resultados. También merece remembranza el probado compromiso y dedicación de los miles de profesionales implicados cada año en la tarea.

Así, de unos pocos investigadores del ICCP, respaldados por un pequeño grupo de *colaboradores*, con que se iniciaron las evaluaciones nacionales hacia finales de la década de los '90, se ha logrado pasar a un estructurado cuerpo de *comisiones de trabajo* que aseguran su desarrollo exitoso. Estas son:

- un *Grupo Nacional de Calidad*, compuesto por Viceministros y Directores Nacionales del Ministerio de Educación, junto a otros especialistas seleccionados; es el órgano que preside la actividad, con el asesoramiento científico-metodológico del ICCP (para lo que sus investigadores se constituyen en jefes de los equipos de *Orientadores Nacionales*),
- las *Comisiones Provinciales*, dirigidas por el Rector del ISP y el Director Provincial de Educación e integrada por funcionarios con rango de *Orientadores Territoriales*,
- los *Directores (Decanos)* de las escuelas (facultades) seleccionadas, quienes ayudan a la ejecución del Operativo Nacional en la *base*, paralelamente a las actividades previstas en el *horario escolar* para el resto de los estudiantes que no pertenecen a la muestra.
- los *Aplicadores*, que son docentes de las propias instituciones educativas seleccionadas que administran los instrumentos de evaluación fuera de su grupo docente, bajo la observancia del *Director* y de uno o varios *Orientadores Territoriales*.
- los *Tribunales de Asignaturas*, compuestos por especialistas destacados en las áreas curriculares evaluadas, los que se encargan de calificar las *preguntas abiertas* de las pruebas de rendimiento cognitivo.
- los *Digitadores*, organizados en dúos que se responsabilizan con la captura de los datos generados por los instrumentos, utilizando Bases de Datos diseñadas por especialistas del ICCP.

TABLA 1. OPERATIVOS NACIONALES DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN

No.	Año	Mes	Escuelas	Muestra ¹	Grados	Instrumentos aplicados
I	1996	Marzo	99	3,726	3	Matemática y Español.
II	1996	Noviembre	100	-	3	Matemática y Español.
III	1997	Marzo	100	3,766	3	Matemática y Español.
IV	1999	Octubre	200	8,000	3 y 6	Matemática y Español, encuestas a docentes e ICEDIM ² .
V	2001	Mayo	280	12,500	4, 6, 7 y 9	Matemática, Español y encuestas a docentes y alumnos.
VI	2002	Septiembre	6,115	34,615 ³	7, 9 y 12	Matemática, Español y encuestas a alumnos.
VII	2003	Noviembre-Diciembre	876 ⁴	19,498	3, 5, 6, 7, 8 y 9	Matemática, Español. (En 7, 8 y 9, además: encuestas a docentes, funcionarios y alumnos).
VIII	2004	Mayo	1,349 ⁵	93,032	2, 4, 6, 7, 8 y 9	En 2, 4 y 6: Matemática, Español y visitas a clases. En 8 y 9: Matemática, Español y encuestas a docentes, funcionarios y alumnos.
IX	2004	Diciembre	384	24,435	3, 6, 10, 11 y 12	Matemática, Español y Ciencias Naturales ⁶ .
X	2005	Mayo-Junio	771	30,642	2, 4, 6, 7, 8, 9, 10 y 11	Matemática, Español, Historia ⁷ y Ciencias Naturales ⁷ , así como encuestas a docentes, a padres y a alumnos.
XI	2006	Mayo-Junio	840 ⁷	73,392	7, 8, 9, 10, 11, 1er. y 2do. año de ETP ⁸ , 1er. año de las carreras de Primaria, Especial, PGI ⁹ y Ciencias Exactas, Ciencias Naturales y Humanidades de los ISP ¹⁰ .	Matemática, Español, Historia y Ciencias Naturales, así como encuestas a alumnos y docentes y revisión de libretas de Matemática y Español a una sub-muestra de alumnos.
XII	2007	Marzo-Abril	738 ¹¹	74,834	4, 6, 9, 11, 2do. año de ETP, 3er. semestre de EDA ¹² y 1er. año de las carreras de Primaria, Especial, PGI y Ciencias Exactas, Ciencias Naturales y Humanidades de los ISP.	Matemática, Español, Historia ⁷ y Ciencias Naturales ⁷ , así como ICEDIM y revisión de libretas de Matemática y Español a una sub-muestra de alumnos.

¹ Salvo unos pocos casos, las aplicaciones han sido muestral para escuelas y alumnos y han estado representadas todas las provincias del país.

² *Índice de Certeza del Diagnóstico del Maestro*: estimación por el docente de la calificación de sus estudiantes en la prueba que se aplicaría, a partir del diagnóstico cognitivo previo.

³ Fue censal en el estrato de escuelas y muestral a nivel de alumnos.

⁴ Se tomó una muestra aleatoria para hacer el seguimiento del nuevo modelo que se estaba proyectando en la Secundaria Básica y en la provincia capital del país, y se aplicó a toda la matrícula de los docentes incluidos en la investigación.

⁵ Muestral a nivel nacional y estratificado a instancia provincial.

⁶ Sólo del 6to. grado en adelante.

⁷ Incluye a 15 Institutos Superiores Pedagógicos, uno por provincia del país.

⁸ Enseñanza Técnico-Profesional.

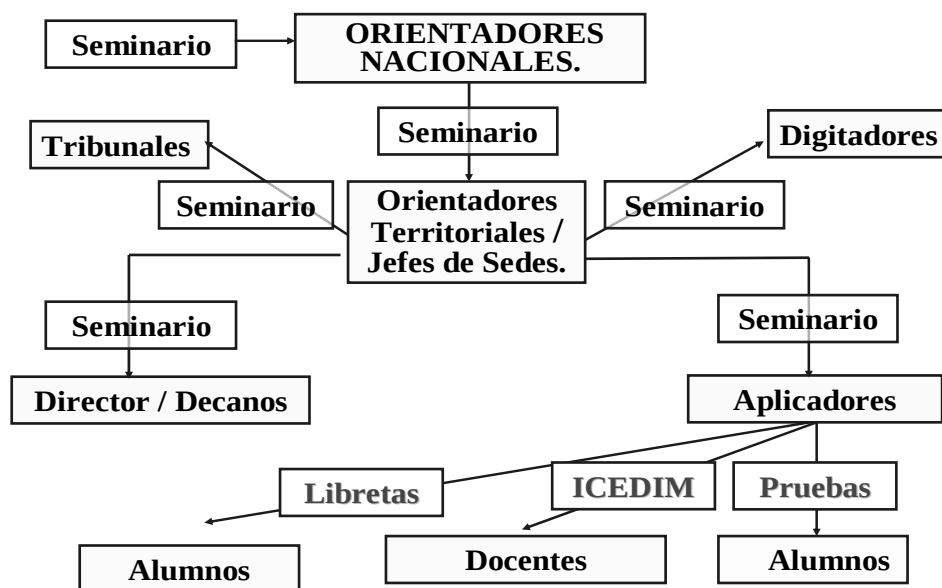
⁹ Profesor General Integral de Secundaria Básica.

¹⁰ Institutos Superiores Pedagógicos.

¹¹ Incluye a 15 Institutos Superiores Pedagógicos, uno por provincia del país.

¹² Educación de Adultos, en su modalidad de Curso de Superación Integral para Jóvenes (desvinculados de los estudios).

FIGURA 1. ESQUEMA ORGANIZATIVO DEL XII OPERATIVO NACIONAL



El grueso de los instrumentos de rendimiento cognitivo que se han estado utilizando en los Operativos Nacionales son del tipo *pruebas en formas paralelas* y se han confeccionado y piloteado de acuerdo con la *Teoría Clásica del Test*. A partir del IX Operativo Nacional se han venido utilizando pruebas del tipo *bloques incompletos balanceados* para la *educación primaria*, con resultados satisfactorios. También, con el asesoramiento del ICCP se dan pasos firmes para elaborar y *calibrar* las *pruebas de rendimiento cognitivo* con ajuste a la *Teoría de Respuesta al Ítems* (TRI), en breve tiempo.

Los resultados de las *pruebas de rendimiento cognitivo* aplicadas se procesan mediante *porcentajes de respuestas correctas* (globalmente y por *tópicos* o *dominios cognitivos*), así como de por cientos de estudiantes por *niveles de desempeño cognitivo*, dos de los siete tipos de *métricas* identificadas por P. Ravela en (Arregui, 2006:196-197)¹.

Las devoluciones de los resultados de los *Operativos Nacionales de Evaluación de la Calidad de la Educación* no son de dominio público, aunque están fuertemente enfocadas hacia la *mejora escolar*. Esa devolución se lleva a efecto a través de una *cascada de cuatro informes*, todos listos antes de comenzar el siguiente curso escolar.

El *Informe I* reporta los resultados más generales, con los por cientos de *respuestas correctas* por *asignaturas* y *tópicos*, abiertos a provincias y sector (urbano, rural); así como los por cientos de estudiantes en los diferentes *niveles de desempeño cognitivo*, también abiertos a asignaturas y provincias. El *Informe II* resume los principales efectos de los *factores asociados al aprendizaje*, resultantes de los otros instrumentos aplicados y, eventualmente, de *estudios de profundización* realizados en sub-muestras de escuelas.

¹ En realidad, las Bases de Datos cubanas están diseñadas para calcular inmediatamente dos tipos de métricas más (el *porcentaje promedio de respuestas correctas* y el *promedio de puntajes brutos*).

Por su parte, el *Informe III* profundiza en los *elementos del conocimiento* más afectados; esto son: los *distractores* de las *preguntas cerradas* y los *créditos-procedimientos* de las *preguntas abiertas* con mayor impacto negativo sobre los resultados generales. Es un informe más técnico y tiene una considerable incidencia sobre las actividades de perfeccionamiento curricular que generan los directivos del MINED y las *Comisiones Nacionales de Asignaturas*, ambos asistidos por investigadores del ICCP.

Como resultado de ese accionar se generan, a instancia nacional, talleres con los directivos de los territorios y se elaboran materiales de carácter metodológico para los docentes, así como se realizan intervenciones de destacados especialistas en el *Seminario Nacional para Educadores*, evento telemático que organiza el MINED hacia el inicio de cada curso escolar y que moviliza prácticamente a todos los directivos, metodólogos y docentes del país.

Desde el pasado Operativo Nacional se ha previsto la confección adicional de un *Informe IV*, individual para cada una de las instituciones educativas de la muestra, el que cuenta con los principales resultados curriculares alcanzados y un análisis técnico de los mismos.

También en las estructuras provinciales de dirección se elaboran informes y se desarrollan análisis y acciones similares, asistidas por los *Grupos Provinciales de Evaluación de la Calidad de la Educación*², una estructura creada en el país desde el 2001, con la aprobación del MINED y la atención metodológica permanente del ICCP.

Precisamente estos *Grupos* se ocupan además de organizar, hacia mediados del curso, estudios provinciales, previos al Operativo Nacional y monitorear permanentemente la marcha del *proceso docente-educativo*, amparados en las *Resoluciones Ministeriales 106/04 y 104/05*.

Para ello reportan trimestralmente los resultados de las *comprobaciones de conocimientos* aplicadas como parte del *sistema de inspección escolar* o las visitas nacionales, provinciales y municipales de *Entrenamiento Metodológico Conjunto* (Gómez-Alonso, 2007). Así, por ejemplo, durante el curso escolar 2006-2007 se informaron al *Grupo Nacional de Calidad de la Educación* un total de 4208795 estudiantes evaluados y 171365 clases observadas. Las evaluaciones de las clases se efectúan con ajusta a una guía de observación estructurada (Valdés, 2006:9-11).

2. MÁS ALLÁ DE LOS OPERATIVOS NACIONALES, EL SECE

Prácticamente desde el inicio de los trabajos en el país de *Evaluación de la Calidad de la Educación*, con el desarrollo de los primeros Operativos Nacionales, se concibió la idea de la creación de un *Sistema de Evaluación de la Calidad de la Educación* (SECE); un esbozo del mismo, en términos de ámbitos e indicadores fundamentales, puede encontrarse en (Valdés-Pérez, 1999) y (Valdés *et al.*, 1999).

Sin embargo, no fue hasta inicios de la presente década que su construcción se asumió con rango de Proyecto de Investigación asociado a un Programa Ramal del MINED. El primer diseño de Proyecto fue presentado en septiembre del 2000 por el Grupo de Investigación de *Evaluación de la Calidad de la*

² Actualmente, estos *Grupos* están integrados por un Jefe, al menos un representante de cada una de las educaciones implicadas en los Operativos Nacionales y un informático; los representantes de las educaciones cuentan con un fondo de tiempo laboral no inferior al 75% de su plan individual, mientras que el Jefe de Grupo y el informático lo hacen al 100%. Responden administrativamente a las *Vice-rectorías de Investigación y Postgrado* de los ISP.

Educación, del ICCP, previéndose entonces como objetos fundamentales del proceso de evaluación los ámbitos siguientes:

- Eficiencia interna y externa del Sistema Educativo,
- Aprendizaje y formación de los alumnos, y
- Dirección del proceso educativo en los centros.

También hacia finales del 2001 se logró instrumentar, dentro del *Plan de Ciencia e Innovación Tecnológica* del MINED, un Programa Ramal afín en el que, además del Proyecto principal a cargo del ICCP (*Un Sistema Integral de la Evaluación de la Calidad de la Educación*), se consideraban un grupo de *tareas por encargo* a diferentes Institutos Superiores Pedagógicos (ISP) del país, que lo complementaban.

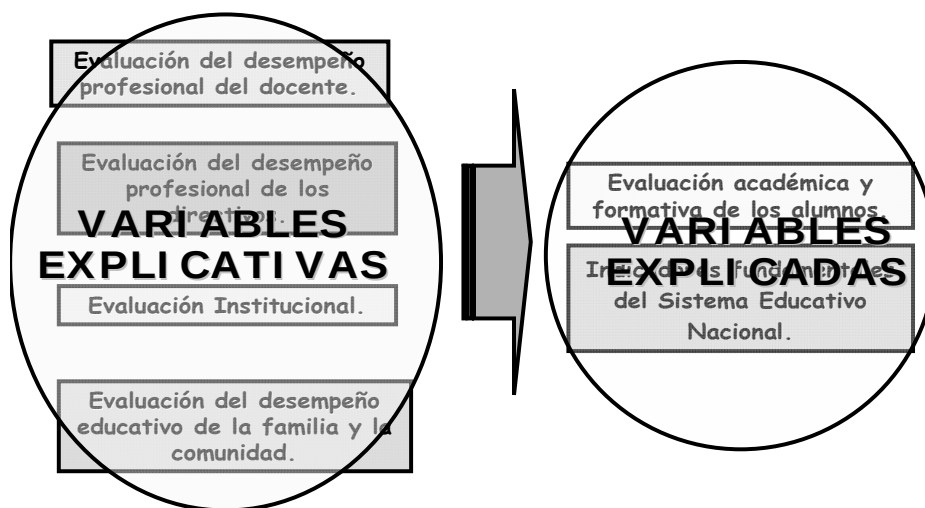
Para el Proyecto principal se habían concebido 5 resultados fundamentales, a ejecutar entre el 2002 y el 2004:

1. Índice General de Calidad Educativa.
2. Línea Base de Evaluación de la Calidad Educativa en Cuba.
3. Sistema de evaluación del desempeño docente.
4. Sistema de Evaluación Institucional.
5. Sistema Integral de Evaluación de la Calidad.

Lamentablemente, no todos los compromisos se pudieron ejecutar durante el período previsto. El Proyecto expiró, aunque antes logró que el estudio realizado en torno al *V Operativo Nacional de Evaluación de la Calidad de la Educación* alcanzara, atendiendo a su aporte social y precisión científica, uno de los Premios de la Academia de Ciencias de Cuba (Campistrous, 2001a).

En la actualidad, tras el ajuste del *Plan de Ciencia y Técnica* del MINED realizado hacia finales de 2005, se logró retomar en el 2006 ese Proyecto (ahora titulado: *Construcción de un Sistema Cubano de Evaluación de la Calidad de la Educación*), también a cargo del ICCP. En el mismo se concibe el *Sistema* como la interrelación de las *metodologías de evaluación del desempeño* de los agentes educativos, originalmente considerados (alumno, docente, directivo y familia), y de las agencias involucradas (escuela y comunidad), así como el control de importantes macro-indicadores de eficiencia del Sistema Educativo Nacional.

FIGURA 2. MODELO DEL CONTENIDO DE EVALUACIÓN DEL SECE



¿Qué representación proyectiva se tiene de esta nueva versión de *Sistema Cubano de Evaluación de la Calidad de la Educación*? ... Sus componentes esenciales son los siguientes:

- **Objetivo general:** *Diagnosticar el estado general del funcionamiento del Sistema Nacional de Educación, sobre la base del establecimiento de juicios valorativos de los niveles de calidad educativa de los componentes evaluados y de la formulación de recomendaciones dirigidas a la mejora escolar.*
- **Contenido de la evaluación:** los ámbitos de *desempeño cognitivo y formativo de los estudiantes, profesional de los docentes y directivos, funcionamiento general de los centros escolares y del desempeño educativo de la familia y la comunidad*, así como de *indicadores de eficiencia interna y externa* del Sistema Nacional de Educación.
- **Metodología y formas de organización de la evaluación:**
 - **Procesos de auto-evaluación:** entre noviembre y diciembre, ejecutados desde las propias escuelas, de acuerdo con las *metodologías de evaluación* diseñadas, con ajustes a las particularidades e intereses adicionales de los colectivos educacionales "*de base*".
 - **Procesos de co-evaluación:** entre enero y febrero, efectuados por equipos evaluadores de escuelas intercambiados y la intervención del *Grupo Municipal de Calidad* en los *Centros de Referencia*³; también con las metodologías de evaluación diseñadas, con adecuación y flexibilidad.

³ Institución educativa que, a instancia municipal y provincial, es seleccionada con vistas a que los cuadros educacionales demuestren las posibilidades de realizar transformaciones educativas, como parte del método de dirección científica conocido como *Entrenamiento Metodológico Conjunto*.

- o **Proceso de evaluación externa:** entre marzo y junio, combinando el Operativo Nacional con *estudios de profundización* desarrollados por los *Grupos Provinciales de Evaluación de la Calidad*. Este evento evaluativo final proporcionaría información, tanto referida a *variables-producto* fundamentales (aprendizaje y orientaciones valorativas), como de valiosas *variables de contexto, entrada y proceso*, propiciando un cruce estadístico de variables que permitirían una aproximación más objetiva a las principales causas de los resultados de las *variables-producto*, como el proporcionado por los *Modelos Jerárquicos Lineales* (Torres *et al.*, 2007), todo lo cual favorecería sensiblemente la toma de decisiones en pos de la *mejora escolar*.

Este último proceso debe complementarse con otros *estudios de profundización* a ejecutar por los *Proyectos de Investigación Asociados* al Programa Ramal. Estos pudieran ser *estudios transversales* (o sea, tratando de penetrar en los resultados del aprendizaje proporcionados por último *Operativo Nacional*), en los que debiera predominar el empleo de técnicas de la investigación cualitativa (*observación participante, entrevista a profundidad, grupos focales, estudios de casos*, etc.), con vistas a detectar estrategias de aprendizaje y de formación axiológica exitosas, así como "*obstáculos epistemológicos*" de los contenidos escolares evaluados, entre otros aspectos, y producir propuestas educativas y metodológicas que los contrarrestan (*guías de ejercicios, otros materiales didácticos y formativos, software educativos, guiones de video-clases*, etc.).

También serían deseables y útiles estudios longitudinales (es decir; utilizando datos de varios *Operativos Nacionales de Evaluación de la Calidad de la Educación*), con vistas a identificar regularidades, *errores cognitivos frecuentes*, etc. y generar productos educativos similares a los arriba señalados (Bernabeu *et al.*, 2007).

Para la correcta implementación del *Sistema de Evaluación*, es importante reconocer también un grupo de **principios generales** que jalonan los múltiples y complejos esfuerzos que ese proceso requiere. Una propuesta inicial es la siguiente:

- **Principio de la unidad de la precisión y la factibilidad de las metodologías de evaluación.**

Esto es, la elaboración de metodologías contentivas de instrumentos e indicadores obtenidos con arreglo a los requerimientos de la investigación científica (especialmente la *validez* y la *confiabilidad* de los instrumentos a emplear), pero a la vez ajustados a una razonable racionalización de los recursos materiales, financieros y humanos disponibles (Torres-Galdós, 2006:51).

- **Principio de la participación masiva de los agentes educativos en la evaluación.**

Como se ha explicado, interesa tanto la generación de procesos de *evaluación externa* (en términos de *Operativos Nacionales de Evaluación de la Calidad de la Educación* y sus correspondientes *estudios de profundización*), como de *evaluación interna* (*auto-evaluaciones* y *co-evaluaciones*) (ICCP, 2006:1).

- **Principio del carácter gradual de la evaluación educativa.**

Se consideran tanto la *evaluación inicial*, como *procesal* y *final*. En la primera se implican los claustros de las instituciones docentes, en la segunda los *Grupos Municipales de Evaluación de la Calidad* y en la tercera los *Orientadores Nacionales y Territoriales* y los *Grupos Provinciales de Evaluación de la Calidad* (Valdés-Pérez, 1999:72).

- **Principio de la unidad de la evaluación de variables-producto y variables-explicativas.**

Al evaluar se deben recolectar datos, tanto del comportamiento de *variables-producto* esenciales (como el aprendizaje y las orientaciones valorativas), como de *variables-explicativas* (de *contexto*, de *entrada* y de *proceso*) que, al ser cruzadas estadísticamente con las primeras, permiten estimar el impacto de las segundas con relación a aquellas.

- **Principio de la unidad de la evaluación de la investigación cuantitativa y la cualitativa.**

La evaluación debe devolver rápidamente un grupo de datos esenciales, en cuya captura y procesamiento predominará el empleo de técnicas de la investigación cuantitativa, pero además debe complementarse con *estudios de profundización* donde intervendrán recursos y técnicas de la investigación cualitativa; estos últimos serán ejecutados por equipos de investigación territoriales, organizados en Proyectos de Investigación asociados al Programa Ramal y extenderá su responsabilidad hasta la producción de propuestas educativas y didácticas especialmente diseñadas para contrarrestar las principales deficiencias cognitivas y formativas detectadas.

Coincidiendo con (Valdés-Pérez; 1999:73), es previsible que semejante empresa no se pueda ejecutar en un plazo breve, requerirá de mucho esfuerzo organizativo, una sostenida capacitación y una rápida generación de resultados concretos y tangibles. Las *metodologías de evaluación educativa* que se han estado diseñando, sus *instrumentos e indicadores* requieren de una repetida y relativamente prolongada *validación práctica*, en escenarios educativos nacionales diferentes y mostrando estabilidad en el tiempo, requisito esencial de la *confiabilidad* de la investigación científica.

3. EL FIN SUPREMO: EL DESARROLLO DE UNA *CULTURA DE LA EVALUACIÓN EDUCATIVA*

Se reconoce que la *evaluación educativa* cumple, cuando menos, tres funciones básicas: *diagnosticar*, *juzgar* y *mejorar*. En efecto, los datos obtenidos sistemáticamente sobre aspectos relevantes de la actuación educativa proyectan una imagen del estado de cumplimiento de los objetivos trazados y de los niveles de calidad alcanzados. La *evaluación educativa* utiliza esa información para contrastarla con el ideal de funcionamiento del Sistema, de modo que se emiten criterios acerca de su estado, enjuiciándolo. Pero la *Evaluación de la Calidad de la Educación* no debe quedarse ahí, debe implicarse en los procesos de transformaciones requeridos y propiciar el cambio.

Esta última función difícilmente podrá lograrse sin la decidida participación de todos. Siguiendo a (Valdés-Pérez, 1999:71-72) e (ICCP, 2006:1), lo adecuado sería implicar a los agentes educativos "*de base*" (docentes y directivos de escuelas) también en la *evaluación*, de lo que se trata es de desarrollar *una cultura de la evaluación educativa*.

¿Qué pasos se han venido dando en esa dirección?... Una estrategia ha sido tratar de crear y movilizar una comunidad nacional de *evaluadores educativos*. Los *Grupos Provinciales y Municipales de Evaluación de la Calidad de la Educación*, así como los *Proyectos de Investigación Asociados* al Programa Ramal, sus reuniones nacionales y eventos científicos anuales, son formidables vehículos para difundir y sembrar ideas.

También resulta muy útil la *Lista de Discusión* creada por la dirección del Programa Ramal en la intranet del MINED (pr-l@listas.rimed.cu), con casi dos años de servicio y más de cien suscriptores actualmente. A través de ella se han podido organizar *seminarios científicos on-line* y se difunden valiosos documentos

para nuestra línea de investigación. Como complemento de la Lista, se está habilitando también dentro de la intranet (<http://ftp.rimed.cu/ICCP>) una *biblioteca virtual* indexada de *Evaluación de la Calidad de la Educación*, donde se depositan numerosas y valiosas obras en soporte digital.

Otro importante paso de avance ha sido la asunción de varios *Grupos Provinciales de Evaluación de la Calidad de la Educación* a *Centros de Estudios* de los Institutos Superiores Pedagógicos, a partir del presente curso escolar, con lo cual se fortalece la producción científica, tanto de corte teórico como práctico, en esa dirección.

FIGURA 3. REPRESENTACIÓN DE LA BIBLIOTECA VIRTUAL



4. CONCLUSIONES

El propósito fundamental de este trabajo ha sido explicar los pasos que se han venido dando en Cuba para fortalecer la línea investigativa de la *evaluación educativa*, ante la evidencia de un conocimiento limitado de ellos en la literatura especializada internacional. El análisis realizado pone en evidencia el prolongado trabajo desplegado en el país con vistas a crear un *Sistema Cubano de Evaluación de la Calidad de la Educación* y, más allá del mismo, el fomento de una *cultura nacional de evaluación* en educación.

Primeramente, se explicaron los esfuerzos que han conducido a la consolidación de los *Operativos Nacionales de Evaluación de la Calidad de la Educación*, como columna vertebral del desarrollo de la *evaluación educativa* en Cuba, para posteriormente describir la visión proyectiva del *Sistema* en cuya construcción se trabaja intensamente, así como los esfuerzos primarios para constituir y fortalecer una comunidad nacional de *evaluadores educativos*.

Tras estos primeros diez años de duro bregar, el Sistema Educativo Nacional ha podido comenzar a apreciar los frutos de la instrumentación científica de la *evaluación educativa* en el país y las amplias

posibilidades que ella ofrece para el necesario y permanente perfeccionamiento del servicio educativo, piedra angular de la *batalla de ideas* que realiza nuestro pueblo, en busca de una plena conquista de la equidad y la justicia social.

REFERENCIAS BIBLIOGRÁFICAS

- Achig, L. (1997). *Auto-evaluación de las instituciones de Educación Superior*. Quito: Universidad de Cuenca.
- Álvarez, B. y M. Ruiz-Casares (Editores) (1997). *Evaluación y Reforma Educativa. Opciones de política. Informe Técnico No.3*. Santiago de Chile: PREAL.
- Arregui, P. (Editora) (2006). *Sobre estándares y evaluaciones en América Latina*. Santiago de Chile: PREAL.
- Bernabeu, M. et al. (2007). *Errores frecuentes de los estudiantes de educación básica en la evaluación del desempeño académico en Matemática y Español*. La Habana: ICCP.
- Campistrous, L. (2001a). *Primeros resultados del Operativo Nacional de Calidad de la Educación*. La Habana: ICCP (02.05.01) (Versión confidencial)
- Campistrous, L. (2001b). *Índice general para evaluar el trabajo educacional*. Informe de investigación. La Habana: ICCP.
- Campistrous, L. (2001c). *Índice Socio-educativo*. Informe de Investigación. ICCP. La Habana: ICCP.
- Campistrous, L. et al. (2002). *Sexto Operativo de la Evaluación de la Calidad de la Educación*. Informe Final. La Habana: ICCP. (02.09.02) (Versión confidencial)
- Castro, O. (1999). *La evaluación de proyectos educativos: un problema teórico y metodológico actual en América Latina. Congreso Pedagogía 99*, La Habana.
- Cerezal, J. et al. (1997). Metodología para la caracterización de los centros internos del nivel medio. *Congreso Pedagogía '97*, La Habana.
- Escudero, T. (2003). Desde los tests hasta la investigación evaluativo actual. Un siglo, el XX, de intenso desarrollo de la evaluación en Educación. *RELIEVE* 9(1); <http://www.uv.es/RELIEVE>
- Fernández Aguerre, T. (2003). *Métodos estadísticos de estimación de los efectos de la escuela y su aplicación al estudio de las escuelas eficaces. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación. Vol. 1(2)*.
- Ferrer, G. (2006). *Sistemas de Evaluación de Aprendizajes en América Latina. Balances y Desafíos*. Santiago de Chile: PREAL.
- Ferrer, G. y Arregui, P. (2003). *Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: criterios para guiar futuras aplicaciones*. Santiago de Chile: PREAL.
- García Ramis L. et al. (1995). *Sistematización del modelo de escuela Cubana*. La Habana: ICCP. (Informe de Investigación).
- García Ramis, L. (1996). *Los retos del Cambio Educativo*. La Habana: Pueblo y Educación.
- Gómez, L. I. (2005). Conferencia Especial: Cuba: una revolución en la Educación. *Congreso Pedagogía 2005*, La Habana.
- Gómez, L. I. y Alonso, S. H. (2007). *El entrenamiento metodológico conjunto. Un método revolucionario de dirección científica educacional*. La Habana: Pueblo y Educación.
- Grupo Editorial Océano (2001). *Enciclopedia General de la Educación*. Barcelona: Océano.

- Guerra, M. (2004). *Modelo para la evaluación integral de la calidad del proceso de dirección de los ISP*. Resultado de Investigación y Desarrollo del Proyecto. Evaluación de la Gestión Directiva. Holguín: ISPJLC. .
- laies, G. et al. (2003). *Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IIPE-UNESCO.
- ICCP (2006). Relación entre evaluación educativa y calidad de la educación: encuentros y desencuentros. ¿Hacia dónde vamos? *Seminario Científico*, La Habana.
- INEE (2004). *Hacia un nuevo paradigma para la Evaluación Educativa*. México: INEE.
- Lorenzo, R. (2004). *Factores asociados al aprendizaje de la Matemática en estudiantes de Secundaria Básica*. Tesis de Maestría. ISPJLC. Holguín.
- LLECE (2001). *Primer Estudio Internacional Comparativo sobre lenguaje, matemática y factores asociados, para alumnos de tercer y cuarto grado de la educación básica. Informe Técnico*. Santiago de Chile: OREALC-UNESCO.
- Milán, M. R. et al. (2004). *La evaluación como un proceso participativo*. *Revista Pedagógica Universitaria* 9(4).
- MINED (1981). *Pedagogía*. La Habana: Pueblo y Educación.
- MINED (1988). *Perfeccionamiento Continuo del Sistema Nacional de Educación: Plan de Estudio de la Educación General Politécnica y Laboral*. La Habana: MINED.
- MINED (2000). *Seminario Nacional para el personal docente*. La Habana: Pueblo y Educación.
- MINED (2001). *Seminario Nacional para Educadores*. La Habana: Pueblo y Educación.
- MINED (2004). Resolución Ministerial 106/04. Gaceta Oficial. República de Cuba.
- MINED (2005). Resolución Ministerial 104/05. Gaceta Oficial. República de Cuba.
- Murillo, J. (1999). Los Modelos Jerárquicos Lineales aplicados a la investigación sobre Eficacia Escolar. *Revista de Investigación Educativa* 17(2), pp.453-460.
- Murillo, J. (2001). *Mejora de la Eficacia Escolar*. Temas del mes de Cuadernos de Pedagogía, 300, pp. 47-74
- Murillo, J. (2002). Aportaciones de la Investigación sobre Eficacia Escolar para el mejoramiento de la calidad y la equidad en América Latina. *IV Simposio Iberoamericano de Investigación y Educación*, La Habana.
- Murillo, J., et al. (2003). *La investigación sobre Eficacia Escolar en Iberoamérica. Revisión Internacional sobre el estado del Arte*. Bogotá: CAB-CIDE.
- Murillo, J. (2006). *Evaluación del desempeño y carrera profesoral docente. Un estudio comparado entre 50 países de América y Europa*. Santiago de Chile: OREALC-UNESCO.
- Organización de Estados Iberoamericanos (1996). Evaluación de la Calidad de la Educación. *Revista Iberoamericana de Educación*, 10.
- Organización de Estados Iberoamericanos (2006). *Curso en Línea en Evaluación Educativa*. Acceso online en: <http://campusoei.org>
- PRELAC-UNESCO (2002). *Modelo de acompañamiento –apoyo, monitoreo y evaluación- del Proyecto Regional de Educación para América Latina y el Caribe. PRELAC. Declaración de la Habana*. La Habana: Andros Impresores.
- Proenza, Y. y L. M. Leyva (2006). Reflexiones acerca de la calidad del aprendizaje y de las competencias matemáticas. *Revista Iberoamericana de Educación*, 41(1).
- Suárez, A. (2004). *Diagnóstico del estado de las investigaciones sobre Evaluación Institucional de la Enseñanza General en Cuba*. Tesis de Maestría en Investigación Educativa. ICCP. Ciudad de la Habana.
- Tiana, A. (2000). *¿Qué variables explican los mejores resultados en los estudios internacionales?* Madrid: INEP.

- Torres, P. *et al.* (1999). Estrategia de trabajo para el mejoramiento del aprendizaje de la Matemática en los preuniversitarios habaneros. *Congreso Pedagogía '99*, La Habana.
- Torres, P. *et al.* (2004). Estrategia cubana para la determinación de una metodología de evaluación internacional. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*. Vol. 1(2).
- Torres, P. *et al.* (2005). Consideraciones para la determinación de una metodología de evaluación institucional en la Educación Cubana. *Congreso Pedagogía '95*, La Habana.
- Torres, P. *et al.* (2007). *Informe de los resultados del estudio de profundización del XII Operativo Nacional de Evaluación de la Calidad de la Educación en centros de referencia provinciales*. La Habana: ICCP. (Inédito)
- Torres, P. y S. A. Galdós (2007). *Evaluación Institucional*. La Habana: Editorial Academia.
- Unidad de Medición de Calidad Educativa. (2002). Sistemas Nacionales de Evaluación del rendimiento escolar en América Latina. *Revista CRECER*, pp.7-10.
- Valdés, H. (1999). *Informe del SECE, 1999*. La Habana: ICCP.
- Valdés, H. (2000). Evaluación del desempeño docente. *Encuentro Iberoamericano sobre evaluación del desempeño docente*, OEI, Ciudad de México.
- Valdés, H. (2004). *Propuesta de un sistema de evaluación del desempeño profesional del docente de la escuela básica cubana, que contribuya a la realización de un proceso evaluativo más objetivo, preciso y desarrollador*. Tesis Doctoral. ICCP. Ciudad de la Habana.
- Valdés, H. (2004). *El desempeño del maestro y su evaluación*. La Habana: Editorial Pueblo y Educación.
- Valdés, H. (2006). La evaluación del desempeño docente: un pilar del Sistema de Evaluación de la Calidad de la Educación en Cuba. *Panel: Sistemas Nacionales de Evaluación: lo que se ha ganado de la experiencia*. (sle)
- Valdés, H. y Pérez, F. (1999). Calidad de la Educación Básica y su Evaluación. La Habana: Editorial Pueblo y Educación.
- Valdés, H. y Torres, P. (2005). Calidad y equidad de la Educación: concepciones teóricas y tendencias metodológicas para su evaluación. *Congreso Pedagogía 2005*, La Habana.
- Valiente, P. y R. Álvarez (2000). Una cultura de la Evaluación. La Evaluación de Sistemas Educativos, Políticas y Programas. IPLAC. Ciudad de la Habana, 2000. (*Material docente del diplomado en Supervisión y Administración Educativa*)
- Valle *et al.* (2003). *La transformación educativa. Consideraciones*. La Habana: Editorial Pueblo y Educación.



JORNADA ESCOLAR COMPLETA EN CHILE. REPRESENTACIONES DE LOS PROFESORES SOBRE SUS EFECTOS EN LOS APRENDIZAJES

Sergio Martinic, David Huepe y Ángela Madrid

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art9.pdf>

* Artículo elaborado en el marco del "Proyecto Representaciones sociales y gestión del tiempo en la sala de clases en Establecimientos con Jornada Escolar Completa". Fondecyt N° 1060715.

Acomienzos de los años 90, el Gobierno de Chile impulsó una reforma de la educación con el fin de mejorar la calidad y equidad de los aprendizajes. Las políticas se orientaron, principalmente, a producir cambios en la cultura de la escuela, el currículum; la gestión directiva y las prácticas pedagógicas.

Entre los programas implementados destaca la Jornada Escolar Completa (JEC). Este se inició en 1997 con el objetivo de lograr que todos los establecimientos que reciben subvención pública aumenten sustantivamente los tiempos de los profesores y de los alumnos en situaciones de aprendizaje. Con esta política la jornada escolar aumenta las horas de clases en un 30% en la Enseñanza Básica y Media. Con ello, el país alcanza un total de 1.100 horas anuales cronológicas de docencia desde 3° a 8° básico y de 1.216 horas en Educación Media superando el promedio de horas de los países de la OECD quienes disponen para un programa curricular típico a las 14 años de 944 horas de instrucción (Brunner, JJ. *et al*, 2006; Valenzuela.J.P., 2006).

Desde 1997 se han incorporado gradualmente las escuelas y liceos a este régimen de jornada. Según cifras actualizadas para el año 2007 existen 7.322 establecimientos que se han incorporado a la Jornada Escolar extendida y que corresponde a más del 80% de los establecimientos del país.

Las evaluaciones realizadas por el propio Ministerio de Educación demuestran que existen resultados positivos en cuanto a la infraestructura, equipamiento; cobertura curricular y trabajo colaborativo de los profesores. Los padres, a su vez, valoran que sus hijos estén más tiempo en la escuela ya que con ello se evite el riesgo de "estar en la calle". (CIDE-PUC,2000, DESUC, 2001 y 2005).

Estos estudios también demuestran que no existen grandes cambios en la forma de organizar y gestionar el tiempo en las escuelas lo que afecta el logro de mejores aprendizajes. En efecto, la mayor parte del tiempo JEC se destina a las clases definidas en el plan común obligatorio sin cambios en el horario tradicionalmente establecido ni en las prácticas pedagógicas. El tiempo adicional o de libre disposición, en la mayor parte de los casos, se destina a la enseñanza de materias.

Durante el año 2006 se realizó una amplia discusión pública sobre el impacto que tiene la Jornada Escolar Extendida en la calidad de los procesos educativos y de los aprendizajes. Esta discusión relevó los problemas pedagógicos que tiene la implementación de la JEC y las dificultades que existen en la cultura escolar para organizar el tiempo y la pedagogía de modo más flexible e innovador.

Las evaluaciones de políticas se han centrado en los procesos de toma de decisiones o en los resultados e impactos de su aplicación. Existe poca preocupación por los procesos y particularmente por lo que piensan sus actores y aquellos que son los responsables de poner en práctica las políticas.

La implementación de la JEC es un buen ejemplo de la importancia que tienen en sus resultados los factores asociados al pensamiento de los profesores y a la cultura de la escuela. Con el fin de abordar la problemática cultural que subyace en los cambios propuestos por las políticas se analizará en este artículo la percepción que tienen los profesores que trabajan en el sistema público municipal sobre los efectos e impactos de este programa.

A partir de información recogida en el último estudio evaluativo del programa y en una investigación de casos realizada durante los años 2006 y 2007 se analizará hasta qué punto la representación que tienen los docentes sobre la realidad social de los niños incide en la apreciación del impacto que tiene este programa sobre la calidad y equidad de los aprendizajes.

1. LA TEORÍA DEL CAMBIO DEL PROGRAMA

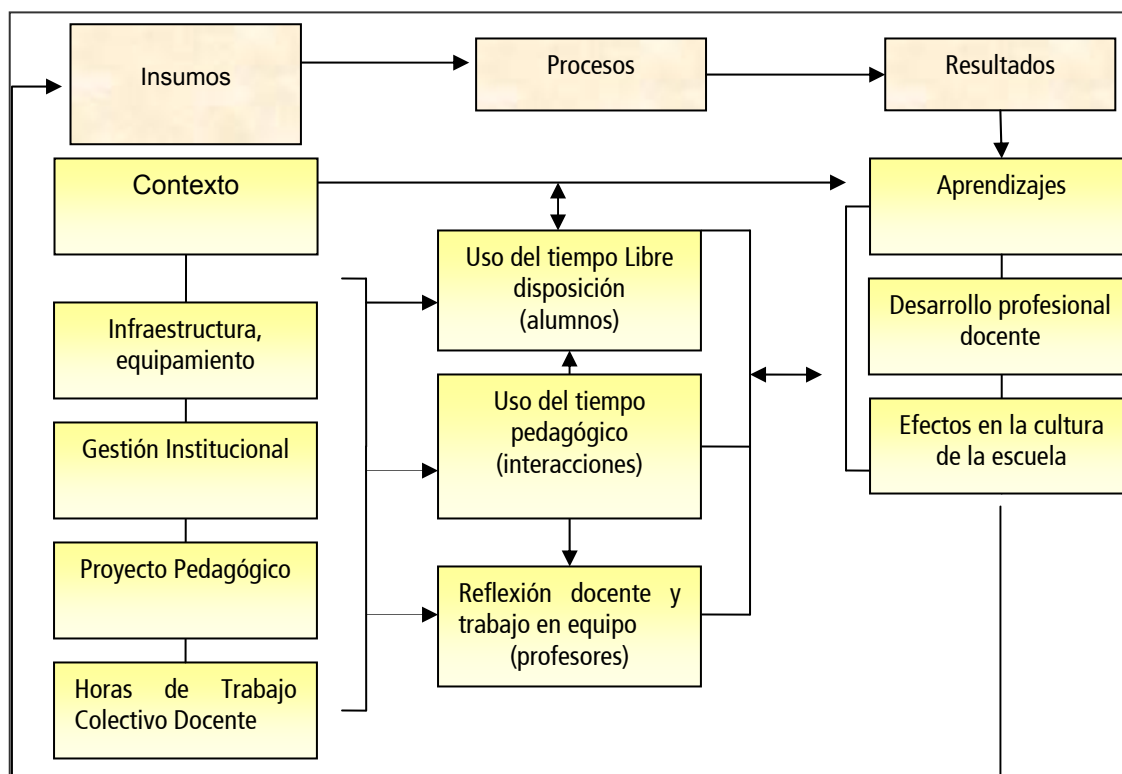
La Jornada Escolar Completa nace con el objetivo de ampliar y reestructurar el uso del tiempo en las escuelas para ponerlo al servicio del mejoramiento de los aprendizajes y de la innovación pedagógica en las escuelas subvencionadas. Así lo demostraba la experiencia internacional y la de los propios establecimientos particulares pagados en Chile (Cox, 2005). Con esta política la jornada escolar aumenta las horas de clases en un 30% en la Enseñanza Básica y Media. Con ello, el país alcanza un total de 1.100 horas anuales cronológicas de docencia desde 3° a 8° básico y de 1.216 horas en Educación Media superando el promedio de horas de los países de la OECD quienes disponen para un programa curricular típico a las 14 años de 944 horas de instrucción (Brunner, J.J. *et al.*, 2006; Valenzuela, J.P., 2006).

Para analizar la teoría de cambio de este programa hemos construido un modelo analítico que representa el núcleo de sus relaciones conceptuales e hipótesis de acción o de intervención.

Los modelos analíticos permiten explicitar la teoría y supuestos de los programas a evaluar; las relaciones hipotéticas y de causalidad que se establecen entre sus componentes y los efectos o cambios que se desean producir. Todo programa descansa en hipótesis que, de cumplirse, permiten lograr los cambios esperados. La evaluación es una verificación de esta hipótesis y de las relaciones establecidas entre sus variables o componentes (Weiss, C., 1992; Martinic, S., 1997; Mokate, K., 2000).

Para organizar la evaluación del Programa y visualizar la interacción de los distintos elementos o componentes de la estrategia del programa se presenta, a continuación, un modelo relacional que identifica insumos, procesos y resultados.

FIGURA 1. MODELO DE RELACIONES PROGRAMA JORNADA ESCOLAR COMPLETA



Los insumos aluden tanto al contexto en el cual se desarrolla el programa como a los recursos de entrada que aporta el sistema educativo. Los procesos dan cuenta del uso y efectos que tienen tales insumos en las prácticas e interacciones de profesores entre sí y con sus alumnos. Por último, los resultados aluden a los logros obtenidos y que se expresan en aprendizajes cognitivos y sociales, cambios en el desarrollo profesional de los docentes y en el clima y cultura de la escuela.

La hipótesis central de Programa es que más tiempo en la escuela significa mayor posibilidad de aprender¹. En otras palabras, se plantea que más tiempo de exposición al aprendizaje de los alumnos tendrá efectos en los resultados de aprendizaje. Para ello, además de la infraestructura y equipamiento, es necesario un proyecto pedagógico y gestión institucional que re-estructure uso del tiempo en la sala de clases y en el establecimiento y la organización de talleres y otras actividades que favorezcan el trabajo conjunto, la reflexión e intercambio de experiencias entre los docentes.

2. PRINCIPALES AVANCES

Desde 1997 se han incorporado gradualmente las escuelas y liceos a este régimen de jornada. Según cifras actualizadas para el año 2007 existen 7.322 establecimientos que se han incorporado y que corresponde a más del 80% de los establecimientos del país.

Las evaluaciones realizadas por el propio Ministerio de Educación demuestran importantes logros de la política (CIDE-PUC, 2000, DESUC, 2001 y 2005). Entre ellos, una importante inversión pública para el mejoramiento de la infraestructura y recursos educativos para los establecimientos; en la asistencialidad a los alumnos (desayunos, almuerzos), generación de equipos de docentes con más tiempo y trabajo conjunto en el establecimiento y mayor tiempo de los alumnos en el establecimiento dedicado al aprendizaje y en el desarrollo de actividades de libre disposición.

En cuanto al uso del tiempo este se clasifica en tres grandes categorías: (i) total de la jornada, (ii) tiempo no pedagógico (recreos y almuerzo) y (iii) tiempo pedagógico el que se divide en dos subcategorías: (a) plan de estudio, referido a los tiempos destinados a los sectores y subsectores curriculares y (b) tiempo de libre disposición y que se entiende como el tiempo que los establecimientos pueden ocupar para actividades definidas en función del proyecto pedagógico e intereses de los estudiantes.

Las evaluaciones realizadas hasta la fecha demuestran que, más del 80% de los establecimientos de Educación General básica y de Enseñanza Media respetan los tiempos totales de la jornada escolar estipulados en la normativa vigente. Lo mismo ocurre con los tiempos pedagógicos los que, en muchos casos, son superiores a la referencia normativa.

Las principales distancias con la norma se encuentran en los tiempos no pedagógicos y en el tiempo de libre disposición. En ambos casos se constata que la mayor parte de los establecimientos ocupan menos del tiempo asignado por la normativa.

¹ "Al aumentar de manera significativa los tiempos pedagógicos con el propósito de desarrollar mejor el nuevo marco curricular, esto repercute significativamente en los aprendizajes y se transforma en un cambio que favorece a los establecimientos más postergados, que pasan a tener mas tiempo, mejor infraestructura, en jornada única para desarrollar su Plan Educativo Institucional" (Mineduc. 2006)

Durante el año 2006 se realizó una amplia discusión pública sobre el impacto que tiene la Jornada Escolar Extendida (JEC) en la calidad de los procesos educativos y de los aprendizajes. Esta discusión relevó los problemas pedagógicos que tiene la implementación de la JEC y la necesidad de orientar sus procesos hacia un mayor impacto en la calidad de los resultados.

En respuesta a esta discusión el Ministerio de Educación realizó una consulta a los establecimientos JEC y el 97% de ellos planteó la necesidad de realizar cambios en el proyecto de Jornada Escolar vigente. Estos están centrados, principalmente, en las horas de libre disposición (97%); espacio escolar (38%) y cambios en los planes de estudios (30%). De este modo, para los actores que respondieron la consulta el principal problema a mejorar se relaciona con las horas de libre disposición las que se entiende, principalmente, como tiempo dedicado a la recreación (58%); deporte (69) o computación (58%). (MINEDUC, 2006).

3. EL PROBLEMA

Pese a la importancia de los cambios observados es evidente que no se ha producido una reestructuración del tiempo pedagógico tal como lo plantea la hipótesis central del Programa. El tiempo se sigue organizando de un modo rígido; el tiempo adicional es "extra" curricular y no se constata una mayor flexibilidad o integración del tiempo "normal" con las actividades del tiempo "extra". Se amplían materias básicas (refuerzos) y los talleres se definen como actividad adicional sin mayor relación con las materias básicas. No se ha reorganizado el tiempo ni tampoco se avanza hacia la flexibilidad del mismo.

Por otra parte, los impactos en el aprendizaje están lejos de las expectativas. Uno de los primeros estudios sobre esta relación es el realizado por el MINEDUC (2002) y que, utilizando los datos SIMCE 2001, estimó la importancia de distintos factores asociados al aprendizaje en lenguaje y matemáticas. En este se demuestra que los alumnos en establecimientos con JEC el efecto en lenguaje tienen un coeficiente de 1,6 y en matemáticas 1,8. ambos son estadísticamente significativos y ello significa que por cada año en JEC los alumnos obtienen dichos puntos en forma adicional².

Resultados similares se encuentran en estudios realizados por J.P.Valenzuela (2005) y C. Bellei (2006). Según J.P. Valenzuela el impacto de la JEC es mayor en lenguaje que en matemáticas y en los establecimientos particulares subvencionados que en los Municipales³ (Valenzuela, 2005).

Las explicaciones de estos resultados, considerados medianos y bajos en relación con la inversión y el impacto de otros programas, pueden atribuirse entre otros, a dos factores: (a) la teoría de base del programa, es decir, la relación que se ha establecido entre tiempo y aprendizaje y (b) el peso de los factores sociales externos y culturales de los propios profesores y de la escuela que inciden en el cambio pedagógico esperado.

En cuanto al primer factor, la investigación educativa –que ha sido importante en el tema en los últimos 20 años- demuestra que no hay una relación directa entre aumento de horas y aprendizaje. Los estudios

² Los cuartos años de Enseñanza Básica de los establecimientos Municipales en el año 2005 obtuvieron en lenguaje un promedio de 243 puntos y en matemáticas un promedio de 235.

³ Los municipales obtienen 0.1-0.2 SD en Lenguaje (4 puntos) y 0.0 en Matemáticas, mientras que las Subvencionadas logran 0.4 SD (7-8 puntos) y 0.2 SD (4-6 puntos), respectivamente.

concluyen que si no se actúa sobre el conjunto del sistema de variables que inciden en el aprendizaje, el efecto de los cambios en el tiempo de la jornada escolar es prácticamente nulo (Cotton, K., 1990; Greenwood, Ch. R 1991; Fuligni, A. J. y H. W. Stevenson, 1995).

El tiempo, por tanto, debe entenderse de un modo relativo y flexible. La ventaja que tiene el contar con mas tiempo para el aprendizaje no depende de su cantidad propiamente tal sino que del compromiso que tienen los estudiantes con la tarea; de la gestión del tiempo en la sala de clases y, muy fuertemente de las representaciones y expectativas que tengan los propios profesores sobre los aprendizajes de sus alumnos y la calidad de su trabajo pedagógico (Levin; Tsung, 1987; Delhaxhe, 1997).

En cuanto al segundo factor, y sin negar el efecto del capital cultural de las familias, diversos estudios destacan la importancia que tiene para la calidad las expectativas que los profesores tienen de los aprendizajes de sus alumnos. En efecto, se ha demostrado que los profesores que tienen bajas expectativas sobre las posibilidades de los estudiantes, tienen prácticas y orientaciones de acción que no estimulan a los niños a obtener mejores logros reproduciendo su clasificación o condición inicial (Gilly, M., 1989; Brophy, J., ed., 1998; Butti, F., 1998, 1999; Martinic, S., 2003; Román, M., 2003, 2007). Por otra parte, también se ha estudiado, que los profesores que apoyan a sus alumnos y creen en su buen desempeño logran alcanzar las metas que se han propuesto (Muñoz-Repiso, M. 2001; Bellei, C. *et al.*, 2003). Como señala M. Fullan (1991), el cambio escolar depende de lo que lo profesores hagan y piensen y por ello sus formas de pensar constituye una variable clave en los modelos de calidad, en la gestión y uso del tiempo en la sala de clases. Toda la tradición de escuelas efectivas demuestra la importancia de esta afirmación. (Scheerens, J. 1992; Murillo, J., 2003, 2005, 2007; Teddlie, C. y D. Reynolds, 2000).

Por lo anterior, el impacto de un programa como la JEC puede depender, fundamentalmente, de la representación que tienen los docentes de las capacidades y posibilidades de aprendizaje de los estudiantes. De este modo el logro de una política pública también es afectado por las representaciones y una cultura de la escuela que tiende a tener bajas expectativas de aprendizaje en los contextos más pobres.

4. MATERIAL Y MÉTODO

El análisis se realizará en base a información reunida en el marco de la evaluación de la Jornada Escolar Completa realizada el año 2005 por DESUC⁴, por encargo del Ministerio de Educación⁵.

Este estudio contó con representatividad nacional y trabajó con muestras representativas de establecimientos, directivos de dichos establecimientos (directores y jefes de unidades técnico-pedagógicas), docentes, padres y apoderados y alumnos. El diseño definió una muestra probabilística de 387 establecimientos de Enseñanza Básica y 340 establecimientos de Enseñanza Media con permanencia en JEC de a lo menos un año, representativas de todo el país. Igualmente, en cada uno de los establecimientos seleccionados para cada nivel, se encuestó al Director(a), al jefe de UTP y a 1 docente de aula⁶.

⁴ Dirección de Estudios Sociológicos de la Universidad Católica de Chile.

⁵ Los autores agradecen al Ministerio de Educación, División de Planificación y presupuesto y a DESUC la posibilidad de acceder a la base de datos señalada.

⁶ Seleccionado aleatoriamente del listado de profesores del grado correspondiente (5° básico ó 3° medio).

En el caso de los docentes, que es precisamente la base de datos que se analizó para efectuar los análisis que se presentan en este artículo, se obtuvieron estimaciones a nivel nacional con un error estimado de + 3.6%, esto bajo supuestos de varianza máxima de una proporción y muestreo aleatorio, y con un nivel de confianza del 95 %. El total de la muestra de docentes está compuesto por 706 docentes de ambos niveles.

Para el análisis se construyeron nuevas variables a partir de los ítemes del instrumento general aplicado a la muestra de profesores (ver ANEXO 1).

5. RESULTADOS

Como se ha señalado, la Jornada Escolar Completa, promueve la generación de mejores condiciones de tiempo y de espacio para el desarrollo de interacciones pedagógicas más efectivas y que mejoran los aprendizajes.

Nuestra hipótesis de trabajo es que no existe una relación directa entre aumento de la jornada escolar y cambio en las prácticas pedagógicas. Por el contrario, estas son mediadas por las representaciones y expectativas que tienen los docentes sobre los aprendizajes de sus alumnos. Las representaciones afectarán, a su vez, las opiniones que los docentes tienen sobre la JEC incidiendo estas en la efectividad de los cambios que pueda tener en los establecimientos.

Para abordar esta hipótesis es necesario, en primer lugar, analizar la opinión que tienen los profesores de los factores sociales del alumno y del establecimientos asociados al aprendizaje. En la tabla siguiente se observa la distribución obtenida.

TABLA 1. SEXO DE PROFESORES Y GRADO DE INFLUENCIA DE REALIDAD SOCIAL DEL ALUMNO Y DEL ESTABLECIMIENTO EN EL MEJORAMIENTO DE LOS APRENDIZAJES (EN %)

		Grado de influencia de realidad social de alumno			Grado de influencia del establecimiento			Total
		baja	media	alta	baja	media	alta	
Sexo	Hombre	10,0	45,3	44,7	47,4	40,8	11,8	287
	Mujer	13,2	40,5	46,2	52,0	39,6	8,4	369
Total		11,8	42,7	45,5	50,0	40,1	9,9	656

Como se aprecia en la tabla, la gran mayoría de los profesores piensa que los factores sociales asociados al alumno tienen una influencia *media* (42,7%) y *alta* (45,5%) en el aprendizaje de los alumnos. Esta valoración cambia al analizar los factores asociados al establecimiento. En efecto, en este caso la mayoría piensa que los factores del establecimiento tienen una influencia *baja* (50%) o *media* (40,1%). Es decir aspectos tales como práctica pedagógica; cantidad de tiempo de los profesores, material educativo, entre otros, no pueden producir cambios sustantivos en los aprendizajes porque el peso de la realidad social de los alumnos es mayor. Esta representación no tiene grandes diferencias según género de los profesores.

En síntesis, la mayoría de los profesores piensa que los problemas o deficiencias de aprendizaje de sus alumnos se explican en un grado medio o alto por los factores sociales y culturales asociados a la

realidad social y cultural de los estudiantes. Estos factores tienen más peso que aquellos que son propios del establecimiento. Este resultados es consistente con el de otros estudios y que demuestran que la mayoría de los profesores atribuyen, principalmente, los problemas de aprendizaje a los factores sociales que afectan a los niños. De este modo, se confirma lo que constatan otros estudios y que se relaciona con una representación donde se atribuye gran importancia a factores externos y que están fuera del control del profesor (Arancibia, V. *et al.*, 1991; Artiles, 1994; Butti, 1999, Martinic, S. 2003, 2006; Román, M., 2003).

Una vez verificada esta relación nos interesó analizar hasta qué punto un programa como la JEC puede romper con esta atribución y fortalecer la importancia de los factores internos del establecimiento y de la propia JEC en el mejoramiento de los aprendizajes. A continuación exploraremos estas relaciones a través de técnicas “no paramétricas” entre las distintas variables propuestas. Esto implicó realizar un análisis de Correspondencia Simple entre las percepciones de los profesores y sus discursos contruidos respecto a la JEC.

5.1. Percepción de la realidad social de los estudiantes

A) JEC y resultados de aprendizajes cognitivos

La primera hipótesis específica que nos planteamos es que cuando el profesor considera que los factores sociales asociados al alumno y su entorno influyen negativamente en el aprendizaje, perciben un impacto positivo de la JEC en las habilidades cognitivas del alumno.

Para explorar esta hipótesis se realizó, en primer lugar, un análisis de contingencia encontrándose una asociación entre las variables⁷; $\alpha = 0.05$, $\chi^2 (4, N = 693) = 25.96$, $p < 0.001$, $V = 0.14$. Luego se realizó un “Análisis de Correspondencia Simple” asociando las variables “Factores asociados al alumno y su entorno que influyen en el aprendizaje” y “Efectos cognitivos positivos sobre el alumno (conocimiento, aprendizaje y motivación)”.

El resultado obtenido se presenta en la siguiente tabla.

TABLA 2. EFECTOS COGNITIVOS DE LA JEC Y Y GRADO DE INFLUENCIA DE LA REALIDAD SOCIAL DE LOS NIÑOS EN EL APRENDIZAJE (DISTANCIAS DE CHI-CUADRADO)

Efectos cognitivos de la JEC	Factores sociales del alumno que afectan el aprendizaje		
	Baja	Media	Alta
Ninguno	-1,44	-0,06	1,79
Media	-5,24	-2,07	6,49
Alta	3,18	1,26	-3,93

En la tabla se observan las distancias, en chi-cuadrado, constatada en esta relación. Los valores positivos indican cercanía entre las variables relacionadas (es decir, a mayor el número positivo, mayor es la correspondencia) y las negativas lejanía.

Se observa que mientras mayor es el peso que se atribuye a los factores sociales del alumno menos es la influencia o efecto cognitivo de la JEC. Esto es, cuando los profesores consideran que los factores

⁷ Solo una celdilla (11,1%) posee valores esperados menores que 5.

asociados al alumno tienen un grado de influencia alta en el aprendizaje, perciben que la JEC no tiene efecto cognitivo y tiene un efecto mediano.. En cambio cuando creen que la realidad social del alumno *influyen en grado medio o Bajo* sobre el aprendizaje, consideran que la JEC afecta fuerte y positivamente dichas habilidades. De este modo, la hipótesis se demuestra solo parcialmente.

B) JEC y efecto en formación valórica y personal del alumno

La segunda hipótesis específica a explorar, propone que cuando el profesor considera que los factores asociados al alumno y su entorno influyen negativamente en el aprendizaje, perciben un impacto positivo de la JEC en la formación valórica y personal del alumno.

Como en el caso anterior, se realizó un análisis de contingencia encontrándose una asociación entre las variables⁸; $\alpha = 0.05$, $\chi^2 (4, N = 688) = 18.25$, $p < 0.001$, $V = 0.12$

Siguiendo el mismo procedimiento, se realizó un "Análisis de Correspondencia Simple" entre la dimensión "Factores asociados al alumno y su entorno que influyen en el aprendizaje" y "Efectos positivos de la JEC sobre la formación valórica y personal del alumno". El resultado obtenido se puede apreciar en la siguiente tabla de distancias.

TABLA 3. IMPACTO JEC EN VALORES Y FACTORES SOCIALES ASOCIADOS (DISTANCIAS DE CHI-CUADRADO)

Impacto de la JEC sobre aspectos Valóricos	Grado influencia realidad socia del alumno		
	Baja	Media	Alta
Ninguno	-0,22	-2,72	3,39
Medio	-3,93	-0,36	2,51
Alto	2,18	0,63	-2,31

En la tabla se observa un patrón similar a la anterior. Cuando el profesor considera que los factores asociados al alumno y su entorno influyen fuerte y negativamente en el aprendizaje, perciben que la JEC no tienen ningún impacto en la formación valórica y personal del alumno. Sin embargo, también se observa una correspondencia con la categoría "media", aunque en menor cuantía. Sólo cuando se percibe que los factores del entorno y del alumno tiene una influencia media o baja en el aprendizaje, se cree que la JEC tiene mucho más impacto sobre la formación valórica y personal de los alumnos. Esta evidencia no confirma en su totalidad la hipótesis propuesta.

C) Impacto JEC sobre las habilidades del alumno

Finalmente, la tercera hipótesis específica indica que cuando el profesor considera que los factores asociados al alumno y su entorno influyen fuerte y negativamente en el aprendizaje, perciben un impacto positivo de la JEC sobre el desarrollo de habilidades en el alumno.

Siguiendo el mismo procedimiento, se realizó un análisis de contingencia encontrándose una asociación entre las variables⁹; $\alpha = 0.05$, $\chi^2 (4, N = 687) = 9.48$, $p = 0.05$, $V = 0.08$ (la asociación es mucho más débil). De igual forma, se aplicó un "Análisis de Correspondencia Simple" entre la dimensión "Factores asociados al alumno y su entorno que influyen en el aprendizaje" y "Efectos positivos de la JEC sobre el desarrollo de habilidades en el alumno". En la tabla siguiente se muestra el resultado.

⁸ Solo una celdilla (11,1%) posee valores esperados menores que 5.

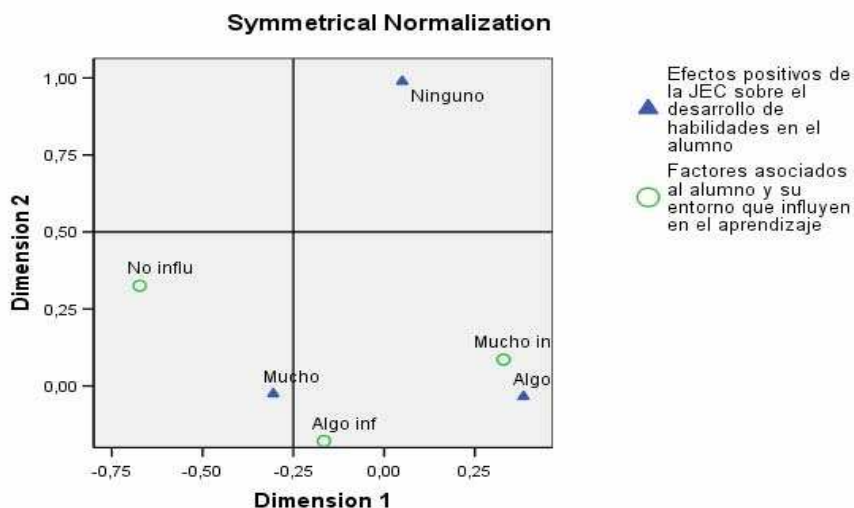
⁹ Solo una celdilla (11,1%) posee valores esperados menores que 5.

TABLA 4. EFECTO JEC EN HABILIDADES ALUMNO Y FACTORES SOCIALES ASOCIADOS (DISTANCIAS DE CHI-CUADRADO)

Desarrollo de habilidades en el alumno	Grado influencia realidad socia del alumno		
	Baja	Baja	Baja
Ninguno	0,19	-0,29	0,09
Medio	-2,49	-0,42	2,03
Alto	1,69	0,49	-1,79

La tabla muestra un patrón similar al observado en las relaciones anteriores. Mientras más alta es la influencia de los factores sociales asociados al alumno menor es el impacto de la JEC sobre el desarrollo de sus habilidades. Esta correspondencia se observa en el siguiente diagrama de dispersión.

FIGURA 2. DIAGRAMA DE DISPERSIÓN BIESPACIAL EN ANÁLISIS DE CORRESPONDENCIA PARA EFECTOS POSITIVOS DE LA JEC SOBRE DESARROLLO DE HABILIDADES EN EL ALUMNO



En el diagrama se observa que cuando los profesores opinan que los factores sociales asociados a los alumnos influyen fuertemente sobre su aprendizaje, consideran que la JEC impacta sólo *en algo* el desarrollo de las habilidades en los alumnos. En cambio, cuando creen que dichos factores influyen en algo o nada el aprendizaje de los alumnos, perciben que la JEC *impacta mucho* en el desarrollo de habilidades de los alumnos. De este modo tampoco hay evidencia suficiente para apoyar la hipótesis alternativa específica.

5.2. Percepción calidad del establecimiento y aprendizaje cognitivo

Una primera hipótesis específica respecto a esta dimensión, es que cuando el profesor considera que los factores asociados al establecimiento (recursos y cuerpo docente) influyen negativamente en el aprendizaje, percibirán que la JEC no impactará en los aprendizajes cognitivos del alumno. Esta

influencia negativa es producto de las carencias que perciben los profesores en el establecimiento en cuanto a recursos; cantidad de profesores; tiempo disponible y calidad de las practicas pedagógicas. Es probable que los profesores consideren que si no se mejoran las condiciones estructurales tanto del establecimiento como el quehacer docente –que constituye la materia prima de la implementación de las políticas- la JEC no tendrá impacto alguno.

Para probar esta hipótesis, al igual que los procedimientos anteriores, se realizó un análisis de contingencia encontrándose una asociación entre las variables¹⁰; $\alpha = 0.05$, $\chi^2 (4, N= 653) = 47.07$, $p < 0.001$, $V = 0.19$ Posteriormente se realizó un “Análisis de Correspondencia Simple” entre la dimensión “Factores asociados al establecimiento que influyen en el aprendizaje” y “Efectos cognitivos positivos sobre el alumno (conocimiento, aprendizaje y motivación)”. El resultado observado es el siguiente.

TABLA 4. EFECTOS COGNITIVOS DE LA JEC Y FACTORES ASOCIADOS AL ESTABLECIMIENTO (DISTANCIAS DE CHI-CUADRADO)

Efectos cognitivos de la JEC	Grado influencia de factores negativos del establecimiento		
	baja	Media	Alta
Ninguno	-6,18	0,61	15,89
Medio	-5,42	2,35	4,50
Alto	4,78	-1,51	-5,83

En la tabla se observa que, cuando los profesores opinan que los factores negativos asociados al establecimiento tienen una alta influencia, consideran que el efecto en los aprendizajes cognitivos de la JEC es prácticamente nulo. Por otra parte, y en un sentido inverso, cuando los docentes piensan que la influencia negativa de la realidad del establecimiento es media o baja el efecto en los aprendizajes cognitivos de la JEC es medio y alto. Esto significa que, para los profesores, el impacto de la JEC requiere de una realidad interna positiva en cuanto a la realidad de los establecimientos, tiempos, prácticas y calidad de los profesores.

6. CONCLUSIONES

En el análisis realizado se constata que, para gran parte de los profesores, el aprendizaje de sus estudiantes depende fuertemente de su realidad social y familiar y de factores asociados a la calidad de los establecimientos.

La atribución a factores externos que inciden en el aprendizaje es tan fuerte que, en la representación de los profesores, la JEC no puede producir grandes cambios en esta determinación social y externa a la calidad de los proceso de aprendizajes. Por el contrario, cuando disminuye el peso externo, para los profesores, la JEC tiene mayor impacto en dichos aprendizajes.

Los resultados confirman las conclusiones de otros estudios y que demuestran que los profesores perciben que la realidad social y cultural de sus alumnos tiene un efecto negativo sobre sus aprendizajes. Perciben

¹⁰ Solo una celdilla (11,1%) posee valores esperados menores que 5.

una realidad social familiar negativa lo que se expresa en indicadores tales como: baja escolaridad del grupo familiar, mal ambiente sociocultural del entorno, mala situación económica, falta de compromiso y apoyo de la familia a la escuela, problemas de salud, insuficiente dedicación al estudio, problemas conductuales e inseguridad en el entorno local. En este sentido, consideran que en "algo" la JEC podría venir a revertir alguno de estos problemas. Sin embargo, la situación estructural, al parecer, tendría mayor peso.

Respecto a la segunda constatación -las condiciones del establecimiento- se demuestra una relación similar. Mientras mas negativa es la realidad del establecimiento menores posibilidades tiene la JEC de impactar en la calidad de resultados en los estudiantes.

De este modo, si se considera que las prácticas pedagógicas son inadecuadas, que a los profesores les falta el tiempo, carencia de docentes de mayor calidad técnica, insuficiente material educativo, falta de incentivos a los profesores, problemas de gestión y administración del establecimiento, deficiente infraestructura y/o equipamiento, insuficiencia de recursos humanos, muy poca capacitación a los profesores y falta de motivación de los docentes entonces la JEC no tendría ningún impacto sobre el desarrollo de habilidades en los alumnos, formación valórica y personal y mejora de las capacidades cognitivas.

El análisis del problema, por lo general, se ha reducido a las expectativas y aspiraciones que los profesores tienen sobre el futuro de sus alumnos. Desde esta perspectiva rápidamente se concluye que las causas radican en el pensamiento y en las actitudes del profesor como un individuo y actor particular. De ello se desprende que a partir de su propio convencimiento y motivación es posible cambiar la situación.

Desde el punto de vista de las representaciones sociales el problema es más complejo. En efecto, el discurso y las prácticas de los profesores sobre los aprendizajes de sus alumnos se producen desde un sistema de referencia construido socialmente y que se hace evidente y normal en los contextos institucionales y sociales en los cuales se desempeñan.

Para Miskel, McDonald y Bloom (1983) los individuos adaptan sus comportamientos, actitudes y creencias al contexto social en que se desenvuelven. Desde este punto de vista, el clima de la escuela y las ideas que predominan en las prácticas que se realizan en la escuela y en el sistema orientan la construcción de significados y creencias socialmente deseables al mismo tiempo que ofrecen razones aceptables para explicar los problemas y lo no deseable que se observan en las acciones cotidianas de profesores y alumnos" (Báez de la Fe, E., 1994:107). Los docentes se desenvuelven en la sala de clases de acuerdo a las representaciones sociales o al conocimiento práctico y que está estrechamente relacionada con la reflexividad y autonomía de su trabajo (Schön, D., 1992, Zeichner, K.M. 1994, Avalos, B., 1994).

Por otra parte, las opiniones que construyen los docentes se hacen efectivas a partir del recurso a múltiples fuentes de información. Algunas son las observaciones directas (las notas, el comportamiento oral del alumno, las entrevistas que tienen), otras son informaciones de segunda mano: opinión de otros profesores, libreta de notas de años anteriores, antecedentes escolares en general, comentarios de la dirección, de los consejeros psicológicos, de los padres" (Coulon, 1993:108-109). Así entonces, la información que leen, los datos que circulan; el tipo de reuniones que se realizan configuran relaciones, modo de circulación de ideas y de información que contribuyen la construcción y circulación de representaciones que orientarán las prácticas de los sujetos.

De este modo, es posible afirmar, que se ha construido un tipo de representación de la realidad social y de las escuelas que atienden a alumnos de los sectores pobres, que incide en la interpretación y en las expectativas que tienen los profesores sobre las políticas y programas que se focalizan en el sector. Mientras no existan cambios en el núcleo duro de esta representación será difícil observar cambios sustantivos y efectivos en las interacciones y prácticas pedagógicas.

Como se señaló anteriormente, la ampliación y mejor uso del tiempo escolar es uno de los cambios más radicales que afectan la cultura y gestión pedagógica de las escuelas. En los conceptos de tiempo y de espacio descansan los principios que separan las materias y regulan las relaciones entre los profesores y de estos con sus estudiantes.

Para llevar a cabo todos estos cambios y las transacciones que ellos implican es necesario, al menos, dos condiciones. En primer lugar, confianzas mutuas entre los actores involucrados y el Director para realizar las coordinaciones y el ajuste requerido y, en segundo lugar, una interpretación compartida de la comunidad educativa sobre los problemas, las estrategias de acción y los objetivos de las medidas que se toman para llevar a cabo los cambios. Para ello los establecimientos deben preparar un proyecto que justifique pedagógicamente su propuesta de organización del tiempo.

Para lograr estas confianzas e interpretaciones compartidas es muy importante la participación de los distintos actores en el diseño mismo del proyecto de Jornada Escolar.

REFERENCIAS BIBLIOGRÁFICAS

- Arancibia, V. y Álvarez, M.I. (1991). *Modelo de variables del profesor y su impacto en rendimiento escolar y autoconcepto académico*. Proyecto Fondecyt N° 1880405. Santiago de Chile: Pontificia Universidad Católica de Chile.
- Artiles, A. (1994). La evaluación de los procesos de pensamiento de los maestros en contextos urbanos un estudio de caso en escuelas primarias de Guatemala. *Ponencia presentada en la Conferencia Anual de la Comparative and International Education Society*, San Diego, California.
- Avalos, B. (1994). Creatividad versus autonomía profesional del profesor. Consideraciones sobre el tema, derivadas de la investigación pedagógica. *Pensamiento Educativo*, 14, pp. 13-50.
- Baez de la Fe, E. (1994). El movimiento de escuelas eficaces: implicaciones para la innovación educativa. *Revista Iberoamericana de Educación*, 4, pp. 93-116.
- Bellei, C., Muñoz, G., Raczynski, D. y Pérez, L.M. (2003). Escuelas Efectivas en Sectores de Pobreza: ¿Quién dijo que no se puede? En: Hevia, R. *La Educación en Chile Hoy*. Santiago de Chile: Ediciones Universidad Diego Portales.
- Bellei, C. (2006). Does Lengthening The School Day Increase Students' Academic Achievement? Results From A Natural Experiment in Chile. Harvard Graduate School. Qualifying Paper (m.s.)
- Brophy, J. (Ed.). (1998). *Advances in Research on Teaching: Expectations in the Classroom* (Vol. 7). Greenwich, Conn: JAI Press, Inc.
- Brunner, J.J. y G. Elacqua (2006). *Calidad de la educación. Claves para el debate*. Santiago de Chile: Ril editores.
- Butti, F. (s.d.) Estudio del éxito y el fracaso escolar. Las representaciones sociales como mediatizador de la práctica docente. En: *Reunión de Comunicaciones científicas y tecnológicas 1998*, UNNE. Publicación de la Sec. Gral de Ciencia y Técnica. Tomo I : Ciencias sociales-Ciencias humanísticas. Edit. Chaco: Universitaria de la UNNE.
- Butti, F. (s.d). Representaciones Sociales de los Maestros y Fracaso Escolar. Hacia una propuesta metodológica. En:

- Reunión de Comunicaciones científicas y tecnológicas 1999, UNNE.* Publicación de la Sec. Gral de Ciencia y Técnica. Tomo I: Ciencias Sociales. Chaco: Edit. Universitaria de la UNNE.
- CIDE-PUC (2000). *Efectos de la Jornada Escolar Completa en el uso del tiempo de profesores y estudiantes en establecimientos básicos y medios.* Santiago de Chile: CIDE-PUC.
- Cotton, K. (1990). *Educational time factors.* Portland: Northwest Regional Educational, Laboratory.
- Coulon, A. (s.d.). *Ethnométhodologie et éducation.* Paris: PUF.
- Delhaxhe, A. (1997). Le temps comme unité d'analyse dans la recherche sur l'enseignement. *Revue Française de Pédagogie*, 118.
- DESUC, Dirección de Estudios Sociológicos P.U.C.CH. (2001). *Informe final estudio Evaluación Jornada Escolar Completa.* Santiago de Chile: PUC.
- DESUC, Dirección de Estudios Sociológicos P.U.C.CH. (2005) *Informe final estudio Evaluación Jornada Escolar Completa.* Santiago de Chile: PUC.
- Fulgini, A.J. y Stevenson H.W. (1995). Time Use and Mathematics Achievement among American, Chinese, and Japanese High Schools Students. *Child Development*, 66(3).
- Fullan, M. (1991). *The new meaning of educational change.* Londres: Casell.
- Gilly, M. (1989). Les représentations sociales dans le champ éducatif". En : Jodelet, D. (ed) *Les représentations Sociales.* Paris: PUF.
- Greenwood, Ch.R. (1991). Longitudinal Analysis of Time, Engagement, and Achievement in At-Risk Versus Non-Risk Students. *Exceptional Children*, 57(6).
- Levin, H. y M. Tsang (1987). The Economics of Student Time. *Economics of Education Review*, 6(4).
- Martinic, S. (1997). *Diseño y evaluación de proyectos sociales.* México: Comexani/Cejuv.
- Martinic, S. (2003). Representaciones de la desigualdad en la cultura escolar. *Revista Persona y Sociedad*, XVII(1), pp. 129-146.
- Martinic, S. (2006). El estudio de las representaciones y el análisis estructural del discurso. En: Canales, M. (ed) *Metodologías de investigación social.* Santiago: Lom.
- MINEDUC (2002). División de Planificación y Presupuesto. *Prueba SIMCE 2º medio 2001. Análisis de resultados.* Santiago de Chile: MINEDUC.
- MINEDUC (2006). El reloj. Sugerencias para la Formulación del Proyecto Pedagógico. Santiago de Chile: MINEDUC.
- Miskel, C, MacDonald, D., y Bloom, S. (1983). Structural and expectancy linkages within schools and organizational effectiveness. *Educational Administration Quarterly*, 19, pp. 49-82.
- Mokate, K.M. (2000). El monitoreo y la evaluación: herramientas indispensables de la gerencia social. Washington: INDES.
- Muñoz-Repiso, M. (2001). Lecciones aprendidas para el sistema español. *Cuadernos de Pedagogía*, 300, pp. 66-73.
- Murillo, F.J. (coord.) (2003). *La investigación sobre Eficacia Escolar en Iberoamérica. Revisión internacional del estado del arte.* Bogotá: Convenio Andrés Bello.
- Murillo, F.J. (coord.) (2004). La mejora de la escuela. *Cuadernos de Pedagogía*, 339, pp. 48-79.
- Murillo, F.J. (2005). La Investigación en Eficacia Escolar y Mejora de la Escuela como motor para el incremento de la Calidad Educativa en Iberoamérica. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 3(2).
- Murillo, F.J. (2007). *Investigación Iberoamericana sobre Eficacia Escolar.* Bogotá: Convenio Andrés Bello.

- Román, M. (2003). ¿Porqué los docentes no pueden desarrollar procesos de enseñanza aprendizaje de calidad en contextos sociales vulnerables? *Revista Persona y Sociedad*, 17(1), pp. 113-128.
- Román, M. (2007). *Investigación Latinoamericana sobre Enseñanza Eficaz*. Ponencia presentada en el Primer Congreso Iberoamericano de Eficacia Escolar y Factores Asociados. Santiago de Chile: UNESCO-MINEDUC.
- Scheerens, J. y Bosker, R.J. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.
- Schön, Donald A. (1992). *La formación de profesionales reflexivos. Hacia un nuevo diseño de la enseñanza y el aprendizaje en las profesiones*. Barcelona: Ediciones Paidós.
- Teddlie, C. y Reynolds, D. (Eds.) (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Press.
- Valenzuela, J.P. (2005) *Evaluación Parcial de una Gran Reforma en el Sistema Educativo Chileno: Desde la Media Jornada a la Jornada Escolar Completa (JEC)*. Santiago de Chile: Universidad de Chile.
- Weiss, C. (1992). *Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción*. México: Editorial Trillas.
- Zeichner, K.M. (1993). El maestro como profesional reflexivo. *Revista Cuadernos de Pedagogía*, 220, pp. 44-49.

ANEXO 1. DEFINICIÓN DE VARIABLES

Las variables construidas de percepción de resultados de la JEC son las siguientes:

1. "Efectos cognitivos positivos de la JEC sobre el alumno (conocimiento, aprendizaje y motivación)" se obtuvo a partir de 6 ítems que medían esta dimensión (Alfa de Cronbach's $\alpha = 0.84$), que surge de la pregunta ¿Cuánto está influyendo la JEC en el mejoramiento de los siguientes ámbitos de desarrollo de los alumnos de este establecimiento?, evaluando como "Nada (=1)", "Algo (=2)" o "Mucho (=3)". Las opciones eran "conocimientos y habilidades en lenguaje", "conocimientos y habilidades en matemáticas", "conocimientos y habilidades en ciencias", "el dominio de idioma extranjero", "el aprendizaje de los alumnos" y "La motivación y el entusiasmo de los alumnos en la tarea escolar".
2. "Efectos positivos de la JEC sobre la formación valórica y personal del alumno" por medio de 3 ítems (Alfa de Cronbach's $\alpha = 0.84$) y bajo la misma pregunta y escala anterior pero con las opciones "La formación valórica", "La capacidad de superación personal" y "la capacidad de usar bien el tiempo libre".
3. "Efectos positivos de la JEC sobre el desarrollo de habilidades en el alumno", vía 4 ítems con similar confiabilidad (Alfa de Cronbach's $\alpha = 0.87$) e igual pregunta y escala previa, presentaba las opciones "el desarrollo de habilidades deportivas", "el desarrollo de habilidades artísticas", "el manejo de tecnología y computación" y "preparación para el trabajo".

Por otra parte las variables construidas de factores asociados a la realidad social del alumno y establecimiento son las siguientes:

1. "Factores asociados al alumno y su entorno que influyen en el aprendizaje". Índice construido a partir de la percepción de profesores sobre la pregunta ¿Cuánto influye en los aprendizajes de los alumnos familias con baja escolaridad, mal ambiente sociocultural del entorno, mala situación económica, falta de compromiso y apoyo de la familia, problemas de salud, baja capacidad de los alumnos, insuficiente dedicación al estudio, problemas conductuales e inseguridad en el entorno local? Cada ítem fue evaluado como "Baja o Nada (=1)", "Media o Algo (=2)" o "Alta o Mucho (=3)":
2. "Factores asociados al establecimiento que influyen en el aprendizaje" construido a partir la pregunta anterior, ¿Cuánto influye en los aprendizajes de los alumnos prácticas pedagógicas inadecuadas, profesores con falta de tiempo, carencia de docentes de mayor calidad técnica, insuficiente material educativo, falta de incentivos a los profesores, problemas de gestión y administración del establecimiento, deficiente infraestructura y/o equipamiento, insuficiencia de recursos humanos, muy poca capacitación a los profesores y falta de motivación de los docentes? Igualmente, cada ítem fue evaluado como "Baja o Nada (=1)", "Media o Algo (=2)" o "Alta o Mucho (=3)":

Estas últimas variables se construyeron por medio de un Análisis Factorial Exploratorio (AFE) alcanzando, de esta manera, un Alfa de Cronbach's $\alpha = 0.91$ y de $\alpha = 0.73$ respectivamente para cada factor. Una vez obtenido estas variables se procedió a aplicar pruebas de asociación entre las variables (Chi-cuadrado, χ^2) y luego "Análisis de Correspondencia Simple"¹¹.

¹¹El análisis de correspondencia (CA) es una técnica de interdependencia y de composición debido a que el mapa perceptual que se construye se basa en la asociación entre objetos y un conjunto de características descriptivas o atributos especificados por el investigador. En su forma más básica, el análisis de correspondencias emplea una tabla de contingencia, que es la tabulación cruzada de dos variables categóricas. A continuación transforma los datos no métricos en un nivel métrico y realiza una reducción dimensional (similar al análisis factorial) y un mapa perceptual (similar al análisis multidimensional) (Hair, Anderson, Tatham & Black, 2001). Esta técnica es la más adecuada puesto que el estudio está basado en variables categóricas y discretas. Por ello, en un sentido genérico, "correspondencia es sinónimo de conexión y de relación recíproca entre dos elementos o conjuntos equivalentes" (Cornejo, 1988).



DOCENTES DE EDUCACIÓN BÁSICA Y SUS CONCEPCIONES ACERCA DE LA EVALUACIÓN EN MATEMÁTICA

Iraima Moreno y José Ortiz

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art10.pdf>



Todo currículo admite diversos niveles de reflexión y en el nivel que está más ligado a la práctica del profesor se concreta en cuatro dimensiones: objetivos, contenidos, metodología y evaluación, es en esta última dimensión que se centra el presente trabajo, haciendo énfasis en el área de Matemática, puesto que según Webb (1992), la evaluación en Matemática involucra aprendizajes, enseñanza, acción docente, currículo y aspectos institucionales, entre otros.

En este sentido, se requiere conocer mejor el pensamiento o concepto que los profesores poseen sobre evaluación en Matemática, ya que con la evaluación se logra una reflexión sobre la práctica docente y al mismo tiempo se adquiere a través de ella, conocimientos no solo respecto al alumno y su proceso de aprendizaje, sino también sobre las características de la asignatura y de la tarea de los docentes como facilitadores del aprendizaje (Gil, 1999).

Asimismo, la evaluación enfatiza la información que mejore la práctica docente, incidiendo directamente sobre el proceso enseñanza y aprendizaje, las tareas, los materiales, la organización, planificación, etc. (Giménez, 1997). En este sentido, la evaluación debería ser algo más que un examen al final del período de enseñanza, la misma debería constituir una parte integral de la enseñanza que informe al profesorado y le sirva de guía para la toma de decisiones, es decir, no hacer evaluación sólo a los alumnos, sino también para los alumnos, de manera que se pueda guiar y mejorar su aprendizaje. (NCTM, 2000).

Desde esta perspectiva, el trabajo tiene como propósito analizar las concepciones de los profesores respecto a evaluación en Matemática.

1. EL PROBLEMA

La evaluación es uno de los aspectos primordiales del trabajo del profesor, puesto que ha de llevarse a cabo de diversas formas y sirve para diversos fines (Cockcroft, 1985; Jimeno, 2006; Moya, 2001; Ruthven, 2003).

En este sentido, Gil (1999), considera que en la evaluación se *"concreta una serie de reglas sociales de validación que tienen que ver con las disciplinas del conocimiento, la forma en que esas prácticas se concretan en la escuela y los fines de socialización que sirven de meta para la actividad escolar"* (p.25)

Por ello, dada la amplitud y los aspectos que abarca el proceso de evaluación, en este estudio se hará referencia a la evaluación dentro del proceso de enseñanza-aprendizaje, particularmente en el área de Matemática, analizando las concepciones que sobre evaluación poseen los profesores de esta asignatura, pues Gil (1999), considera que es importante conocer qué tipo de evaluación se lleva a cabo y cuál es el sistema de ideas sobre el que se sostiene; esto implica entender la evaluación como parte integral del currículo (Moya, 2001). Lo que permite deducir, que la evaluación en Matemática debe ser efectuada tomando en cuenta tanto el aspecto cualitativo como el cuantitativo, buscando así que se valore y comprenda las consideraciones, interpretaciones, interés y aspiraciones de quienes actúan en el proceso de enseñanza-aprendizaje, a fin de ofrecer información pertinente y oportuna a cada uno de estos, la cual les permitiría observar y orientar el proceso, reconocer y analizar la eficacia y planificar mejor (Flores, 2001).

Asimismo, Calderón y Deiros (2003), señalan que algunas de las insuficiencias que pueden presentar los exámenes elaborados por los profesores de Matemática son: *"...se exige poco, en evaluaciones escritas*

que los estudiantes expresen sus ideas utilizando el lenguaje matemático. La evaluación generalmente se limita a la reproducción de los procedimientos desarrollados por el profesor..." (p.332).

En este sentido, la evaluación tal como lo dice Gil (1999), se suele identificar con examen, prueba terminal y calificación, lo cual repercute en los alumnos puesto que éstos también identifican el término evaluación con examen, promoción y control.

Por su parte, los estándares del National Council of Teachers of Mathematics (NCTM, 2000), especifican que los profesores necesitan tener una idea clara de lo que se debe enseñar y aprender, de tal manera que la evaluación garantice que cada alumno avance productivamente en la dirección apropiada.

De allí pues, que la importancia de este trabajo radica en que además de conocer las concepciones que poseen los profesores respecto a evaluación en Matemática, las mismas serán caracterizadas y contrastadas con lo establecido en las normativas legales vigentes en Venezuela, lo cual permitirá constatar si esas concepciones llevan al docente a desarrollar prácticas evaluativas cónsonas con las que al respecto señala la normativa de Educación Básica en la actualidad.

Respecto a la normativa legal vigente sobre evaluación en Venezuela, la evaluación es una parte esencial del proceso educativo ya que permite determinar en qué medida se logran los objetivos educativos (Ministerio de Educación, 1987). De allí que la evaluación tal como lo establece el artículo 63 de la Ley Orgánica de Educación (1999) y el artículo 89 del Reglamento General de la Ley Orgánica de Educación, debe ser continua, integral y cooperativa.

Se conjetura, en este trabajo, que los profesores tienen concepciones diferenciadas sobre la evaluación en Matemática; y que las concepciones que poseen los docentes respecto a evaluación en Matemática, son congruentes con lo establecido en las normativas legales vigentes.

El objetivo general de la investigación consiste en analizar las concepciones de los profesores respecto a la evaluación en Matemática; para lo cual se persigue: identificar las concepciones que poseen los profesores respecto a evaluación en Matemática; caracterizar las concepciones de los profesores respecto a evaluación en Matemática; e identificar posibles articulaciones entre las concepciones de los profesores respecto a evaluación en Matemática y lo establecido en las normativas legales vigentes.

2. METODOLOGÍA

Se recurrió a la complementariedad de técnicas cualitativas, pues según recomendaciones de Valles (1999), "el uso combinado de distintos procedimientos metódicos y técnicos es con el fin común de mejorar la investigación social en la práctica" (p. 202). De hecho, en el campo educativo la complementariedad constituye una vía para lograr explicar situaciones concretas (Ortiz, 2002). Se acude al *estudio de caso* como un abordaje que permite la incorporación de varias técnicas combinadas según la naturaleza del estudio (Stake, 2005; Yin, 2003). Se recurre al *grupo de discusión* como una técnica, que según Canales y Peinado (1995), es comúnmente usada en los estudios cualitativos. Esta técnica favoreció la creación de un espacio de "opinión grupal" que permitió la verificación de opiniones pertinentes, adecuadas o válidas en torno al tema de la evaluación en Matemática.

Para los instrumentos se consideró un instrumento elaborado por Gil (1999) que permite estudiar las concepciones y creencias que, sobre la evaluación, tienen los profesores de Matemática en ejercicio, éste

servió de guía para la elaboración del guión orientador de las conversaciones, tanto en los grupos de discusión como la entrevista en profundidad.

Con base en este instrumento, se consideró pertinente elaborar un guión que constaba en un principio de trece temas, que luego de ser aplicado al primer grupo de discusión se redujo a diez temas, dado que algunas eran muy similares. Las mismas fueron organizadas de acuerdo a las siguientes dimensiones: evaluación de objetivos, evaluación de contenidos, evaluación docente-alumno, evaluación de proceso-producto, evaluación para la toma de decisiones.

Los sujetos del estudio fueron siete profesores de matemática de la tercera etapa de Educación Básica (7º, 8º y 9º grados) que laboran en la Unidad Educativa Amantina de Sucre y el Liceo Francisco Lazo Martí, ubicados en el Municipio Biruaca y San Fernando, del Estado Apure, Venezuela. Los mismos participaron tanto en los grupos de discusión como en las entrevistas a profundidad en forma voluntaria.

La investigación fue estructurada en tres fases: la primera fase, está referida a la sensibilización de los docentes participantes en el estudio, para lo cual se realizó la presentación del proyecto y sus objetivos. La segunda fase, corresponde a la realización de dos grupos de discusión, cada uno con tres profesores de Matemática de la tercera etapa de Educación Básica, tomando como criterio de selección el grado en el cual se desempeñan los mismos, así el grupo 1, lo formaron 3 profesores de 7º grado y 8º grado y el otro grupo 8º grado y 9º grado. En la tercera fase, se aplicó una entrevista en profundidad, en la cual se consideró la información resultante de la segunda fase, con tres profesores de los que participaron en los grupos de discusión y un profesor que no asistió a estos.

La discusión estuvo orientada por una serie de temas previamente establecidos por los investigadores. Con ello se persiguió la emergencia de aspectos novedosos respecto a la concepción de los profesores sobre la evaluación en matemática. Cada grupo de discusión se inició con una presentación donde se expuso el objetivo de la reunión, luego se procedió a la presentación de cada uno de los profesores quienes, a pesar que trabajaban en grados diferentes, se conocían entre sí. El uso de esta técnica, permitió revelar la opinión de los docentes en cuanto a la concepción de la evaluación en Matemática. El primer grupo de discusión, lo conformaron tres profesores de Matemática, de 7º y 8º grado de Educación Básica, cuyos años de experiencia docente oscilaban entre uno y diez años. La reunión tuvo una duración aproximada de hora y media. Los profesores pertenecían a la Unidad Educativa "Amantina de Sucre", ubicada en el Municipio Biruaca, Estado Apure; todos de sexo masculino y se utilizó como medio de registro material audiovisual: un grabador y una filmadora. En la transcripción de este grupo, la intervención de los profesores se antecede con las letras R, I y C, para respetar el anonimato de los mismos. El segundo grupo de discusión, estuvo conformado por tres profesoras de Matemática de 8º y 9º grado de Educación Básica, cuyos años de experiencia oscilaban entre uno y seis años. La reunión se desarrolló con docentes que laboran en la Unidad Educativa "Lazo Martí", ubicada en el Municipio San Fernando, Estado Apure. Se utilizó como medio de registro un grabador. En la transcripción, la intervención de los profesores se antecede con las letras A, Z y M, respectivamente.

La entrevistas en profundidad, favorecieron la obtención de información valiosa a través de los procesos comunicativos (Valles, 1999), es decir, permitió la captación de una gran riqueza informativa. Para la aplicación de la entrevista, se elaboró un guión de dimensiones o temas, en función de los cuales giró la conversación, con la finalidad de contrastar y complementar la información obtenida en los grupos de discusión respecto a evaluación en matemática. Las cuestiones formuladas estuvieron focalizadas en las preguntas siguientes: 1. ¿Qué es evaluar?, 2. ¿Para qué evaluar a los alumnos en Matemática?, 3. ¿Qué

evaluamos en Matemática?, 4. ¿Cómo evaluar a los alumnos en Matemática?, 5. ¿Quién debe evaluar a los alumnos en Matemática?, 6. ¿Qué instrumentos se utilizan para evaluar en matemática?, 7. ¿Qué importancia tiene la evaluación en Matemática?, 8. ¿Para qué se usa la evaluación?, 9. ¿Cómo se expresan los resultados de la evaluación?, 10. ¿Y será que lo más conveniente es evaluar a los alumnos de esa manera: cuantitativamente, a través de una nota? ¿Con eso decimos todo? y 11. ¿Qué dificultades plantea la evaluación en matemática? Los informantes fueron profesores de Matemática en ejercicio de ambos sexos, no se estableció un lugar fijo para realizar las entrevistas, todas las entrevistas fueron grabadas.

Se realizaron cuatro entrevistas, cada una con una duración de 30 minutos aproximadamente. La primera entrevista fue realizada a una profesora de Matemática de octavo grado de Educación Básica, cuyo nombre se identificó con la letra **Z**. La segunda entrevista, se realizó a una profesora de Matemática de octavo grado de Educación Básica, identificada con la letra **A**, para garantizar la confidencialidad. La tercera entrevista, fue realizada a un profesor de Matemática de octavo y noveno grado de Educación Básica. Para tomar sus respuestas a cada pregunta lo identificamos con la letra **R**. La cuarta entrevista fue realizada a un profesor de octavo grado, con un año de experiencia. Se tomó la letra **F** para su identificación en las transcripciones.

Para el análisis cualitativo de los datos se parte del discurso, tomando en consideración las elaboraciones verbales de los profesores participantes, y como unidad de análisis la frase, tanto en los grupos de discusión como en las entrevistas a profundidad. Esta técnica de análisis (Ortiz, 2002) permite la extracción de información “desde dentro”, emergente del discurso de los sujetos, de allí que se justifica la utilización de la misma. El procesamiento de la información se realizó en forma manual. Para ello, luego del registro, mediante la grabación, de las conversaciones, tanto de los grupos de discusión como de las entrevistas, se procedió a la transcripción textual de las mismas, respetando en lo posible las expresiones y pausas que ocurrían en cada conversación. Se realizó la lectura de la transcripción del primer grupo de discusión para identificar unidades de información: frases, oraciones, párrafos y se asignaron descriptores provisionales. Después se inició la lectura y relectura del segundo grupo de discusión y de las cuatro entrevistas realizadas, y se identificaron las unidades de análisis en el texto de cada uno. Se organizaron las unidades de información (frases u oraciones) en tablas de acuerdo a la similitud de su contenido y se ubicaron en función de las dimensiones objeto de estudio, según los temas que se abordaron. El análisis concluyó cuando se llegó al punto de saturación de las categorías y dimensiones.

En lo concerniente a la validación de los instrumentos, se hizo tomando en consideración las sugerencias realizadas en cada reunión con el grupo de investigación, así como también la revisión y discusión realizada con varios educadores matemáticos.

Las respuestas dadas por los profesores, tanto en los grupos de discusión como en la entrevista a profundidad, fueron recopiladas en materiales audiovisuales como la filmadora y grabadora. Posteriormente se procedió a la transcripción de las mismas de manera textual, utilizando una letra que identificaba a cada uno de los informantes.

3. RESULTADOS

La información obtenida tras la realización de los dos grupos de discusión y la aplicación del guión para la entrevista en profundidad consiste en una serie de respuestas dadas por los profesores informantes a cada uno de los temas planteados. Cada respuesta se concreta en un enunciado que puede presentarse una o más veces según la opinión de los profesores, en las preguntas planteadas tanto en los grupos de discusión como en las entrevistas en profundidad.

Con el propósito de presentar la visualización de los resultados obtenido, se muestra el cuadro 1, que contiene las respuestas dadas por los profesores, a cuatro de los temas planteados, en el primer grupo de discusión, en el que se reflejan algunos enunciados de los docentes en forma sintetizada.

CUADRO 1. RESPUESTAS DEL GRUPO DE DISCUSIÓN 1

Dimensión explorada	Grupo de discusión 1 (GD1)
Percepción de la evaluación en matemática	R: ... es verificar el logro de algo... I: ...es un proceso... C: ...procedimiento para verificar si se está logrando el objetivo que se planteó...
Propósito de la evaluación	R:... Para ver el rendimiento académico... logró el objetivo... domina la competencia I:... Verificar si el objetivo que estas dando llegó al alumno. C:... está logrando el objetivo...
Aspectos a evaluar	R:...procedimientos de los alumnos, comportamientos dentro del aula... las habilidades que tiene el muchacho... su comportamiento. I:... procedimientos, destrezas, resultados... El comportamiento del alumno... Evaluamos a los alumnos... los alumnos nos evalúan a nosotros... C:... La responsabilidad.
Instrumentos de evaluación	R:... va a depender también del objetivo que se plantee el profesor... en una prueba yo verifico si el muchacho alcanzó las competencias.... las fallas que tuvo el muchacho...Nos interesa más lo que el muchacho aprenda dentro del aula. I:... a través de exámenes, talleres... pruebas orales. C:... Evaluaciones individuales... talleres para que aprendan a trabajar en grupo... La pizarra. R:...las escalas de estimación, la lista de cotejo...las pruebas... C: ...elaborar un material...como un taller...

Del cuadro 1, y considerando los demás aportes de los profesores al resto de las dimensiones exploradas o temas propuestos; en lo que respecta a la percepción de la evaluación, importancia de la evaluación y uso de la evaluación, se concluye que los docentes consideran a la evaluación como el logro, por parte de los alumnos, de un contenido desarrollado en el aula de clases, y, por lo tanto, si es logrado el alumno es promovido, tal como lo manifiestan: C (GD1) "Es un procedimiento para verificar si se está logrando el objetivo que se planteó, porque normalmente colocamos un objetivo que se quiere lograr con el estudiante y con la evaluación verificamos si se está cumpliendo o no"; I (GD1) "...la evaluación nos permite ver el rendimiento del alumno, para ver si llegaron esos objetivos que se planteo el profesor..." y R (GD1) "La evaluación yo pienso que se usa para ver si el alumno esta capacitado a seguir un grado superior."

En cuanto al propósito de la evaluación y la expresión de sus resultados, se refleja la connotación cuantitativa que le dan los profesores a la evaluación en matemática, por cuanto consideran que: "se evalúa para ver el rendimiento académico. Si el muchacho en verdad logró el objetivo, si domina la competencia que plantea uno como profesor" R (GD1). "La mayoría de las cosas creo que tienen un instrumento de medición, nuestro instrumento de medición es una evaluación, ponderada a diez (10), si pasa de diez se está logrando el objetivo si es menos no se está logrando". C (GD1). "De manera cuantitativa, a través de resultados." I (GD1).

Los aspectos que los profesores consideran deben ser evaluados en matemática (ver cuadro 1) son los procedimientos realizados por los alumnos, habilidades, destrezas y comportamiento en el aula.

Asimismo, los docentes manifiestan que los instrumentos mayormente utilizados en la evaluación en matemática (ver cuadro 1) son pruebas y talleres (pruebas en grupo), aunque reconocen la existencia de otros instrumentos, como las escalas de estimación, listas de cotejo, las cuales permiten aplicar la técnica de observación. Esto se evidencia en expresiones como: R (GD1) "...en una prueba yo verifico si el muchacho alcanzó las competencias y además de eso voy a verificar las fallas que tuvo el muchacho..." y Respecto al tema de los entes evaluadores en matemática, los docentes responden que el único responsable de evaluar debe ser el docente del área. Esto se evidencia en los siguientes enunciados: R (GD1) ", los docentes que están a cargo de dichas cátedras..." e I (GD1) "El docente, sí, el docente que esté capacitado..."

Asimismo, los docentes consideran que alguna de las dificultades que plantea la evaluación en matemática son: el tiempo para desarrollar los objetivos del área, pérdida de tiempo en explicación a los alumnos, mala base, errores ortográficos, considerando que estas dificultades se deben a: C (GD1) "...yo creo que deben aumentar las horas, las horas de Matemática... repetición de pruebas". R (GD1) "...ministerio de educación debe hacer una reconsideración..."

R (GD1) "...Se pueden utilizar las escalas de estimación, la lista de cotejo..."

A partir de estos enunciados se podría afirmar que estos docentes conciben la evaluación como un proceso mediante el cual se puede verificar el logro de los objetivos y contenidos planificados por el docente con la finalidad de dar una ubicación a los alumnos en el grado inmediato superior, mediante una calificación determinada.

Respecto al otro grupo de discusión, se observó una congruencia entre las opiniones de los docentes y lo que establece el principio de evaluación del NCTM (2000) respecto a la importancia de la evaluación como herramienta para la toma de decisiones, donde establece que la evaluación y la enseñanza deben estar integradas de manera que constituya una parte rutinaria de la actividad docente. Esta relación se evidencia en el siguiente enunciado: A (GD2) "La evaluación es eje fundamental, o sea está inmersa entre la enseñanza y el aprendizaje".

Para el caso de las cuatro entrevistas a profundidad, realizadas a tres profesores que participaron en los grupos de discusión y un profesor externo a los grupos, pero que es profesor activo en una de las instituciones seleccionadas en la investigación, se presenta en el cuadro 2, las respuestas dadas al propósito de la evaluación en matemática.

CUADRO 2. RESPUESTAS DADAS EN LAS CUATRO ENTREVISTAS E1, E2, E3 Y E4

Dimensión explorada	E1	E2	E3	E4
Propósito de la evaluación en matemática	Z:... Para ver si han logrado obtener las habilidades necesarias para resolver problemas que se les puedan presentar más adelante... para ver si el maestro es bueno o malo...	A:... Saber el nivel de conocimiento que tienen los alumnos...percibir si el alumno aprendió... Los conocimientos que uno quiere que aprenda.	R:...con la finalidad de verificar los resultados en cuanto a los objetivos que se dan... ...a través de la aprobatoria de ese objetivo, ellos pasen a un grado superior.	F:...Para ver donde el alumno se siente más deficiente...

En esta primera pregunta se pudo observar que no existe mucha variación en cuanto a las respuestas obtenidas en los grupos de discusión, ya que las opiniones de los profesores muestran que la evaluación se centra más que todo en el alumno respecto a sus conocimientos, habilidades, fallas o deficiencias, verificar los logros alcanzados respecto de los objetivos. Sin embargo, un profesor considera la posibilidad de la evaluación centrada hacia el docente: Z (E1) "...para ver si el maestro es bueno o malo... porque cuando un alumno tiene diez, es porque también el profesor tiene diez". Es decir, considera que no solamente se debe evaluar al alumno sino también a otros agentes del proceso de enseñanza-aprendizaje de la matemática, y comienza por la labor del profesor.

Al revisar las demás respuestas, a los otros temas presentados en las entrevistas, se encontró, entre otros hallazgos, que la evaluación en matemática se usa para: "...obtener toda la información tanto para la adquisición del conocimiento del alumno como para el desenvolvimiento del docente..." Z (E1), "...saber si el alumno aprobó un contenido..." A (E2). Asimismo, la evaluación permite tomar decisiones sobre la promoción de un alumno al grado inmediato superior. Lo cual refleja una concepción de la evaluación como lo afirma García (2003) un instrumento que dota de cierto "poder objetivo" a los actores encargados de evaluar para actuar como jueces que deciden desde posiciones aparentemente competentes, quién pasa la materia o quién se gradúa.

El análisis de las respuestas emitidas por los profesores informantes, permitió establecer el cuadro 3 sobre las concepciones de la evaluación y su relación con lo establecido en la normativa legal vigente.

En el cuadro 3 se evidencia una discrepancia entre los constructos o conceptos extraídos de las opiniones de los profesores informantes respecto a la evaluación en matemática, y los lineamientos emanados de la legislación vigente sobre este tema.

Se puede decir entonces que los docentes entrevistados mantienen una visión generalizada de la evaluación en la que se tiene por prioridad o centro al alumno, pues manifiestan que se debe evaluar su actuación, trabajo, conocimientos, deficiencias e incluso su comportamiento; mientras que la legislación establece que en el proceso educativo debe ser objeto de evaluación tanto el alumno como el profesor, así como los medios, métodos o instrumentos utilizados en el proceso.

Asimismo, los profesores utilizan generalmente como instrumento de evaluación, las pruebas, por ser, según ellos, la forma más cómoda de obtener los resultados y muchas veces sin el diseño adecuado de las mismas. Lo estipulado en la normativa legal indica que los métodos y procedimientos utilizados en la evaluación deben ser planificados, aplicados y comprobados durante el proceso de aprendizaje.

Los docentes evalúan utilizando pruebas o exámenes, pero ¿qué pasaría si esas pruebas o exámenes presentan una estructuración no adecuada? y se convierte en mera repetición de lo que el profesor explica en clase, en las que se plantea una serie de ítems que según Webb (1992), se puntúan como malos o buenos o simplemente el profesor revisa si la tarea realizada por el alumno es "*correcta o incorrecta*", (NCTM, 2000), para luego asignar una nota. Esto se evidencia en repuestas como: A (E2): "...a veces... unas pruebas mal diseñadas..." En este sentido mostramos, a manera de ejemplo, en el gráfico 1, unas preguntas de un examen aplicado a estudiantes de octavo grado en la Escuela Básica "Amantina de Sucre".

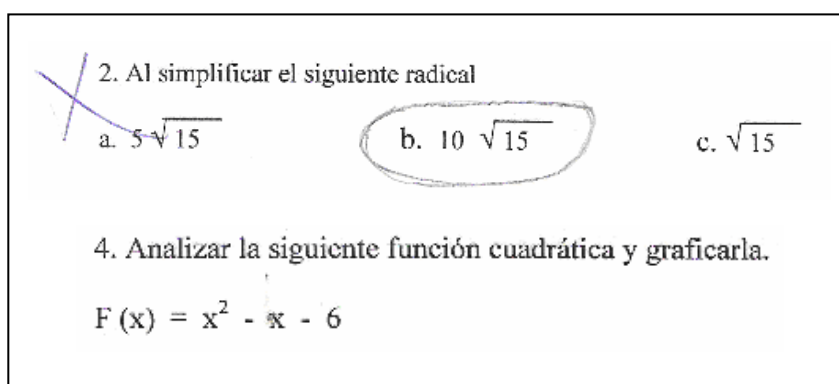
CUADRO 3. CONTRASTE ENTRE CONCEPCIONES DE LOS PROFESORES Y LO ESTABLECIDO EN LA NORMATIVA LEGAL

Concepciones	Normativa Legal
La evaluación permite: • Detectar fallas de los alumnos. • Verificar el logro de los objetivos. • Obtener información sobre el rendimiento de los alumnos. • Tomar decisiones en cuanto a la promoción y orientación de los alumnos. • Verificar el comportamiento de los alumnos. • Ver si el alumno esta capacitado a seguir un grado superior. • Saber si el alumno aprobó un contenido. Los resultados de la evaluación se expresan: • de manera cuantitativa, a través de una nota. • con una calificación. • Se evalúa a través de exámenes o pruebas. A veces unas pruebas mal diseñadas. • Obtener información sobre el trabajo del docente. Se usa: • Solamente la evaluación sumativa. • Se evalúa al maestro y ellos mismos con la auto-evaluación y la coevaluación.	Ley Orgánica de Educación (1999): • La evaluación, como parte del proceso educativo, será continua, integral y cooperativa. • Determinará de modo sistemático en qué medida se han logrado los objetivos educ. • Deberá apreciar y registrar de manera permanente mediante procedimientos apropiados, el rendimiento del educando, tomando en cuenta los factores que integran su personalidad. • Valorará la actuación del educador y, en general, todos los elementos que constituyen dicho proceso. (Art. 63). Reglamento General de la Ley Orgánica De Educación: • La evaluación constituye un proceso permanente dirigido a identificar y analizar tanto las potencialidades para el aprendizaje, los valores, los intereses y las actitudes del alumno para estimular su desarrollo, como los aspectos que requieran ser corregidos. • Apreciar y registrar en forma cuantitativa el progreso en el aprendizaje y dominio de competencias del alumno. (Art. 88). • Los métodos y procedimientos que se utilicen en el proceso de evaluación se planificarán, aplicarán y comprobarán (Art. 90). • La evaluación que realice el Ministerio de Educación y Deportes incluirá tanto a los docentes como a los alumnos. (Art. 91). • La actuación general del alumno será evaluada a través de los siguientes tipos de evaluación: diagnóstica, formativa y sumativa. (Art. 92). • En la evaluación de la actuación del alumno participaran: el docente; el alumno, mediante la autoevaluación y la sección o grupo, por medio de la coevaluación. (Art. 97)

Fuente: Datos de la Investigación.

En la primera pregunta se pide simplificar un radical que no se muestra y se le dan opciones de respuesta para que indique la correcta, la cual podría ser cualquiera de ellas. Como el alumno no respondió la respuesta que esperaba el profesor, entonces éste le marcó con una equis. En la segunda pregunta, se le pide al alumno que analice y grafique una función, para lo cual no se especifica de qué forma debe ser ese análisis. Al revisar el examen, observamos que el alumno resuelve la ecuación de segundo grado $x^2 - x - 6 = 0$ aplicando la fórmula respectiva, para luego buscar las coordenadas del vértice y finalmente graficar la parábola. De acuerdo a lo que corrigió el profesor, se supone entonces que ese es el "análisis" que pedía hacer. Se evidencia también que las respuestas incorrectas son indicadas con una equis.

GRÁFICO 1. PREGUNTAS DE EXAMEN



De acuerdo con lo observado y según el tratamiento que los profesores dan a las respuestas emitidas por los alumnos en las pruebas, el informe Cockcroft (1985) afirma que al realizar la corrección de los trabajos de los alumnos, una cruz no es de gran utilidad para ellos si no va acompañada de una indicación del punto donde ha cometido el error, al igual que una explicación de la naturaleza de la equivocación, de esta manera el profesor toma conciencia de los errores que se cometen frecuentemente y en consecuencia preparar las clases posteriores.

Por ello, es necesario que los profesores superen la consideración superficial de tarea "*correcta o incorrecta*" y se centren en cómo piensan los alumnos al hacer las tareas, esforzándose por identificar las ideas válidas de los estudiantes, (NCTM, 2000). En este sentido, las evaluaciones deberían dar ocasión a múltiples enfoques ya que los alumnos muestran lo que saben y pueden hacer de distintos modos.

En cuanto a la aplicación de los tipos de evaluación, los docentes afirman que solamente se utiliza la evaluación sumativa, por cuanto se busca es asignar una nota o calificación con la finalidad de promover o no al alumno; mientras que en la legislación se establecen tres tipos de evaluación: la diagnóstica, para identificar las actitudes, conocimientos, habilidades, destrezas, intereses y motivaciones de los alumnos; la formativa, con la finalidad de determinar en qué medida se logran las competencias, contenidos y objetivos programáticos; y la sumativa, cuya finalidad es determinar el logro de las competencias, los contenidos y objetivos programáticos, logros que serán expresados de forma cuantitativa en la tercera etapa de Educación Básica.

Respecto al objeto y fines de la evaluación en matemática, se observa una concordancia entre los conceptos emitidos por los profesores y lo establecido en la reglamentación vigente. Lo mismo sucede cuando se refieren a los agentes internos encargados de la evaluación: el profesor; así como en las

formas de participación en el proceso de evaluación: mediante la autoevaluación y la coevaluación. Sin embargo, en la práctica esto no se lleva a cabalidad, ya que solo se utilizan las pruebas para obtener la información requerida en el proceso de enseñanza – aprendizaje de la matemática.

4. CONCLUSIONES

Teniendo en cuenta los objetivos de esta investigación, se presentan las conclusiones relacionadas con las concepciones emergentes del pensamiento de los profesores respecto a la evaluación en matemática según las dimensiones estudiadas, su caracterización y su congruencia con lo establecido en las normativas legales vigentes.

Respecto a las concepciones de los profesores, acerca de la evaluación, tenemos que:

- La evaluación se centra más que todo en el alumno respecto a sus conocimientos, habilidades, fallas o deficiencias, verificar los logros alcanzados respecto de los objetivos y los contenidos.
- Se refleja la connotación cuantitativa que le dan los profesores a la evaluación en Matemática, por cuanto consideran que se realiza para colocar una nota.
- La evaluación es importante, porque su finalidad es la de obtener información sobre los alumnos y su aprendizaje.
- La evaluación de objetivos se usa para verificar el logro de los objetivos planeados.
- Los profesores consideran que se debe evaluar los contenidos, pero se centran más que todo en el aspecto conceptual, aunque algunos hacen alguna mención somera a lo actitudinal y procedimental.
- Las evaluaciones en Matemática se hacen a través de pruebas, las cuales son elaboradas por los profesores de acuerdo con los contenidos planificados.
- No se hace referencia a la evaluación de procesos, pues generalmente las evaluaciones se limitan a una repetición de lo explicado por el profesor en clase, por lo tanto se hace referencia solo a la evaluación de producto o evaluación sumativa.
- Dado que la evaluación esta centrada en el trabajo del alumno, los profesores hacen poca referencia a la evaluación del docente, sin embargo algunos consideran que también debe ser evaluado su trabajo.
- En cuanto a la evaluación para la toma de decisiones, la aplican mayormente para asignar una calificación a los alumnos y decidir de acuerdo a la puntuación establecida si son promovidos o no al grado inmediato superior.
- Los profesores se asignan cierta responsabilidad en la complejidad de la evaluación y dan mayor importancia a que los evaluadores sean internos, es decir, que debe ser el profesor, pero consideran la posibilidad de que puede hacerse una evaluación externa.

En el cuadro 4 se sintetizan las concepciones captadas en colectivo (grupos de discusión) y de manera individual (entrevistas). Se aprecia que en grupo la opinión de los docentes tiende a ser más general que cuando se expresan de manera individual, sin embargo el contenido de las expresiones refleja opiniones

coincidentes. Por ejemplo, en lo referente al propósito de la evaluación, en el grupo emerge el contenido matemático pero sin profundizar cuales aspectos consideran del mismo; mientras que en las entrevistas, el discurso contempla habilidades, deficiencias, logros de los alumnos y además abren la posibilidad de evaluar al docente. Asimismo, respecto al tema de la importancia de la evaluación, en el grupo mencionan la verificación del logro de los alumnos, en tanto que en la entrevista, el logro está explícitamente asociado a los conocimientos adquiridos por los alumnos. De lo antes expuesto, se podría afirmar que las dos técnicas para recabar información, respecto a las concepciones acerca de la evaluación en matemática, se complementan y aportan información para dar respuesta a los objetivos planteados.

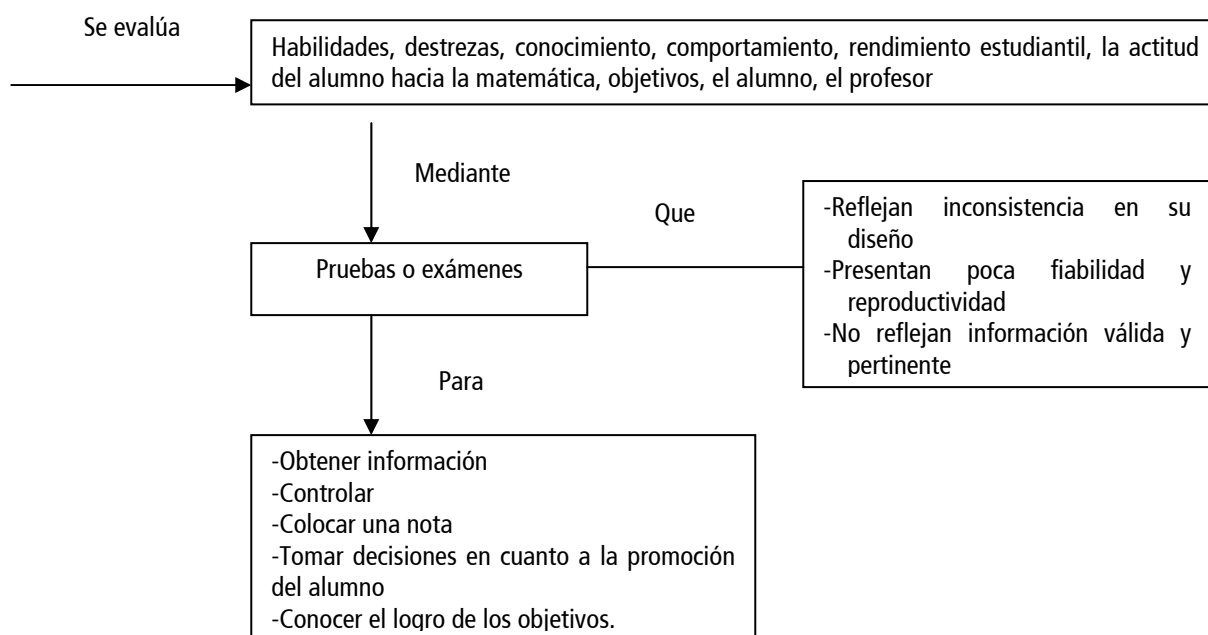
CUADRO 4. CONCEPCIONES DE LOS PROFESORES DESDE LOS GRUPOS Y DE MANERA INDIVIDUAL

Dimensiones exploradas	Grupos de discusión	Entrevistas
Propósito de la evaluación en matemática	Evaluar al alumno respecto a los contenidos matemáticos desarrollados en el aula.	Evaluar al alumno, sus habilidades, deficiencias y logros. Evaluar al docente.
Aspectos a evaluar	Conocimientos, procedimientos, actitudes, habilidades y destrezas de los alumnos.	Se evalúa el conocimiento adquirido, el trabajo realizado y los logros alcanzados.
Estrategias de evaluación	Observación del trabajo del alumno en actividades individuales y grupales.	Trabajo del alumno mediante pruebas, aun cuando contemplan otras alternativas.
Responsable de la evaluación	El docente.	El docente, dando la posibilidad a los alumnos y algún evaluador externo que también participe.
Instrumentos utilizados	Pruebas y trabajo en grupo.	Los instrumentos más utilizados son las pruebas. También se acude a trabajo en grupo e interrogatorios orales.
Importancia de la evaluación	Verificación de logros de los alumnos.	Permite verificar el logro de los alumnos y los conocimientos adquiridos
Uso de la evaluación	Promoción del alumno. Diagnosticar dificultades de los alumnos.	Tomar decisiones sobre la promoción de un alumno al grado inmediato superior.
Expresión de los resultados	Cuantitativa.	Cuantitativa.
Dificultades que surgen de la evaluación	Falta de tiempo para desarrollar los objetivos. Carencias formativas del docente.	Carencias de formación del docente en el uso de los instrumentos de evaluación. Falta de tiempo para desarrollar los objetivos. Planificación de la evaluación..
Origen de las dificultades en la evaluación	Poca preparación previa de los alumnos. Número de alumnos por aula. Falta de recursos institucionales. Condición económica del docente.	Falta de supervisión educativa, insuficiente preparación del profesor. Poca preparación previa del alumno. Poco tiempo semanal dedicado a la matemática. Inadecuadas condiciones de remuneración del profesorado. Carencia de recursos institucionales y número excesivo de alumnos.

Sobre la caracterización de las concepciones de la evaluación, se presenta el gráfico 2, donde se muestra la visión que presentan los profesores informantes respecto a la evaluación en Matemática, es decir, qué se evalúa, cómo se evalúa y para qué se evalúa.

En cuanto a esta caracterización de la evaluación en Matemática, algunos autores como Van Den Heuvel-Panhuizen y Becker (2003), Webb (1992) y Flores (2001) plantean puntos de vista sobre la evaluación en matemática que guardan estrecha relación con los enunciados extraídos de las opiniones de los profesores informantes, que en su mayoría mostraron vinculación con la evaluación de objetivos.

GRÁFICO 2. LA EVALUACIÓN EN MATEMÁTICA, EN LA EDUCACIÓN BÁSICA, DESDE LOS DOCENTES



Respecto a la correspondencia de las concepciones con lo establecido en la normativa legal vigente, se puede decir que, en Venezuela para el año 2006, existe un nivel medio de correspondencia entre las concepciones de los profesores y lo establecido en la normativa sobre evaluación, evidenciado en cuanto a:

- Objeto y fines de la evaluación en Matemática.
- Los agentes internos encargados de la evaluación: el profesor.
- Las formas de participación en el proceso de evaluación: mediante la autoevaluación y la coevaluación a las cuales solo hicieron referencia.
- La forma de expresar los resultados de la evaluación: cuantitativa.

Observándose discrepancia en cuanto a:

- Evaluación centrada mayormente en el alumno.
- Los instrumentos: Se evalúa generalmente a través de exámenes o pruebas y veces unas pruebas mal diseñadas. Aun cuando hacen referencia a otros tipos de instrumentos.
- La aplicación de los tipos de evaluación: solo se usa la evaluación sumativa o de producto.

Finalmente, los docentes entrevistados mantienen una visión de la evaluación en la que se tiene por prioridad o centro el alumno, pues manifiestan que se debe evaluar su actuación, trabajo, conocimientos, deficiencias e incluso su comportamiento; mientras que la legislación establece que en el proceso

educativo debe ser objeto de evaluación tanto el alumno como el profesor, así como los medios, métodos o instrumentos utilizados en el proceso. Aunque algunos consideran que el docente también debe ser evaluado.

Desde una mirada prospectiva del problema abordado en este escrito, se considera que la experiencia puede ser mejorada, considerando recomendaciones tales como realizar otros estudios que impliquen diseñar, implementar y evaluar programas de formación que incluyan la evaluación en educación matemática en temas específicos de las matemáticas escolares.

REFERENCIAS BIBLIOGRÁFICAS

- Calderón, R. M y Deiros, B. (2003). Evaluación del Aprendizaje de las Matemáticas. En: J. Delgado Rubí (Ed.), *Acta Latinoamericana de Matemática Educativa*. México: Comité Latinoamericano de Matemática Educativa.
- Canales, M. y Peinado, A. (1995). Grupos de Discusión. En: J. M. Delgado y J. Gutiérrez (Eds.), *Métodos y Técnicas Cualitativas de Investigación en Ciencias Sociales*. Madrid: Síntesis.
- Cockcroft, W.H. (1985). *Las Matemáticas sí Cuentan. Informe Cockcroft*. Madrid: Ministerio de Educación y Ciencia.
- Flores, P. (2001). Aprendizaje y Evaluación. En: E. Castro (Ed.), *Didáctica de la Matemática en la Educación Primaria* (Capítulo 2). Madrid: Síntesis.
- García, G. (2003). Currículo y evaluación en Matemáticas. Un estudio en tres décadas de cambio en la educación básica. Bogotá: Magisterio.
- Gil, F. (1999). *Marco conceptual y creencias de los profesores sobre evaluación en matemáticas*. (Tesis doctoral). Granada, España: Universidad de Granada.
- Giménez, J. (1997). *Evaluación en Matemática. Una integración de Perspectivas*. Madrid: Síntesis.
- Jimeno, M. (2006). *¿Por qué las niñas y los niños no aprenden matemáticas?* Barcelona: Octaedro.
- Ley Orgánica de Educación. (1980, Julio 28): Gaceta Oficial de la Republica de Venezuela, N° 2.635 (Extraordinario) y su Reglamento (1986, Enero 22). Decreto N° 975. Reforma parcial del Reglamento General de la Ley Orgánica de Educación (1999, Septiembre 15). Gaceta Oficial N° 36787, Septiembre 15, 1999.
- Ministerio de Educación. (1987). *Modelo Normativo y Plan de Estudio y Evaluación del Rendimiento Escolar*. Caracas: Ministerios de Educación.
- Moya, A. (2001). Reflexiones sobre la Teoría y Práctica de la Evaluación en Educación Matemática. *Boletín de Investigación, UPEL, J. M. Siso Martinez, 1*.
- NCTM (2000). *Principios y Estándares para la Educación Matemática*. (Primera Edición en Castellano). Sevilla: National Council of Teachers of Mathematics/SAEM.
- Ortiz, J. (2002). Modelización y Calculadora Gráfica en la enseñanza del Álgebra. Estudio Evaluativo de un Programa de Formación. (Tesis doctoral). Granada, España: Universidad de Granada.
- Ruthven, K. (2003). Alternativas a la evaluación con exámenes: expectativas y dificultades. En N. Gorgorió, J. Deulofeu y A. Bishop (coords.), *Matemáticas y Educación. Retos y Cambios desde una Perspectiva Internacional*. Barcelona, España: ICE/Graó
- Stake, R.E. (2005). Qualitative Case Studies. En N. Denzin y Y. Lincoln (eds.), *The Sage Handbook of Qualitative Research* (Third Edition). Thousand Oaks, CA: Sage.
- Valles, M. (1999). *Técnicas Cualitativas de Investigación Social. Reflexión Metodológica y Práctica Profesional*. Madrid: Síntesis.

- Van den Heuvel-Panhuizen, M. y Becker, J. (2003). Towards a Didactic Model for Assessment Design in Mathematics Education. En A. Bishop, M.A. Clements, C. Keitel, J. Kilpatrick y F. Leung (Eds.), *Second International Handbook of Mathematics Education*. Dordrecht, Holanda: Kluwer.
- Webb, N. (1992). Assessment of Students Knowledge of Mathematics: Steps Toward a Theory. En D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning*. New York: Macmillan.
- Yin, R. (2003). *Case Study Research. Designs and Methods* (Third Edition). Thousand Oaks, CA: Sage.



PROGRAMAS DE AUDITORIA INSTITUCIONAL UNIVERSITARIA. COMPARACIÓN DE LA PROPUESTA ESPAÑOLA CON EL SISTEMA BRITÁNICO

Reina Ferrández

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art11.pdf>



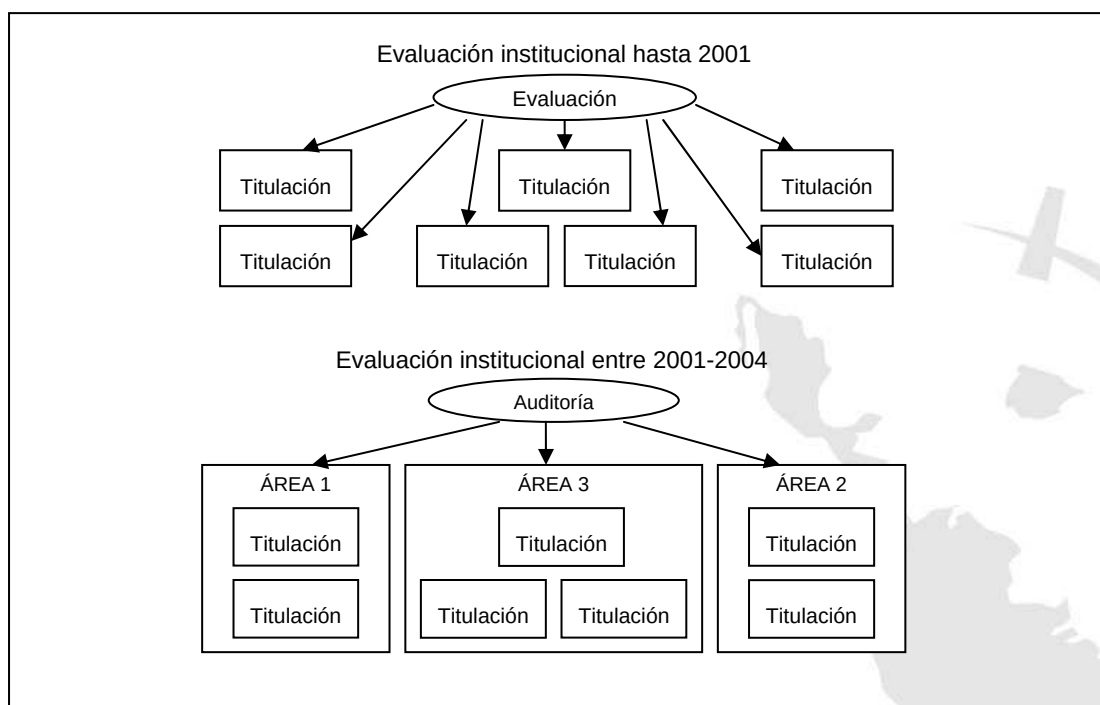
1. LA EVALUACIÓN INSTITUCIONAL EN EL REINO UNIDO

El proceso de evaluación institucional en el Reino Unido ha ido evolucionando, desde un sistema de acreditación basado en indicadores cuantitativos y centrado en titulaciones hasta centrado en la actuación global de la institución, bajo la filosofía de que si una institución es capaz de demostrar que tiene y pone en práctica los mecanismos necesarios para asegurar la calidad de su oferta educativa, no es necesario descender a niveles inferiores de análisis. Esta evolución ha venido dada por una serie de hechos que hicieron plantearse a las autoridades la necesidad de simplificar el modelo existente (JM Consulting, 2005).

1. La práctica totalidad de los títulos obtenía buenos resultados
2. Elevado coste económico
3. Continuas interferencias con la vida académica ordinaria

De esta manera, el sistema existente, fue reemplazado en 2001 por un marco para la garantía de calidad (QAF: *Quality Assurance Framework*) basado no ya en titulaciones (*Subjects*), sino en lo podríamos denominar Áreas (*Disciplines*), en un intento de simplificación del sistema (HEFCE, 2005) (ver Figura 1).

FIGURA 1. ESQUEMA ILUSTRATIVO DEL CAMBIO EN EL FOCO DE LA EVALUACIÓN INSTITUCIONAL EN EL REINO UNIDO ENTRE LOS AÑOS 2000 Y 2004

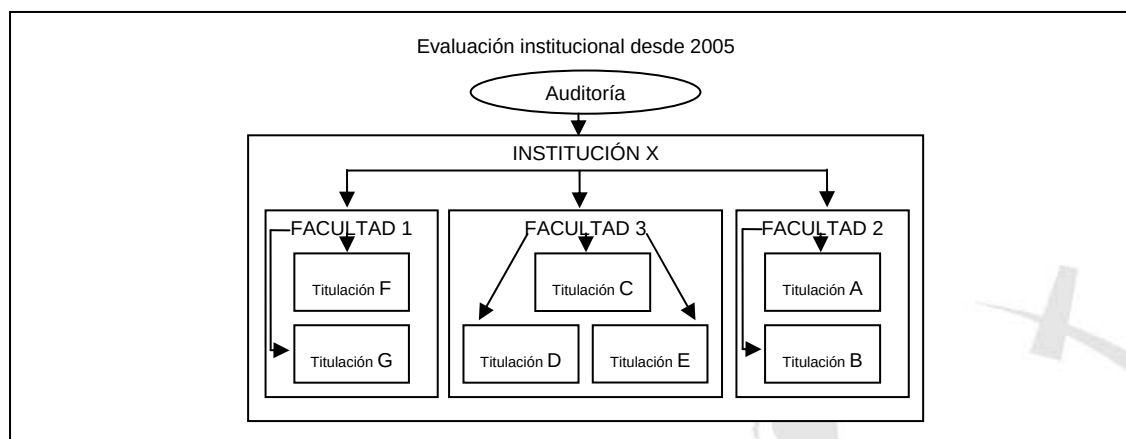


Pero además surge un cambio aún de mayor trascendencia, necesario, a mi modo de ver por la ampliación del punto de mira de la evaluación. Este cambio hace que los indicadores cuantitativos pasen a un segundo plano, centrando toda la atención en una serie de documentos elaborados a partir de la experiencia acumulada y el consenso de todas las instituciones de HE. Estos documentos, en su conjunto, se denominan Infraestructura Académica (AI: *Academic Infrastructure*) (ver QAA, 1999, 2000, 2001,

2004a, 2004b, 2004c, 2004d, 2004e, 2006a, 2006b, 2006c, 2006d, 2007a, 2007b) que junto con las estadísticas cuantitativas (TQI: *Teaching Quality Information*) (ver HERO, 2007a y TQI, 2007) y los resultados sobre la encuesta nacional de estudiantes (NSS: *National Students Survey*) (ver HERO, 2007b) se han convertido en la base de información con la que cuentan los evaluadores para poder elaborar sus juicios.

Sin embargo, el modelo basado en Disciplinas, también ha sido revisado en los últimos años (QAA, 2006e) y el nuevo sistema de auditoría se centra ya no en los procesos singulares, sino que atienden hacia macroprocesos, es decir, hacia cómo organiza y gestiona la institución la calidad de su oferta y sus niveles académicos, de tal modo que si se demuestra su capacidad, y los resultados, comparando con los diferentes documentos de contraste (AI, TQI, NSS) se mantienen o mejoran, no hay por qué dudar de que los procesos singulares son también de calidad. De esta forma, la responsabilidad de la calidad de los títulos no depende necesariamente de agentes externos sino que depende directamente de las instituciones (ver figura 2), que deberán poner en marcha los mecanismos suficientes para que los criterios de calidad establecidos se cumplan.

FIGURA 2. ESQUEMA ILUSTRATIVO DEL CAMBIO EN EL FOCO DE LA EVALUACIÓN INSTITUCIONAL EN EL REINO UNIDO A PARTIR DE 2005



2. LA EVALUACIÓN INSTITUCIONAL EN ESPAÑA

Dentro de nuestro contexto la primera experiencia en Evaluación Institucional promovida por el Consejo de Universidades (Consejo de Universidades, 1992, 1994) tiene lugar durante el curso 93-94, mediante un plan piloto que se desarrolló ya en fase no experimental durante el curso 96-97 (Consejo de Universidades, 1995, 1996a, 1996b, 1997a, 1997b) y que volvió a ponerse en marcha durante el curso 98-99 (Consejo de Universidades, 1998a, 1998b), continuando con el II Plan de Calidad de las Universidades y en la actualidad el Plan de Evaluación Institucional (PEI) coordinado e impulsado desde la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA, 2007a). Todos estos planes, basados en el sistema de revisión por pares siempre han tenido una finalidad encaminada a la promoción de la cultura de la calidad y la mejora continua y el nivel de análisis se ha centrado en las titulaciones o programas si bien se realizaban algunas aproximaciones hacia procesos generales de la institución en cuanto que algunos de los servicios ofertados, por ejemplo la biblioteca, no podían limitarse a un programa concreto.

Con una finalidad más centrada en la rendición de cuentas y bajo la perspectiva de la Convergencia Europea, simultáneamente al Plan de Evaluación Institucional gestionado por la ANECA, aparece también el Plan Piloto de Acreditación de títulos así como distintos programas de Certificación de Servicios como las Bibliotecas o la Mención de Calidad de los Programas de Doctorado. Todas estas iniciativas igualmente centran su atención en procesos singulares cercanos al esquema estructural mostrado en primera parte de la Figura 1 que ejemplificaba la primera etapa del sistema británico.

Sin embargo, y al igual que ocurría en el Reino Unido así como en otros países europeos, la polémica sobre la utilidad y justificación del elevado coste de un sistema de aproximación a la calidad institucional basado en títulos y servicios o procesos aislados individuales ha sido un punto continuo de debate, que si bien en el Reino Unido se ha solventado, en este momento con una ampliación del foco hacia la institución como una unidad, en nuestro contexto, sin embargo continúan apareciendo iniciativas que no acaban de decidirse hacia una opción clara.

Así, en el mes de junio de este año, ha aparecido una nueva alternativa; el Programa audit. (ANECA, 2007b) cercano a la filosofía mostrada ya en el sistema británico en el que el punto de interés cambia desde el proceso en sí hacia cómo se gestiona dicho proceso, es decir, lo importante ya no es sólo lo que se hace, sino que se pretende ir más allá evaluando lo que se hace para controlar y mejorar lo que se hace, es decir pasamos de una evaluación de procesos hacia una evaluación de la gestión de los procesos, lo que sigue sin quedar claro a partir de los documentos publicados hasta ahora es el foco de atención, que en algunos casos hace clara referencia hacia los títulos, a veces hacia los centros y en otros casos hacia la institución.

Las siguientes secciones de este artículo van a ir detallando el funcionamiento del sistema británico así como los principales elementos que lo integran y se irá comparando, en la medida de lo posible, con la propuesta española a partir de la documentación existente sobre dicho proceso.

Asimismo, dado que el modelo de Auditoría Institucional británica lleva algunos años en funcionamiento, en aquellos elementos en los que se considere adecuado o pertinente se hará referencia a las repercusiones concretas a nivel institucional que dicho sistema está conllevando a partir del estudio de un caso realizado durante los meses de mayo y junio sobre la Universidad Metropolitana de Leeds¹

Debe tenerse presente, además, que el Programa Audit está en una fase inicial, habiéndose publicado hasta el momento únicamente tres documentos que pretenden iniciar en la filosofía de la gestión de la calidad a las instituciones universitarias que aún no lo han hecho. Además presenta una serie de herramientas y guías que pretenden ilustrar a las instituciones cómo elaborar y diseñar sus sistemas de garantía de calidad. Pero además, estos documentos presentan la utilidad de unificar sistemas de manera que, las agencias de calidad puedan llevar a cabo el reconocimiento y evaluación de éstos. Éste último punto resulta de sumo interés, bajo mi punto de vista puesto que ha sido, y está siendo uno de los grandes problemas existentes en el Reino Unido. Las instituciones tienen serios problemas para recoger las evidencias necesarias para demostrar el funcionamiento de sus mecanismos de gestión de la calidad dado que al evolucionar de un sistema basado en Disciplinas a un sistema global, no estaban estructuralmente preparadas para ello.

¹ Este estudio se realizó mediante una estancia de investigación subvencionada por la Conselleria de Empresa Universitat i Ciencia de la Generalitat Valenciana.

3. AUDITORÍA INSTITUCIONAL

3.1. Los fines y objetivos

Los fines principales de este sistema de evaluación se centran tanto en el control como en la mejora. En el control por cuanto persiguen del cumplimiento de determinados elementos establecidos, como son el ajuste de los títulos y calificaciones al documento marco creado para tal fin (*FHEQ: The Framework for Higher Education Qualifications in England, Wales and N. Ireland*) (ver QAA, 2000; 2006e). Mediante este control se espera que las instituciones proporcionen a sus estudiantes la oportunidad de obtener un aprendizaje de calidad en todos sus programas.

Pero además, también existe una finalidad formativa. Puesto que se espera que este proceso impulse la mejora de la calidad al basarse en "información obtenida a través de la evaluación, revisión interna y externa y feedback de los evaluadores" (QAA, 2006e:3).

En cuanto a los objetivos concretos, encontramos que se relacionan con la evaluación de la gestión para asegurar el mantenimiento de los niveles académicos de manera que los usuarios del sistema tengan plena confianza en el mismo, así como la rendición de cuentas ante la administración que financia. Pero además, resulta destacable el objetivo de controlar la mejora, es decir, el sistema no se conforma con el mantenimiento de dichos niveles mínimos, que sería el propio de un sistema meramente acreditativo, sino que busca controlar que se establezcan y pongan en marcha los mecanismos necesarios para la mejora continua de los procesos institucionales (QAA, 2006e:3).

La propuesta española del programa Audit parece responder a un objetivo muy semejante (ANECA, 2007:4) si bien se está trabajando en los documentos que desarrollarán la evaluación final de dichos mecanismos (procesos de verificación y de certificación (ANECA, 2007:5) por lo que en este momento no puede pormenorizarse más a este respecto.

3.2. El punto de mira: Los criterios de evaluación

Se comentaba en la primera sección de este artículo que este proceso, en el Reino Unido, se dirige al análisis de macroprocesos, es decir, hacia cómo organiza y gestiona la institución la calidad de su oferta y sus niveles académicos. La premisa de dicho análisis reside en que si la institución es capaz de demostrar su capacidad de gestión, y los resultados, comparando con los diferentes documentos de contraste (AI, TQI, NSS) así lo verifican, no hay por qué dudar de que los procesos singulares sean también de calidad.

De este modo, la auditoría institucional centra su atención en seis criterios que abarcan la totalidad de las actuaciones referidas a la enseñanza (QAA 2006e:4-5):

1. **La Gestión institucional de los niveles académicos (tanto en docencia de grado y postgrado como de postgrados de investigación²).** Se refiere a cómo se utilizan los diferentes recursos normativas y documentos para mantener los niveles académicos. Así, son de especial interés la selección de los examinadores externos (agentes externos encargados de velar por la adecuación de los sistemas de evaluación de los estudiantes así como los niveles de las calificaciones), la realización de evaluaciones internas y externas, el establecimiento de normativas internas de

² Sería el equivalente español a los estudios de tercer ciclo.

evaluación, el seguimiento de los estándares nacionales a partir de los documentos base (AI) u otros puntos de referencia.

2. **Gestión institucional de las oportunidades de aprendizaje.** A todo lo anterior, se añaden elementos como la inclusión de estudiantes como participantes en los sistemas de calidad, el practicum, o la gestión/modificación de la oferta educativa basada en resultados de investigación, cómo se utilizan los recursos de aprendizaje, cuáles son las políticas de admisión, qué tipo de apoyo se proporciona a los estudiantes y al personal de la institución.
3. **Acercamiento institucional a la mejora de la calidad.** Hace referencia a elementos como los incentivos y desarrollo del personal, colaboración de los estudiantes en la mejora, y difusión de buenas prácticas.
4. **Colaboraciones.** Hace referencia a cómo se gestiona la calidad y adecuación de la colaboración con otras instituciones.
5. **Organización institucional de la oferta investigadora para los estudiantes de postgrado.** Evaluación de la investigación, participación de los estudiantes de postgrado en la gestión de la calidad, etc.
6. **Información publicada.** Ajuste y adecuación de la información que se publica, satisfacción de los estudiantes con dicha información. (Parte de esa información se publica en el TQI).

El programa Audit presenta siete criterios que en algunos casos pueden resultar semejantes aunque en otros, los diferentes marcos normativos hacen que se marquen diferencias (ANECA, 2007b):

1. **Política y objetivos de calidad.** Se trata de un criterio muy amplio que hace referencia a que la política del centro responda a una cultura de toda la institución. No encontramos un paralelo en el referente británico pero cuando el nivel de análisis es la propia institución no parece tener demasiado sentido evaluar la existencia de dicha política: el resultado de la propia auditoría informará si la política y objetivos de calidad de las instituciones son adecuados o no.
2. **Diseño de la oferta formativa.** Se refiere a la existencia de mecanismos que garanticen la calidad y actualización de dicha oferta. Sería el paralelo, en cierta manera al segundo criterio del sistema británico.
3. **Desarrollo de la enseñanza y otras actuaciones orientadas a los estudiantes.** Salvando los diferentes referentes formativos, éste sería el equivalente al primer criterio en el programa británico. Se refiere a la existencia de procedimientos para garantizar y promover el aprendizaje del estudiante.
4. **Personal académico y de apoyo a la docencia.** Hace referencia a la existencia de mecanismos de control y mejora del personal destinado a la docencia. No existe mención explícita a este respecto en el modelo británico, que incluye este aspecto dentro del segundo criterio.
5. **Recursos materiales y servicios.** Se refiere a la existencia de mecanismos de gestión y mejora de servicios y recursos para el aprendizaje. No existe tampoco mención explícita a este respecto en el modelo británico. De nuevo se incluye este aspecto dentro del segundo criterio.
6. **Resultados de la formación.** Se refiere a la existencia de mecanismos de medida, análisis y utilización de datos de resultado de aprendizaje, inserción y satisfacción. No existe tampoco un

criterio específico, sin embargo, es uno de los elementos de contraste. Es decir, la información se recoge a nivel nacional y es publicada en varias bases de datos, como *Teaching Quality Information (TQI)*, *National Student Survey (NSS)*, e informes públicos de los examinadores externos. (HERO, 2007a; 2007b). Los auditores contrastarán los resultados con el resto de criterios, y éstos serán la base para emitir el juicio final. No se considera necesario establecer un criterio explícito puesto que la institución "*necesita*" esa información para dirigir su política en materia de gestión de la calidad, que es la que va a ser auditada.

7. **Información pública.** Al igual que en el caso británico, este criterio pretende asegurar la existencia de mecanismos para publicar y actualizar la información relativa a las titulaciones y sus programas. Si bien en el criterio británico se hace un mayor énfasis en la publicación también de los resultados.

3.3. La información necesaria: Evidencias

Para llevar a cabo el proceso auditor, al igual que en cualquier otro proceso de revisión por pares, la institución debe proporcionar información sobre sus acciones. En este caso, las evidencias son la parte fundamental del proceso y que, mientras en el caso británico está proporcionando grandes problemas a las instituciones por la poca sistematización existente a este respecto³, parece adivinarse que en la propuesta española, esa sistematización, comienza desde los primeros documentos publicados (ANECA, 2007, Documentos 02 y 03)

En el sistema británico, el equipo auditor utiliza la técnica del "*trailing*" para seleccionar y recoger las evidencias de una manera organizada de documentos y reuniones sobre los procesos y procedimientos institucionales (las auditorías no se detienen en titulaciones o disciplinas específicas, si no que el enfoque se realiza sobre el proceso de decisión e implementación institucional y sus resultados). Esta técnica consiste en la recogida de una muestra de los documentos que ejemplifican o evidencias cómo se llevan a cabo los diferentes procesos. Estos fragmentos de seguimiento (*sampling trails*) se utilizan para examinar los procedimientos o políticas que utiliza la institución y probar su eficacia. Se trata de un recorrido a través de la implementación y resultado de políticas y procedimientos desde el nivel de programación institucional.

La finalidad de las auditorías es trabajar con la documentación ya existente en la institución, es decir, que la institución no necesita producir ningún documento adicional. De hecho, este tipo de fragmentos de seguimiento los construye la propia institución para sus procesos de revisiones internas periódicas (QAA, 2006e:13). No obstante, debo hacer constar a este respecto, que al realizar diversas entrevistas con los responsables de procesos clave de la *Leeds Metropolitan University*, todos coincidieron en que este punto no era realista, puesto que todos los procesos de revisión internos afectan a unidades menores (facultades, titulaciones, etc.) por lo que, al ser el proceso de auditoría un proceso institucional, se ven obligados a construir dichos fragmentos desde el punto de vista global de toda la institución.

En la propuesta española (ANECA, 2007b, Documento 02:5-11), se define claramente como construir dichas evidencias para cada uno de los siete criterios mencionados en el apartado 3.2. Sin duda, este hecho es de gran ayuda para la elaboración u homogeneización de dichas evidencias, y además resulta más realista que el sistema británico. Por supuesto la redacción de dichos procesos no debe

³ En algunos casos se está utilizando el *brainstorming* para encontrar las evidencias que puedan avalar la gestión de algunos procesos.

confundirse con las evidencias en sí, que se trataría no de la descripción de un procedimiento de gestión de la calidad para un proceso determinado, sino el demostrar que ese procedimiento se ha llevado a cabo.

3.4. Criterios de contraste: los indicadores

Una vez se ha visto cuáles son los criterios y qué información debe presentar la institución, se precisa conocer los elementos de contraste. El equipo auditor debe emitir un juicio de valor sobre los diferentes procesos clave identificados en la sección anterior. Estos juicios se emiten tras el contraste de las evidencias con una serie de indicadores que, en el caso británico, no solo van a hacer referencia a resultados como ocurría en los sistemas anteriores a 2001 (ver Ferrández, 1999) sino que el cambio que se ha producido radica en que los juicios se van a emitir sobre procesos

Para ello, los criterios seguidos son, por una parte, de corte cualitativo, centrados en una serie de documentos agrupados bajo el nombre de *Academic Infrastructure (AI)* (cuatro en total: Código de Prácticas; Descripción de títulos; Especificidad de títulos y Criterios de calificación) elaborados a partir de la experiencia recogida en todos los años de trayectoria evaluadora y bajo el consenso de todas las instituciones de Educación Superior del Reino Unido.

Por otra parte, como apoyo para la emisión del juicio, se cuenta, adicionalmente con una serie de datos cuantitativos: *Teaching Quality Information (TQI)* y *National Student Survey (NSS)*

Dado que en nuestro país no existe, hasta el momento, ningún documento de estas características, y considerando que se trata de elementos imprescindibles para ayudar en la elaboración de los juicios, considero oportuno realizar una breve descripción de los mismos, no con el espíritu de imitación, pero sí como descripción de buenas prácticas en el sentido de que se trata de documentos contextualizados y elaborados sobre elementos que las instituciones han acordado, consensuado y considerado valiosos,

3.4.1. Código de prácticas. (Code of Practice)

En este código se identifican una serie de principios/preceptos del sistema que cubren temas relacionados con la gestión de la calidad y niveles académicos en la Educación Superior. Está formado por 10 secciones que proporcionan un punto de referencia a las instituciones para asegurar consciente, activa y sistemáticamente la calidad y niveles académicos de sus títulos y calificaciones. No se trata de una normativa de corte legal sino que se trata de una mera referencia. Se asume que las instituciones se aseguran de cumplir con la legislación.

Las secciones son:

1. Postgrados de investigación (QAA, 2004c).
2. Oferta de estudios en colaboración; aprendizaje flexible y distribuido (incluido e-learning) (QAA, 2004d).
3. Estudiantes con discapacidad (QAA, 1999).
4. Examinadores externos (QAA, 2004e).
5. Reclamaciones y quejas de estudiantes sobre asuntos académicos (QAA, 2007a).
6. Valoración de los estudiantes (QAA, 2006c).
7. Diseño, aprobación, control y evaluación de un título (QAA, 2006d)

8. Educación, información y guía sobre la titulación (QAA, 2001)
9. Prácticas en empresas (QAA, 2007b)
10. Admisión de estudiantes (QAA, 2006b)

La estructura de este código, en cada una de sus secciones incluye preceptos y explicaciones de éstos:

- a) Los *preceptos* expresan los puntos clave que la Educación Superior ha identificado como importantes para asegurar la calidad y niveles académicos. Las instituciones deben ser capaces de demostrar que llevan a cabo acciones para desarrollar dichos preceptos de una manera eficaz a través de su propia gestión y procesos organizativos, teniendo en cuenta su cultura, necesidades, tradiciones y toma de decisiones.
- b) Las *explicaciones* muestran por qué dichos preceptos son importantes.

3.4.2. Descripción de títulos ("Subject" Benchmarks) (QAA, 2004a)

Son un medio para proporcionar la comunidad académica la descripción de la naturaleza y características de los títulos en un área específica. Representan también las expectativas generales sobre los niveles exigidos para otorgar el título con una determinada calificación (honor, master...). Describen lo que proporciona coherencia e identidad a una disciplina, y definen lo que se espera de un titulado en términos de habilidades y destrezas necesarias para el desarrollo de su labor dentro de dicha área.

No se trata de un currículo nacional para un área determinada sino de un marco conceptual establecido por la comunidad académica de dicho área. Se han diseñado para ayudar a los que diseñan, ejecutan y evalúan un título. También pueden ser de interés para los estudiantes y empleadores potenciales, puesto que les permite obtener información sobre la naturaleza y niveles de un título en un área determinada.

3.4.3. Especificidad de títulos (Programme specifications) (QAA, 2006a)

Se trata de descripciones concretas sobre los resultados de aprendizaje esperados de un título de HE, y el medio por el que dichos resultados son logrados y demostrados.

Este documento, lo que pretenden es mostrar como se han combinado las materias en un programa determinado de una institución, pero no se trata sólo de eso, le da un valor añadido unificando las expectativas del "Subject" Benchmarks y las cualificaciones del HE quality standards estableciendo los resultados esperados si se finaliza con éxito un programa determinado. Son una fuente de información más específica, particularmente para futuros y actuales estudiantes.

En resumen, una especificación de título se puede utilizar para explicar los objetivos y resultados de aprendizaje que se pretenden con un programa, demostrando que puntos de referencia se han utilizado para diseñarlo y cuáles son los resultados de aprendizaje esperados.

3.4.4. Criterios de Cualificación (HE Qualification standards (Statements)) (QAA, 2000)

Pretende ayudar a los empleadores, escuelas, padres, estudiantes, etc. a entender los logros y atributos representados por los principales certificados de conocimiento acreditado (títulos). El esquema tiene cinco niveles, tres pregraduados y dos postgraduados. La Tabla 1 resume los principales criterios de cada una de las certificaciones académicas válidas.

TABLA 1. RESUMEN DE LAS PRINCIPALES CUALIFICACIONES OTORGADAS EN LA EDUCACIÓN SUPERIOR DEL REINO UNIDO

1 Certificate	Nivel C	<i>Certificates of Higher Education</i> Tiene un buen conocimiento de los conceptos básicos de un área, y ha aprendido como tomar diferentes aproximaciones para resolver problemas. Es capaz de comunicarse ajustadamente y tiene las cualidades necesarias para el ejercicio básico de un empleo con alguna responsabilidad personal.
2 Intermediate	Nivel I	<i>Foundation degrees, ordinary (Bachelors) degrees, Diplomas of Higher Education and other higher diplomas</i> Han desarrollado un buen conocimiento y comprensión de los principios del campo de estudio y han aprendido a aplicar dichos principios de manera amplia. A través de esto son capaces de evaluar la adecuación de los diferentes acercamientos para resolver problemas. Sus estudios pueden tener una orientación profesional que les capacite para trabajar eficazmente en el campo elegido. Tienen las cualidades necesarias para ser empleados en situaciones que requieran el ejercicio de responsabilidad personal y toma de decisiones.
3 Honours	Nivel H	<i>Bachelors degrees with Honours, Graduate Certificates and Graduate Diplomas</i> Ha desarrollado comprensión de un complejo cuerpo de conocimiento. Alguno de ellos dentro de los límites de una disciplina académica. A través de esto el graduado desarrolla destrezas técnicas analíticas para resolver problemas que le capacitan para muchos tipos de trabajo. El graduado es capaz de evaluar la evidencia, argumentos y asunciones para alcanzar juicios importantes y comunicarlos eficazmente. Puede desarrollar su trabajo en situaciones que requieran responsabilidad personal y toma de decisiones en circunstancias complejas e impredecibles.
4 Masters	Nivel M	<i>Masters degrees, Postgraduate Certificates and Postgraduate Diplomas</i> Muchos de los estudios con este título llevan el estandarte de una disciplina académica o profesional. Los estudiantes han mostrado originalidad en la aplicación del conocimiento y entienden como dicho conocimiento avanza con la investigación. Son capaces de tratar cuestiones complejas de manera sistemática y creativa. Muestran originalidad manejando y resolviendo problemas. Reunen las cualidades necesarias para un empleo que requiera realizar juicios importante en algunas circunstancias, iniciativa personal y responsabilidad en ambientes profesionales complejos e impredecibles.
5 Doctoral	Nivel D	<i>Doctorates</i> Se otorga para la creación e interpretación del conocimiento. Que extiende el límite de una disciplina, normalmente mediante una investigación original. Los que obtienen este nivel son capaces de conceptualizar, diseñar y llevar a cabo proyectos para la generación de nuevo conocimiento y/o entendimiento. Tienen las cualidades necesarias para empleos que requieran la habilidad de realizar juicios informados sobre cuestiones complejas en campos específicos e innovación en el manejo y resolución de problemas.

3.4.5. Información Cuantitativa: Teaching Quality Information (TQI), National Student Survey (NSS), e informes públicos de los examinadores externos. (HERO, 2007a; 2007b)

Se trata de una base de datos oficial en la que se encuentra ubicada la información en cifras sobre los resultados de la Educación Superior financiada con fondos públicos en el Reino Unido, de manera que dicha información sea accesible a la sociedad en general.

También proporciona una serie de informes e información publicada por las instituciones sobre la calidad y niveles de sus cursos.

Todos los datos actualizados incluyen información sobre las calificaciones de entrada, cuántos continúan los estudios, cuántos graduados y con qué calificación y qué tipo de empleo o estudios posteriores han seguido.

3.5. El juicio

Por último, y aunque tampoco contamos, de momento, con el sistema completo, recuérdese que faltan por definir los procesos de verificación y certificación, faltaría, en este recorrido, cerrar el ciclo de la auditoría y para ello, definir como se construye el juicio.

Así, en el sistema británico, en su forma global, puede decirse que el proceso de auditoría sigue el formato de una revisión por pares, si bien, se presentan algunas salvedades, como son el alto grado de sistematización y planificación de todos los pasos (ver QAA, 2006e anexo A); el número de personas de dentro de la institución afectadas directamente durante el periodo de las visitas, que se reduce, normalmente a los encargados de los procesos clave y los estudiantes y; el número de visitas que recibe la institución antes de la firma del informe final, que son un mínimo de tres: la reunión preliminar; la visita documental (*Briefing visit*) y la auditoría propiamente dicha (*Audit Visit*).

Una vez contrastada la información de la institución con los criterios expuestos anteriormente, el equipo auditor emitirá unos *juicios* de valor categorizados en tres niveles: confianza, confianza limitada y no confianza, así como una serie de *recomendaciones* también categorizadas en tres niveles: esenciales, aconsejables, recomendables. Todos ellos van encaminados hacia la confianza que razonablemente puede tenerse en la adecuación de la gestión de los niveles académicos de los títulos y en la adecuación de la gestión de las oportunidades de aprendizaje disponibles para los estudiantes tanto en el momento en el que se emite el juicio como en un futuro próximo (QAA, 2006e). Es decir, el juicio no se basa en los resultados obtenidos, sino en el modo en el que la institución asegura el mantenimiento y mejora de esos niveles:

El juicio de los equipos auditores no se centra en los niveles en sí, sino en el modo en el que la institución se asegura de que sus niveles académicos están asegurados mediante el trabajo de sus examinadores, internos y externos, comparados con los niveles de referencia del Academic Infrastructure. Tampoco se trata del juicio sobre las oportunidades de aprendizaje en sí, sino en el modo en el que la institución asegura que estas oportunidades son de una adecuada calidad, en contraste con el Academic Infrastructure. (p.7)

Cerrado el proceso, y transcurridos normalmente 3 años después de la auditoría se lleva a cabo un seguimiento para evidenciar los progresos llevados a cabo por la institución desde que la auditoría tuvo lugar.

4. DISCUSIÓN

El sistema de evaluación institucional británico tiene una gran tradición e historia, y ha resultado modelo de referencia obligado para todos los sistemas de evaluación de la calidad de la educación superior que se han ido instaurando en los diversos países, sobre todo europeos.

Sin embargo, lejos de tratarse de un modelo estático, parece quedar claro que ha ido evolucionando como fruto del análisis de esa trayectoria, y esa evolución ha derivado en una simplificación del sistema, lo que no significa una relajación en los objetivos de control, sino que más bien parece responder a una

adaptación de las nuevas circunstancias. Es decir, después de casi 15 años de trayectoria (las primeras evaluaciones acreditadas a nivel nacional datan de 1993) las universidades ya contaban con una infraestructura suficiente que le permitía abordar el tema de la calidad desde su propio seno. Por lo que parece lógico que el sistema se desarrolle hacia el aseguramiento de que esa infraestructura funciona adecuadamente.

En este sentido, en nuestro país también podemos hablar de evolución, pero únicamente en términos de mejora del proceso de evaluación y en la aparición de nuevos procesos, que añaden elementos nuevos, pero en ningún momento parecen sustituir a los sistemas existentes, por lo que este nuevo programa Audit, podría convertirse en un programa más, y creo que el riesgo de sobrecarga de las instituciones debería empezar a plantearse. Con ello no quiero decir que en otros sistemas como el británico, analizado aquí, no se lleven a cabo otros procesos de evaluación, sino que esos procesos se han dejado en manos de las propias universidades que deciden quién y como se han de evaluar⁴, y son objeto de análisis en la auditoría institucional. La ventaja de este acercamiento es que es la evaluación deja de verse como algo impuesto, sino que es algo necesario dentro del sistema.

En segundo lugar, debe destacarse la evolución del punto de vista en el control, que ya no pasa únicamente por la comparación de resultados a partir de indicadores de rendimiento cuantitativos. Sino que se hace una incursión hacia los procesos clave en forma de criterios de comparación de corte cualitativo. Es decir, si cualquier evaluación de calidad debía tener en cuenta los objetivos de la institución (Ginés-Mora, 1998; IBHE, 2003; Johnes y Taylor, 1990b; Moravcsik, 1986; Tejedor, 1997) en el caso británico, en la actualidad, las instituciones han consensuado cuáles son esos objetivos valiosos comunes a todas ellas y a partir de ahí se han elaborado los diferentes documentos a los que se ha denominado *Academic Infrastructure* que sirven de base a los auditores para emitir sus juicios.

Aquí es donde considero que tenemos nuestra gran asignatura pendiente, hasta la fecha no ha aparecido ningún documento ni hay iniciativa conocida aparte de las que estén relacionadas con los planes de estudio, que anime a la discusión conjunta sobre qué consideramos valioso en nuestras instituciones de tal manera que del resultado de dicho debate pueda construirse un modelo de referencia para la evaluación.

Por último, un tercer elemento digno de mención, y que considero que deberíamos tener presente a la hora de establecer alguna modificación en los sistemas existentes es el del control del cambio, es decir, aun tratándose de un sistema de rendición de cuentas, sin lugar a dudas, la auditoría institucional en el sistema británico no se conforma con el aseguramiento de los mínimos exigidos, sino que se controla la planificación y ejecución del planes de mejora con un sistema de seguimiento más propio de una evaluación continua de corte formativo que de una evaluación final de corte meramente sumativo.

Sin embargo, la evolución de todos estos aspectos no ha ido exenta de problemas y resistencias por parte de las instituciones, que han visto como en pocos años, cinco, para ser exactos, el sistema ha dado un giro radical que ha obligado y está obligando a reestructurar la infraestructura existente para adaptarla a los nuevos cambios.

⁴ A excepción de Ciencias de la Salud y Maestro que continúan llevando a cabo, debido a sus específicas características evaluaciones de calidad como programas individuales (denominados Major Review y Offsed respectivamente).

Así, el primer cambio necesario es el de la centralización de procesos, con los sistemas anteriores, centrados en titulaciones o áreas, cada unidad era la responsable del aseguramiento de sus niveles académicos y su oferta. Con el sistema basado en la Institución las Facultades pierden autonomía, y con ello la oportunidad de resaltar aquellos aspectos de sus programas que podrían considerarse como puntos fuertes.

El segundo cambio, y como consecuencia del anterior, radica en la necesidad de unificar los diferentes procesos. Resulta altamente complicado construir una evidencia documental de ámbito institucional cuando cada facultad ha desarrollado procesos diferentes para asegurar su calidad. De hecho, en el estudio particular realizado en la Universidad Metropolitana de Leeds, que será objeto de una auditoría centrada en la Institución en el año 2009, las facultades están en este momento inmersas en un proceso de identificación de evidencias (algunas utilizan la técnica de *Brainstroming* para ello) de los diferentes procesos de tal manera que puedan localizarse aquellas comunes a todas ellas y construir de este modo la historia global de cada proceso clave.

Llegados a este punto, y ya dentro de nuestro contexto, en los últimos años, sobre todo desde la puesta en marcha de la LOU en 2001 y la creación de la Agencia Nacional (ANECA) y con el telón de fondo de la Convergencia Europea, el debate se ha centrado en la necesidad de establecer un sistema de acreditación de calidad en cuanto a si ésta debería focalizarse hacia titulaciones o hacia instituciones.

Por supuesto no voy aquí a proporcionar una respuesta a este debate, simplemente, considerando la experiencia en este tema de un país como el Reino Unido con una tradición e historia muy superior a la nuestra, plantearé una serie de presupuestos que podrían ayudar a redirigir éste hacia otros puntos de interés.

En primer lugar, nuestras infraestructuras de calidad se encuentran básicamente centralizadas, en cuanto a que en todas las universidades existen Oficinas de Calidad o equivalentes, al mismo tiempo que las evaluaciones realizadas, aunque basadas en titulaciones, siempre han contado con estas oficinas como facilitadoras del proceso. Este hecho parecería avalar la conveniencia u oportunidad de una auditoría a nivel institucional. Las ventajas de esta opción ya se han mencionado a lo largo de este artículo y al no contar con los inconvenientes surgidos en Gran Bretaña parece la opción más obvia.

Dejando a un lado el foco de la auditoría, otro punto a nuestro favor se ubicaría en la unificación de los procedimientos para recabar las evidencias que sin duda alguna, mejora el inexistente procedimiento británico.

Sin embargo, en contrapartida, la tarea pendiente sería la del consenso, es decir, lejos de debatir la importancia o no de que las titulaciones sean las mismas o equivalentes, lo que no puede dejar lugar a dudas es la necesidad de determinar cuáles son los elementos básicos de calidad que toda institución debe garantizar, es decir, cuáles son los objetivos clave comunes a todas las universidades y, a partir de ahí consensuar los criterios para su evaluación.

En el programa Audit, aun sin quedar claro el nivel de análisis, y siguiendo la línea anterior de otros programas como el de evaluación institucional o el de acreditación (ANECA, 2007a), se enmarcan una serie de criterios básicos que, sin dudar de su importancia, se reducen, básicamente a una lista de elementos que basta señalar si existen o no. Es decir, hasta ahora están faltos de contenido, falta el consenso entre las instituciones que dé un significado exacto a cada uno de los criterios junto con un conjunto de buenas prácticas asociadas a cada uno de ellos, de tal manera que el propio documento se

convierta en una herramienta útil de trabajo para la mejora continua de la calidad de nuestras universidades, además de proporcionar una base sólida para el aseguramiento de la calidad.

Por último, una preocupación y una cuestión en el aire ¿Qué ocurre con la investigación? En ambos sistemas, el británico y el propuesto en nuestro país se hace referencia explícita a los grupos de interés, pero en ningún caso esos grupos de interés parecen mostrar preocupación alguna por la investigación. Dentro del sistema británico la filosofía está claramente definida (ello no quiere decir que yo la comparta) es el organismo que financia el que evalúa, y para ello tienen la RAE (*Research Assessment Exercise*) En nuestro país, tradicionalmente la investigación ha sido siempre incluida en los planes de evaluación de las titulaciones ¿Es posible que no exista un espacio para ella en el programa audit?

REFERENCIAS BIBLIOGRÁFICAS

- Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) (2007a). Plan de Evaluación Institucional. http://www.aneca.es/active/active_ense_pei.asp/
- Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) (2007b). Programa Audit. Documentos de trabajo 1, 2 y 3. http://www.aneca.es/active/active_audit.asp/
- Consejo De Universidades (1992). *Programa Experimental de Evaluación*. (versión 0). Madrid: Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1994). *Programa Experimental de Evaluación de la Calidad del Sistema Universitario*. (Documento 13). Madrid: Ministerio de Educación y Ciencia, Comité Técnico de la Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1995). *Programa de Evaluación Institucional de la Calidad de las Universidades*. Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1996a). *Plan Nacional de Evaluación de la Calidad de las Universidades*. Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1996b). *Plan Nacional de Evaluación de la Calidad de las Universidades. Guía de Evaluación*. Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1997a). *Plan Nacional de Evaluación de la Calidad de las Universidades. Guía de los evaluadores externos*. Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1997b). *Evaluación de la Calidad de las Universidades. Informe*. Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Consejo De Universidades (1998a). Plan Nacional de Evaluación de las Universidades. *Jornada de formación* del 15 de julio 1998.
- Consejo De Universidades (1998b). *Plan Nacional de Evaluación de la Calidad de las Universidades. Guía de Evaluación*. Ministerio de Educación y Ciencia, Secretaría General del Consejo de Universidades.
- Ferrández, R. (1999). *Indicadores de calidad en evaluación institucional universitaria. Análisis de un indicador para la universitat Jaume I*. Valencia: Servicio de Publicaciones de la Universitat de València.
- Ginés-Mora, J. (1998). Evaluación Institucional de la Universidad. *Revista de Educación*, 315. pp. 29-44.
- Higher Education and Research Opportunities (HERO) (2007a). Education Quality and standards: Teaching Quality Information. En: http://www.hero.ac.uk/uk/inside_he/education_quality_and_standards/teaching_quality_information__tqi_.cfm (última visita el 30 de mayo de 2007).

- Higher Education and Research Opportunities (HERO) (2007b). Education Quality and standards: National Students Survey. En: http://www.hero.ac.uk/uk/inside_he/education_quality_and_standards/national_student_survey__nss_.cfm (última visita el 30 de mayo de 2007).
- Higher Education Funding Council for England (HEFCE) (2005) *Review of the Quality Assurance Framework*. (Burslem Report). Phase one outcomes. London: HEFCE. <http://www.hefce.ac.uk/pubs/hefce/2005/>
- Illinois Board of Higher Education (IBHE) (2003) *Establishing Performance Indicators to Assess Progress toward meeting the goals of the Illinois Commitment. Final recommendations*. Educational Resources of Information Center (ERIC).
- Johnes, J. y Taylor, J. (1990). Performance Indicators in Higher Education. *Higher Education*, 23(2), pp. 209-210.
- JM Consulting (2005). *The costs and benefits of external review of quality assurance in higher education*. http://www.hefce.ac.uk/pubs/rereports/2005/rd17_05/rd17_05.pdf
- Ling, P. (2005). Autonomy versus accountability: Measuring university teaching performance. En: Association for Teacher Education in Europe (ATEE), *30e Conférence annuelle de l'ATEE*, Amsterdam, 22-26 octobre 2005
- Moravcsik, (1986). Assessing the Methodology for Finding a Methodology for Assessment. *Social Studies of Science*, 26, pp.534-39.
- Quality Assurance Agency for Higher Education (QAA) (1999). *Section 3: Students with disabilities*. En *Code of practice for the assurance of academic quality and standards in higher education*. http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section3/COP_disab.pdf
- Quality Assurance Agency for Higher Education (QAA) (2000). *The framework for higher education qualifications in England, Wales and Northern Ireland*. <http://www.qaa.ac.uk/academicinfrastructure/FHEQ/EWNI/default.asp>
- Quality Assurance Agency for Higher Education (QAA) (2001). Section 8: Career education, information and guidance. En: *Code of practice for the assurance of academic quality and standards in higher education*. <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section8/careereducation.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2004a). *Foundation Degree qualification benchmark*. <http://www.qaa.ac.uk/reviews/foundationDegree/benchmark/FDQB.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2004b). *Guidelines on the accreditation of prior learning* <http://www.qaa.ac.uk/academicinfrastructure/apl/APL.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2004c). Section 1: Postgraduate research programmes. En *Code of practice for the assurance of academic quality and standards in higher education*. <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section1/postgrad2004.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2004d). Section 2: Collaborative provision and flexible and distributed learning (including e-learning). En *Code of practice for the assurance of academic quality and standards in higher education*. <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section2/collab2004.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2004e). Section 4: External examining. En *Code of practice for the assurance of academic quality and standards in higher education*. http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section4/COP_external.pdf
- Quality Assurance Agency for Higher Education (QAA) (2005). *Handbook for academic review: For review of directly funded higher education in further education colleges* <http://www.qaa.ac.uk/reviews/academicReview/acrevhbook2004/HandbookAcademicReview.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2006a). *Guidelines for preparing programme specifications*. <http://www.qaa.ac.uk/academicinfrastructure/programSpec/guidelines06.pdf>

- Quality Assurance Agency for Higher Education (QAA) (2006b). Section 10: Admissions to higher education. En *Code of practice for the assurance of academic quality and standards in higher education*. <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section10/RecruitmentandAdmissions.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2006c). Section 6: Assessment of students. En *Code of practice for the assurance of academic quality and standards in higher education*. http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section6/COP_AOS.pdf
- Quality Assurance Agency for Higher Education (QAA) (2006d). Section 7: Programme design, approval, monitoring and review. En *Code of practice for the assurance of academic quality and standards in higher education*. <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section7/programmedesign.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2006e). *Handbook for institutional audit: England and Northern Ireland* <http://www.qaa.ac.uk/reviews/institutionalAudit/handbook2006/handbookComments.pdf>
- Quality Assurance Agency for Higher Education (QAA) (2007a). Section 5: Academic appeals and student complaints on academic matters En *Code of practice for the assurance of academic quality and standards in higher education*. http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section5/COP_complaints.pdf
- Quality Assurance Agency for Higher Education (QAA) (2007b). Section 9: Placement Learning. En *Code of practice for the assurance of academic quality and standards in higher education*. <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section9/PlacementLearning.pdf>
- Tejedor, F.J. (1997). Informes de Evaluación Institucional. En *Revista Española de Pedagogía*, 208. pp. 471-498.



USO DA TEORÍA DE RESPONSA AO ÍTEM (TRI) PARA ANALIZAR A EQÜIDADE DO PROCESSO DE AVALIAÇÃO DO APRENDIZADO DISCENTE*

Wagner Bandeira Andriola

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 1

<http://www.rinace.net/riee/numeros/vol1-num1/art12.pdf>

* Este texto é parte de um capítulo da tese doutoral do autor, intitulada Detección del Funcionamiento Diferencial del Ítem (DIF) en Tests de Rendimiento. Aportaciones Teóricas y Metodológicas, defendida publicamente em 27 de novembro de 2002 no Departamento de Métodos de Investigación y Medidas en Educación (MIDE) da Universidad Complutense de Madrid (UCM).

J á que os fins educativos consistem, essencialmente, em determinar as mudanças qualitativas por que passam os aprendizes, em termos de aquisição de aprendizagens, podemos garantir, portanto, que *a avaliação da aprendizagem é o processo de determinar em que medida ou grau se conseguem tais mudanças, possibilitando, assim, um juízo de valor acerca da qualidade dessas supostas mudanças*. Nesta definição, há duas idéias básicas:

- *O objeto da avaliação é o comportamento*, que é aqui entendido como sendo determinado por atributos intrapsíquicos ou organísmicos, tais como: motivação, compreensão, raciocínio, memorização, entre outros, pelo qual a tarefa avaliadora consiste na busca das manifestações externas que sirvam de indicadores dos aspectos internos, não avaliados diretamente (Andriola, 1998a; 2002). As relações entre os atributos internos e as manifestações externas —p. ex. as respostas aos itens de um teste de rendimento— têm por base o modelo correlacional;
- *A avaliação deve ser um processo sistemático de coleta e análise de informações*, que deve proporcionar pelo menos duas medidas: uma no início do processo de ensino e outra, no final, pois sua finalidade última é a emissão de um juízo de valor acerca das mudanças de comportamento decorrentes da aprendizagem. É interessante dizer que todo *testemunho* —p.ex.: conversas informais, debates e seminários em sala de aula, frequência e participação dos alunos, dentre outras formas— é válido para avaliar a aprendizagem (Mc Donald, 2003). Os testes de rendimento constituem uma dessas inumeráveis formas de *dar testemunhos*, que permitem a obtenção de informações úteis, válidas e fidedignas sobre a aprendizagem de estudantes. Não obstante, em qualquer *testemunho*, existem algumas etapas que devem ser seguidas:
 - a) seleção de comportamentos ou ações que sejam indicadores dos objetivos curriculares que se deseja avaliar —tipo de objetivo educativo a ser avaliado;
 - b) apresentação de tarefas pedagógicas que possibilitem aos alunos expressar o que sabem sobre um determinado conteúdo —tipo de tarefa pedagógica a ser utilizada.

Como em toda e qualquer atividade científica, porém, a avaliação do rendimento possui vulnerabilidades ou limitações, que são descritas a seguir.

1. ORIGEM DOS ESTUDOS SOBRE O VIÉS DOS TESTES

O viés dos instrumentos de medida psicológica e educacional é um tópico que aparece tratado tardiamente, no seio da psicometria moderna (Muñiz, 1994). De acordo com Angoff (1993), seu estudo sistemático se iniciou nos Estados Unidos, no final dos anos 1960, numa época em que estavam em moda os debates sobre direitos civis e desigualdades de oportunidade entre brancos e outras minorias étnicas (Rossi, Freeman e Lipsey, 1999).

Os resultados dos processos de avaliação educacional, executados por instituições reconhecidas, tais como o *Educational Testing Service (ETS)*, foram discutidos entre diversos intelectuais, para quem as diferenças de rendimento educativo observadas entre os diversos grupos étnicos e socioeconômicos refletiam, na realidade, disparidades nas oportunidades educacionais e discriminação contra grupos

minoritários, tais como negros, hispano-americanos, judeus e árabes. Foi, portanto, a discussão social, alheia em grande parte ao círculo psicométrico especializado, que obrigou os especialistas da área a produzirem outros procedimentos estatísticos, com o objetivo de provar que seus testes ou instrumentos de medida não tinham nenhum tipo de viés (Cole, 1993).

Nessa mesma época, os investigadores começaram a preocupar-se com o estudo sistemático das diferenças entre grupos demográficos. Estavam interessados em buscar explicações a respeito das suas verdadeiras causas. Martínez Arias (1997) destaca que a investigação sobre o viés dos itens pode remontar aos estudos realizados por A. Binet, em 1910, a respeito das diferenças de *status* socioeconômico no rendimento dos sujeitos submetidos a alguns testes desenvolvidos por ele próprio. Os resultados possibilitaram a proposição da hipótese de que o rendimento mais baixo destes sujeitos, em alguns itens, poderia decorrer do efeito do treinamento cultural, em vez de reais diferenças na capacidade mental ou construto latente medido pelo teste. Também W. Stern, o introdutor da expressão *quociente intelectual*, pode ser considerado como um dos primeiros investigadores da área, visto que ele estudou as diferenças relacionadas com a classe social, na Alemanha (Andriola, 2002).

Apesar desses autores pioneiros, o começo da moderna investigação sobre o viés encontra-se nos trabalhos de K. Eells, A. Davis, R. J. Havighurst, V. E. Herrick e R. W. Tyler, realizados na Universidade de Chicago, em 1951. Nesses estudos, os investigadores encontraram variações nos itens, em alguns aspectos muito peculiares, tais como conteúdo e formato, que reduziam ou exageravam as diferenças observadas entre os grupos comparados. Surgiram, assim, os primeiros dados a respeito dos problemas técnicos que possuíam alguns itens dos testes de rendimento, então utilizados na avaliação da aprendizagem. Eram informações sobre os problemas referentes ao uso indevido da linguagem escrita, que possibilitava certa vantagem de um grupo de sujeitos sobre outro, isto é, muitos dos termos empregados nos testes eram mais familiares a alguns grupos específicos de estudantes, tais como os norte-americanos brancos, originários da classe média. Em consequência, os sujeitos pertencentes aos grupos minoritários, que não conheciam ou não empregavam cotidianamente esses termos, tinham rendimento mais baixo. Apareceu, então, o interesse pela investigação sistemática pelo *funcionamento diferencial do item* (DIF).

2. PRESENÇA DE DIF: FENÔMENO QUE SE CHOCA COM A EQUIDADE DA MEDIDA

No âmbito da Teoria de Resposta ao Item (TRI), o item não tem DIF, quando a sua curva característica (CCI) é idêntica para os grupos comparados, considerando-se um mesmo nível ou magnitude da variável latente medida (LORD, 1980; MELENBERGH, 1989). Em linguagem matemática podemos dizer que o item não tem DIF com respeito à variável G (grupo) dado Z (nível de θ) se, e somente se, $F(X | g, z) = F(X | z)$, onde:

- X é a pontuação no item;
- g é o valor obtido dado a variável G ;
- z é o valor obtido dado a variável Z .

Nesse contexto, os valores esperados (E) são dados por $E(X | g, \theta) = E(X | \theta)$ para todo g e θ . No caso de itens dicotômicos, os valores esperados são as probabilidades de acerto ao item, que podem ser

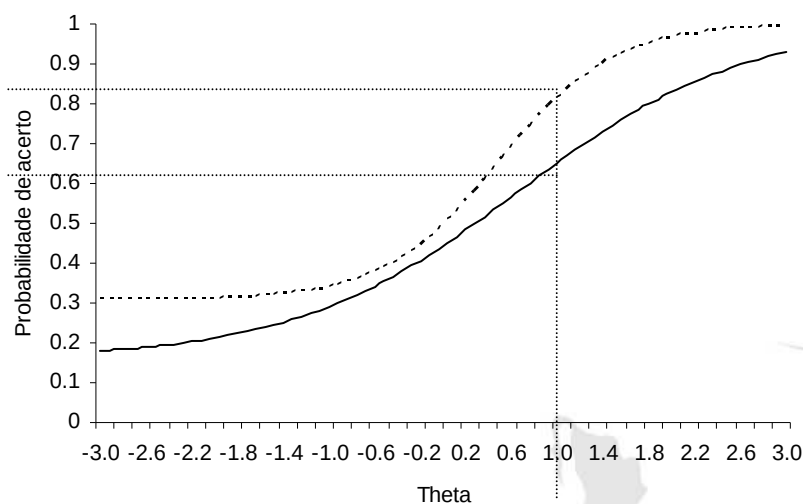
expressas nos seguintes termos: $P(X=1 | g, \theta) = P(X=1 | \theta)$ para todo g e θ . No segundo caso, a equação expressa, em realidade, a curva característica do item (CCI).

Geralmente, os estudos para a determinação do DIF utilizam dois grupos, denominados: *grupo de referência (GR)* e *grupo focal (GF)*. Como já enfatizamos, no âmbito da TRI um item tem DIF se para valores iguais de θ não correspondem valores iguais de $P(\theta)$ nas CCI's dos grupos considerados, isto é, quando $T_{jGR}(\theta) \neq T_{jGF}(\theta)$, onde:

- T_{jGR} é a pontuação verdadeira do sujeito j pertencente ao grupo de referência e que possui certa magnitude na variável latente θ ;
- T_{jGF} é a pontuação verdadeira do sujeito j pertencente ao grupo focal e que possui certa magnitude na variável latente θ .

Para visualizar o DIF de um hipotético item, apresentamos, a seguir, a Figura 1.

FIGURA 1. REPRESENTAÇÃO DAS CCI'S DE UM SUPOSTO ITEM COM DIF



Legenda: Linha descontinua - CCI das mulheres (GR); Linha contínua - CCI dos homens (GF).

Na Figura 1, observamos que, para uma mesma magnitude de θ , a probabilidade de acerto ao item $P(\theta)$ é sempre superior para as mulheres. Isto indica que em níveis iguais de competência na variável medida θ não há igualdade na probabilidade de acertar o item. Neste caso, o item está enviesado *contra* os homens (GR), pois os valores de $P(\theta)$ para um mesmo nível de θ são sempre maiores para as mulheres (GF). Por exemplo, para $\theta = 1,1$ temos valores aproximados de $P(\theta) = 0,69$ para os homens e $P(\theta) = 0,87$ para as mulheres.

Como consequência de resultados dessa natureza, Douglas, Roussos e Stout (1996) propuseram os conceitos DIF benigno e DIF adverso. No caso do DIF beneficiar o grupo de referência (GR), isto é, quando $T_{jGR}(\theta) > T_{jGF}(\theta)$, caracteriza-se a existência de DIF benigno (ANDRIOLA, 2006). O DIF adverso ocorre no

caso do DIF beneficiar ao grupo focal (GF), ou seja, quando $T_{jGR}(\theta) < T_{jGF}(\theta)$. No exemplo da Figura 1, temos um caso de DIF adverso.

3. DIFERENÇAS ENTRE CENTROS EDUCACIONAIS: CONTROVÉRSIAS E DADOS

Nessa mesma época, os estudiosos da área educacional demonstraram que as pessimistas conclusões de J. S. Coleman e colaboradores, em 1966, e de C. S. Jencks e colaboradores, em 1972, a respeito da pouca ou nula influência dos centros educacionais sobre os resultados acadêmicos dos estudantes, são totalmente equivocadas (Andriola, 2002). Os dados de estudos executados nos últimos 20 anos permitem-nos anotar: o efeito que um centro educacional tem sobre a aprendizagem dos seus estudantes pode ser identificado e, até certo ponto, medido (OCDE, 1995; Nuttall, Goldstein, Prosser e Rabash, 1989).

Ademais, como destaca Orden Hoz (1993), as investigações a respeito das influências dos centros educacionais sobre o rendimento acadêmico têm um papel preponderante em muitos países. Vários estudos foram executados com o objetivo de tentar caracterizar os fatores que diferenciam uns centros educacionais de outros, no que se refere a associação dos mesmos ao rendimento acadêmico dos estudantes (Soares, 2002).

Por exemplo, Miles (1974) destacou dez características de um centro educacional são: possuir objetivos claros; ter um bom sistema de comunicação; regras claras de hierarquia; utilização racional dos recursos; coesão entre seus membros; moral elevada de seus membros; preocupação com a inovação; autonomia; adaptação e equilíbrio nas técnicas de resolução de problemas. Estas características têm elevado grau de interdependência e, ademais, definem um marco apropriado de índices indiretos de qualidade, pois são aspectos que, indubitavelmente, condicionam em um sentido ou em outro o funcionamento do centro educacional (De Miguel, 1989).

Sammons, Hillman e Mortimore (1998) destacaram onze fatores responsáveis pela alta eficácia de alguns centros educacionais: capacidade de liderança dos diretores; visão e objetivos comuns; ambiente adequado à aprendizagem; ênfase no processo ensino-aprendizagem; preocupação pela qualidade do ensino; existência de expectativas positivas elevadas; uso de reforços explícitos; acompanhamento regular do progresso dos alunos; explicitação dos direitos e deveres dos alunos; colaboração lar-centro educacional; e, por fim, preocupação pela qualificação profissional.

Segundo Moura Castro (1999), os principais fatores que caracterizam os centros educativos de qualidade são:

- Destinar a maior quantidade de horas possíveis ao envolvimento dos estudantes com as suas tarefas ou atividades escolares;
- Selecionar de bons professores;
- Preocupar-se pela formação e qualificação dos professores;
- Fazer com que os professores sintam-se responsáveis pelos êxitos dos seus alunos;
- Fazer com que os professores utilizem metodologias adequadas às características sociais e cognitivas dos seus aprendizes;
- Valorar o papel social do educador.

Muitas outras investigações sobre as diferenças entre os centros educativos põem ênfase sobre os processos instrutivos relacionados, fundamentalmente, com o contexto de ensino-aprendizagem (Dunkin, 1978; Horsburgh, 1999; Moura Castro, 1999; Rego, 2000, 2001). Nesse âmbito, existem estudos a demonstrar que o modo como os estudantes estejam envolvidos nas atividades educativas é um fator muito importante para explicar suas aprendizagens. De acordo com Alexander e Judy (1988), os estudantes mais envolvidos nas atividades escolares demonstram maior capacidade para organizar e associar as novas informações com as antigas, gerando, assim, outros conhecimentos. Nuthall (1999) destaca o fato de que o conhecimento é resultado da utilização de um conjunto de processos cognitivos reforçados nos âmbitos social, cultural e educativo. No caso do contexto educativo, a dinâmica utilizada na sala de aula é um dos fatores mais determinantes para que o aprendiz adquira esses processos cognitivos.

Tobias (1994), Nuthall e Alton-Lee (1995) identificaram quatro fatores primários, que têm grande poder explicativo para as diferenças individuais do rendimento:

- A compreensão dos objetivos das disciplinas;
- A participação nas atividades acadêmicas de grupo;
- Os conhecimentos anteriores e as crenças no êxito pessoal; e
- O interesse e motivação pessoal.

Dos 10 aspectos destacados por Miles (1974) e dos 11 enunciados por Sammons, Hillman e Mortimore (1998), alguns têm grande similitude com os *fatores primários* pinçados por Tobias (1994), Nuthall e Alton-Lee (1995), quais sejam: *objetivos claros e compartilhados*, *boa comunicação entre os membros do centro educativo*, *coesão*, *moral elevada* e, finalmente, *preocupação com a inovação*.

Os *objetivos escolares* devem possuir as desejáveis características de serem claros e aceitos pelos membros do centro educativo. Ademais, devem ser alcançáveis com os recursos disponíveis e apropriados para as demandas do contexto. As *boas comunicações* devem sofrer o mínimo de distorção no percurso que vai do emissor ao destinatário, isto é, as tensões e problemas devem ser rapidamente identificados em virtude do uso de uma boa comunicação. A comunicação tem efeito sobre a *coesão*, já que este último tem efetiva ligação com o autoconhecimento do centro educativo, em seu conjunto e sobre as partes constituintes. A *moral elevada*, no âmbito organizativo, está associada à idéia de *soma de sentimentos individuais de satisfação*, que apóiam os desejos de realizar esforços para alcançar os objetivos planejados. Finalmente, a *preocupação com a inovação* é a característica desejável de mover-se em direção de novos objetivos e procedimentos.

Várias características da escola são fundamentais para lograr que seus alunos alcancem os propósitos educativos que lhes permitam continuar desenvolvendo-se e aprendendo com autonomia. Não obstante, de acordo com Sammons, Hillman e Mortimore (1998), características socioeconômicas, de gênero, etnia e linguagem também exercem influência sobre os resultados acadêmicos. Mortimore et al. (1988) demonstraram que, em termos do progresso estudantil (valor agregado pela instituição educacional), os efeitos da escola são muito mais importantes do que fatores como idade, gênero e classe social.

Finalmente, Scheerens (1992), Fuller e Clark (1994) fazem algumas reflexões a esse respeito, ressaltando que os efeitos das escolas podem variar para diferentes tipos de conteúdos, sendo maiores para Matemática e Ciências, que são ensinadas basicamente no ambiente educacional, do que para leitura ou línguas estrangeiras, mais suscetíveis a influências do lar.

As diversas investigações aportam informações válidas sobre a influência que exercem os centros educativos sobre o rendimento escolar dos aprendizes (Soares, 2002). Nesse âmbito, cabe mencionar que as profundas diferenças entre as escolas públicas e privadas brasileiras, em muitos dos fatores destacados por Miles (1974), Sammons, Hillman e Mortimore (1998), Tobias (1994), Nuthall e Alton-Lee (1995), têm reflexo sobre aspectos facilmente perceptíveis. Assim, por exemplo, Barreto, Trompieri Filho e Andriola (1999) assinalam que as diferenças no rendimento acadêmico dos alunos desses dois tipos de escolas podem repercutir sobre suas crenças pessoais de êxito escolar.

Já o estudo executado por Andriola (1997b) demonstrou que os alunos desses dois tipos de escolas têm distintas expectativas sobre a universidade brasileira e isso reflete, em certo sentido, as crenças no êxito pessoal. Enquanto 70% dos alunos das escolas particulares crêem que cursar uma carreira universitária lhes garantirá obter uma profissão de qualidade, apenas 11% dos alunos de escolas públicas concordam com essa opinião. De forma análoga, 89% dos alunos de escolas públicas crêem que outras variáveis, tais como: o esforço pessoal e o fato de cursar uma carreira universitária, lhes permitirão obter uma profissão de qualidade. Por outro lado, tão-somente 30% dos estudantes de escolas particulares compartilham essa mesma opinião.

Com base no exposto, pode-se inferir que existem distinções, qualitativas e quantitativas, nas experiências sociais e educacionais dos alunos de escolas particulares e públicas (Ramos, 1999; Nuthall, 1999). No caso brasileiro, as diferenças nessas experiências influem, entre outras coisas, nas opções de trabalho e no grau de aprendizagem em algumas disciplinas, tais como línguas e matemática (Andriola, 1995; 1997ab; 1998c, 1999; Ramos, 1999). Nesse contexto, é quase inevitável que alguns itens presentes em ambos os testes reflitam essas experiências em seus conteúdos e, desse modo, venham a possibilitar a vantagem de um grupo sobre o outro, isto é, determinem a presença do DIF.

Hamilton (1999) observou que, em alguns casos de DIF favorável às mulheres são exigidos determinados conhecimentos obtidos fora da escola, sobretudo em itens utilizados na avaliação do conhecimento em ciências. Também Andriola (2001 d) encontrou seis itens com DIF em um banco de itens organizado para a avaliação do raciocínio verbal de estudantes brasileiros do ensino médio (Andriola, 1998b), sendo três deles benignos aos alunos de escolas públicas (GR) e os outros três benignos aos alunos de escolas particulares (GF).

Por outro lado, Clauser, Nungester e Swaminathan (1996) estudaram a presença de DIF em 440 itens de um teste para medir a aptidão para a Medicina Clínica utilizado pelo *National Board of Medical Examiners*. Os autores supuseram que os itens com DIF poderiam medir alguma aptidão secundária que, de algum modo, pudera possibilitar que certos grupos de indivíduos, com características demográficas muito específicas, avantassem outros indivíduos de grupos com características demográficas distintas. O fato de haverem realizado residência médica pode, supostamente, proporcionar aos estudantes a aquisição de habilidades que lhes possibilita terem vantagem na resolução das tarefas presentes nos itens utilizados. Assim, em comparação com os alunos com a mesma habilidade, os que fizeram residência médica tiveram maior probabilidade de acertar o mesmo item em até 30% dos casos.

4. HIPÓTESES DA INVESTIGAÇÃO

Com fundamento nos dados e nas diversas informações resultantes dos estudos relatados, elaboramos duas hipóteses para serem testadas na presente investigação. A primeira (H_1) indica que,

"Considerando-se o tipo de escola dos alunos, existirão itens, nos Testes de Português e Matemática, com DIF favorável aos alunos de escolas particulares (GR)".

Expressando a hipótese em notação matemática, temos:

$H_0: P_{GRi}(\theta_s) = P_{GF i}(\theta_s)$ [indica a ausência de DIF].

$H_1: P_{GRi}(\theta_s) > P_{GF i}(\theta_s)$ [indica a presença de DIF benigno al GR].

Onde:

$P_{GRi}(\theta_s)$ e $P_{GF i}(\theta_s)$ são, respectivamente, as probabilidades do grupo de referência – GR (alunos de escolas particulares) e focal – GF (alunos de escolas públicas) de acertar o item i , dado certa magnitude (s) da variável latente medida (θ).

Nesse contexto, há que se destacar que para detectar a presença de itens com DIF foi utilizado o método Mantel-Haenszel, que foi desenvolvido por N. Mantel e W. Haenszel no ano 1959, e aplicado ao estudo do DIF por P. W. Holland e D. T. Thayer em 1988 (Angoff, 1993; Dorans y Holland, 1993). Consiste, basicamente, na comparação das frequências observadas e esperadas de acertos e erros nos grupos de referência e focal, de acordo com os distintos níveis de habilidade (j) escolhidos pelo investigador.

A segunda hipótese (H_2) da investigação aponta que

"Os alunos de escolas particulares (GR) terão melhores rendimentos que os de escolas públicas (GF) nos itens dos Testes de Português e Matemática".

Em notação matemática, temos:

$H_0: \mu_1 = \mu_2$ [indica a ausência de diferença entre o GR e o GF].

$H_2: \mu_1 > \mu_2$ [indica a existência de diferença entre o GR e o GF, favorável ao GR].

Onde:

μ_1 e μ_2 são, respectivamente, as pontuações médias dos alunos de escolas particulares (GR) e de escolas públicas (GF) em cada um dos testes.

Os alunos que estudaram todo o ensino fundamental (I e II) e médio em escolas particulares compõem o GR ($n = 17.763$), enquanto aqueles que o estudaram em centros educacionais públicos compõem o GF ($n = 4.441$).

5. MÉTODO EMPREGADO

A investigação executada foi do tipo correlacional (*ex post-facto*), já que os dados foram coletados *in situ* e, ademais, não houve nenhum tipo de manipulação de variáveis (Bryman e Cramer, 1992; Kvanli, 1988).

5.1. Descrição da amostra de aprendizes

Com as respostas dos 29.777 candidatos inscritos no processo seletivo da Universidade Federal do Ceará (UFC) nas provas de Língua Portuguesa e Matemática, foram efetuadas as análises para verificar a existência do DIF nos itens dos mencionados testes de rendimento. Não obstante, antes disso, é aconselhável descrever as principais características demográficas da amostra de estudantes.

O gênero feminino foi majoritário entre os respondentes, correspondendo a 16.581 casos (55,7%), sendo que 14.147 tinham idades entre 19 e 24 anos; 94,9% viviam na zona metropolitana de Fortaleza ($n = 28.224$); 64,2% haviam estudado todo o ensino fundamental em escolas particulares ($n = 19.119$) e tão-somente 15,9% em escolas públicas ($n = 4.719$). Esse dado nos faz constatar que um reduzido percentual de alunos do ensino fundamental público chega a tentar ingressar em uma universidade pública. A diferença entre o percentual de egressos procedentes de escolas particulares (64,2%) e de escolas públicas (15,9%) permite-nos obter uma imagem que demonstra a baixa qualidade do ensino público no Ceará.

5.2. Instrumentos utilizados

Foram empregados dois testes de rendimento das disciplinas de Língua Portuguesa e Matemática, ambos destinados a avaliar o grau de aprendizagem dos candidatos a ingressar na Universidade Federal do Ceará para o ano acadêmico de 2000. Essas provas foram aplicadas na primeira etapa do processo seletivo (vestibular) e compunham-se, respectivamente, de 18 e 15 questões fechadas, com cinco opções propostas como respostas. Para identificar as características demográficas dos sujeitos da amostra, foram utilizados os dados resultantes da aplicação do questionário para a caracterização socioeconômica e cultural dos candidatos.

6. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

6.1. Análise do DIF no Teste de Português segundo o tipo de escola

A Tabela 1 contém os valores do coeficiente α_{MH} , da prova de contraste (χ^2_{MH}), das probabilidades de que α_{MH} deva-se ao acaso (p) e as respectivas classificações para o DIF, nos 18 itens do Teste de Português.

TABELA 1. VALORES α_{MH} PARA OS ITENS DO TESTE DE PORTUGUÊS

ITENS	α_{MH}	χ^2_{MH}	p	TIPO DE DIF
1	0,30	7,51	0,05	Moderado
2	-0,02	0,07	n.s.	Inexistente
3	-0,35	16,07	0,05	Moderado
4	-0,11	2,03	n.s.	Inexistente
5	-0,18	3,76	0,05	Moderado
6	-0,07	0,37	n.s.	Inexistente
7	0,35	11,93	0,05	Moderado
8	0,87	45,94	0,05	Moderado
9	0,74	49,55	0,05	Moderado
10	-0,58	42,01	0,05	Moderado
11	0,20	4,74	0,05	Moderado
12	-0,18	3,04	n.s.	Inexistente
13	0,50	29,62	0,05	Moderado
14	-0,60	29,77	0,05	Moderado
15	-0,07	0,62	n.s.	Inexistente
16	0,35	15,91	0,05	Moderado
17	-0,37	9,03	0,05	Moderado
18	-0,31	12,70	0,05	Moderado

Fonte: pesquisa própria.

Os dados apresentados na Tabela 1 demonstram que tão só os itens 2, 4, 6, 12 e 15 não possuem DIF (27,8% do total); nos demais 13 itens o DIF é moderado (72,2% do total), sendo que nos itens 3, 5, 10, 14, 17 e 18 o DIF é benigno ao GR (alunos de escolas particulares), pois os valores do índice α_{MH} são negativos. Para esses casos a hipótese H_1 [$P_{GRI}(\theta_s) > P_{GFI}(\theta_s)$] foi corroborada. Nos itens 1, 7, 8, 9, 11, 13 e 16 o DIF é adverso ao GR, já que os valores do índice α_{MH} são positivos. Para esses itens, a hipótese H_1 [$P_{GRI}(\theta_s) > P_{GFI}(\theta_s)$] não foi corroborada.

6.2. Análise do DIF no Teste de Matemática segundo o tipo de escola

A Tabela 2 contém os valores do coeficiente α_{MH} , da prova de contraste (χ^2_{MH}), as probabilidades de que os valores dos α_{MH} decorrem do acaso (p) e as respectivas classificações para o DIF.

TABELA 2. VALORES α_{MH} PARA OS ITENS DO TESTE DE MATEMÁTICA

ITENS	α_{MH}	χ^2_{MH}	p	TIPO DE DIF
1	0,17	3,05	n.s.	Inexistente
2	-0,41	10,76	0,05	Moderado
3	0,74	53,17	0,05	Moderado
4	0,94	80,91	0,05	Moderado
5	0,38	19,11	0,05	Moderado
6	-0,02	0,02	n.s.	Inexistente
7	-0,14	2,00	n.s.	Inexistente
8	0,55	22,63	0,05	Moderado
9	0,27	7,36	0,05	Moderado
10	-0,84	64,38	0,05	Moderado
11	-0,63	49,71	0,05	Moderado
12	-0,37	11,01	0,05	Moderado
13	-0,37	13,46	0,05	Moderado
14	-0,22	4,44	0,05	Moderado
15	-0,25	6,68	0,05	Moderado

Fonte: pesquisa própria.

Os valores obtidos para o coeficiente α_{MH} e para a prova de contraste (χ^2_{MH}) apresentados na Tabela 2, demonstram que tão-só os itens 1, 6 e 7 não possuem DIF; nos demais 12 itens o DIF é moderado, entre os quais há cinco itens com DIF benigno ao GR (3, 4, 5, 8 e 9), pois possuem valores positivos para o coeficiente α_{MH} . Para esses itens, a hipótese $H_1 [P_{GR}(\theta_s) > P_{GF}(\theta_s)]$ foi corroborada. Para os demais sete itens (2, 10, 11, 12, 13, 14 e 15) o DIF é adverso ao GR, já que os valores negativos do coeficiente α_{MH} assim o indicam. Nesses casos, a hipótese $H_1 [P_{GR}(\theta_s) > P_{GF}(\theta_s)]$ não foi corroborada.

6.3. Análise do impacto no Teste de Português segundo o tipo de escola

O estudo do impacto em cada um dos 18 itens da prova de Português é importante para que possamos identificar e compreender as possíveis diferenças entre os alunos de tipos distintos de escolas, em termos de suas pontuações médias. Assim, poderemos saber se há alguma relação entre possíveis diferenças no impacto e a presença do DIF nos itens 3, 5, 10, 14, 17 e 18 (que possuem DIF benigno ao GR). A Tabela 3 apresenta os resultados da Análise de Variância (*one-way*) para os 18 itens do teste de Português.

TABELA 3. COMPARAÇÃO ENTRE AS PONTUAÇÕES MÉDIAS DO GR E DO GF NOS 18 ITENS DO TESTE DE PORTUGUÊS

Itens	GR			GF			F	p
	μ	(σ)	n	μ	(σ)	n		
1	112,18	(19,84)	6.103	101,12	(21,06)	965	254,97	0,000
2	111,76	(17,43)	9.884	101,87	(18,12)	1.751	473,66	0,000
3	106,38	(19,80)	9.856	95,09	(19,75)	2.055	563,16	0,000
4	107,87	(19,67)	6.917	95,76	(19,69)	1.324	421,04	0,000
5	108,34	(18,03)	12.104	98,58	(17,78)	2.388	587,69	0,000
6	110,91	(17,00)	11.593	101,59	(17,28)	2.119	536,12	0,000
7	108,89	(16,80)	14.207	99,96	(17,06)	2.655	627,87	0,000
8	105,70	(17,96)	16.572	97,09	(16,89)	3.316	646,90	0,000
9	107,61	(17,36)	14.839	99,50	(16,69)	2.759	513,46	0,000
10	108,14	(18,90)	8.086	97,49	(19,23)	1.708	444,43	0,000
11	110,14	(17,51)	12.341	100,30	(18,56)	2.224	583,89	0,000
12	105,04	(20,62)	4.424	92,85	(21,49)	898	257,34	0,000
13	109,32	(18,34)	9.377	99,92	(18,76)	1.550	350,38	0,000
14	108,41	(21,04)	3.611	96,51	(21,04)	779	204,72	0,000
15	112,48	(20,69)	6.269	98,27	(21,90)	1.101	433,46	0,000
16	110,34	(17,52)	11.349	100,46	(18,28)	1.951	523,48	0,000
17	110,37	(22,32)	2.964	95,19	(21,41)	589	230,20	0,000
18	106,31	(19,06)	9.250	95,23	(19,11)	1.915	535,49	0,000

Fonte: pesquisa própria.

Observando a Tabela 3, nos damos conta de que os 18 itens possuem diferenças significativas entre os alunos do GR e os do GF, já que os valores de F assim o indicam. Nesses itens as médias das pontuações dos componentes do GR foram superiores às médias dos componentes do GF. Em outras palavras, há indícios de que os alunos do GR têm maior grau de conhecimento de Língua Portuguesa do que os do GF.

Nesse contexto, a segunda hipótese da investigação (H_2 : $\mu_1 > \mu_2$), afirmando que “os alunos de escolas particulares (GR) terão melhores rendimentos que os de escolas públicas (GF) nos itens dos testes de Português e Matemática” foi corroborada.

6.4. Análise do impacto no Teste de Matemática segundo o tipo de escola

Apresentamos, a seguir, a Tabela 4, contendo os resultados do teste da Análise de Variância (*one-way*) para cada um dos 15 itens do teste de Matemática.

TABELA 4. COMPARAÇÃO ENTRE AS PONTUAÇÕES MÉDIAS DO GR E DO GF NOS 15 ITENS DO TESTE DE MATEMÁTICA

Itens	GR			GF			F	p
	μ	(σ)	n	μ	(σ)	n		
1	101,01	(20,43)	9.349	101,68	(21,22)	1.661	1,51	n.s.
2	101,26	(20,44)	3.302	102,05	(21,18)	593	0,74	n.s.
3	100,79	(20,32)	6.321	101,42	(20,24)	879	0,75	n.s.
4	101,09	(20,59)	6.279	100,75	(21,19)	828	0,19	n.s.
5	100,95	(20,32)	10.540	101,70	(21,23)	1.899	2,15	0,000
6	100,43	(19,95)	4.273	102,49	(22,58)	753	6,55	0,011
7	100,65	(20,38)	4.671	100,45	(20,14)	917	0,75	n.s.
8	101,10	(20,36)	4.649	102,11	(21,39)	672	1,44	n.s.
9	100,47	(20,20)	6.204	100,92	(21,14)	1.060	0,44	n.s.
10	101,54	(20,36)	3.568	101,08	(21,33)	859	0,35	n.s.
11	100,82	(20,28)	5.912	100,90	(21,51)	1.376	0,02	n.s.
12	100,73	(20,43)	3.547	101,97	(21,07)	710	2,14	n.s.
13	101,07	(20,64)	4.635	101,79	(21,43)	973	0,96	n.s.
14	101,03	(20,27)	4.661	101,24	(20,99)	887	0,08	n.s.
15	101,66	(20,46)	5.264	101,61	(20,93)	1.063	0,01	n.s.

Fonte: pesquisa própria.

Como podemos observar na Tabela 4, tão só os itens 5 e 6 possuem diferenças significativas entre as pontuações médias do GR (alunos que estudaram todo o ensino fundamental e médio em escolas particulares) e o GF (alunos que estudaram todo o ensino fundamental e médio em escolas públicas), já que os valores *F* assim o indicam; nos demais 12 itens, não há qualquer diferença entre as pontuações médias de ambos os grupos.

Nos dois únicos casos de diferenças significativas entre as pontuações médias do GR e do GF (itens 5 e 6), os valores das pontuações dos componentes do GF foram superiores às médias dos componentes do GR, ou seja, considerando-se os alunos que acertaram os itens 5 e 6 do Teste de Matemática, existem

indícios de que os sujeitos do GF têm maior grau de conhecimento sobre os conteúdos de matrizes, determinantes e equações do 2º grau (item 5) e sistemas de equações, fatorização e equação modular (item 6) que os alunos do GR.

Há uma explicação plausível para essas diferenças favoráveis aos alunos de escolas públicas em ambos os itens. Existe um considerável número de alunos de escolas públicas que estudaram em dois colégios que lhes proporcionam excelente formação acadêmica e profissional, que se chamam *Centro Federal de Ensino Tecnológico do Ceará (CEFET/CE)* e *Colégio Militar*, cujo ensino nas áreas de Matemática, Física e Química é reconhecido como de alta qualidade. Assim, são compreensíveis as diferenças detectadas em ambos os itens, favoráveis aos que estudaram todo o ensino fundamental e médio em escolas públicas (GF).

Nesse contexto, a segunda hipótese da investigação ($H_2: \mu_1 > \mu_2$), afirmando que “os alunos de escolas particulares (GR) terão melhores rendimentos que os de escolas públicas (GF) nos itens dos Testes de Português e Matemática”, não foi corroborada.

6.5. Ponderações finais acerca dos resultados

Considerando-se o *tipo de escola dos estudantes*, detectamos 13 casos de DIF (72,2% do total) no teste de Português (itens 1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17 e 18), sendo que nos itens 3, 5, 10, 14, 17 e 18, o DIF é benigno ao GR (alunos de escolas particulares). Identificamos 12 casos de DIF (80% do total) no teste de Matemática (itens 2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14 e 15), tendo cinco deles DIF benigno ao GR (itens 3, 4, 5, 8 e 9).

Esses dados demonstram que a afirmação de Muñiz (1997) é verdadeira. Segundo o autor, não existem provas inteiramente isentas de vieses. Nesse âmbito, temos que detectar a quantidade de vieses que pode ser aceitável em um determinado teste ou item.

Com respeito ao impacto em ambos os testes, foram detectadas diferenças significativas entre as pontuações médias dos grupos em todos os 18 itens do teste de Português favoráveis, em sua totalidade, aos alunos de escolas particulares. Em apenas dois itens do teste de Matemática (5 e 6), as diferenças verificadas foram favoráveis aos alunos de escolas públicas.

Como destacamos antes, muitas investigações aportam informações válidas sobre a influência que exercem os centros educativos sobre o rendimento escolar dos aprendizes, dando-nos conta de que as profundas diferenças entre as escolas públicas e privadas brasileiras têm reflexo sobre aspectos facilmente perceptíveis, tais como o grau de aprendizagem dos discentes. Efetivamente, pelo menos quanto ao domínio da língua materna, os alunos de escolas particulares demonstraram ser muito superiores aos aprendizes de escolas públicas. Quanto ao domínio da Matemática, podemos acentuar que houve um “*empate técnico*”.

Através da análise qualitativa dos itens com DIF, constatamos que termos “*pouco conhecidos*” ou “*pouco utilizados*” pelos estudantes estiveram presentes em 25% desses casos no teste de Português, ascendendo a 36,4% no teste de Matemática. Esses resultados indicam certa “*pobreza no vocabulário*” dos estudantes, apontando para uma importante disfunção das instituições de ensino. Como afirmou o Apóstolo São Mateus: *arbor ex fructu cognoscitur*.

6.6. Ponderações finais acerca da investigação

O propósito de nosso trabalho foi contextualizar o estudo do funcionamento diferencial do item (DIF) no amplo e complexo campo da avaliação da qualidade educacional. Para isso, tentamos demonstrar o aspecto que une as investigações sobre o DIF aos processos de avaliação educacional. Afirmamos que a qualidade educacional, independentemente do seu significado, é um construto intimamente associado ao uso de procedimentos sistemáticos de avaliação (Andriola, 1999a, 2000, 2001). Ademais, sabemos que um poderoso indicador da qualidade de uma instituição educacional é o grau em que esta consegue alcançar os objetivos curriculares propostos *a priori*, isto é, o grau de efetividade ou eficácia escolar que possui (Sammons, Hillman e Mortimore, 1998). Posteriormente, destacamos que o procedimento mais adequado para medir o grau de efetividade escolar é através da comparação dos resultados educativos observados com aqueles esperados (Orden Hoz, 1992), isto é, comparar o grau de aprendizagem obtido pelos alunos com os objetivos educativos propostos. Para tanto, devemos usar os conhecidos testes de rendimento ou referidos a um critério, já que são considerados os instrumentos mais adequados a esse tipo de estudo.

Não obstante, é bastante comum que os itens utilizados nesse tipo de teste tenham uma característica conhecida como *DIF*. Como enfatizamos nos tópicos anteriores, o DIF ocasiona sérios problemas às avaliações educacionais (Hartle e Bataglia, 1993). Trata-se de um fator de injustiça para alguns grupos de respondentes, já que os alunos que possuem o mesmo grau de aprendizagem, mas que provêm de grupos demográficos distintos, têm diferentes probabilidades de acertar um mesmo item. Portanto, devemos reconhecer a relevância dessas investigações, pois podem proporcionar maior equidade aos processos de avaliação educacional, através da identificação e não utilização daqueles itens que possuam algum tipo de DIF. Por fim, devemos enfatizar que *o estudo do DIF deve preceder qualquer processo de avaliação educacional sério, ou seja, esse tipo de estudo deve, necessariamente, compor a fase de pré-teste dos itens antes de sua definitiva utilização* (Andriola, 2001, 2002, 2006).

7. À GUIA DE CONCLUSÃO

Constatamos que a área de investigação do DIF no âmbito educativo e psicológico é recente e, ademais, necessita de boas hipóteses, fundamentadas em teorias científicas, que tentem “abrir perspectivas” aos estudos do DIF (Hambleton, 1997; Roznowski e Reith, 1999; Scheuneman & Gerritz, 1990). Como opina Bond (1993):

In general, however, theories about why items behave differentially across groups can be described only as primitive¹ (pág. 278).

Schmitt, Holland e Dorans (1993) acreditam que a área que investiga o DIF não tem progredido no grau desejado em virtude de três fatores:

- Pelo fato de as investigações acerca do DIF são relativamente recentes e, atualmente, a ênfase está no desenvolvimento de métodos estatísticos para sua identificação. Por exemplo, o desenvolvimento das modernas técnicas para a detecção do *funcionamento diferencial das*

¹ De um modo geral, as teorias sobre o porquê de alguns itens funcionarem diferentemente para certos grupos podem ser descritas somente como primitivas.

alternativas (DAF) tem o mesmo objetivo das técnicas DIF (Thissen, Steinberg e Wainer, 1993; Thissen, Steinberg e Fitzpatrick, 1989);

- Porque a identificação do DIF e os fatores a ele relacionados necessitam boas teorias sobre a dificuldade diferencial dos itens. Há que destacar que, esse é um campo, no qual as teorias acerca dos processos cognitivos presentes na resolução dos itens não se encontram, todavia, minimamente avançadas;
- Porque a identificação e a descrição dos citados processos cognitivos é muito complexa, já que intervêm múltiplos fatores. Ademais, é um campo de investigação que exige o trabalho multidisciplinar de psicólogos, pedagogos e matemáticos, algo difícil de ser obtido no estágio atual de desenvolvimento investigativo brasileiro.

Devemos dizer que o processo de criação de boas hipóteses explicativas do DIF deverá, logicamente, ser árduo, difícil e frustrante. As hipóteses deverão ser corroboradas ou de rejeitadas, algo bastante comum à atividade científica. Como nos lembra Júlio Verne, a ciência é composta de erros, que são os passos em direção à verdade. Mais recentemente, o economista norte-americano Paul Samuelson, ganhador do prêmio Nobel de Economia, saiu-se com a seguinte frase: de funeral em funeral a ciência avança (Wilson, 1999). Todavia, muito tempo antes, na Roma antiga, o orador Cícero sentenciava: *Vivere est cogitare*.

Devemos ter claro o fato de que a presença do DIF em itens de instrumentos de medida psicológica e pedagógica é um grave problema que atenta contra o suposto da padronização ou uniformização das condições de avaliação. É uma fonte de injustiça, já que produz falta de equidade aos processos avaliativos; permite aos sujeitos que possuem o mesmo grau na variável latente ou construto medido pelo item obter melhores resultados, já que esses têm maiores probabilidades de acertá-lo (Douglas, Roussos e Stout, 1996).

Nesse âmbito, caberá aos responsáveis pela construção, administração e comercialização de testes psicológicos e pedagógicos, a verificação da presença do DIF em seus instrumentos, já que a sua existência é um fator de invalidação dos resultados. Também os psicometristas que começam a organizar bancos de itens necessitam verificar a presença do DIF e, assim, evitar utilizá-los em processos avaliativos (Andriola, 1998b, 2001, 2002).

Para finalizar, mencionaremos uma célebre frase latina, que é muito sugestiva e sintetiza em nossa opinião, a importância dos estudos sobre o DIF no âmbito da avaliação psicológica e educacional: *fiat justitia, pereat mundus*.

REFERENCIAS BIBLIOGRÁFICAS

- Allalouf, A., Hambleton, R.K. e Siresi, S.G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement*, 36 (3), pp. 185-198.
- Alexander, P.A. e Judy, J.E. (1998). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58, pp. 375-404.
- Andriola, W.B. (1995). Avaliação do raciocínio numérico em estudantes do 2º grau. *Educação em Debate*, 29/32, pp. 95-99.

- Andriola, W.B. (1997a). Avaliação do raciocínio verbal em estudantes do 2º grau. *Estudos de Psicologia*, 2(2), pp. 277-285.
- Andriola, W.B. (1997b). Expectativas de estudantes do 2º grau sobre a Universidade. *Educação em Debate*, 33, pp. 39-45.
- Andriola, W.B. (1998a). Apresentação de um Modelo Teórico destinado à avaliação dos Programas Estaduais de Qualificação Profissional (PEQ's). *Ensaio: Avaliação de Políticas Públicas em Educação*, 19(6), pp. 259-266.
- Andriola, W.B. (1998b). Utilização da teoria da resposta ao item (TRI) para a organização de um banco de itens destinados à avaliação do raciocínio verbal. *Psicologia: Reflexão e Crítica*, 11(3), pp. 295-308.
- Andriola, W.B. (1998c). Inteligência, aprendizagem e rendimento escolar segundo a Teoria Triárquica da Inteligência (TTI). *Educação em Debate*, 35, pp. 75-80.
- Andriola, W.B. (1999a). Evaluación: La vía para la calidad educativa. *Ensaio: Avaliação e Políticas Públicas em Educação*, 25, pp. 355-368.
- Andriola, W.B. (1999). Avaliação do raciocínio abstrato em estudantes do ensino médio. *Estudos de Psicologia*, 4(1), pp. 23-37.
- Andriola, W.B. (2000). Calidad educativa y efectividad escolar: conceptos y características. *Educação em Debate*, 39(1), p. 7-14.
- Andriola, W.B. (2002). Determinación del funcionamiento diferencial de los ítems (DIF) destinados a la evaluación del razonamiento verbal a partir del tipo de escuela. *Bordón: Revista de Pedagogía*, 53(4), pp. 473-484, 2002.
- Andriola, W.B. (2002). Detección del funcionamiento diferencial del ítem (DIF) em tests de rendimento. Aportaciones teóricas e metodológicas. *Tese Doutoral*. Madrid: Universidad Complutense de Madrid.
- Andriola, W.B. (2006). Estudos sobre o viés de itens em testes de rendimento: uma retrospectiva. *Estudos em Avaliação Educacional*, 17(35), pp. 115-134.
- Angoff, W.H. (1993). Perspectives on differential item functioning. En: P. W. Holland e H. Wainer (Eds.), *Differential Item Functioning*. New Jersey: Lawrence Erlbaum Associate.
- Barreto, J.A.E., Trompieri Filho, N. e Andriola, W.B. (1997). Desenvolvimento da estrutura cognitiva de alunos da 4ª e 8ª séries. *Educação em Debate*, 37, pp. 101-113.
- Bond, L. (1993). Comments on the O'Neill and McPeck paper. En: P. W. Holland e H. Wainer (Eds.), *Differential Item Functioning*. New Jersey: Lawrence Erlbaum Associates.
- Bryman, A. e Cramer, D. (1992). *Análise de dados em Ciências Sociais*. Oeiras: Celta Editora.
- Cluser, B.E., Nungester, R.J. e Swaminathan, H. (1996). Improving the matching for DIF analysis by conditioning on both test score and an educational background variable. *Journal of Educational Measurement*, 33(4), pp. 453-464.
- Cole, N.S. (1993). History and development of DIF. En: P.W. Holland e H. Wainer (Eds.), *Differential Item Functioning*. New Jersey: Lawrence Erlbaum Associates.
- De Miguel, M. (1989). Modelos de investigación sobre organizaciones educativas. *Revista de Investigación Educativa*, 1(13), pp. 21-56.
- Douglas, J.A., Roussos, L.A. e Stout, W. (1996). Item-Bundle DIF hypothesis testing: identifying suspect bundles and assessing their differential functioning. *Journal of Educational Measurement* 33(4), pp. 465-484.
- Dunkin, M. J. (1998). Student characteristics, classroom process and student achievement. *Journal of Educational Psychology*, 70 (6), pp. 998-1009.
- Fuller, B. y Clark, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of

- classroom tools, rules and pedagogy. *Review of Educational Research*, 64, pp. 119-157.
- Hambleton, R. K. (1997). Perspectivas futuras y aplicaciones. En: J. Muñiz, *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Ediciones Psicología Pirámide.
- Hamilton, L. S. (1999). Detecting gender-based differential item functioning on a constructed-response science test. *Applied Measurement in Education*, 12(3), pp. 211-235.
- Hartle, T.W. Y Battaglia, P.A. (1993). The Federal Role in standardized Testing (p. 291-311). En: R. E. Bennett y W. C. Ward (Org.), *Construction versus Multiple Choice Items in Cognitive Measurement*. New Jersey: Lawrence Erlbaum Associates.
- Horsburgh, M. (1999). Quality monitoring in higher education: the impact on student learning. *Quality in Higher Education*, 5(1), pp. 9-25.
- Kvanli, A.H. (1988). *Statistics. A Computer Integrated Approach*. Saint Paul: West Publishing Company.
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates.
- Martínez Arias, R. (1997). *Psicometría. Teoría de los Tests Psicológicos y Educativos*. Madrid: Ediciones Síntesis.
- Mc Donald, B.C. (2003). (org.). *Esboços em avaliação educacional*. Fortaleza: Editora da UFC.
- Mellenbergh, G.J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), pp. 127-143.
- Miles, M.B. (1974). Las diez características del centro docente "sano". *La Educación Hoy*, 5, pp. 197-198.
- Moura Castro, C. (1999). Escolas feias? Escolas boas? *Ensaio: Avaliação e Políticas Públicas em Educação*, 7(25), pp. 343-354.
- Muñiz, J. (1994). *Teoría Clásica de los Tests*. Madrid: Ediciones Pirámide.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Nuthall, G. (1999). Relating learning to individual differences in ability. *Journal of Educational Research*, 31(3), pp. 212-255.
- Nuthall, G. y Alton-Lee, A.G. (1995). Assessing classroom learning: how students use their knowledge and experience to answer achievement test questions in science and social studies. *American Educational Research Journal*, 32, pp. 185-223.
- Nuthall, D.L., Goldstein, H., Prosser, R. y Rabash, J. (1989). Differential scholl effectiveness. *International Journal of Educational Research*, 13(7), pp. 769-776.
- OCDE – Organización para la Cooperación y el Desarrollo Europeo (1995). *Schools under scrutiny*. Paris: Head of Publication Service.
- Orden Hoz, A. (1992). Calidad y evaluación de la enseñanza universitaria. *Resúmenes del Congreso Internacional de Universidades*. Madrid, julio, pp. 531-539.
- Orden Hoz, A. (1993). La escuela en la perspectiva del producto educativo. Reflexiones sobre evaluación de centros docentes, *Bordón*, 45(3), pp. 263-270.
- Ramos, E.A. (1999). Aprendizagem humana. *Cadernos de Educação*, 23, pp. 37-49.
- Rego, A. (2000). Impactos dos comportamentos de cidadania docente sobre os alunos universitários. A perspectiva dos estudantes e dos professores. *Linhas Críticas*, 6(10), pp. 9-30.
- Rego, A. (2001). Eficácia comunicacional dos docentes do Ensino Superior: evidência confirmatória do constructo. *Psicologia: Reflexão e Crítica*, 14(3), pp. 563-568.

- Rossi, P.H., Freeman, H.E. y Lipsey, M.W. (1999). *Evaluation. A systematic approach*. London: Sage Publications.
- Roznowski, M. e Reith, J. Examining the measurement quality of tests containing differentially functioning items: Do biased items result in poor measurement? *Educational and Psychological Measurement*, 52 (2), p. 248-269, 1999.
- Sammons, P., Hillman, J. e Mortimore, P. (1998). *Características Clave de las Escuelas Efectivas*. México: Secretaría de Educación Pública.
- Scheerens, J. (1992). *Effective schooling: research, theory and practice*. Londres: Cassell.
- Scheuneman, J. D. Y Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, 27(2), pp. 109-131.
- Schmitt, A.P., Holland, P.W. y Dorans, N.J. (1993). Evaluating hypotheses about differential item functioning (p. 281-319). En: P. W. Holland y H. Wainer (Eds.), *Differential Item Functioning*. New Jersey: Lawrence Erlbaum Associates.
- Soares, J.F. (Coord.) (2002). *Escola eficaz: um estudo de caso em três escolas da rede pública do Estado de Minas Gerais*. Belo Horizonte: GAME/UFGM.
- Thissen, D., Steinberg, L. Y Fitzpatrick, A.R. (1989). Multiple-choice models: the distractors are also part of the item. *Journal of Educational Measurement*, 26(2), pp. 161-176.
- Thissen, D., Steinberg, L. y Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models (p. 67-113). En: P. W. Holland y H. Wainer (Eds.), *Differential Item Functioning*. New Jersey: Lawrence Erlbaum Associates.
- Tobias, S. (1994). Interest, prior knowledge and learning. *Review of Educational Research*, 64(1), pp. 37-54.
- Wilson, E.O. (1999). *Consilience. La Unidad del Conocimiento*. Barcelona: Ediciones Galaxia Gutemberg.
- Zumbo, B.D. (1999). *A Handbook on the theory and methods of differential item functioning (DIF). Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense of Canadá.