

Aproximaciones al análisis de *accountability* e implicancias altas y bajas de evaluaciones a gran escala en cinco países de América Latina

Aproximações à análise de *accountability* e implicâncias altas e baixas de avaliações a grande escala em cinco países da América Latina

Approaches to the analysis of *accountability* and high and low implications of large scale assessments in five countries of Latin America

Silvia Radulovich, Maria Rueda y Cristina Basualdo
Universidad Nacional de Salta. Argentina
sradulov@unsa.edu.ar - alejrueda@hotmail.com - crisbas@unsa.edu.ar

75

/ RESUMEN

Este trabajo forma parte de algunas discusiones y análisis en curso a propósito de nuestro trabajo en el Proyecto de Investigación que llevamos adelante sobre las funciones políticas de las evaluaciones nacionales en América Latina, con el ánimo de un intercambio crítico para la revisión y enriquecimiento de la progresión analítica.

En el marco de las reformas educativas de los '80 y '90 en América Latina surgen las evaluaciones nacionales en un buen número de países al igual que las aplicaciones de pruebas internacionales que, según concluyen numerosos estudios, no han incidido significativamente en mejores prácticas educativas en el nivel docente ni en mejores aprendizajes en el nivel de los alumnos.

Aun así, las evaluaciones nacionales persisten y parecen atender a funciones políticas primarias, secundarias y terciarias, así denominadas por Angulo Rasco (1995: 203-204) atento al grado de visibilidad social que presentan. Desde el ángulo comparativo, resulta necesario identificarlas en cada país, identificando también aquellas *equivalencias funcionales* en tanto sistemas nacionales de evaluación, expresión que –de acuerdo a Reijo Rivarola, quien cita a Marton– designaría los sentidos y configuraciones de significado en el discurso y en la acción, haya o no convergencia o identidad en los hechos descriptibles entre los casos en comparación (en Altbach y Kelly, 1990: 300).

Palabras claves: análisis, reformas educativas, evaluaciones educativas.

// RESUMO

Este trabajo forma parte de algunas discusiones e análisis en curso relacionadas al nuestro proyecto de investigación que realizamos donde se estudia las funciones políticas de las evaluaciones nacionales en América Latina, con la intención de realizar un intercambio crítico para el enriquecimiento de la progresión analítica.

No contexto de las reformas educativas de los años 80 y 90 en América Latina surgen las evaluaciones nacionales en un buen número de países del mismo modo que las aplicaciones de exámenes internacionales que, según las conclusiones de numerosos estudios, no influenciaron significativamente en mejoras prácticas educativas, en lo que concierne a la docencia, ni en mejores aprendizajes, en lo que concierne a los alumnos.

Aun así, las evaluaciones nacionales persisten y parecen atender a las funciones políticas primarias, secundarias y terciarias, así denominadas por Angulo Rasco (1995: 203-204) de acuerdo al grado de visibilidad social que presentan. Desde el ángulo comparativo es necesario identificarlas en cada país, situando también las *equivalencias funcionales* en los sistemas nacionales de evaluaciones, expresión que – de acuerdo a Reijo Rivaola, citando a Marton – designaría los sentidos y configuraciones de significado en el discurso y en la acción, haya o no convergencia o identidad en los hechos descriptivos de los casos en comparación (en Altbach y Kelly, 1990: 300).

Palabras clave: análisis, reformas educativas, evaluaciones educativas.

/// ABSTRACT

This work is part of some ongoing discussions and analysis about our research project on the political functions of national assessments in Latin America, with the aim of a critical exchange for the revision and enhancement of the analytical progression.

As part of the educational reforms of the '80s and '90s in Latin America, national assessments as well as international testing applications are carried on in a number of countries, that, according to the conclusions drawn from many studies, have not had a significant impact on better educational practices in teaching processes neither in learning ones. Still, national assessments persist and seem to respond to political primary, secondary and tertiary functions, as Angulo Rasco called them (1995 : 203-204), given the degree of social visibility they present. From a comparative perspective, it is necessary to identify them in each country, considering also those *functional equivalents* as national evaluation systems, expression that, according to Reijo Rivaola, quoting Marton- produce the meanings and configurations of meaning in speech and in action, whether there is or not convergence or identity in the facts described between the cases compared (Altbach and Kelly , 1990 : 300).

Keywords: analysis, educational reforms, educational evaluations.

PRE-TEXT

Este trabajo forma parte de algunas discusiones y análisis en curso a propósito de nuestro trabajo en el Proyecto de Investigación que llevamos adelante sobre las funciones políticas de las evaluaciones nacionales en América Latina, con el ánimo de un intercambio crítico para la revisión y enriquecimiento de la progresión analítica.

PRUEBAS Y POLÍTICA

En el marco de las reformas educativas de los '80 y '90 en América Latina¹ surgen las evaluaciones nacionales en un buen número de países al igual que las aplicaciones de pruebas internacionales que, según concluyen numerosos estudios², no han incidido significativamente en mejores prácticas educativas en el nivel docente ni en mejores aprendizajes en el nivel de los alumnos.

Aun así, las evaluaciones nacionales persisten y parecen atender a funciones políticas primarias, secundarias y terciarias³, así denominadas por Angulo Rasco (1995: 203-204) atento al grado de visibilidad social que presentan. Desde el ángulo comparativo, resulta necesario identificarlas en cada país, identificando también aquellas *equivalencias funcionales* en tanto sistemas nacionales de evaluación, expresión que –de acuerdo a Reijo Rivarola, quien cita a Marton– designaría los sentidos y configuraciones de significado en el discurso y en la acción, haya o no convergencia o identidad en los hechos descriptibles entre los casos en comparación (en Altbach y Kelly, 1990: 300).

Estas funciones políticas de las evaluaciones nacionales parecen presentar potencialidad heurística para una aproximación comprensiva a nuestro objeto de indagación, es decir, sin pretensión de explicaciones causales ni afán de generalización nomotética.

Sintéticamente, podemos decir que las funciones **terciarias**, *explícitas* y *retóricas*, se dirigen a la búsqueda de aceptación general y legitimidad política bajo el sustento de la legitimidad técnica proporcionada por el conocimiento experto para la elaboración de las pruebas y su interpretación de resultados. Dentro de éstas se identifican tres áreas de interés para el público: **(a)** la motivación individual para el aprendizaje y el logro para el ascenso social, a través de la obtención de certificaciones, al igual que la creación de expectativas de seguridad y bienestar económico y profesional; **(b)** la movilidad social como posibilidad y la confirmación o legitimación de la posición social adquirida; y, por último la función más enunciada, **(c)** proporcionar información sobre el sistema educativo, orientando la política curricular, apreciando la calidad educativa del sistema, al igual que informando sobre problemas, logros y límites del mismo. Cada una de estas tres áreas, en sus producciones discursivas, opera como núcleos articuladores de significados donde el signifiante 'calidad' es el más relevante, adjudicando sentido y significado a estas evaluaciones nacionales al operar con justificación legitimadora.

Las funciones **secundarias** de las evaluaciones del sistema educativo, tienen un carácter de intermedias en cuanto al nivel de explicitación: parte de su temática es presentada, defendida y aceptada oficialmente ante el público, aunque sin presentar la otra cara de cada una de ellas. Un ejemplo es la argumentación oficial (y de agencias económicas diversas) que manifiesta la importancia de un determinando *saber igual* para todos y señala también la importancia de la obtención de *iguales resultados* para todos. Y **no** expresan el implícito reforzamiento de **(a)** la homogeneidad cultural (en clave globalizante), de **(b)** la valoración de determinados aprendizajes y de **(c)** determinado tipo de contenidos curriculares, que dejan fuera de consideración tanto las consecuencias de exclusión cultural y social a través de la estandarización (Radulovich S.: 2003) como el limitado nivel de exigencia cognitiva de las pruebas.

¹ Para una caracterización más integral y precisa de la relación entre las políticas educativas de los '90 y los sucesivos cambios en la década en curso, ver Betancur (2007).

² Brown, 2001; Dorn, 1998; Bradfoot, 1984, sin exhaustividad de menciones.

³ Con independencia de la voluntad y/o conciencia de los actores concretos.

Por último, las funciones **primarias** son funciones ocultas y esenciales de los sistemas nacionales de evaluación de **(a)** la gestión productivista del sistema, **(b)** el control administrativo y **(c)** la selección de individuos. Si bien *eficiencia, eficacia, productividad* son términos frecuentes y –conjeturalmente– comprendidos en el discurso institucional y social general, resultan más sutiles sus implicancias en relación con la vida cotidiana de las instituciones educativas y de los sujetos institucionales individuales de autorresponsabilización, control y autocontrol docente y de la gestión institucional (Radulovich, 2000, 2008; Fernández, 2007). También se advierte un efecto de desvalorización de logros o aprendizajes que, aun siendo sustantivos, no son evaluados por las pruebas estandarizadas, la competencia entre instituciones y la instrumentación de resultados para la puja dentro del ‘mercado’ educativo.

ANALIZANDO...

Nos centraremos en las pruebas nacionales en cinco países de América Latina en relación con la operación de algunos conceptos que, insertos en el discurso educativo durante la puesta en marcha de las reformas educativas, *encubren* sentidos más económicos que propiamente educativos. Sobre todo cuando documentos oficiales toman a la ‘calidad’ como un significante articulador de un conjunto de decisiones de política educativa de reforma, calidad que aparece como expectativa deseable y generalizada en la opinión pública, la que, además, parece aceptar la existencia de una necesidad de alguna valoración –fundada técnicamente– del nivel o el estado de conocimiento que los alumnos alcanzan durante y al final de su ‘pasaje’ por el sistema educativo. Y las evaluaciones nacionales estandarizadas aparecen como un vehículo confiable para esta valoración de resultados por cuanto gozan de suficiente credibilidad pública, producto de la legitimación que ofrece el conocimiento experto que avala su elaboración, además de los ‘refuerzos’ significantes que agregan los medios de difusión al publicar los resultados.

78

El significante **calidad** de la educación presenta relevancia también como *articulador* de discursos normativos –leyes, decretos, reglamentos–, de discursos periodísticos, de discursos institucionales y áulicos.

Sin embargo, si bien leyes y documentos escritos parecen otorgar un significado unívoco a ‘calidad’, en entrevistas con funcionarios políticos de varios ministerios nacionales se nos expresó con todas las letras que **no** existe tal convergencia de significados. Se observa en la mayoría de los casos que se asocia ‘calidad’ con ‘excelencia’⁴ y en otros –los menos– se asocia ‘calidad’ con ‘inclusión’ y ‘pertinencia social’⁵. Sin embargo aparece un predominio de la relación significante de equivalencia entre costo-beneficio por un lado e insumos-producto por otro que, combinadas, conducen al concepto sustentador de este enfoque economicista de las reformas en la región: *producción educativa* (Mancebón Terrubia, 1999; Radulovich, 2007). Más allá de lo que el discurso explícito y las bienintencionadas acciones de algunos funcionarios pretendan, las evaluaciones resultan un instrumento de control para llegar a determinar la función de producción educativa, concepto económico que traduce claramente la lógica de fondo implícita de las evaluaciones nacionales (cfr. Radulovich, 2007 y Basualdo, 2007).

⁴ Entendida como alta competencia en ciertos estándares de conocimiento pretendidamente universales.

⁵ Bianchetti (2007) analiza exhaustivamente estas distintas ideas de ‘calidad’.

I. RESULTADO(S) EDUCATIVO 'CONTANTE Y SONANTE'

Se presenta como razonable para los gobiernos, entonces, pensar las evaluaciones macro como un modo de establecer e informar al 'público' cómo *funciona* el sistema educativo consecuentemente sostenido por dineros públicos y así dar cuenta de la 'calidad' de sus resultados. Se ha denominado *accountability* a esta función y es un concepto proveniente de la cultura anglosajona asociado con operaciones contables, que se instala en el campo educativo y es traducido inicialmente como 'rendición de cuentas' en su sentido cuasi literal y reformulado posteriormente como 'responsabilidad', como *responsabilidad por los resultados* (Ravela, 2006) aunque es frecuente la utilización indistinta de ambas expresiones –cual sinónimos– en documentos oficiales y académicos varios. ¿Es acaso la administración gubernamental quien rinde cuentas o se responsabiliza de los resultados ante el público?

Si bien con frecuencia documentos oficiales de distintos países –y hasta en sus leyes de educación– dan razón de las 'evaluaciones' macro como una instancia de responsabilización política y administrativa, en la práctica la difusión pública de resultados se presenta o bien con escasa inteligibilidad –formatos técnicos que requieren 'traducción', o se ofrecen sobreesimplificados en una única cifra– o bien como efecto demostración de las deficiencias docentes o de las instituciones en general.

En la literatura que se define a favor de la rendición de cuentas, en general, encontramos que son las instituciones educativas concretas quienes *deben* hacerlo ya que son las que concretamente realizan la actividad educativa y porque para ello tienen la prerrogativa de 'autonomía', aunque no incluyen del mismo modo concreto en sus análisis las *condiciones* socioeconómicas, laborales, organizativas, normativas, socioculturales concretas en que tal actividad se desarrolla. Además, las diferentes líneas descentralizadoras de las reformas educativas han descentralizado las responsabilidades y centralizado los 'controles de calidad' a través de las evaluaciones macro estandarizadas, que, desde luego son externas a los efectivos procesos educativos desarrollados.

Esta idea de 'rendición de cuentas' o *accountability* se interpreta, en el nivel de formulación y ejecución de políticas, al menos desde dos **posiciones** sobre el **para qué** instrumentar evaluaciones nacionales, relacionando este 'para qué' con el tipo de decisiones que suponen los resultados.

La primera posición (**a**) concierne a la toma de decisiones 'duras' a partir de los resultados, tanto en relación con los sujetos educativos (docentes y alumnos) como con las unidades institucionales de la educación, o ambos. En esta posición, estas evaluaciones nacionales son de *altas implicancias* o alto impacto o altas consecuencias, para sujetos y/o para instituciones. La preocupación principal –no siempre explícita– es optimizar la asignación de recursos, hacer más eficiente el gasto en educación... o no incrementar el presupuesto en educación 'ineficiente'. Las evaluaciones censales (totalidad de la población escolar dentro del nivel educativo considerado), la publicación en la prensa de los resultados obtenidos por cada escuela ordenados de mayor a menor (*ranking*), los incentivos económicos de carácter individual (a los docentes) o colectivo (a la institución) son algunas expresiones de esta postura. Donde no se logra el premio-incentivo se plantea –de hecho– la *sanción social* de las familias de retirar los hijos de tales instituciones 'no premiadas' porque las escuelas son responsables ante sus clientes o usuarios de la

'calidad'⁶ que ofrecen y logran, o bien de la reducción de posibilidades de ampliación de subsidios por parte de las agencias gubernamentales correspondientes.

Esto se inscribe en planteos de 'mercado educativo' y 'control de los padres sobre las escuelas', que, desde luego promueve la competencia entre instituciones desde razones no precisamente educativas, como expresamos *ut supra*.

Las consecuencias o implicancias fuertes en relación con los alumnos son la acreditación de nivel y/o la preselección para el siguiente nivel educativo, lo que representa una presión temporal y un efecto subjetivo de 'irreversibilidad' al final de un proceso que pretende ser educativo...

Esta posición respecto de la 'rendición de cuentas' está sustentada, entre otras, en el supuesto de la educación como una organización que, principalmente, tiene la función social de apoyo al sistema productivo por lo que, entonces, interesa determinar en última instancia, la función de *producción educativa*⁷ del sistema, ya mencionada, a la que se llega a través de complejos algoritmos. Por su propia naturaleza, estos algoritmos nada tienen que ver con desarrollos curriculares contextualizados ni con los procesos educativos reales. Sobre todo, debido a la intangibilidad del 'producto' educativo (Mancebón Terrubia, 1999, Radulovich, 2007), la dimensión temporal de los 'procesos de producción', el carácter acumulativo del aprendizaje humano y la incidencia de factores exógenos al contexto escolar, por lo que, entonces, cada "unidad producida" es diferente a las demás.

De los cinco casos-países considerados podemos ubicar en esta posición a **Chile**. En este país, la instrumentación de pruebas nacionales de alto impacto comienza en 1982 (Prueba de Evaluación del Rendimiento Escolar: PER:) mientras que el Sistema de Medición de la Calidad Educativa (SIMCE) se inicia en 1988. Las consecuencias o implicancias son altas tanto para los docentes como para las instituciones. Además de las pruebas SIMCE aplicadas a los alumnos, el Ministerio de Educación desarrolla el Sistema Nacional de Incentivos Educativos (SNED) focalizado en las instituciones que considera de excelencia y en los docentes. Ahora bien, la subvención que otorga el SNED es asignada según los desempeños indicados en los resultados de las pruebas SIMCE, en base a indicadores de *efectividad* y *superación* que comprenden el 65% del total de indicadores (Mizala [2003] en Ravela, 2006). Esta subvención es entregada cada dos años y en total comprende el 25% de la matrícula regional. Del monto total asignado a cada establecimiento, el 90% corresponde a los profesores –según la cantidad de horas cronológicas de cada cual– y es entregado trimestralmente durante dos años. Cada institución destina el 10% restante según criterios propios. Dado que los resultados de los tests (las pruebas SIMCE) presentan dificultad de comparabilidad entre instituciones tan diversas, éstas son agrupadas según características geográficas, socioeconómicas y 'tipo de enseñanza' construyendo 'grupos homogéneos' para presentar la información al difundir los resultados en un diario de circulación nacional.

Como puede apreciarse, en ambos instrumentos de política sistémica (SIMCE y SNED) se trata de incentivar la *oferta* de lo que produce el sistema para sus 'clientes' o 'usuarios' y subvencionar la demanda.

⁶ No es factible condensar la calidad de la acción educativa de una institución concreta en un número y por lo tanto no puede ser reducida a un único indicador que, por otra parte, no resulta válido por su imposibilidad de generalización entre unidades de base no homogénea (ver Angulo Rasco, 1992, 1995, 1999; Radulovich, 2003, 2008).

⁷ Relación que vincula los *inputs* con los *outputs* de la institución educativa en términos de eficiencia. (Mancebón Terrubia, 1999:116).

Además de esta intencionalidad explícita en los documentos, se han relevado algunos efectos que Ravela (197) llama 'perversos' como, el denominado 'goma de mascar' que consiste en el

(...) incremento o 'estiramiento' que se produce de las distancias en los resultados entre los extremos del sistema educativo (...), la brecha en calidad es creciente entre la educación que reciben los más pobres y la que reciben los más ricos (...) (2006: 200).

Otro 'efecto colateral' es el de condicionar el desarrollo curricular –las clases concretas– a lo que miden las pruebas y así reducir el campo de preocupaciones y el tiempo dedicado a actividades de alta complejidad cognitiva, dando mayor alcance al entrenamiento dentro del formato de las pruebas de opción múltiple constituyendo así, una 'enseñanza para la evaluación' más que 'una enseñanza para la comprensión' (Elliott, 1990: 212–233).

La competencia entre instituciones y el sistema de incentivos o premios y sanciones o 'castigos', provoca 'ganadores' y 'perdedores' lo que potencializa a unos y desmoraliza a otros. Se ha constatado, además, que las consecuencias o implicancias fuertes para los docentes y las instituciones aumentan las probabilidades de medidas defensivas, como por ejemplo retirar del aula a los alumnos de menor rendimiento al momento de realizar la prueba, lo que aumenta también la probabilidad de incrementar sesgos en la medición...

Podemos afirmar que estas referencias confieren plausibilidad a la presencia y predominio de funciones políticas primarias de las evaluaciones nacionales en Chile: la gestión productivista del sistema, el control administrativo y la selección de individuos. Desde luego las funciones secundarias y las explícitas terciarias de proporcionar información sobre el estado del sistema también están presentes.

También **Bolivia** podría ser incluida dentro de esta posición y para este caso tomaremos como referencia el trabajo técnico de Guillermo Ferrer (2006) del Programa de Promoción de la Reforma Educativa en América Latina y el Caribe PREAL por no disponer de actualizaciones oficiales respecto de las evaluaciones nacionales en ese país, y también algunos análisis realizados en otra investigación que hemos llevado a cabo con anterioridad.

Las evaluaciones nacionales son desarrolladas por el Sistema de Medición de la Calidad de la Educación (SIMECAL) que depende del Ministerio de Educación. La Ley de Reforma Educativa de 1994 plantea la institucionalización de inquietudes y experiencias propias de una sociedad con el 66.2% de población indígena que conforma un país multiétnico, pluricultural y multilingüístico. Es así que la *interculturalidad* y la *participación* son algunos de los principios rectores para la política educativa establecidos en la Ley. Estos principios parecen ser tenidos en cuenta en las evaluaciones nacionales cuando son elaborados los cuatro modelos de prueba, para cada área de conocimiento, en tres de los idiomas aborígenes –aimara, guaraní, quechua– y en español, según el lugar sociocultural y lingüístico de aplicación. El SIMECAL se aplica desde 1997 a través de pruebas de Matemática y Lenguaje. A pesar de la adecuación idiomática de las pruebas, la similitud con el resto de Latinoamérica persiste en cuanto a la pretensión de que éstas proporcionan indicadores de *eficacia* del Sistema Educativo y de la reforma, confiando en la metáfora estadística como garantía de validez de la información que el instrumento de evaluación recoge. Los indicadores de eficacia considerados son los promedios de puntajes totales por área de contenido y puntajes totales por habilidad; porcentaje de alumnos con rendimiento escolar

satisfactorio (puntaje > a 55 puntos); porcentaje de alumnos con rendimiento escolar regular (puntaje entre 45 y 55 puntos); de alumnos con rendimiento escolar en riesgo (puntaje < a 55 puntos).

El SIMECAL progresivamente fue teniendo otras utilidades además de la medición del aprendizaje escolar: acreditación de finalización de nivel en el bachillerato (Prueba de Aptitud Académica), evaluación de docentes (maestros interinos, bachillerato pedagógico, asesores pedagógicos) para el ascenso y asignación dentro del sistema de bonos y también para la preselección de candidatos a directores distritales, directores de escuela e ingreso a institutos de formación de maestros. Se aprecia aquí, entonces, la relación con la rendición de cuentas dadas las altas implicancias de estas evaluaciones, hasta 2004 inclusive.

En cuanto a las funciones políticas aparece un predominio de las secundarias; aunque el rasgo de la homogeneización cultural se presente mitigado por la adecuación lingüística, al mantener la misma *forma* de las pruebas, el efecto homogeneizante persiste al igual que se impone tanto el contenido como la valoración de un mismo tipo de aprendizaje para todos los sujetos que realicen la prueba.

II. RESULTADOS QUE NO SUENAN TANTO...

La segunda posición **(b)** acerca de la contribución de las evaluaciones nacionales a la 'rendición de cuentas' o responsabilización por los resultados, es aquella que las concibe como insumo para la toma de decisiones 'blandas' o en otras palabras, las referidas al mejoramiento de las escuelas y/o de los procesos educativos en la práctica áulica, es decir, cumplirían una *función formativa*. En este caso, las pruebas nacionales tienen *bajas implicancias* o mínimas consecuencias. La información sobre resultados se difunde a la prensa y también se *disemina*, es decir, su distribución es planificada en relación con destinatarios específicos, cada institución la recibe en distinto nivel de desagregación: por grado/curso, por alumno, por nivel socioeconómico, relaciones estadísticas entre desempeños individuales y normativos por grado/curso, entre otros. Cada tipo de información es proporcionada en relación con cada una de las áreas o campos de conocimiento abarcados por las pruebas, principalmente Matemática y Lengua.

En general, estamos encontrando que en Argentina y Uruguay, las pruebas nacionales podrían ser ubicadas hoy dentro de esta posición orientada a la función formativa.

En **Argentina** se aplica, desde 1993, el Operativo Nacional de Evaluación (ONE) con sucesivas modificaciones tanto técnicas como en la periodicidad y cobertura de aplicación y posterior comunicación de los resultados. Atentas al tipo y características de la instrumentación del ONE, hemos identificado tres momentos que se corresponden con tres gobiernos distintos: **(a)** 1993, primer año de aplicación y de administración anual; **(b)** año 2000, a partir del cual se van produciendo algunas modificaciones de naturaleza cuantitativa: disminución de cobertura, reformulación de muestras, disminución de aplicaciones censales, disminución del número de actividades en la prueba, entre otras, y con aplicación bianual. Por último **(c)**, en 2004, marca el inicio de otra fase a partir de la cual no hay aplicaciones censales sino sólo muestrales y continúa la aplicación bianual; se producen modificaciones en las pruebas y en la forma de devolución de los resultados a las instituciones.

Las variaciones son: de pruebas referidas a normas como en las dos fases anteriores se replantean **(1)** como pruebas referidas a criterios con tres niveles de desempeño; **(2)** se modifican los criterios para la

selección de las muestras; **(3)** las actividades o ítems de las pruebas son seleccionados de las actividades docentes de las distintas provincias; **(4)** para cada área de conocimiento se aplican cuatro modelos de prueba y **(5)** notable coordinación entre el ministerio nacional y las provincias.

La información de los resultados presenta un cambio respecto de las fases anteriores ya que éstos son entregados por el ministro nacional a los ministros provinciales en reuniones del Consejo Federal de Educación, a partir de lo cual, cada provincia establece las estrategias de diseminación en las instituciones, con los equipos técnicos.

Es en este punto donde puede advertirse la intención de asignar una función formativa a los resultados macro para producir una mejora en las prácticas micro de aula. A pesar de ser pruebas de *bajas implicancias*, paradójicamente, aún hoy docentes y directivos viven la aplicación de las evaluaciones nacionales como un tiempo perdido o como una amenaza latente ¿Dependerá esta percepción de la historia de quince años de aplicación durante los '90 con la posibilidad de eventuales altas implicancias para docentes? (Pacto Federal II que no se concretó por resistencia de los docentes). ¿De eventuales consecuencias para la acreditación de nivel durante la aplicación censal de finalización de la educación media, durante la primera fase? ¿De la forma de comunicación de los resultados a cada institución que hubieran planificado las provincias?

A lo largo de esas tres etapas, el proceso comprende las siguientes tendencias-rasgos: de mayor a menor incidencia en los Contenidos Básicos Comunes nacionales; de mayor a menor publicidad de los resultados a través de la prensa y de menor a mayor incidencia de los Ministerios Provinciales y del Consejo Federal de Educación.

Respecto de las funciones políticas, en estos quince años (de 1993 a 2008 en que fue la última aplicación) parece haber un desplazamiento del predominio de las funciones primarias hacia las terciarias. En su inicio (noviembre de 1993), la difusión a través de la prensa de los primeros resultados del ONE tuvo efectos deslegitimizantes para la actividad docente por los reducidos porcentajes de aciertos. Sin embargo, esas cifras reflejaban la falta de validez de contenido de las pruebas, no eran más que una muestra de la gran distancia entre la enseñanza ideal y formal supuesta en la confección de las pruebas (basada en los Contenidos Básicos Comunes –CBC– aprobados poco antes de la primera aplicación, ese mismo año) y la enseñanza real basada en los lineamientos curriculares de cada provincia. Justamente esa primera aplicación produjo en general resultados mínimos en porcentajes (60%) y aún más bajos. También eran fruto de notables defectos técnicos de elaboración de las pruebas referidas a normas. Por otro lado, los resultados en porcentajes se publicaron ordenando jerárquicamente las provincias –y las escuelas dentro de cada una de ellas– bajo la forma de *rankings*.

Las funciones políticas secundarias de las evaluaciones nacionales están vigentes en las tres etapas, con predominio de las terciarias en el presente, sobre todo para las instituciones y en relación con la información que estas mediciones producen. La *accountability* **no** presenta manifestaciones ni de naturaleza retórica ni en la acción, por eso ubicamos a la Argentina en esta posición. Pero la vigencia e instrumentación de la función formativa de estas evaluaciones queda en manos de las provincias mientras que la validez curricular y el ajuste técnico de las pruebas, en manos del Ministerio Nacional.

Uruguay también realiza la devolución de la información de resultados de las pruebas macro a cada institución desde una interpretación de utilidad para decisiones 'blandas', **sin** relación con la rendición de

cuentas o *accountability*. La pretensión oficial –al menos retórica– es que las evaluaciones macro puedan cumplir una función formativa en el nivel micro, es decir, en las aulas.

Desde 1996, la Unidad de Medición de Resultados Educativos (UMRE) comienza las acciones de evaluación a través de la aplicación de pruebas “referidas a normas”, en colaboración con otras unidades administrativas de la Administración Nacional de Educación Pública (ANEP) y tiene a su cargo las evaluaciones en educación primaria y media, las internacionales y las de ingreso a la institución de formación docente. Respecto del Ministerio de Educación, la UMRE goza de autonomía técnica pero no administrativa y los resultados se presentan a cada institución con fuertes recaudos de confidencialidad, clasificados según una escala de cuatro niveles predeterminados de suficiencia, especificando el porcentaje de alumnos que ha obtenido cada nivel y comparados en tablas con resultados de escuelas de características socioculturales similares del mismo contexto sociocultural del departamento y del país, sin establecer *ranking* entre ellas. Las escuelas están agrupadas en cinco categorías en cuya construcción se otorgó preeminencia a la dimensión cultural por sobre la dimensión económica, a partir de relevamientos de información sociocultural de acuerdo a la composición de su alumnado. Las cinco categorías del contexto sociocultural son: muy favorable, favorable, medio, desfavorable y muy desfavorable (ANEP: 2000).

Personal técnico de la UMRE presenta estos resultados en talleres con los docentes de cada institución de la muestra de aplicación, presentando también explicaciones de las actividades (o ítems) de mayor dificultad en las pruebas y al mismo tiempo, propuestas de mejora entregando material impreso al respecto. Los docentes solicitan la multiplicación de estos talleres que encuentran oportunos y valorables, aunque sólo se desarrollan en las escuelas que participaron de la muestra. Los resultados se entregan en tiempo real para que el valor para las escuelas sea inmediato. Las actividades se desarrollan los días sábados y, por su participación, los maestros reciben una compensación económica equivalente al 25% del salario de 1º grado cubriéndose además los gastos de traslado y almuerzo (Ferrer, 142).

Por otro lado, se pretende que la evaluación nacional permita identificar escuelas con *experiencias relevantes* en cuanto a su capacidad de promover el aprendizaje de los alumnos, sobre todo en contextos socioculturales desfavorables o muy desfavorables.

Respecto de la rendición de cuentas o *accountability*, las autoridades sólo proporcionan información muy global a la prensa. No es un objetivo pero está en estudio transformar la evaluación del 12º grado en requisito de graduación (Benveniste, 2004: 466) por lo que, mientras tanto, las funciones políticas predominantes son las terciarias y secundarias.

En Brasil encontramos un sistema de evaluación nacional desde 1990 y subsistemas de evaluación en varios estados federales⁸: Paraná desde 1995, Sao Paulo desde 1996 y Minas Gerais desde 2000, dos subsistemas de los estados de Ceará y de Río de Janeiro en 1992 y el del estado de Bahía establecido más recientemente.

El Sistema Nacional para la Evaluación de la Educación Básica (SAEB) es implementado por el Ministerio Federal de Educación desde 1990 y ha sido objeto de varias modificaciones desde su creación en cuanto a la estructura de la muestra nacional, la inclusión de muestras de instituciones privadas y

⁸ Los subsistemas de evaluación estadual requerirían un estudio particularizado ya que la situación de cada uno respecto de la rendición de cuentas o *accountability* difiere acorde sean sus objetivos de política educativa.

elaboración de cuestionarios a docentes, alumnos y directivos en relación con 'factores asociados al rendimiento' (socioambientales) basados en el concepto de *efectividad* de las escuelas.

La agencia evaluadora es el SAEB que está administrada por el Instituto Nacional de Estudios e Investigaciones Educativas (INEP), órgano dependiente del Ministerio Federal de Educación, aunque con cierto grado de autonomía administrativa. Como subsistema dentro del INEP se encuentra el Examen Nacional de Enseñanza Media (ENEM) que desde una relación técnica con el SAEB, en forma independiente aplica y administra sus pruebas a los alumnos del último curso del nivel medio de escolaridad que, al ser de participación voluntaria, requiere de inscripción y registro previo de quienes acepten ser evaluados.

Como la aplicación del SAEB es muestral no se realizan informes individuales de y a las escuelas sino que la INEP elabora informes específicos de tipo pedagógico *general* referido a algunos ítems (o actividades) de la prueba, desempeños promedio y algunas explicaciones posibles de la incorrecta selección de respuestas en cada uno de ellos.

Una característica distintiva de las pruebas del SAEB es que luego de su aplicación se realizan estudios de seguimiento cualitativo en aula para comprender qué factores pedagógicos y qué aspectos interactivos pueden incidir en los resultados. Con esta información y la de los resultados, se realizan reuniones informativas y asambleas en todos los estados en las que participan equipos técnicos estatales y locales, docentes, directores, universidades y medios de prensa. Los docentes han requerido que se coloque el acento en las características técnicas y en la matriz de referencia de las pruebas en análisis, en cuanto a los contenidos y las competencias consideradas.

En relación con nuestro análisis podemos afirmar que los resultados del SAEB **no** tienen altas implicancias para docentes, alumnos o las instituciones. Los resultados del ENEM, de administración voluntaria, son utilizados como datos para la obtención de becas integrales o parciales para estudiantes de sectores sociales desfavorecidos o como parte de procesos selectivos por instituciones de nivel superior. No consideraríamos de alta implicancia esta prueba pues las probables consecuencias que pudieran derivarse de los resultados del ENEM son conocidas y promovidas por los propios sujetos.

En cuanto a las funciones políticas del SAEB, habría un predominio de las funciones secundarias en tanto las matrices de contenidos y de estándares de aprendizaje; son comunes para todos, independientemente de sus características socioculturales.

Hemos de profundizar estas iniciales articulaciones interpretativas a fin de elaborar la comparación propiamente dicha, más allá de la yuxtaposición que hemos obtenido hasta el momento...

UNA ÚLTIMA INTERROGACIÓN: ¿POR QUÉ LA EVALUACIÓN?

No es casual que el 'vendaval evaluativo' marque el paso de las reformas educativas de manera más o menos firme, que cual 'control remoto' condicione siempre en alguna medida las macro y microdecisiones institucionales, que trastorne las interacciones educativas de relaciones entre sujetos a relaciones entre juicios. Las mediciones de rendimiento estudiantil a escala macro están conformadas según una lógica económica de adopción de estándares, uso de indicadores de desempeño, competencia y rendición de cuentas, condensando una metodología racionalizadora de la asignación y distribución de recursos. ¿Hasta cuándo?

REFERENCIAS BIBLIOGRÁFICAS

- Angulo Rasco, J.F. (1999) "El neoliberalismo o el surgimiento del mercado educativo", en ANGULO RASCO, J.F. y otros *Escuela pública y sociedad neoliberal*. Madrid, Buenos Aires: Miño y Dávila Editores.
- _____ (1995) "La evaluación del sistema educativo: algunas respuestas críticas al por qué y al cómo", en *Volver a pensar la educación. Prácticas y discursos educativos*, Vol. 2. Madrid: Morata.
- _____ (1992) *Descentralización y evaluación en el sistema educativo español. Algunas claves para el pesimismo*. Publicación de la Universidad de Cádiz.
- ANEP (2000) *Evaluaciones Nacionales de Aprendizajes en Educación Primaria en el Uruguay (1995-1999)*. En <http://www.anep.edu.uy/observatorio/paginas/enaepu.pdf>
- Benveniste, L. (2004) "La evaluación estudiantil como reflejo de la relación estado-sociedad" en Carnoy, M., Cosse, G. y Cox, C. *Las reformas educativas en la década de 1990: un estudio comparado de Argentina, Chile y Uruguay*. BID, Ministerios de Educación de Argentina, Chile y Uruguay, Grupo Asesor de la Universidad de Stanford.
- Betancur, N. (2007) "¿Hacia un nuevo paradigma en las políticas educativas? Las reformas de las reformas en Argentina, Chile y Uruguay (2005-2007)" en *Revista Uruguaya de Ciencia política*, N° 16. Montevideo: ICP.
- Broadfoot, Patricia (1984) 'From public examinations to profile assessment': the French experience' en Broadfoot, P. (Ed.) *Selection, Certification and Control*. London: Falmer Press.
- Brooke, N. (2007) *Accountability Educativa en Brasil: una visión general*. Universidad Federal de Minas Gerais. En [http://www.oei.es/evaluación educativa/Brooke.pdf](http://www.oei.es/evaluación%20educativa/Brooke.pdf) .
- Brown, M. (2001) "La tiranía de las carreras de caballos internacionales" en SLEE, R. y G. WEINER *¿Eficacia para quién? Crítica de los movimientos de las escuelas eficaces y de la mejora escolar*. Madrid: Akal.
- Corvalán, J. y R.W. Mcmeekin (2006) 'Accountability' educativa: posibilidades y desafíos para la América Latina a partir de la experiencia internacional. Santiago de Chile: PREAL, CIDE.
- Dorn, Sh. (1998) *The political legacy of school accountability systems*. Educational Policy Analysis Archives. En <http://epaa.asu.edu/epaa/v6n1.html>
- Elliott, J. (1990) *La investigación-acción en educación*. Madrid: Morata.
- Fernández, S. (2007) *Evaluación de la calidad educativa en Chile. Escuelas pobres pero efectivas... efectivas pero pobres...* Ponencia presentada en las Primeras Jornadas Nacionales de Investigación Educativa. Mendoza, 3 y 4 de mayo de 2007.
- Ferrer, G. (2006) *Sistemas de evaluación de aprendizajes en América Latina. Balances y desafíos*. Santiago de Chile: PREAL.
- Mancebón Terrubia, M. J. (1999) "La función de producción educativa: algunas conclusiones de interés en la especificación de los modelos de evaluación de la eficiencia productiva de los centros escolares" en *Revista de Educación* N° 318. Madrid: Ministerio de Educación y Ciencia.
- Mizala, A. y Romaguera, P. (2006) "Desafíos metodológicos de los sistemas de evaluación e incentivos en educación. El caso del SNED en Chile. PREAL/GDN", en RAVELA, P. "¿Qué es la rendición de cuentas? Hacia la responsabilidad compartida", *Fichas didácticas*. Ficha N° 11. PREAL-GTEE.

Radulovich, S. (2008) *Tan lejos y tan cerca... o cómo crear ficciones educativas...* La Plata, Ponencia presentada en las V Jornadas de Sociología de la UNLP.

_____ (2003) "Evaluaciones nacionales y homogeneización cultural: ¿qué sucede con las evaluaciones de Lengua?" en Fernández, A. y Rodas, J (Comp.) *Historia y sociolingüística del español en el noroeste argentino*. Salta, Universidad Nacional de Salta.

_____ (2000) *Evaluación en contextos de Reforma educativa... o una parte de lo que nos aqueja*. Ponencia presentada en el Encuentro 2000 Educación y Reforma ¿ficción o realidad? Tartagal, Salta.

Raivola, R. (1990) "¿Qué es la comparación? Consideraciones metodológicas y filosóficas" en Altbach, Ph. y G. Kelly, G. (Comp.) *Nuevos enfoques en Educación Comparada*. Madrid: Mondadori.

Ravela, P. (2006) "¿Qué es la rendición de cuentas? Hacia la responsabilidad compartida" en *Fichas didácticas*. Ficha N° 11, PREAL-GTEE.
