

Archivos Analíticos de Políticas Educativas

Revista Académica evaluada por pares

Editor: Sherman Dorn

College of Education

University of South Florida

El Copyright es retenido por el autor (o primer coautor) quien otorga el derecho a la primera publicación a **Archivos Analíticos de Políticas Educativas**. Los artículos que aparecen en **AAPE** son indexados en H.G. Wilson & Co. y en el Directory of Open Access Journals (<http://www.doaj.org>).

Editores Asociados para Español y Portugués

Gustavo Fischman
Arizona State University

Pablo Gentili
Laboratorio de Políticas Públicas
Universidade do Estado do Rio de Janeiro

Volumen 14 Número 18

Julio 5, 2006

ISSN 1068-2341

Problemas Técnicos y Usos Políticos de las Evaluaciones Nacionales en el Sistema Educativo Argentino

Silvina Gvirtz
Silvina Larripa
Universidad de San Andrés, Argentina
y
Verónica Oelsner
Universidad Humboldt, Berlín, Alemania

Citación: Gvirtz, S., Larripa, S., y Oelsner, V. (2006). Problemas Técnicos y Usos Políticos de las Evaluaciones Nacionales en el Sistema Educativo Argentino. *Archivos Analíticos de Políticas Educativas*, 14(18). Recuperado [fecha] de <http://epaa.asu.edu/epaa/>

Resumen

Este artículo presenta resultados de investigaciones que han explorado las relaciones entre la dimensión técnica y la dimensión política de la evaluación de sistemas educativos, tomando como caso el sistema de exámenes nacionales vigente en la Argentina desde 1993. En la primera parte del trabajo exponemos algunos problemas técnicos que ha presentado la puesta en marcha del sistema de evaluación nacional en el mencionado país. En la segunda parte realizamos una lectura de estos inconvenientes técnicos en el marco del contexto político de reforma educativa en el que el sistema de evaluación surge y se desarrolla. Asimismo, analizamos los usos efectivos que se han hecho de la información provista por las evaluaciones. Por último, en las conclusiones, presentamos reflexiones acerca de la función de las

evaluaciones nacionales en contextos de reforma educativa, y de las posibilidades de su consolidación como sistemas que informen de forma válida y confiable sobre la marcha de la educación en el mediano y largo plazo.

Abstract

This article presents results from different research investigations which have explored the relations between the technical and the political dimension of the assessment of educational systems. The case study taken on for this matter is the national evaluation system in force in Argentina since 1993. In the first part we present some technical problems which the implementation of this system has encountered in this country. In the second part we carry out an analysis of these technical inconveniences, within the political context of educational reform in which the evaluation system arises and develops. In addition, we present an analysis of the effective use of the information provided by the evaluations. Finally, in the conclusions, we present some considerations on the role of national evaluations in educational reform contexts, and on the prospects of their consolidation as systems which inform in a valid and reliable form about the course of education in the mid and long term.

Contexto de Surgimiento Y Caracterización del Sistema Nacional de Evaluación de la Calidad de la Educación en la Argentina (SINEC)¹

El surgimiento del sistema nacional de evaluación en la Argentina no es un hecho aislado ni un fenómeno original. Responde a tendencias más globales en política educativa. En el nivel internacional, el desarrollo y la propagación de los sistemas nacionales de evaluación de la educación se originan en las preocupaciones por la eficacia y la eficiencia de los sistemas educativos. Esto se observa a partir de la década del '60, principalmente en los Estados Unidos y un poco más tarde, en algunos países europeos.

Preocupaciones similares, fundamentalmente en la década del '90, dan impulso a la implementación de sistemas nacionales de evaluación en América Latina. Estos sistemas de evaluación se desarrollan en el marco de reformas educativas más amplias que comienzan con la gestación de leyes de educación a comienzos de esa década. A partir de la sanción de estas leyes, se operan cambios en tres aspectos fundamentales de la educación: se modifican *los modos de gobierno y gestión* de los sistemas de educación (tendiente a una descentralización o desconcentración del poder de dirección de los órganos centrales hacia los estados, provincias o municipio, dependiendo del país); se produce un cambio en la *estructura* general de los sistemas educativos (que involucra en la mayoría de los casos la ampliación de la obligatoriedad escolar además de modificaciones en los propósitos de la formación inicial, básica y postbásica de los niños y jóvenes); y como consecuencia de los puntos anteriores, se renueva y modifica el *currículum* escolar para cada uno de los niveles educativos².

¹Silvina Larripa y Verónica Oelsner desean agradecer a la Fundación Luminis, institución de la que han sido becarias, por el apoyo otorgado para la realización de las investigaciones, cuyos resultados presentan en este trabajo.

² Cabe señalar que si bien estos cambios caracterizan las transformaciones operadas en la región, no en todos los países se verificaron todos ellos. En Uruguay, por ejemplo, El Estado Nacional continúa siendo el órgano principal de gobierno del sistema educativo del país.

Como parte de de estas transformaciones, se ponen en marcha sistemas de evaluación en Honduras, Colombia, República Dominicana, Argentina, Brasil y El Salvador; países a los que siguen más adelante otros como Uruguay, México, Costa Rica, Bolivia, Paraguay, Nicaragua o Venezuela (Tiana Ferrer, 2000). Cuba y Chile los implantan con anterioridad. En general, todos estos sistemas se insertan dentro de las burocracias ministeriales:³ en Chile, el Ministerio de Educación establece el Sistema de Medición de la Calidad de la Educación (SIMCE); en México se crea el Sistema Nacional de Evaluación Educativa (SNEE); y en Brasil, el Sistema Nacional de Evaluación de la Educación Básica (Braslavsky, 1999).

Tabla 1

Año de implementación de los sistemas de evaluación por país

País	Año
Argentina	1993
Bolivia	1996
Brasil	1993
Chile	1988
Colombia	1991
Costa Rica	1995
Cuba	1975
El Salvador	1993
Honduras	1990
México	1994
Nicaragua	1998
Paraguay	1996
República Dominicana	1992
Uruguay	1996
Venezuela	1995

Fuente: Arancibia (1996).

Sin embargo, si se consideran las características de los sistemas de evaluación implementados en los distintos países, se ve que éstos se diferencian entre sí en aspectos relativos a su diseño, implementación y desarrollo. Cada país define sus objetivos y diseña sus instrumentos, enfatizando distintos usos y funciones para sus evaluaciones, llegando a responder, según el caso, a paradigmas de evaluación y de usos de los resultados muy diferentes. Los sistemas de evaluación de Chile y Uruguay representan un claro ejemplo en tal sentido. Mientras que el primero tiene un sistema de evaluación más orientado hacia la rendición de cuentas y a la conformación de rankings de instituciones, el segundo enfatiza la necesidad de contar con información acerca del sistema educativo, a modo de diagnóstico, para guiar e informar políticas compensatorias del gobierno central (Benveniste, 2000a; 2000b).

Entre estos diferentes sistemas de evaluación de escala nacional implementados en los 90 en la región se encuentra el Sistema Nacional de Evaluación de la Calidad de la Educación (SINEC) de la Argentina, el cual se crea en 1993 bajo el ámbito del Ministerio de Cultura y Educación de la Nación. El SINEC surge en el marco de un amplio proceso de reforma del sistema educativo, impulsado por dos leyes: la Ley de Transferencia de 1991 (N° 24.049) por la cual se termina de

³ En otros países, en cambio, agencias que no dependen en forma directa del Estado o son totalmente independientes, son las que están a cargo de la evaluación del sistema educativo a escala nacional.

transferir el gobierno de las escuelas desde la Nación hacia las jurisdicciones,⁴ y la Ley Federal de Educación de 1993 (N° 24.195), a partir de la cual se inicia una transformación profunda del sistema educativo argentino. Este proceso de reforma implicó, fundamentalmente, la redefinición de las atribuciones y responsabilidades del Estado nacional y los Estados provinciales en materia educativa, un cambio en la estructura de ciclos y niveles educativos, y la aparición de nuevos marcos curriculares.

En este nuevo contexto, el Ministerio de Educación nacional debe convertirse en articulador de políticas educativas. El sistema de exámenes nacionales, dependiente de dicho Ministerio, se presenta entonces desde su surgimiento como un instrumento para garantizar dicha articulación y mejorar la calidad de la educación en todo el país. Así, desde su formulación, el SINEC:

...se orienta a desarrollar el proceso de evaluación para ser aplicado en todas las jurisdicciones, ciclos, niveles y regímenes, con el fin de proponer los ajustes que corresponden a efectos de mejorar la calidad educativa. (Ministerio de Educación, s/f, p. 6)

El Sistema Nacional de Evaluación de la Calidad (...) busca a través de la evaluación de logros académicos de los estudiantes suministrar información válida y confiable sobre cuánto es lo que aprende y qué es lo que aprende el alumno en su recorrido por el Sistema Educativo y acerca de cuáles son los factores asociados a ese aprendizaje. Este sistema proporcionará los insumos fundamentales para la toma de decisiones relacionada con la formulación de políticas educativas de mejoramiento de la calidad. (Ministerio de Educación, s/f, p. 2)

Entre sus objetivos prioritarios, el SINEC se propone:

...mejorar la calidad de la educación a través de la producción sistemática y permanente de información acerca del desempeño y los resultados del sistema educativo. (MCyE, 1997, p. 13)

Y entre los específicos, el SINEC busca:

...implementar operativos nacionales de evaluación de la calidad a fin de producir información sobre el rendimiento de los alumnos para la elaboración de políticas de mejoramiento. (Ministerio de Educación, s/f, p. 6)

Desde 1993 comienzan a implementarse anualmente los Operativos Nacionales de Evaluación (ONE). Estos operativos consisten en la aplicación de pruebas de conocimiento y cuestionarios complementarios a una amplia muestra de secciones escolares de establecimientos públicos y privados de todo el país. Durante el año 1997 y sólo para el último año del nivel medio, esta evaluación se administró en todas las escuelas (adquiriendo, así, carácter de censo). Lo mismo sucedió en el año 2000 con las evaluaciones aplicadas en el sexto grado de la Educación General Básica (EGB). Las pruebas de conocimiento se proponen medir los niveles de logro académico de los alumnos, mientras que los cuestionarios complementarios recaban información que permite explicar tales resultados (por ejemplo, relativa al nivel socioeconómico del alumnado).

⁴ Argentina es un país federal que cuenta con 24 jurisdicciones: 23 provincias y la Ciudad Autónoma de Buenos Aires. Históricamente, el Estado nacional había tenido un papel central en el sostenimiento y gobierno directo del sistema de educativo en todo el país. Esto comienza a cambiar hacia fines de la década del 70, cuando el gobierno de las escuelas de nivel primario comienza a transferirse del ámbito nacional a las provincias y a la Ciudad de Buenos Aires. Con la Ley de Transferencia de 1991, culmina el traspaso de las escuelas secundarias nacionales a las jurisdicciones.

A lo largo de los años, el presupuesto asignado al Sistema Nacional de Evaluación de la Calidad se ha ido incrementando. Según estimaciones de Andrada, Narodowski y Nores (2002), en 1997 prácticamente se duplica el presupuesto total (\$4.285.000) respecto de los años anteriores (\$2.280.823 en 1995 y \$2.540.600 en 1996), lo cual coincide con la primera aplicación censal del Operativo en el último año del Nivel Medio. La cifra se mantiene sin variaciones importantes hasta el año 2000, en que se triplica (\$12.000.000), al extender el universo y las áreas evaluadas. Se agrega el sexto año de EGB a la evaluación censal y el país comienza a participar de evaluaciones internacionales.⁵ De este modo, la evaluación del sistema educativo ha alcanzado un lugar central en la agenda político-educativa argentina de la última década.

Objetivos, Aspectos Metodológicos y Enfoque Teórico del Trabajo

En este artículo nos centramos en el Sistema Nacional de Evaluación de la Calidad educativa de la Argentina y analizamos su solvencia y los usos dados a la información producida desde una perspectiva tanto técnica como política. Para ello, nos basamos en la reunión de los principales resultados de dos tesis de maestría que han tomado como objeto de estudio al SINEC; una, focalizando en los problemas técnicos, la otra, en los aspectos políticos.⁶

Entre los problemas técnicos, se ha dado prioridad al análisis del grado de comparabilidad longitudinal de las medidas de rendimiento educativo provistas por el SINEC, es decir, en qué medida el sistema de evaluación nacional respondía a su propósito central de proveer—año a año—información comparable sobre el rendimiento de los alumnos, la cual es necesaria para conocer el estado y los cambios en el tiempo que sufre el desempeño académico de los estudiantes. Para ello, se realizó un análisis de la composición de las pruebas nacionales aplicadas entre los años 1995 y 2000 en el tercer año de la educación básica, en las áreas de Lengua y Matemática. También fueron objeto de análisis las tablas de especificaciones (instrumento que se emplea para la construcción de las pruebas) así como informes de resultados y otros documentos de difusión producidos por el SINEC durante el mismo período. El estudio documental se complementó con entrevistas no estructuradas a funcionarios y ex funcionarios directamente vinculados al SINEC, así como a especialistas en evaluación externos.

En cuanto a los aspectos políticos, se han indagado las funciones no declaradas del sistema de evaluación. Partiendo de una concepción de la evaluación como actividad política (Benveniste, 2000a y 2000b; House, 1994; MacDonald, 1989), se han considerado, por un lado, los postulados que sostienen que más allá de lo formalmente proclamado, los sistemas de evaluación suelen ser utilizados por los gobiernos para legitimar, informar y a veces controlar, las actividades de los diferentes actores del sistema escolar (véase por ejemplo Weiss, 1998 y House, 1994). Por otro lado, se han tomado como referentes trabajos que asocian la evaluación con nuevas formas de ejercer el control o dar dirección al sistema educativo desde el nivel central de gobierno, en el marco de sistemas descentralizados (entre ellos, Lundgren, 1992 y Weiler, 1996). Desde este encuadre teórico,

⁵ En el año 2000 Argentina participa por primera vez de estudios internacionales de dos organismos: la IEA (Asociación Internacional para la Evaluación del Logro Educativo) y la OCDE (Organización para la Cooperación y el Desarrollo Económico).

⁶ El trabajo que ha abordado el estudio del Sistema Nacional de Evaluación de la Calidad en Argentina desde un enfoque técnico es el de Larripa (2003), *El Sistema Nacional de Evaluación de la Calidad educativa: acerca de la comparabilidad de sus resultados, Argentina: 1995-2000*. Tesis de Maestría. Victoria: Universidad de San Andrés; mientras que la investigación que ha explorado el SINEC desde una perspectiva política es la de Oelsner (2002). *Las evaluaciones nacionales de la calidad de la educación: sus usos y funciones en el sistema educativo argentino (1994-2001)*. Tesis de maestría. Victoria: Universidad de San Andrés.

se han estudiado para el período 1994-2001 los usos que el Ministerio de Educación de la Nación ha dado al SINEC, a través de las acciones que ha desarrollado a partir de cada Operativo Nacional de Evaluación (ONE).⁷ A este respecto, se identificaron las líneas de acción desarrolladas por el Ministerio a partir de la aplicación de cada ONE, se describió cada una de las líneas de acción identificadas y, a partir de la articulación de las diferentes líneas de acción y de su relación con el contexto político, se propusieron las funciones políticas cumplidas por el SINEC a lo largo del período (1994-2001). La técnica de recolección de datos consistió en la revisión de fuentes secundarias: materiales elaborados por el Ministerio de Educación y legislación producida durante el período (leyes, decretos y resoluciones). La revisión de periódicos de difusión masiva y las entrevistas con actores clave cumplieron una función indispensable para completar la indagación, pero se consideraron como material de apoyo.

A continuación, exponemos primero un análisis de la construcción de los instrumentos y sus problemas. Luego, desde una mirada de corte político, profundizamos en los diversos usos dados al sistema de evaluación por parte del Ministerio nacional. En la última sección, presentamos las conclusiones que se desprenden de ambas líneas de análisis.

Problemas en la Construcción e Interpretación de la Información Provista por el SINEC

A una década de la implementación de programas de evaluación de escala nacional en Latinoamérica—en la que los países dirigieron sus acciones al logro de un funcionamiento sistemático de los mismos—, los esfuerzos parecen concentrarse ahora en la superación de problemas y deficiencias constatados en su funcionamiento. Siguiendo a Ravela (2000), los problemas fundamentales detectados en muchos de los sistemas de evaluación de la región son: la falta de definiciones precisas acerca del rol del sistema de evaluación y de los impactos esperados sobre el sistema educativo; y la escasa conciencia por parte de los órganos responsables del sistema, de que las decisiones técnicas adoptadas en el diseño de la evaluación pueden servir a unos fines y no a otros. Como consecuencia de ello, se verifican deficiencias, como los “malos usos” o interpretaciones poco válidas de los resultados obtenidos, así como problemas en el diseño del sistema y de los instrumentos de evaluación que impiden el aprovechamiento de la información relevada.

A continuación, presentamos algunos problemas técnicos detectados en el Sistema Nacional de Evaluación de la Calidad (SINEC) que funciona en la Argentina. Confiamos en que el tratamiento de los mismos pueda ser de utilidad para reflexionar acerca de las vinculaciones entre la dimensión técnica y la dimensión política que atraviesa a todo sistema de evaluación. Como intentaremos mostrar a continuación, las decisiones técnicas se toman en el marco de un contexto político particular. Al mismo tiempo, los resultados que se obtienen bajo un determinado modelo de evaluación posibilitan ciertos usos de los resultados y no otros.

Específicamente, haremos referencia a dos problemas fundamentales detectados en el SINEC: los problemas en la interpretación de los resultados de rendimiento académico (debidos a un desconocimiento de las implicancias que impone el enfoque de evaluación bajo el cual se diseñan los instrumentos de medición), y los problemas en la construcción de los instrumentos que impiden la generación de información comparable a lo largo del tiempo.

⁷ Centramos la mirada en las acciones desarrolladas a partir de los resultados, entendiendo que son dichas acciones las que determinan las funciones efectivamente cumplidas, más allá de las que se proclaman y de las que permiten técnicamente los instrumentos diseñados.

Problemas en la interpretación de los resultados de rendimiento

Tanto las definiciones clásicas como las más actuales del término *evaluación* coinciden en un aspecto central: se evalúa con la finalidad de tomar algún tipo de decisión (Gronlund, 1973; Martínez Arias, 1996). Como consecuencia de ello, las decisiones técnicas implicadas en el desarrollo de un sistema de evaluación deberían orientarse de acuerdo con el tipo de decisiones o intervenciones que deseen realizarse con los resultados que esperan obtenerse. Siguiendo a Martínez Arias (1996), los propósitos para los cuales suelen emplearse las puntuaciones de un test o sistema de tests se incluyen en alguna de las siguientes categorías: a) decisiones relativas al nivel de instrucción, donde se busca conocer si una persona o grupo de personas dominan o no ciertas capacidades y/o conocimientos predeterminados de antemano; b) decisiones de diagnóstico, en las que el objetivo es identificar deficiencias o errores particulares en el rendimiento de una persona o grupo, con objeto de emprender algún tipo de intervención; c) decisiones de selección, donde el objetivo es admitir o excluir sujetos, por lo que el test se orienta a seleccionar, de entre los examinados, a aquellos más “capaces” en algún sentido; d) decisiones de asignación, en las que el objetivo es determinar en qué lugar debe colocarse a una persona dentro de una jerarquía de puestos; y e) decisiones de clasificación, donde el objetivo también es asignar personas, pero ya no a un nivel sino a un tipo de ocupación o puesto.

De acuerdo con el tipo de decisión a tomar, es posible optar por un determinado enfoque de evaluación. Actualmente, los enfoques conceptuales en los que puede basarse cualquier sistema de tests son dos: la aproximación de los rasgos latentes o medida de los atributos de las personas, interesada en situar a los distintos examinados a lo largo de un continuo que define o representa el atributo medido, y que se conoce como Tests Referidos a las Normas (TRN); y la aproximación del dominio, que pone el acento en conocer el grado o nivel de ejecución del examinado en un conjunto de competencias referidas a un dominio o criterio definido con anterioridad a la prueba, y que se conoce como Tests Referidos al Criterio (TRC).

Volviendo a la tipología de decisiones, podemos decir que en los dos primeros tipos de decisiones (diagnosticar y conocer el nivel de instrucción), la aproximación de los TRC resulta más útil, ya que en este caso el resultado del test brindará información sobre el grado en que los alumnos dominan un determinado universo de contenidos. En cambio, tres últimos tipos (selección, asignación y clasificación) la aproximación de los TRN resulta más acorde, ya que lo que se precisa obtener es una medida diferencial del atributo medido (Martínez Arias, 1996).

En el caso del sistema nacional de evaluación argentino, el objetivo de las pruebas se enmarca en los dos primeros tipos de decisiones. Así lo expresan los documentos oficiales:

El objetivo fundamental de estas pruebas es brindar información sobre el qué y cuánto aprenden los alumnos durante su paso por el sistema educativo, así como indagar en qué medida van adquiriendo las capacidades y los contenidos que su propio desarrollo, los diseños curriculares y la sociedad misma requieren (Ministerio de Cultura y Educación [MCyE], 1998, p. 17).

No obstante estas finalidades, la aproximación que predomina en las etapas de construcción de las pruebas no es de tipo criterial sino que está mucho más próxima a la normativa. El énfasis está colocado en procurar que los instrumentos tengan suficiente potencialidad para discriminar entre alumnos conforme a su rendimiento académico, antes que en indagar aquello que los estudiantes son capaces (y no) de hacer en las diversas áreas de contenido del currículum.

Por ejemplo, para la construcción del instrumento definitivo se tienen en cuenta sólo aquellas preguntas que son contestadas en las evaluaciones piloto por más del 20% y por menos del 80% de la población. Este criterio de selección optimiza el poder discriminativo del instrumento

(Ebel, 1977) a fin de poder establecer con mayor precisión comparaciones al interior de la población evaluada. Pero al mismo tiempo impide lecturas absolutas de los resultados que se obtienen, ya que al haber descartado las preguntas dominadas por más del 80% de la población examinada en los pilotajes, los resultados no pueden ser interpretados como si fuera posible alcanzar un rendimiento del 100%.

Sin embargo, los resultados de rendimiento suelen interpretarse en términos absolutos (propio del enfoque de evaluación conforme a un criterio):

En este sentido, las aspiraciones y expectativas en términos de logro deberían estar cercanas al 100% de los resultados. Los rendimientos de un 50% están dando cuenta de muchos déficits importantes respecto de lo básico y lo central, esto es, lo indispensable que la enseñanza pública tiene que garantizar, lo cual muestra una situación deficitaria (MCyE, 1997, p. 28).

Asimismo, en los informes anuales de resultados los porcentajes de respuestas correctas se traducen en “puntajes”, lo que también promueve en el público una interpretación de los resultados en términos absolutos (ver Figura 1).

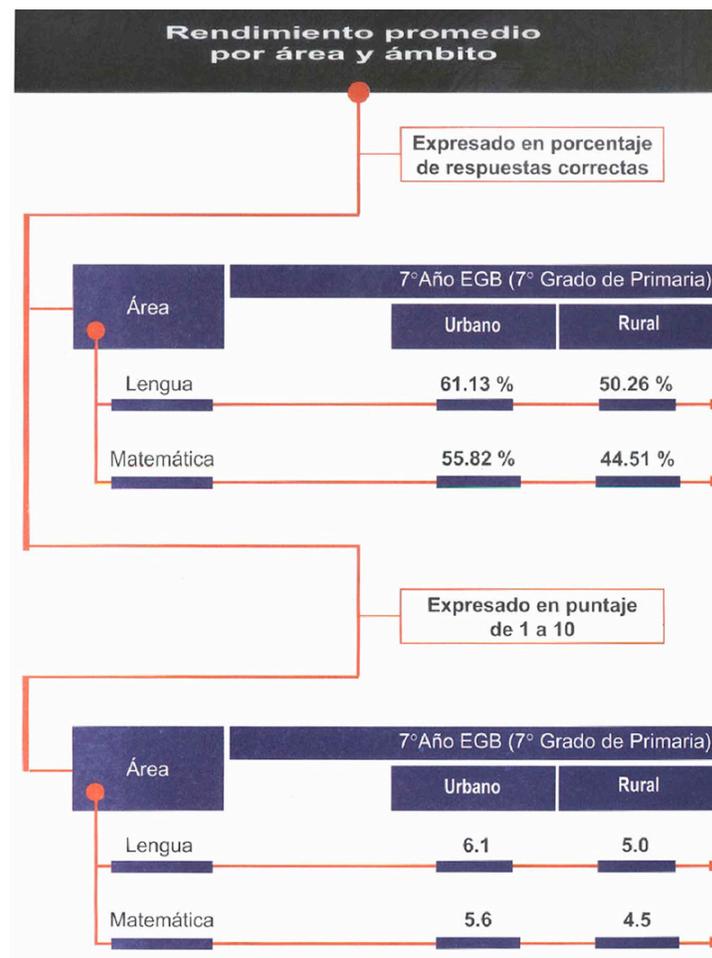


Figura 1.
Fuente: MCyE (s/f).

Este tipo de interpretación de los porcentajes de respuestas correctas resulta erróneo ya que no tiene en cuenta el hecho de que la dificultad de cada ítem de la prueba ha sido un factor previamente controlado en la selección de los ítems definitivos. En consecuencia no es posible esperar que los alumnos respondan cerca del 100% del test, ni se debería interpretar su puntuación sobre la base de 10 puntos.

En segundo lugar, como el poder de discriminación del ítem es el criterio predominante en la selección de ítems, muchas buenas preguntas desde el punto de vista curricular pueden ser descartadas. Como consecuencia de ello, los resultados de este tipo de pruebas no brindan en realidad mucha información sobre lo que los alumnos efectivamente son capaces de hacer en cada área curricular. Siguiendo a Ravela (2001):

...como la selección de los ítems que forman parte de las pruebas en general se efectúa privilegiando los que tienen niveles de dificultad media y no a partir de una determinación sustantiva de cuál es el nivel de logro que se espera que alcancen los alumnos, en realidad el resultado no dice nada acerca de en qué grado los estudiantes son capaces de hacer lo que se espera de ellos. En otras palabras, los resultados usualmente observados de entre 50% y 60% de respuestas correctas en realidad están determinados a priori al seleccionar los ítems, pero no significan que se esté logrando ese porcentaje de lo esperado como resultado del paso de los alumnos por el sistema educativo”. (p.21)

Por esta razón, Popham (1999) concluye que las pruebas de rendimiento construidas de este modo no deberían utilizarse para evaluar la calidad de la educación sino “para hacer las interpretaciones comparativas [entre estudiantes o grupos de estudiantes] que se supone deben suministrar” (p. 4), las cuales están referidas a una norma y a un grupo relativamente pequeño y particular de ítems. Por ello, propone concebir pedagógicamente los resultados de las pruebas en términos de “aproximaciones gruesas al status de un estudiante respecto a la esfera de contenidos representada en la prueba” (p. 5).

Problemas en la Comparación Longitudinal de Resultados de Rendimiento

Una de las funciones principales de los sistemas de evaluación de escala nacional es el monitoreo del estado y las variaciones en la “calidad” de la educación, la cual es usualmente medida en términos del rendimiento académico que los alumnos alcanzan en los tests. El cumplimiento de esta tarea implica poder establecer comparaciones diacrónicas entre los resultados que se obtienen en cada relevamiento. Para ello, tanto la dificultad como los contenidos de los tests deben ser equivalentes o equiparables (De Landsheere, 1996).

En el caso argentino, los instrumentos aplicados para medir el rendimiento académico han sufrido variaciones de importancia de un año a otro. Estas variaciones se han dado en dos aspectos fundamentales: el tipo de habilidades y contenidos medidos en cada una y la dificultad de las tareas propuestas en los ítems. En vistas de ello y a fin de poder emplear la información recabada, el Ministerio de Educación encomendó a la agencia australiana A.C.E.R. (Australian Council for Educational Research) la realización de un estudio de equiparación de las puntuaciones correspondientes a las pruebas de Lengua y Matemática aplicadas entre 1995 y 1999 en los distintos años escolares evaluados.

Si bien este estudio logró neutralizar las diferencias de dificultad de las pruebas, una de nuestras investigaciones (Larripa, 2003) señala que dicho estudio no tuvo a su alcance la posibilidad de resolver las variaciones interanuales en las habilidades y contenidos evaluados, las cuales fueron importantes. A modo de ejemplo, la Figura 2 muestra diferencias en los tópicos de ortografía y

gramática evaluados en las pruebas administradas a alumnos de 3er. año (8/9 años) entre 1995 y 2000:

LENGUA, Pruebas de 3er. año EGB (1995-2000): porcentaje de ítems que evalúan Reflexión sobre los hechos del lenguaje, según temas que conforman dicha subárea y año.

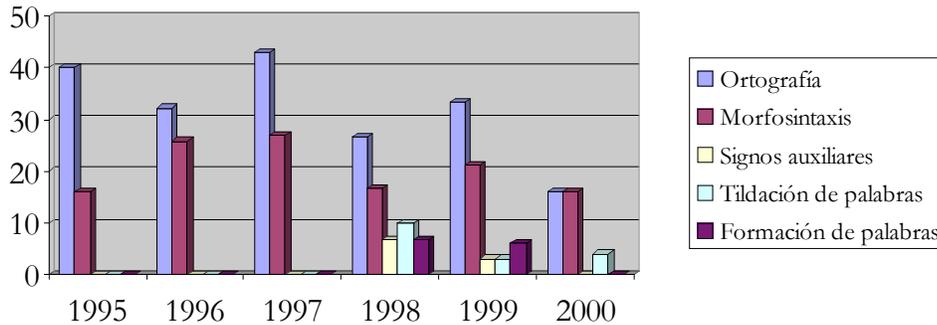


Figura 2. Porcentaje de ítems que evalúan Reflexión sobre los hechos del lenguaje de la Pruebas de 3er. año EGB (1995-2000), según temas y año.

Fuente: Larripa (2003)

Los datos presentados en la Figura 2 señalan, en primer lugar, que las categorías temáticas medidas por las pruebas administradas por el SINEC entre 1995 y 2000 en el 3er. año de escolaridad básica para el área de Lengua/subárea “Reflexión acerca de los hechos del Lenguaje”, han variado considerablemente. Se observa que las únicas dos categorías temáticas que se han mantenido presentes en todas las pruebas, desde 1995 hasta el año 2000, han sido la de *Ortografía* y la de *Morfosintaxis*. En cambio, los temas vinculados a la *Formación de palabras* y *Signos auxiliares* han estado presentes sólo en las evaluaciones de 1998 y 1999, para dejar de ser considerados en la prueba del año 2000. Respecto de los contenidos referidos a *Tildación de palabras*, los datos señalan que estuvieron presentes en las pruebas sólo desde 1998 (y hasta el año 2000 inclusive).

En segundo lugar, y considerando ahora los porcentajes de ítems asignados a cada una de las categorías temáticas, se observa que éstos han presentado una gran variabilidad a lo largo del tiempo. *Ortografía* registra variaciones importantes año tras año: mientras que en 1995 y 1997 los valores son altos (40 y 43% respectivamente), en 1996, 1998 y 1999 rondan el 30%, para finalmente caer en el 2000 a sólo un 16% de los ítems de la prueba. *Morfosintaxis* también presenta variaciones sucesivas en el tiempo: de un 16% de los ítems en la prueba de 1995, pasa a un 26% y 27% en los dos años siguientes, para en 1998 y 2000 volver acercarse al valor de 1995.

Respecto de los temas que se evalúan sólo a partir de 1998 cabe señalar que la que presenta mayor variabilidad entre 1998 y 1999 es la categoría temática *Tildación de palabras* (7%), seguida de *Signos auxiliares* (4.7%). En cambio, *Formación de palabras* se mantiene prácticamente sin variaciones entre 1998 y 1999. En el año 2000, como ya se señaló, estas dos últimas categorías no son evaluadas por ningún ítem, mientras que *Tildación* presenta una asignación de ítems prácticamente igual a la de 1999.

Las pruebas que se asemejan más en cuanto a los temas que efectivamente se evaluaron son la de 1998 y 1999. Por otra parte, las pruebas de 1996 y 1997 son las que presentan menor variación

en las proporciones de ítems que evalúan *Morfosintaxis*, aunque entre ambas existe una diferencia del 10% en los ítems asignados a *Ortografía*. Asimismo, los dos pares de evaluaciones mencionados no presentan una distribución del número de ítems que evalúan los diversos temas de la subárea Reflexión sobre el Lenguaje que pueda ser caracterizada como similar. Y lo mismo sucede con la evaluación del año 2000, que resulta muy diferente a la del resto de las pruebas, y la de 1995, donde la proporción de ítems que evalúan contenidos de Ortografía es considerablemente más alta que la de las demás evaluaciones.

Este es sólo un ejemplo de los cambios significativos que la investigación reveló en cuanto a los contenidos evaluados por los operativos implementados por el SINEC para un mismo año escolar y área curricular, hecho que pone en cuestión la posibilidad de realizar comparaciones longitudinales de las medidas de rendimiento con alto grado de validez.⁸

Los cambios en los contenidos medidos año a año por las pruebas del SINEC se deben, en parte, a la decisión política de introducir nuevos tópicos de acuerdo con el cambio curricular que comenzó a operarse en la Argentina en el marco de la reforma educativa puesta en marcha por la Ley Federal de Educación. Sin embargo, nuestra investigación ha constatado irregularidades en los procedimientos de construcción de los instrumentos, que estarían vinculadas a un escaso conocimiento técnico de los límites dentro de los cuales estas variaciones podían ser introducidas sin afectar la comparabilidad de las mediciones. La más importante de estas falencias se evidencia en las variaciones halladas en la composición de la tabla de especificaciones.

La tabla de especificaciones (también llamada “marco conceptual” de la prueba) es un instrumento que se construye a fin de definir el contenido de un test de rendimiento. Suele asumir el formato de un cuadro de doble entrada donde se consignan, en uno de los ejes, los contenidos o subtemas a evaluar y, en el otro, el tipo de capacidad o grado de dificultad con que dichos contenidos serán medidos. Asimismo, en la tabla se determinan el porcentaje o número de ítems que se contemplarán para medir cada una de las intersecciones que se considerarán para construir el test definitivo. La construcción de la tabla de especificaciones persigue fundamentalmente dos propósitos: a) lograr un ajuste entre las capacidades y contenidos que se desean medir, y los elementos incluidos en las pruebas (es decir, aumentar la validez de contenido de las pruebas); y b) contar con la posibilidad de establecer con mayor grado de validez comparaciones interanuales del rendimiento de los alumnos evaluados en un mismo año escolar y área del conocimiento (comparaciones que efectivamente el SINEC llevó a cabo y plasmó en los informes de resultados). El primer objetivo supone la construcción de instrumentos que respeten las capacidades y contenidos considerados en las tablas de especificaciones, así como los pesos relativos o importancia otorgada a cada uno de ellos en las tablas. El segundo objetivo implica mantener estable la estructura de las tablas de especificaciones a lo largo del tiempo, lo cual, bajo el cumplimiento del primer objetivo, garantizará la posibilidad de establecer comparaciones interanuales del rendimiento del sistema educativo en términos de la “calidad” de los resultados que el mismo produce.

Nuestra investigación muestra, por el contrario, que la composición de la tabla de especificaciones (para evaluar un mismo año escolar en una misma área curricular) ha sufrido variaciones importantes a lo largo del tiempo. A continuación presentamos un ejemplo.

Considerando que el rendimiento académico es definido por el SINEC como la manifestación (o no) por parte del alumno de un conjunto de “*capacidades* determinadas en relación con un recorte del saber” (ME, 1999, p.3),⁹ se tomó la distribución general de ítems según el tipo de capacidad a medir establecido en la tabla de especificaciones de 1995 y la de 1996 para tercer año de

⁸ Para más ejemplos sobre estos problemas, ver Larripa (2003).

⁹ La definición operacional del rendimiento académico fue expuesta con mayor detenimiento en el Capítulo III. Aquí, el destacado en la misma es nuestro.

la educación básica en el área de Matemática, y se compararon entre sí, con las proporciones establecidas por la tabla de 1998 (y sus versiones equivalentes de 1999 y 2000). Se observó que la importancia otorgada a la medición de los diversos tipos de capacidades en Matemática resulta marcadamente diferente, tal y como muestra la Figura 3.

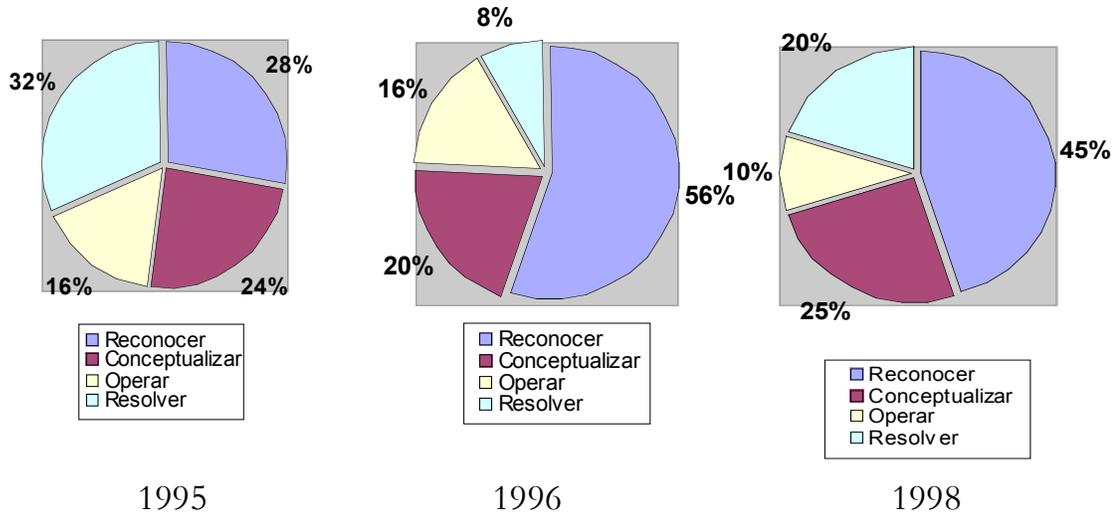


Figura 3. Porcentaje de ítems por tipo de *capacidad* medida según Tabla de Especificaciones de Matemática, 3er. año, 1995, 1996 y 1998.

Fuente: Larripa (2003)

En síntesis, para el período 1995-1998 los datos presentados no avalan las afirmaciones oficiales que sostienen que la estructura general de las tablas de especificaciones ha permanecido estable.¹⁰ Por el contrario, las marcadas variaciones en la estructura de las tablas de especificaciones observadas en los primeros cuatro años (en un lapso total de seis) ponen de manifiesto cambios importantes en la concepción acerca de qué tipo de capacidades debían ser priorizadas en la evaluación nacional de los alumnos del 3er. año de la EGB. Asimismo, denotan una falta de consenso en el ámbito del Ministerio de Educación Nacional (y en particular, en la Dirección Nacional de Evaluación, responsable inmediata del SINEC) respecto de la necesidad (ineludible en un programa de medición y evaluación de esta escala) de emplear las tablas de especificaciones como herramientas para aumentar la similitud en la composición de las pruebas definitivas, condición básica para realizar comparaciones interanuales entre los resultados académicos obtenidos luego de cada Operativo Nacional de Evaluación.

En otras palabras, y aún habiéndose equiparado la *dificultad* de los instrumentos, nuestro trabajo de investigación ha arrojado evidencias claras de variaciones importantes en las *capacidades y contenidos* evaluados por las pruebas del SINEC. Hay temas que no han sido medidos en todos los operativos; asimismo, el énfasis con que han sido medidas ciertas capacidades y contenidos ha variado significativamente de un año a otro. Estas variaciones impiden hacer comparaciones interanuales de los resultados provistos por cada una de las pruebas. Debido a estos problemas, no es posible entonces dar cumplimiento a una de las funciones principales del sistema de evaluación:

¹⁰ “Las tablas de especificaciones mantienen básicamente la estructura acordada por consenso en los años anteriores. Se han introducido en ellas algunas modificaciones que no alteran su forma básica ni su contenido, para precisarlas y aclararlas” (MCyE, s/f, p. 3).

monitorear las variaciones que se producen en la calidad de la educación que provee el sistema educativo

En síntesis, hemos mostrado hasta aquí, en primer lugar, la existencia de una *disparidad entre la finalidad* asignada al sistema de evaluación *y el enfoque* que predomina en el diseño de los tests, problema que se pone de manifiesto cuando se intenta utilizar la información recabada. Este problema ya fue señalado por Esquivel (2000) para otros sistemas de evaluación de la región. El autor explica que la usual elección del enfoque normativo, aún cuando se buscaba obtener otro tipo de información, estuvo condicionada por el hecho de que había más conocimiento disponible en los equipos técnicos sobre elaboración de pruebas en el marco de dicho enfoque que la que existía sobre la evaluación conforme a criterio. En segundo lugar, nos referimos a otro tipo de deficiencias técnicas en el diseño del conjunto de tests que hacen imposible la *comparación longitudinal* de las medidas de rendimiento entre cohortes. Nuevamente aquí, la falta de experiencia y conocimiento especializado en los cuadros a cargo del diseño del programa de evaluación podría explicar los problemas encontrados.

Nosotros presentaremos una hipótesis más compleja: en la siguiente sección buscaremos mostrar que los problemas arriba reseñados no pueden explicarse sólo con la falta de conocimiento técnico en los cuadros de funcionarios a cargo del sistema de evaluación. Antes bien, sostendremos que el SINEC ha presentado dichos problemas técnicos porque surge en un contexto de transformación profunda del sistema educativo, en el que la obtención de información de corto plazo predominó sobre la necesidad de obtener información confiable en el largo plazo. En un contexto de este tipo, la preocupación por los aspectos técnicos del sistema de evaluación quedó soslayada ante la preeminencia de “otros usos” más inmediatos de la información, vinculados con el contexto de reforma educativa en el que el sistema de exámenes se desarrolló.

Los Usos Políticos del Sistema Nacional de Evaluación de la Calidad en el Marco del Proceso de la Reforma Educativa

En este apartado nos proponemos presentar usos del Sistema Nacional de Evaluación relacionados con los objetivos y las tensiones políticas en torno a la implementación de la reforma educativa. Mostramos primero cómo, más allá de los objetivos de mejora declarados, en el caso argentino, la información provista por el sistema de evaluación contribuyó a construir un diagnóstico crítico del sistema educativo a fin de proveer legitimación a las acciones de reforma educativa que estaban iniciándose. En segundo lugar, señalamos cómo desde el nivel central se buscó emplear los resultados de la evaluación para contribuir al poder de dirección del Estado nacional sobre el sistema de educación. Por último, presentamos datos que señalan el intento de emplear la evaluación como una herramienta para orientar, desde instancias de gobierno central, las prácticas pedagógicas de los actores escolares en un contexto en que la gestión del curriculum se encuentra crecientemente provincializada.

La Provisión de Legitimidad a una Reforma Polémica

Durante los primeros años del proceso de reforma educativa (1994-1996), el SINEC presenta una función informativa: provee a la audiencia educativa y a la opinión pública información acerca del funcionamiento general del sistema educativo. Luego de cada Operativo se divulgan los resultados alcanzados por los alumnos de todo el país mediante informes generales, informes para las escuelas y por medio de la prensa. También se elaboran y distribuyen cuadernillos para los

docentes, las llamadas “Recomendaciones metodológicas para la enseñanza”, en los que se destacan las principales dificultades detectadas en las pruebas y se incluyen sugerencias para superarlas.¹¹

Sin embargo, en sus inicios, la puesta en marcha de la reforma del sistema educativo que marca la Ley Federal de Educación encuentra fuertes resistencias en algunas jurisdicciones y en los sectores docentes sociales, fundamentalmente. A nivel jurisdiccional, la resistencia se vincula con que no todos los ministerios cuentan con las mismas capacidades y con recursos semejantes para asumir exitosamente las nuevas funciones, y el acuerdo parlamentario nacional no funciona como condición suficiente para comprometerlos en su implementación. El proceso de transformación también genera conflicto entre los docentes. Respecto del Sistema Nacional de Evaluación, por ejemplo, dos de los principales sindicatos (CTERA y SUTEBA) denuncian la falta de participación de los docentes en las distintas instancias del sistema. Asimismo, califican el uso de la información por parte del Ministerio de Educación nacional como “una búsqueda de culpables” antes que como instrumento de mejoramiento de la calidad de la educación (Nores, 2001). En consecuencia, el gobierno central requiere de fuentes de legitimación para las profundas transformaciones que se intentaban extender a las jurisdicciones.

En el marco de estas preocupaciones, la información provista por los Operativos de Evaluación parece prestarse para tal búsqueda de legitimación. Los resultados de los Operativos permiten definir “objetivamente” la situación crítica de la educación. Siguiendo a House (1994), “la evaluación conduce a una opinión fundada de que algo es de cierto modo” (p. 20). Los bajos niveles de rendimiento se exponen dejando de lado información sobre factores contextuales que lo afectan (como ser el estado edilicio de la escuela, los recursos materiales disponibles o la situación socioeconómica de los alumnos). En base a ellos, se realizan propuestas de intervención dirigidas a los docentes. Así, en los comienzos del proceso de reforma, el SINEC ayuda a instalar “el sentido de crisis”¹² del sistema educativo y, más puntualmente, enmarca el problema en el plano de la enseñanza. En un contexto político controvertido, los resultados de la evaluación contribuyen a “armar la escena”¹³ (Weiss, 1998) para introducir los nuevos lineamientos, presentando un estado de situación que provee de legitimidad a las acciones de intervención.

La Presión sobre los Actores del Sistema para la Implementación de la Reforma

A tres años de iniciada la reforma educativa, siguen existiendo divergencias entre diferentes actores y sectores educativos en torno a su implementación.¹⁴ Sin embargo, el proceso de reforma no disminuye su marcha y las políticas nacionales comienzan a orientarse a lograr un compromiso mayor por parte de los responsables de los niveles provinciales y escolares del sistema educativo para que aboquen su tarea a las disposiciones consideradas en la nueva ley (y en los acuerdos federales que tuvieron lugar a partir de ella). En este contexto, los datos provistos por el SINEC se constituyen en elementos importantes.

¹¹ Cabe señalar que con la sanción de la Ley Federal de Educación, cada jurisdicción goza del derecho de elaborar su propio diseño curricular conforme con un marco de contenidos básicos comunes a todo el país. En este contexto, resulta entonces muy problemático el hecho de que desde la Nación se elaboraran y distribuyeran en forma directa a las escuelas, sin consulta con las provincias, estas recomendaciones metodológicas.

¹² Desde la perspectiva de la administración de empresas se ha destacado que para introducir un cambio importante en cualquier organización resulta fundamental instalar entre sus actores el “sentido de crisis”, generando así un mayor grado de aceptabilidad de tal cambio (Kotter, 1996).

¹³ El término en inglés utilizado por la autora es *window dressing*.

¹⁴ La Ciudad de Buenos Aires, por ejemplo, no aplicaba la nueva estructura educativa definida en la Ley 24.195 y, en diferentes ocasiones, algunas provincias decidían no participar en los Operativos Nacionales de Evaluación.

En el caso argentino, pasados algunos años de la introducción de la “Transformación Educativa”, la comunicación de los resultados de las evaluaciones comienza a adoptar rasgos que involucran directamente a las provincias y a las escuelas. Como afirma Benveniste (2004):

evaluar per se no es una condición suficiente para que las escuelas o regiones se alineen con directivas del Estado central. El gobierno central debe ligar sus prácticas evaluativas a esquemas que conlleven a los actores educativos a aceptar y cumplir directrices gubernamentales. (p. 18)

En primer lugar, en los informes de los Operativos de Evaluación comienza a exponerse una tabla con los resultados promedio de cada una de las provincias, ordenadas de mayor a menor según sus valores de rendimiento. En segundo lugar, se reportan en la prensa, además de los puntajes nacionales, los de las provincias que han obtenido los resultados más bajos. En tercer lugar, ya desde 1997, vía la prensa, se publican los nombres de las escuelas primarias y secundarias que han obtenido los mejores logros, a la vez que se instaura transitoriamente un sistema de premios para tales escuelas. Por último, sobre la base de la información recabada de manera censal para el último año de Nivel Medio, a partir de 1997, se publican en Internet *rankings*¹⁵ de los establecimientos según su rendimiento promedio (ver Tabla 2).

Tabla 2

Principales acciones vinculadas con la diseminación de la información (1998-1999)

Audiencia	Tipo de información	Medio
Opinión pública	Resultados nacionales de rendimiento académico	Prensa
	Jurisdicciones con los más bajos resultados	Prensa
	Escuelas con los más altos resultados	Prensa
	Resultados escolares (promedio por área evaluada)	Internet
Escuelas, ministerios jurisdiccionales y Congreso de la Nación	Resultados nacionales y jurisdiccionales de rendimiento académico	Informe de resultados

Fuente: Oelsner (2002)

Mediante la difusión de este tipo de información se responsabiliza públicamente a las jurisdicciones por la gestión educativa realizada; tipo de difusión que responde a una de las posibles funciones de un sistema de evaluación nacional: promover la rendición de cuentas o *accountability*. En el caso argentino, teniendo en cuenta el decidido compromiso del gobierno central de avanzar con la reforma a pesar de las tensiones persistentes, creemos que este tipo de difusión—junto con la

¹⁵ Sin cuestionar aquí la pertinencia de la construcción de listados o *rankings* como medio de democratización de la información o como modo de incentivar el mejoramiento escolar a partir de la competencia, debe aclararse que en 1998 y 1999 estos listados comparan a las escuelas exclusivamente sobre la base de su rendimiento en las pruebas de Matemática y de Lengua. No se consideran otras variables escolares ni extra escolares. Como se aclaró en el capítulo conceptual, los *rankings* pueden ser contruidos sobre la base de criterios y medidas diferentes en función de lo que se desee comparar. Por ejemplo, los *rankings* de excelencia académica de las escuelas de negocios, de leyes, de medicina, etcétera, que publica la revista norteamericana *U.S. News & World Report*, se basan sobre una combinación de distintas medidas. Algunas de ellas se refieren a la reputación académica, a las tasas de retención y graduación, a los recursos de las instituciones y al rendimiento de los alumnos.

omisión de información sobre factores contextuales asociados al rendimiento y el sistema de incentivos—fue utilizado por parte de las autoridades centrales para instar a los actores locales a que alinearan sus políticas con las políticas nacionales.

En cuanto a la información recabada sobre los factores contextuales que inciden en el rendimiento escolar, al igual que en los primeros años de funcionamiento del SINEC, no se incluye en la comunicación de resultados. El análisis de las diferencias socioeconómicas, regionales o culturales entre las poblaciones escolares provee información sobre su incidencia en el rendimiento, poniendo en relación el peso de la actividad estrictamente didáctica. Por el contrario, su omisión evita que la responsabilidad de los distintos actores por los resultados obtenidos sea relativizada por la influencia de otros factores, manteniendo a la vez la definición del problema en el plano exclusivamente didáctico.

Por su parte, los sistemas de incentivos que trajeron aparejados los mecanismos de *accountability* (temporalmente y sólo para algunos años) incurrieron en faltas metodológicas. Específicamente, en la Argentina se otorgaron premios materiales a “las mejores escuelas” según los resultados alcanzados en la evaluación nacional sin tener en cuenta que el operativo no había sido censal, sino muestral. Por definición, cualquier sistema de incentivos debe basarse sobre una toma censal (Ravela, 2001). Un funcionario entrevistado está consciente de este problema y lo explica del siguiente modo:

El premio era honorífico o material didáctico. Salió en la tapa de los diarios, pero era macanas totales. No porque los datos no fueran ciertos, sino porque vos no podés decir que la escuela primaria de Las Pirquitas de Catamarca es la mejor escuela del país. Porque fue muestral, porque a lo mejor tenía cuatro séptimos grados y se evaluó a uno. Pero la directora de la escuela de Las Pirquitas estaba ahí y Menem¹⁶ le daba un premio. (E2)

Cabe señalar que este tipo de decisiones fue de corte político y no técnico. Es decir, no fueron incentivos diseñados por los equipos técnicos de la unidad de evaluación. Estos equipos identificaron las escuelas con más altos rendimientos, mas la decisión de premiarlas públicamente fue de la cúpula del gobierno. Como explica otro ex funcionario respecto a los actos y a las entregas de premios:

Nosotros con eso no teníamos nada que ver. No era parte de la tarea de evaluación para nada. El acto no lo armábamos nosotros. Era ceremonial, el área ministro. Nosotros lo que poníamos era la información de base. (...) Yo jamás estuve de acuerdo. Me llevó discusiones, peleas. (...) Eso fue una decisión política, no técnica. (E1)

La Orientación Pedagógico-Didáctica de Acuerdo con la Reforma

A partir de la aplicación de cada ONE, el Ministerio de Educación distribuye entre los profesores (además de los resultados de las pruebas): 1) las ya mencionadas recomendaciones metodológicas para la enseñanza y 2) los instrumentos de evaluación de 5to/6to año del secundario.

Estas líneas de acción surgen del Ministerio nacional y se dirigen directamente a las escuelas. Se trata de propuestas sin obligación de aplicación, ya que, una vez completada la provincialización del sistema educativo, el Ministerio deja de tener injerencia directa sobre las escuelas, siendo los ministerios provinciales quienes se ocupan de ellas. Sin embargo, señalábamos el mapa político de

¹⁶ Carlos Saúl Menem era el Presidente de la República Argentina.

controversias entre los distintos niveles y actores de gestión educativa en torno a la implementación de la Ley, así como la preocupación del gobierno central por lograr los compromisos necesarios que facilitaran el avance de la reforma. Dada esta situación, los materiales distribuidos, a la vez de ofrecerse como elaboradas herramientas de apoyo a la actividad docente, parecen también un medio para hacer llegar a las escuelas las pautas curriculares establecidas a partir de nueva Ley.

Los cuadernillos con recomendaciones para los docentes se refieren al aprendizaje y a la enseñanza de ciertos temas, incluyen desarrollos conceptuales disciplinares y sugieren actividades para trabajar en el aula los puntos detectados como problemáticos. Pero los elementos incluidos en los materiales no sólo apuntan a ofrecer alternativas para superar las dificultades identificadas en el aprendizaje. Se refieren a la vez a los nuevos contenidos definidos nacionalmente (CBC)—uno de los ejes de la reforma—y también a lo que será evaluado al año siguiente. Así, las propuestas que se distribuyen para "ofrecer a los docentes herramientas para que reflexionen sobre su práctica", "plantear algunas alternativas metodológicas para la enseñanza" o "realizar un aporte a la tarea cotidiana de cada uno de los docentes", difunden paralelamente los nuevos contenidos de enseñanza en las escuelas.

La distribución de instrumentos de evaluación constituye otro modo de establecer pautas para la enseñanza. Quienes postulan el uso de la evaluación como medio para reformar la enseñanza, enfatizan que los instrumentos de evaluación son medios efectivos para comunicar qué es lo que debe ser enseñado y cómo (Chapman & Mählck, 1997; Haertel, 1997). Repartir las pruebas no se limitaría entonces a "la finalidad de que sean conocidas, analizadas y utilizadas" por parte de los actores escolares "para confrontar resultados nacionales y jurisdiccionales con los alcanzados por sus propios cursos" (Operativo Nacional de Evaluación 1995, s/f, p. 5), como propone el Ministerio. Como explica un alto ex funcionario, la distribución de las pruebas apunta también a transmitir a los docentes, con ejemplos concretos, la nueva orientación didáctico-pedagógica:

cuando bajábamos un instrumento de medición de evaluación estábamos dando una pauta. (...) los instrumentos de evaluación de secundaria están muy bien hechos. Tanto los de Lengua como los de Matemática están elaborados en base a competencias. Entonces estás dando una señal al sistema acerca de qué es lo deben hacer. (...) Teníamos que dar una señal contundente al sistema acerca de cuál es la línea didáctico-pedagógica que se tiene que seguir. (E1)

El caso de esta medida, sin embargo, es más delicado que el de las Recomendaciones Metodológicas. Al igual que el sistema de premios para las escuelas y siguiendo nuestra línea de argumentación, muestra cómo, en la preocupación por alcanzar determinados objetivos políticos, se descuidan algunas pautas técnicas básicas. Las pruebas contienen ítems de anclaje, que son ejercicios que se repiten año tras año a fin de poder realizar comparaciones longitudinales de las puntuaciones. Dichos ítems se utilizan para reducir posibles sesgos debidos a las diferencias de aptitud entre los grupos comparados o a la variabilidad de la dificultad de las pruebas a lo largo del tiempo. Su función es proporcionar puntos de referencia comunes para la comparación (Lord, 1980). Para que los ítems de anclaje puedan cumplir esta función, es condición que sean confidenciales. Al hacerlos públicos íntegramente, la validez de los instrumentos siguientes queda afectada. El testimonio del mismo ex funcionario revela una clara conciencia de esta falta. Cuando es interrogado acerca de la publicación de instrumentos completos, responde que eso nunca habría sucedido. En sus términos:

Nunca se dio publicidad de la totalidad de las pruebas. Se dio publicidad parcial a los instrumentos *de manera tal de conservar...* Se bajaban a las escuelas modelos de pruebas, que eran

muy parecidas a las pruebas reales pero no eran reales. Porque *la prueba tenía que ser confidencial*. Se modelizaban pruebas. (E1; las itálicas son propias).

Síntesis y Conclusiones

Desde comienzos de la década de los '90, casi la totalidad de los países latinoamericanos ponen en funcionamiento algún tipo de sistema nacional de evaluación de la educación. En el nivel de la proclama, estos programas asumen como finalidad principal suministrar la información necesaria para profundizar el conocimiento sobre el sistema educativo y proporcionar los insumos fundamentales para la formulación de políticas de intervención focalizadas. La evaluación se presenta así como un instrumento técnico al servicio de la mejora de la *calidad* de la educación provista en todo el sistema. Al mismo tiempo, los sistemas nacionales de evaluación se vinculan, a través de las acciones que posibilitan, con políticas que buscan aumentar la *equidad* con que se distribuye el servicio educativo.

A más de una década de la etapa de implementación, se han constatado falencias en el funcionamiento de muchos de los sistemas de evaluación de la región. Estas deficiencias suelen atribuirse a la falta de experiencia y conocimiento técnico por parte de los cuadros a cargo del diseño y desarrollo de estos sistemas.

En este artículo, nos hemos referido al caso del Sistema Nacional de Evaluación de la Calidad de la educación (SINEC) vigente en la Argentina desde 1993. Al respecto, hemos dado cuenta de dos de sus principales falencias técnicas: a) la disparidad entre los objetivos de la evaluación y el enfoque adoptado para la construcción de los instrumentos de prueba; y b) la imposibilidad de contar con información confiable sobre la evolución del rendimiento de los alumnos en el tiempo. Estos problemas suelen atribuirse a la falta de experiencia y de conocimiento acumulado en la región sobre estos temas, factor que no hemos descartado. Sin embargo, hemos propuesto una explicación más compleja para los problemas hallados. A lo largo de este trabajo, presentamos hallazgos de investigación que muestran cómo el SINEC aparece vinculado no sólo al cumplimiento de los objetivos formales de mediano y largo plazo (medición del estado y la evolución de los aprendizajes de los alumnos) sino, también, a otros usos más inmediatos que se le han dado a los resultados de las evaluaciones nacionales (tales como la búsqueda de legitimación y apoyo a la reforma educativa en curso). En el marco de estos usos más inmediatos de la evaluación, no resultaría tan importante preservar cierto tipo de recaudos técnicos que sí resultan indispensables al considerar el sistema de evaluación en tanto proveedor de información sobre la evolución del rendimiento académico en el largo plazo.

Desde ya, no creemos que los problemas señalados en el ámbito de las evaluaciones nacionales deban apoyar argumentos a favor del fin de las políticas de evaluación en la región. Antes bien, pensamos que la lectura crítica y contextualizada de la evaluación—como la de otras políticas—debe traducirse en una comprensión más profunda de las funciones que asume la evaluación en nuestros países.

Asimismo, esperamos que el trabajo presentado ejemplifique la importancia de atender a los modos en que la evaluación a escala nacional se articula con otras políticas educativas (como las de reforma en los modos de gobierno de la educación, las de transformaciones en la estructura de niveles y ciclos escolares, o las políticas de cambio curricular) a fin de profundizar en el estudio del funcionamiento y del impacto que estos sistemas de evaluación están teniendo en la región.

Por último, esperamos que este artículo aporte a la construcción de una mayor conciencia sobre la importancia de exigir y procurar mayor conocimiento teórico y técnico para el desarrollo de las políticas educativas, entre ellas las de evaluación.

Referencias

- Andrada, M., Narodowski, M. y Nores, M. (2002). *La evaluación de la educación en la Argentina: Desde los operativos nacionales hasta los boletines escolares: 1993-2001*. Buenos Aires: Editorial Prometeo.
- Arancibia, V. (1996) Los sistemas de medición y evaluación de la calidad de la educación, *Documentos* (2), Laboratorio Latinoamericano de Evaluación de la calidad de la educación/ UNESCO.
- Benveniste, L. (2000a). *The politics of student testing: a comparative analysis of national assessment systems in southern cone countries*. Stanford, CA: Stanford University.
- Benveniste, L. (2000b). La evaluación del rendimiento académico y la construcción de consensos en Uruguay. Grupo de Trabajo sobre Estándares y Evaluación. PREAL - GRADE. publicación digital de preal. bajado de la página de preal: <http://www.preal.cl/>
- Benveniste, L. (2004). La evaluación estudiantil como reflejo de la relación estado-sociedad. En Las Reformas Educativas en la Década de 1990: Un Estudio Comparado de Argentina, Chile y Uruguay. Proyecto Alcance y Resultados de las Reforma Educativas en Argentina, Chile y Uruguay. Buenos Aires: Banco Interamericano de Desarrollo, Ministerios de Educación de Argentina, Chile y Uruguay, Grupo Asesor de la Universidad de Stanford, pp. 447-482.
- Braslavsky, C. (1999). *Re-haciendo escuelas: Hacia un nuevo paradigma en la educación latinoamericana*. Buenos Aires: Santillana.
- Chapman, D. W. & Mählick, L. O. (1997). Changing what happens in schools: central-level initiatives to improve school practice. En D. Chapman, L. Mählick, A. E. M. Smulders (Eds.) *From Planning to Action: government initiatives for improving school-level practice*. Paris: UNESCO-International Institute for Educational Planning, Pergamon.
- De Landsheere, G. (1996). *El pilotaje de los sistemas educativos*. Madrid: La Muralla.
- Ebel, R. (1977). *Fundamentos de la medición educacional*. Buenos Aires: Guadalupe.
- Esquivel, J. M. (2000). El diseño de las pruebas para medir logro académico: ¿referencia a normas o a criterios? En P. Ravela(Ed.), *Los Próximos Pasos: ¿hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?* PREAL/ GRADE: Lima.
- Gronlund, N. (1973) *Medición y evaluación de la enseñanza*. México, D.F.: Pax..
- Haertel, E. H. (2002). *Pruebas de desempeño y reforma educativa*. Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL).
- Hanson, M. (1997). *La descentralización educacional: problemas y desafíos*. Documento N° 9. Santiago de Chile: PREAL.
- House, E. R. (1994). *Evaluación, ética y poder*. Madrid: Morata.

- Larripa, S. (2003). *El sistema nacional de evaluación de la calidad educativa: acerca de la comparabilidad de sus resultados. Argentina: 1995-2000*. Tesis de maestría no publicada. Victoria: Universidad de San Andrés.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbawn.
- Lundgren, U. (1992). *Teoría del curriculum y escolarización*. Madrid: Ed. Morata.
- MacDonald, B. (1989). La evaluación y el control de la educación. En G. Sacristán, Jimeno y A. Pérez Gómez (Comps.) *La enseñanza: su teoría y su práctica*. Madrid: Akal.
- Martínez Arias, R. (1996) *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.
- Ministerio de Cultura y Educación (1997). *Evaluación y uso de estrategias*.
- Ministerio de Cultura y Educación (1997). El sistema nacional de evaluación de la calidad en Argentina, OEL.
- Ministerio de Cultura y Educación (s/f). Dirección Nacional de Evaluación, *Operativo Nacional de Evaluación 1997. Informe de Resultados*.
- Ministerio de Educación (s/f). *Evaluación de los resultados del SINEC 1993/1999*.
- Monseur, C. (2000) *Item response theory scaling of literacy and mathematics achievement data from 1995 to 1999 in five grades*. Informe N°13, Argentina, Ministerio de Educación.
- Nores, M. (2001). *El Sistema Nacional de Evaluación de la Calidad –SINEC*. Informe preliminar. Caso Educación. BID/CEDI.
- Oelsner, V. (2002). *Las evaluaciones nacionales de la calidad de la educación: sus usos y funciones en el sistema educativo argentino (1994-2001)*. Tesis de maestría no publicada. Victoria: Universidad de San Andrés.
- Perusia, J. C. (2001). *Evolución de los resultados de los Operativos de Evaluación de la Calidad Educativa 1995-1999, Puntuaciones equiparadas de las pruebas de Lengua y Matemática*. Informe N°13. Argentina: Ministerio de Educación.
- Popham, J. (1999) Why Standardized Tests Dont Measure Educational Quality?, *Educational Leadership*, vol 56, (6), Alexandria: ASCD (Association for Supervision and Curriculum Development).
- Ravela, P. (2001). *¿Cómo presentan sus resultados los sistemas nacionales de evaluación de la calidad educativa en América Latina?* Santiago de Chile: PREAL/ GRADE.
- Rossi, P. H., Freeman, H. E., Lipsey, M. W. (1999) *Evaluation: A Systematic Approach* (6a. ed.). Thousand Oaks: SAGE Publications.

Tiana Ferrer, A. (2000). *Cooperación internacional en evaluación de la educación en América Latina y el Caribe: análisis de la situación y propuestas de actuación*. Serie de informes técnicos del Departamento de Desarrollo Sostenible. Washington, D.C.: Banco Interamericano de Desarrollo.

Weiler, H. (1996). Enfoques comparados en descentralización educativa. En: Pereyra, M., García Mínguez, J., Gómez, A. & Beas, M. (Comps.) *Globalización y descentralización de los sistemas educativos*. Barcelona: Pomares.

Weiss, C. H. (1998). *Evaluation. Methods for Studying Programs and Policies* (2a. ed.). New Jersey: Prentice-Hall.

Fuentes documentales citadas

República Argentina, Ley N° 24. 049.

República Argentina, Ley N° 24.195.

República Argentina, Ministerio de Cultura y Educación (1997) *Evaluación y uso de estrategias*.

República Argentina, Ministerio de Cultura y Educación de la Nación (1997) *El Sistema Nacional de Evaluación de la Calidad en Argentina*, Buenos Aires, OEI.

República Argentina, Ministerio de Cultura y Educación de la Nación (s/f) *6° Operativo Nacional de Evaluación 1998. Propuesta de Tablas de Especificaciones*.

República Argentina, Ministerio de Educación (s/f) *Evaluación de los resultados del SINEC 1993/1999*.

Fuentes documentales analizadas

Materiales elaborados por el Ministerio de Educación de la Nación

Informes de resultados nacionales:

Operativo Nacional de Evaluación 1994. Primer Informe. S/f.

Operativo Nacional de Evaluación 1995. S/f.

Operativo Nacional de Evaluación 1996. S/f.

Operativo Nacional de Evaluación 1997. S/f.

1° Operativo Nacional de Evaluación de Finalización del Nivel Secundario 97. S/f.

Operativo Nacional de Evaluación 1998. S/f.

2° Operativo Nacional de Evaluación de Finalización del Nivel Secundario 98. S/f.

7° Operativo Nacional de Evaluación Muestral EGB 99. S/f.

3° Operativo Nacional de Evaluación de Finalización del Nivel Secundario 99. S/f.

Recomendaciones e interpretaciones didácticas de los resultados para los docentes:

Recomendaciones Metodológicas para la Enseñanza. Matemática. Nivel Primario. 1994.

Recomendaciones Metodológicas para la Enseñanza. Lengua. Nivel Primario. 1994.

Recomendaciones Metodológicas para la Enseñanza. Lengua. Nivel Medio. 1994.

Recomendaciones Metodológicas para la Enseñanza. 3er Operativo Nacional de Evaluación (1995). Lengua. 1997.

Recomendaciones Metodológicas para la Enseñanza. 3er Operativo Nacional de Evaluación (1995). Matemática. 1997.

Recomendaciones Metodológicas para la Enseñanza. 3er Operativo Nacional de Evaluación (1995). Ciencias Sociales. 1997.

Recomendaciones Metodológicas para la Enseñanza. 3er Operativo Nacional de Evaluación (1995). Ciencias Naturales. 1997.

Recomendaciones Metodológicas para la Enseñanza. Algunas consideraciones sobre la producción de textos escritos. Lengua. 1998.

Recomendaciones Metodológicas para la Enseñanza. La comprensión lectora de textos argumentativos. Lengua. 1998.

- Recomendaciones Metodológicas para la Enseñanza. 5° Operativo Nacional de Evaluación. Cursos de Finalización de Ciclos de E.G.B.. Matemática. 1998.*
- Recomendaciones Metodológicas para la Enseñanza. Geometría de Coordenadas. 1° Censo Nacional de Finalización del Nivel Secundario (V/VI Año, Nov. 1997). Matemática. 1998.*
- Recomendaciones Metodológicas para la Enseñanza. Vectores en el plano y en el espacio. 1° Censo Nacional de Finalización del Nivel Secundario (V/VI Año, Nov. 1997). Matemática. 1998.*
- Recomendaciones Metodológicas para la Enseñanza. Representación gráfica de funciones. 1° Censo Nacional de Finalización del Nivel Secundario (V/VI Año, Nov. 1997). Matemática. 1998.*
- Recomendaciones Metodológicas para la Enseñanza. 5° Operativo Nacional de Evaluación. 1° y 2° Ciclos de E.G.B.. Lengua. 1998.*
- Recomendaciones Metodológicas para la Enseñanza. 5° Operativo Nacional de Evaluación. 3° Ciclo de E.G.B.. Lengua. 1998.*

Instrumentos de evaluación

Orientaciones para la Evaluación, de 5° y 6° nivel secundario. Contenidos, competencias, instrumentos. S/f. Pruebas de Lengua y Matemática de 3er. año de la Educación General Básica administradas en los Operativos Anuales de Evaluación durante el período 1995-2000.

Sobre las autoras

Silvina Gvirtz: sgvirtz@udesa.edu.ar

Silvina Larripa: silvilarripa@yahoo.com.ar

Verónica Oelsner: veronica.oelsner@rz.hu-berlin.de .

Universidad de San Andrés

Silvina Gvirtz es Doctora en Educación por la Universidad de Buenos Aires, investigadora del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y directora de la Escuela de Educación de la Universidad de San Andrés. Es autora de varios libros y numerosos artículos, el último de ellos, “National evaluation system in Argentina: Problematic present and uncertain future”, en: *Assessment in Education: Principles, Policy & Practice*, Vol. 11, N°3, Londres, Carfax, 2004. En el año 2003 fue designada fellow de la Fundación John Simon Guggenheim. Es Directora del Anuario de la Sociedad Argentina de Historia de la Educación.

Silvina Larripa es Licenciada y Profesora en Ciencias de la Educación por la Universidad de Buenos Aires y Especialista y Magíster en Educación por la Universidad de San Andrés. Actualmente realiza su doctorado en la Universidad de San Andrés, es becaria doctoral del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y se desempeña como docente en la Universidad de San Andrés y en la Universidad del Salvador.

Verónica Oelsner es Licenciada y Profesora en Ciencias de la Educación por la Universidad de Buenos Aires y Especialista y Magíster en Educación por la Universidad de San Andrés. Actualmente realiza su doctorado en el Departamento de Educación Comparada de la Universidad Humboldt de Berlín (Alemania) y se desempeña como asistente de investigación en un proyecto del mismo departamento, “Representaciones del ‘Beruf’ y estructuras de formación profesional en comparación. Estudios sobre el caso hispanoamericano (Argentina) y el asiático-oriental (China)”, en el marco del programa interdisciplinario “Representaciones de órdenes sociales en transformación”.

AAPE Comité Editorial

Editores Asociados

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco,
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Brasil

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

**Romualdo Portella do
Oliveira**

Universidade de São Paulo,
Brasil

Daniel Schugurensky

Ontario Institute for Studies
in Education, Canada

Daniel Suarez

LPP-UBA, Argentina

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación,
Chile

Sigfredo Chiroque

Instituto de Pedagogía
Popular, Perú

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
USA

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez

Gómez
Universidad de Málaga,
España

Diana Rhoten

Social Science Research
Council, USA

Susan Street

CIESAS Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona
Lisboa

Lilian do Valle

Universidade Estadual do Rio
de Janeiro, Brasil

Dalila Andrade de Oliveira

Universidade Federal de
Minas Gerais, Brasil

Mariano Fernández

Enguita Universidad de
Salamanca. España

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

María Loreto Egaña

Programa Interdisciplinario
de Investigación en
Educación, Chile

Vanilda Paiva

Universidade Estadual do Rio
de Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, USA

Jurjo Torres Santomé

Universidad de la Coruña,
España

***EPAA* Editorial Board**

Editor: Sherman Dorn University of South Florida

Production Assistant: Chris Murrell, Arizona State University

[Michael W. Apple](#)

University of Wisconsin

[Greg Camilli](#)

Rutgers University

[Mark E. Fetler](#)

California Commission on Teacher
Credentialing

[Richard Garlikov](#)

Birmingham, Alabama

[Thomas F. Green](#)

Syracuse University

[Craig B. Howley](#)

Appalachia Educational Laboratory

[Patricia Fey Jarvis](#)

Seattle, Washington

[Benjamin Levin](#)

University of Manitoba

[Les McLean](#)

University of Toronto

[Michele Moses](#)

Arizona State University

[Anthony G. Rud Jr.](#)

Purdue University

[Michael Scriven](#)

University of Auckland

[Robert E. Stake](#)

University of Illinois—UC

[Terrence G. Wiley](#)

Arizona State University

[David C. Berliner](#)

Arizona State University

[Linda Darling-Hammond](#)

Stanford University

[Gustavo E. Fischman](#)

Arizona State University

[Gene V. Glass](#)

Arizona State University

[Aimee Howley](#)

Ohio University

[William Hunter](#)

University of Ontario Institute of
Technology

[Daniel Kallós](#)

Umeå University

[Thomas Mauhs-Pugh](#)

Green Mountain College

[Heinrich Mintrop](#)

University of California, Los Angeles

[Gary Orfield](#)

Harvard University

[Jay Paredes Scribner](#)

University of Missouri

[Lorrie A. Shepard](#)

University of Colorado, Boulder

[Kevin Welner](#)

University of Colorado, Boulder

[John Willinsky](#)

University of British Columbia