

---

# archivos analíticos de políticas educativas

Revista académica evaluada por pares,  
independiente, de acceso abierto  
y multilingüe



Arizona State University

---

Volumen 18 Número 8

10 de abril 2010

ISSN 1068-2341

---

## Evaluaciones Estandarizadas: Seis Reflexiones Críticas

**Ignacio Barrenechea**

George Washington University (GWU).

Citación: Barrenechea, I. (2010) “Evaluaciones estandarizadas: seis reflexiones críticas”  
*Archivos Analíticos de Políticas Educativas*, 18 (8). Recuperado [fecha] de <http://epaa.asu.edu/ojs/article/751>

**Resumen:** Este trabajo presenta una revisión de la literatura que analiza seis limitaciones claves que tienen las evaluaciones estandarizadas. Las seis críticas que se presentan son: la tensión que existe entre el concepto de inteligencias múltiples y las evaluaciones estandarizadas, la desatención de componentes del currículum real – el cual no necesariamente se agota en el currículum prescripto, los riesgos de enseñar para el test, los incentivos que se generan de hacer trampa con los resultados, la falta de consideración de las diferencias socio económicas de los alumnos que son evaluados, y, por último, las limitaciones de los resultados de las evaluaciones estandarizadas para predecir el éxito laboral de los estudiantes. Este trabajo concluye que este tipo de evaluaciones, si bien pueden tener ciertas ventajas, cuando son usadas como el único instrumento para tomar decisiones educativas, las limitaciones son mayores que sus beneficios

**Palabras-clave:** evaluación; estandarización; currículum; inteligencias múltiples; contextos educativos.

**Abstract:** This paper presents a review of the literature analyzing six key limitations that standardized evaluations. The six criticisms are: the tension that exists between the concept of multiple intelligences and the standardized evaluations, the disregard of components of the real curriculum - which is not necessarily exhausted in the prescribed curriculum, the risks of

teaching for the test, the incentives that are generated to cheat or inflate the results of the tests, the lack of consideration of the economic differences of the pupils who are evaluated, and, finally, the limitations of the results of standardized evaluations to predict the labour success of the students. This paper concludes that this type of evaluations, though they have certain advantages, when used as the only instrument to make educational decisions, the costs outweigh the benefits

**Keywords:** evaluation; standardization; curriculum; multiple intelligences; educational contexts.

**Resumo:** Este trabalho apresenta uma revisão da literatura para a análise de seis limitações chave que têm testes padronizados. As seis limitações apresentados são: tensão entre o conceito de inteligências múltiplas e avaliações padronizadas, componente desatenção do currículo reais - o que não significa necessariamente acabar com o currículo prescrito, o riscos de ensino para o teste, os incentivos para gerar armadilha com os resultados, a falta de consideração das diferenças sócio econômico dos alunos que são avaliados e, finalmente, a limitações dos resultados de testes padronizados para prever o sucesso do trabalho dos alunos. O documento conclui que tais avaliações, embora possam ter algumas vantagens, se como o único instrumento usado para tomar decisões educacionais, o limitações superam seus benefícios.

**Palavras-chave:** avaliação; padronização do currículo; inteligências múltiplas; contextos educativos.

## Introducción

Las evaluaciones estandarizadas han sido utilizadas para satisfacer diferentes necesidades en el campo educativo, aunque mas recientemente este tipo de evaluaciones se han convertido en una pieza importante, tal vez la más relevante, según Koretz, (2002) para delimitar la responsabilidad de las escuelas y de los educadores. En los Estados Unidos, por ejemplo, las primeras utilizations de tests estandarizados fueron para cuestiones de análisis psicológico, y datan del año 1840. Quienes defienden con más ímpetu este tipo de evaluaciones sostienen que las mismas motivan a los alumnos a que quieran superarse, dado que los estudiantes encontrarían en la fácil comparación de sus resultados, un incentivo para estudiar, por lo tanto mejorando sus logros y rendimiento escolar. Según esta percepción las evaluaciones estandarizadas están positivamente correlacionadas con rendimiento académico (Jacob, 2001). Sin embargo, esta postura también es criticada, ya que se considera que las evaluaciones sólo ejercen un efecto positivo en un porcentaje mínimo de la población (en aquellos que pueden obtener buenos resultados); la mayoría de los estudiantes perciben las evaluaciones como una presión excesiva, en algunos casos inmanejables. Por ejemplo, Paris (1991) considero que a medida que las edades de los estudiantes ascienden, también lo hace el resentimiento, la ansiedad, el cinismo y la desconfianza con respecto a las evaluaciones estandarizadas.

En este trabajo pretendo demostrar que este tipo de evaluaciones, si bien pueden ser útiles en cuanto ofrecen un mecanismo de evaluación rápido y de relativamente bajo costo, tienen también una serie de limitaciones, las cuales ponen en duda el supuesto que este tipo de evaluaciones sean una herramienta fundamental a la hora de medir los logros de un sistema educativo, o las posibilidades de los estudiantes para acceder a una instancia superior de formación académica.

Con el motivo de organizar el desarrollo de este trabajo he decidido dividir mis críticas en seis partes. En la primera sostendré que existen diferentes tipos de inteligencia, con lo cual la estandarización de las evaluaciones implica evaluar solamente una parcialidad del desarrollo intelectual. En la segunda de las criticas demostraré que existe un *corpus* de

conocimiento, contenido tanto en el currículo prescripto como también en el currículo real (que muchas difiere del prescripto), que no es evaluado en estas pruebas estandarizadas. En la parte siguiente desarrollaré la postura que sostiene que este tipo de evaluación incentiva a los docentes a que dediquen gran parte de su tiempo en la preparación exclusiva para estas evaluaciones, lo cual implica una pérdida de tiempo que podría ser invertido en la enseñanza de contenidos o aptitudes. A continuación describiré la estructura de incentivos que genera corrupción en los resultados de este tipo de evaluaciones. En la quinta crítica enfatizaré el rol que juega el contexto socio económico del educando, a la hora de enfrentar las evaluaciones estandarizadas. Por último, la sexta crítica estará basada en la ineficacia de los *tests* para predecir el éxito de los estudiantes en un mercado laboral tan cambiante.

Antes de comenzar con las críticas mencionadas, describiré algunos ejemplos de implementaciones de este tipo de evaluaciones en diferentes regiones del mundo. En la región asiática, por ejemplo, las evaluaciones estandarizadas han crecido en importancia durante los últimos años. Por ejemplo, en China, desde 1977 los exámenes denominados *gaokao* han sido el foco central del sistema de admisión a los *colleges* y los resultados han sido el único criterio para determinar el status de admisión del estudiante (Suen, 2006, Pág. 115). De hecho, en este país la presión que estos exámenes ejerce no sólo incumbe a los alumnos, sino que a la familia en su conjunto: los padres no están menos presionados que los estudiantes mismos. El rendimiento de los estudiantes no es sólo un asunto personal, pero también un tema de honor familiar, gloria o vergüenza. Historias de padres sacrificando sus propias carreras para ayudar a que sus hijos se preparen para los *gaokaos* no son para nada poco comunes (Suen, 2006, Pág. 115). También vale la pena mencionar el caso de Japón. Los alumnos deben someterse a una evaluación estandarizada sumamente importante. De su resultado, dependerán las posibilidades de los alumnos de poder acceder al circuito universitario. La presión ha crecido tanto en los últimos años que ha surgido un circuito de colegios suplementarios, denominados *juku*, al cual los alumnos más pudientes atienden, con el fin de obtener preparación adicional para esos exámenes (Rotberg, 2006).

El caso de América latina, en especial Chile, merece un párrafo aparte. Como sugieren Narowdoski y Nores (2002) dentro de América Latina, el sistema educativo chileno ha sido muy estudiado, por ser considerado pionero en el camino hacia la descentralización, El modelo actual de provisión de educación y administración en Chile es uno de un cuasi mercado<sup>1</sup>, (Narowdoski y Nores, 2002; Narodowski y Andrada, 2004), pero es importante remontarnos a la década del 80 para terminar de entender las características de la transición de un modelo de monopolio estatal, con un estado muy presente en todas las decisiones concernientes a la financiación, a la provisión y a la regulación en materia de políticas educativas, a un modelo de cuasi mercado.. De hecho, como explica Cox:

El sistema escolar de Chile es heredero de un modelo altamente centralizado de provisión estatal de educación, que se remonta a mediados del siglo XIX, interrumpido por una reforma descentralizadora y privatizadora profunda aplicada por el régimen militar a lo largo de la década de 1980, la que no es revertida por el Gobierno democrático que asume en 1990 (Cox, 1997, Pág. 4).

---

<sup>1</sup> Puede ser denominado como un modelo cuasi mercado de provisión educativa, ya que esto cumple con las características generales de tal mercado, es decir el financiamiento público del servicio con su provisión en un mercado competitivo donde organizaciones públicas y/o privadas participan (ver Narowdoski y Nores, 2002).

En Chile, todos los alumnos deben rendir evaluaciones estandarizadas de alcance nacional. Por ejemplo, en el sitio oficial de Universidad de Chile se establece que las Pruebas de Selección Universitarias (en adelante, PSU) son evaluaciones que en los casos de Lenguaje y Comunicación, y Matemática, deben rendirse de *manera obligatoria*. El alcance del resultado de estas evaluaciones no es menor para el porvenir de los estudiantes: de hecho, gran parte del universo de lo posible, en cuanto a sus elecciones de carreras e instituciones universitarias, estará determinado por el resultado que obtengan en estas evaluaciones. Es decir, si bien en principio Chile introdujo una política de descentralización profunda, que implica mucho más que la mera *provincialización o municipalización* que existió en otros países latinoamericanos como ser la Argentina, el Estado nacional se ha reservado la facultad de ser evaluador del rendimiento de los estudiantes, ya sea de aquellos que asisten a escuelas públicas o privadas, subvencionadas o no subvencionadas por el Estado.

En el caso de Europa, las evaluaciones estandarizadas también son utilizadas. Por ejemplo, en el caso de Inglaterra, los alumnos son expuestos a la edad de dieciséis a una evaluación estandarizada para determinar aquellos alumnos que tendrán la posibilidad de entrar al “A-Level” (nivel avanzado) (Fowler, 2004). A su vez, al finalizar su escuela secundaria, los alumnos deben someterse a otra evaluación estandarizada que determinará a qué universidades pueden asistir. El caso de Turquía también es importante dado que, dada la sobredemanda que existe de alumnos con intenciones de ingresar al circuito universitario (solamente el 20% puede ingresar, por cuestiones de cupo), el resultado de las evaluaciones estandarizadas se ha convertido en el instrumento más importante para permitir el ingreso a las universidades (Rotberg, 2006).

La importancia de la medición de las habilidades pareciera un elemento constitutivo de la idiosincrasia sobre la cual descansa el sistema educativo de los Estados Unidos. En aquel país, la división de los alumnos en grupos según las habilidades de los niños se realiza desde una muy temprana edad. Por el contrario, en muchos países europeos, la asignación a un grupo específico, en función de las habilidades de los estudiantes, está prohibida hasta los catorce años de los alumnos. Si bien cada gobierno impuso sus propias variables para conducir la suerte política, económica y social de la nación, una de ellas pareciera ser constante a lo largo de las diferentes presidencias: la estandarización de las evaluaciones educativas. Es decir, desde la publicación de *A Nation at Risk* en el año 1983 hasta el año 2001, cuando se estructuró *No Child Left Behind* (NCLB, en adelante, por sus siglas en inglés), la estandarización fue ocupando cada vez un lugar más importante en el sistema educativo de ese país.

Las evaluaciones estandarizadas se han convertido en una especie de panacea a la hora de implementar el proceso de evaluación en el sistema educativo de los Estados Unidos. Si bien durante la década del ochenta algunos estados le dieron mucha importancia a la estandarización, por ejemplo Langer & Pradl (1984) y Firestone (1992) recalcan el hecho que más de cuarenta estados se volcaron a la estandarización, NCLB, en el 2001, marcó un punto de inflexión en cuanto a la utilización de la estandarización en el sistema educativo de los Estados Unidos. Entre otros autores, Elmore (2003, Pág. 6), por ejemplo, sostiene que esta ley constituye el acto más claro, en la historia de los Estados Unidos, de la nacionalización del sistema educativo. No obstante ello, las evaluaciones estandarizadas, fuertemente respaldadas por NCLB, han sido fuertemente criticadas por no ser útiles a la hora de evaluar el nivel de pensamiento crítico y de razonamientos de alto orden de los estudiantes.

A diferencia de la tendencia de descentralización que se ha vivido en varios países de América Latina durante los últimos veinte años, como ser Argentina, Brasil, Colombia y

Chile, se podría sostener, en principio que la NCLB implica una nacionalización de gran parte de las políticas educativas en los Estados Unidos<sup>2</sup>. Si bien constitucionalmente son los gobiernos estatales, e incluso locales de menor envergadura, los que tradicionalmente han controlado la administración y la regulación del sistema educativo, la iniciativa NCLB obliga a que todos los estados desarrollen estándares, y evaluaciones a través de las cuales se determine el grado de cumplimiento de dichos objetivos.

Hursh (2005) sostiene que desde la publicación, en 1983 de *A Nation at Risk*, en la que se sugería que la responsabilidad de los problemas económicos de los Estados Unidos se debían al mal funcionamiento del sistema educativo, los gobiernos estatales se han preocupado más en posibilitar mecanismos de evaluación más “objetivos”. En un estudio similar, realizado años después de la publicación de este documento, Apple (1996) relata que sectores vinculados al mundo empresarial y corporativo presionaron para que se implemente un sistema de políticas educativas uniformes a nivel nacional, basadas en estándares. Si bien es cierto que durante la década del ochenta y del noventa muchos estados decidieron autónomamente incluir la estandarización como mecanismo de evaluación, no todos debían hacerlo en la misma medida. De hecho, Iowa era un estado que a diferencia de prácticamente la totalidad de los estados restantes, abogaba por la no utilización de la estandarización como mecanismo de evaluación. La obligatoriedad de la estandarización, no sólo implica un claro paso hacia el fortalecimiento del gobierno nacional en la regulación de la educación, sino que además implica un debilitamiento del “juicio” del docente a la hora de determinar el rendimiento de los alumnos. Es decir, la valoración del docente, proveniente de su juicio profesional, se ve minimizado por un examen diseñado y administrado por autoridades estatales.

La minimización de la capacidad del docente de evaluar el nivel de sus alumnos, surge una inquietud: así como los profesores pueden ser subjetivos en su evaluación integral del alumno, e incluso tener un interés particular en que el alumno obtenga buenos resultados, ¿hasta qué punto no tienen las autoridades estatales un incentivo “subjetivo” similar? Es decir, podría argumentarse que, por ejemplo, las autoridades pueden incrementar la proporción de alumnos que se gradúan a través de una simplificación de las evaluaciones estandarizadas; es decir, las autoridades estatales pueden adaptar la rigurosidad de la evaluación en función de sus propias necesidades. Hay numerosos trabajos que sustentan estas perspectivas. Hursh (2005) menciona el caso de Paige quien en su cargo de superintendente en Texas (luego fue Secretario de Educación en el gobierno de George W. H. Bush), le ordenó a los directores de las escuelas de Texas, con el fin de bajar el porcentaje de deserción escolar, que los alumnos abandonaron la institución no a causa de la deserción escolar, sino por otros motivos, como ser haber elegido otra institución. Si bien es cierto que en su trabajo Hursh no ofrece evidencia sólida para justificar dicha denuncia, existen otros trabajos que parecieran, al menos, justificar la existencia de una presunción en contra de Paige. Por ejemplo, Capello (2004) presenta el caso de Robert Kimbal, quien se desempeñaba como asistente del director de una escuela secundaria de Houston. Kimbal, relata Capello, se sorprendió cuando a pesar que de 1000 alumnos en primer año, solamente 300 llegaron a completar el último año, la institución en la que él desempeñaba su cargo no reportó ningún caso de deserción escolar. En la misma línea Winerip (2003) denunció que de los 5000 estudiantes que dejaron su institución educativa en 16 escuelas, 2999 deberían haber sido considerados casos de deserción escolar, sin embargo, ello no sucedió.

---

<sup>2</sup> Es importante detenerse en el término “en principio”, dado que más adelante en este trabajo se discutirá que, probablemente la implementación de NCLB conduzca a un fortalecimiento del sector privado en la provisión del bien educación.

La iniciativa conocida como NCLB obliga a que los estados cumplan con un progreso académico anual adecuado (AYP, por sus siglas en inglés). A través de esta iniciativa se buscó proteger a los alumnos que provienen de sectores económicos más desprotegidos. No obstante ello, es importante mencionar que las instituciones académicas que no cumplan con AYP sufren sanciones por parte del estado. Por ejemplo, según publicara el Departamento de Educación de los Estados Unidos (US Department of Education, 2002, Pág. 6), las instituciones que no cumplan con los objetivos planteados, deben permitirle a sus alumnos que elijan otra institución, y a su vez, costear los costos de transporte que el alumno podría llegar a tener. En suma, en caso de no cumplir los objetivos planteados por tercer año consecutivo, la institución no sólo debe hacer frente a costos de transporte, sino que además debe ofrecerles a los alumnos servicios “gratuitos” de preparación fuera del horario escolar, para lograr mejorar los resultados de sus evaluaciones. Dado que los gastos de transporte, como también los costos que lógicamente surgen de la provisión “gratuita” de esta preparación adicional, la institución posee una baja en sus recursos. Con lo cual, surge la inquietud, ¿puede una institución mejorar sus resultados si cada año recibe menos financiamiento que el anterior.

### **Las inteligencias múltiples y las pruebas estandarizadas**

En el modelo de Gardner las inteligencias se pueden dividir en siete áreas principales, es importante recordar, no obstante, que Gardner (1983) puntualiza que ese listado no es taxativo, es decir, pueden surgir nuevos tipos de inteligencia; ello demuestra que desde la postura de Gardner la inteligencia deja de ser un concepto estático para convertirse en un modelo dinámico. En la misma línea de pensamiento, Campbell y Campbell (1999) describieron que según Gardner, la inteligencia implica la habilidad de resolver problemas, de hacer contribuciones a la comunidad y de identificar nuevos desafíos por perseguir. Los autores sostienen que la definición de Gardner se centra en un proceso dinámico (resolver problemas y contribuir a la comunidad a la que uno pertenece), no limitando la inteligencia a un valor estático y cuantificable (Campbell y Campbell, 1999).

En una entrevista, Gardner sostuvo que: “Estoy siendo un tanto provocativo intencionalmente. Si hubiera dicho que hay siete clases de competencias, la gente hubiera bostezado y dicho ‘sí, sí’. Pero llamándolas ‘inteligencias’ estoy diciendo que nos hemos inclinados a colocar en un pedestal una variedad llamada inteligencia, y que en realidad hay una pluralidad de estas, y algunas son cosas en las que nunca hemos pensado como ‘inteligencia’ de manera absoluta” (Weinreich-Haste, 1985, Pág. 48). Armstrong (1999) luego explica que en definitiva, lo que Gardner postuló es una versión más sofisticada y mejorada de la concepción de la inteligencia dividida entre ‘cerebro izquierdo’ y ‘cerebro derecho’ (Armstrong, 1999). El alumno, teniendo en cuenta el concepto de inteligencias múltiples, ya no es visto como un agente que absorbe datos que le son suministrados, sino que construye su propio aprendizaje; el aprendizaje ha dejado de ser entendido como un mecanismo para la acumulación de hechos y verdades, por el contrario, el conocimiento es visto como el producto de las interpretaciones que los propios estudiantes generan cada vez que se enfrentan a la información que les es presentada en diferentes contextos (ver García y Pearson, 1994).

Ahora bien, ¿Cómo puede el docente describir las inteligencias múltiples de sus alumnos? Armstrong (1999) reconoce que cuando sus alumnos empiezan la escuela, ya cuentan con formas de aprender que se vinculan más con un tipo de inteligencia específico que con otro. Armstrong menciona este punto con lo que “Howard Gardner llama

proclividades o inclinaciones hacia inteligencias específicas, desde una muy temprana edad” (Armstrong, 1999, Pág. 45). Tanto Gardner (1983) como Armstrong (1999) concuerdan en que la mejor herramienta con la que cuentan los docentes para identificar estas áreas de fortalezas de los alumnos es la observación. Con lo cual, la inquietud que surge es ¿hasta qué punto las evaluaciones estandarizadas priorizan la observación y el criterio del docente a la hora de evaluar el nivel de aprendizaje de los alumnos? Más allá de la observación, Armstrong sugiere otras herramientas adicionales que el docente debería utilizar para evaluar el rendimiento de sus alumnos. Por ejemplo, establece que “si los alumnos muestran una capacidad particular para contar historias o cantar canciones, grábelos y guarde el casete como documento. Si los alumnos tienen habilidad para dibujar o pintar, guarde algunos ejemplos de su trabajo o sáqueles fotografías o diapositivas (...) En última instancia, la información evaluativa de las IM consistirá de varias clases de documentos, incluyendo fotos, dibujos, muestras de trabajo escolar, casetes de audio, videos, fotocopias color y otros mas” (Armstrong, 1999, Págs. 48 – 53). En otro trabajo, Hackney Gray (2005) mencionó otra desventaja de las evaluaciones estandarizadas para considerar diferentes tipos de aprendizaje. Es decir, por ejemplo, sostuvo que el estilo de aprendizaje de los nacidos entre 1987 y 1998 (los *millennials* – ver Hackney Gray, 2005, Pág. 100), tienen un estilo aprendizaje basado en los medios. Este tipo de aprendizaje implica una construcción activa del conocimiento a través del aprendizaje colaborativo, y por ejemplo, el aprendizaje de múltiples perspectivas para lograr a través del pensamiento crítico, un entendimiento de fenómenos sociales. No obstante ello, el enfoque de elección múltiple, que sirve de estructura de la gran mayoría de evaluaciones estandarizadas, no permite a que los alumnos vuelquen múltiples perspectivas, y rara vez pueden manifestar libremente el resultado de la puesta en marcha de su pensamiento crítico.

Si bien es cierto que existen por separados tests estandarizados que permiten evaluar diferentes tipos de inteligencias, como ser la Escala de Inteligencia de Weschler para niños, el cual incluye sub tests que requieren inteligencia lingüística, inteligencia lógico-matemática e inteligencia espacial, “no hay en el mercado un ‘megatest’ que pueda ofrecer un panorama comprehensivo de las inteligencias múltiples de los alumnos” (Armstrong, 1999, Pág. 47). De ello se desprende un problema central de las evaluaciones estandarizadas, dado que no pareciera existir una relación coherente entre el tipo de enseñanza que se pretende que los alumnos reciban y el tipo de evaluación que se les impone. Por ejemplo, en el Estado de Nueva York un millón doscientos mil estudiantes vieron cambios en la manera en la cual se les debía enseñar a leer. Si bien existen debates pedagógicos acerca de cual es la mejor manera de enseñar a leer, no es incoherente afirmar, que probablemente existan casos en los que el mejor método para aprender, no se corresponde con el mejor para la mayoría de las personas. Es decir, como bien sugiere Meier (2004, Pág. 66), tanto en la educación como en la medicina, no siempre una solución es la mejor para todos. En esos casos, precisamente, es el ojo clínico del buen medico, y por lo tanto de buen docente, el que debe adaptar el tratamiento a las necesidades específicas del paciente. No obstante ello, el afán de estandarizar persigue, al menos implícitamente, minimizar el rol de ese ojo clínico del docente. En una postura un tanto más radical, Dworkin (2005) sostiene que las evaluaciones estandarizadas pueden ser utilizadas como un mecanismo de discriminación negativa, dado que, en definitiva, perjudica claramente a aquellos alumnos que tienen problemas con el tipo de evaluaciones de elección múltiple (Dworkin, 2005, Pág. 170). Si bien es cierto que no es estrictamente necesario que todas las evaluaciones estandarizadas sean de elecciones múltiples, es un hecho de la realidad que por lo general, más aún si se tiene en cuenta la cantidad de exámenes que se deben corregir a nivel estatal, lo son.

Es importante también tener en cuenta la importancia de reconocer que una educación basada en el concepto de inteligencias múltiples se basa en las fortalezas del educando. Tanto la motivación como el entusiasmo no son ajenos a los resultados que los alumnos obtienen a través del proceso educativo. Paris (1991) sostuvo que la motivación y la perseverancia del estudiante, así como el contenido y el ajuste de la tarea, tienen efectos fuertes sobre el aprendizaje (Paris et al., 1991). Similarmente, Armstrong (1999) afirmó que las evaluaciones estandarizadas estimulan “el aprendizaje extrínseco (por ejemplo, aprender a aprobar un test con notas altas [mientras que las evaluaciones auténticas] estimula el aprendizaje por sí solos. [A su vez, las evaluaciones auténticas] ofrecen experiencias interesantes, activas, vividas y excitantes [mientras que los tests estandarizados] crean tensiones que afectan negativamente el desempeño del niño” (Armstrong, 1999, Pág. 157). Por último, Ruggieri (2002) describió como al utilizar libros de cómics, sus estudiantes pudieron utilizar métodos novedosos para hacer un análisis de las críticas a cuestiones de importancia social. Lo novedoso del recurso elevó el interés de los alumnos; de hecho, el 99% de los alumnos cumplió con el deber de buscar cómics y hacer un análisis al respecto, mientras que el promedio habitual de cumplimiento del grupo era 15% inferior. En suma, el promedio en los resultados de los alumnos creció un 5 %. Es decir, a través de la implementación de estos métodos alternativos de evaluación, un sector importante de la literatura ha entendido que se logra aumentar el grado de motivación del alumnado.

En resumen, a través de las evaluaciones estandarizadas se busca lograr una alta calidad de enseñanza para todos los alumnos. Ello, sin embargo, requiere que el docente se adapte a las necesidades específicas del alumno (lo cual concuerda con una enseñanza basada en las inteligencias múltiples), sin embargo, por otro lado el concepto de estandarización se suele materializar en una exigencia a los alumnos para que demuestren lo que han aprendido por medio de tests concentrados principalmente en dos tipos de inteligencia: la inteligencia lógico-matemática y la inteligencia lingüística. Concordantemente Gardner y Hatch, en un trabajo publicado en 1989 demostraron que en las escuelas se utiliza prácticamente dos tipos de simbología: simbolización lingüística y simbolización lógica matemática, dejando de lado muchos otros tipos de simbologías que son importantes en la actividad cognitiva del ser humano (ver Gardner y Hatch, 1989). Lógicamente uno podría decir que es imposible contemplar todos los tipos de inteligencias de los alumnos en estas evaluaciones; justamente esa es la crítica central: la estandarización no contempla el hecho que los estudiantes tienen diferentes habilidades, y por lo tanto, no son “bienes” homogéneos.

## **Curriculum oculto y real en relación a las evaluaciones estandarizadas**

El ministerio de educación de Victoria de Canadá aportó una muy útil definición del término currículo al establecer que éste era: “todas las medidas que adopta un centro para hacer realidad el aprendizaje y el desarrollo de los alumnos, incluidas la secuenciación, formato y contenidos de las asignaturas; los métodos de enseñanza y la forma en la que están organizados los profesores y alumnos (en Darling-Hammond, 2001, Pág. 295). Autores como Gallart y Darling Hammond entre tantos otros sostienen que el currículo que se enseña no es una copia fiel del plan de estudios que recibe la escuela por parte del ente que la regula. Es decir, existen diferentes tipos de *currícula* que se desarrollan paralelamente: el currículo prescripto y el currículo oculto, los cuales conforman el currículo real (ver Gallart, 2006). Es decir, lo que se enseña en las escuelas implica en cierto grado, una adaptación de lo pautado a la realidad de esa escuela. De este concepto podemos inferir que, por lo tanto, los

*currícula* real en las escuelas de un mismo país no son necesariamente homogéneos. Si bien es cierto que puede haber un núcleo que se repite en todos los *currícula*, también es cierto que los estudiantes de diferentes escuelas aprenden diferentes conocimientos y, especialmente, aptitudes. Por razones obvias, las evaluaciones estandarizadas no contemplan estas diferencias, con lo cual, es evidente que lo que se evalúa, en la mayoría de los casos, no es necesariamente lo que los alumnos *realmente* saben, tampoco reflejan los métodos mediante los cuales los alumnos aprenden (ver Paris et al., 1981).

Uno de los objetivos explícitos de los procesos de evaluación es motivar a los alumnos a que puedan demostrar lo que saben. De esta idea se desprende la noción que es óbice: que la evaluación condiga con aquello que fue enseñado (Smith, 1990). Si en los exámenes se hace hincapié en una metodología de enseñanza que no fue utilizada por los profesores dentro de la escuela se pierde ese supuesto efecto motivador que tienen las evaluaciones. Por lo tanto, este tipo de evaluación que expone a los alumnos a una situación de presión y ansiedad, en la mayoría de los casos no sirve como indicador del currículo real, es decir, de lo que los estudiantes realmente aprenden en las escuelas. No debemos olvidar, que por sus características estructurales, las evaluaciones estandarizadas solamente evalúan una ínfima parte de la asignatura (Koretz, 2002). Ello, sumado al hecho que gran parte de lo que se enseña no está necesariamente prescripto en el currículo, determina que el contenido del examen solamente refleja una ínfima parte de lo aprendido.

Robert Beck, entre otros autores (ver Barickman, 1978) sostiene que las evaluaciones estandarizadas deben caer, por lo general, en la metodología de elección múltiple por una cuestión de economía, ya que como bien se dice el tiempo vale oro, con lo cual, cuanto más tiempo lleve corregir esos exámenes mas costoso será el sistema. Ello hace que los métodos de evaluación sobre los cuales reposan estos *tests* sean defectuosos, ya que no son capaces de medir todas las habilidades aprendidas, dado que no todas las habilidades son cuantificables con tanta facilidad. (ver Beck, 1976). Es importante reflexionar acerca de la sugerencia de Darling Hammond (2001) con respecto al rol que la estandarización debe tener en el currículo: “los estándares comunes y externos deberían ocupar un 50% del tiempo lectivo y el curriculum, dejando el 50% restante para que los profesores desarrollen una enseñanza basada en los intereses y experiencias de los alumnos, tomando en cuenta así sus necesidades específicas de aprendizaje” (Darling Hammond, 2001, Pág. 298). Tal como sugiere Pophan (2004) los docentes deben invertir su tiempo en tareas orientadas a cumplir con los estándares dictados estatalmente, las cuales son muchas veces incompatibles con la instrucción de alta calidad y con su propia realidad áulica; en definitiva la creatividad cede su espacio para la excesiva preparación de las evaluaciones estandarizadas (Pophan, 2004, Pág. 1).

En la misma línea de pensamiento, casi treinta años atrás, Ben Peretz (1975) mencionó que la idea es que sea el profesor quien logre hacer realidad el curriculum al utilizar diferentes materiales y recursos para poder así construir ideas junto con los estudiantes (Ben Peretz, 1975). A su vez, Darling Hammond (2001) sostiene que: “las orientaciones curriculares que prescriben en exceso los contenidos y los métodos impiden que los profesores construyan los puentes necesarios entre las experiencias de los alumnos y las metas del aprendizaje” (Darling Hammond, 2001, Págs. 297 y 298). Las experiencias de los alumnos permiten que ellos construyan sus propios aprendizajes, convirtiéndose no solamente en sujetos, sino en actores de conocimiento. Es importante una vez más otorgarle al profesor la posibilidad de utilizar su ojo crítico para adaptar ese currículo prescripto a uno real, que condiga con los intereses y las problemáticas de su grupo de alumnos, en particular.

En contraposición a lo que sucede con las evaluaciones estandarizadas, el concepto de una evaluación auténtica, basada en principios de inteligencia múltiple, pareciera contribuir en el fortalecimiento del currículum real. Campbell y Campbell (1999) sostuvieron que la ventaja del concepto de las inteligencias múltiples no prescribe un currículum fijo e inviolable. Es decir, cuando las instituciones abrazan el concepto de las inteligencias múltiples gozan de un mayor grado de libertad para adaptar el currículum a las necesidades específicas de sus alumnos. Coincidentemente, Armstrong (1999) sostiene que las evaluaciones auténticas, basadas en un concepto de inteligencias múltiples “permite a los maestros desarrollar currículos significativos y evalúa en el contexto enseñado [mientras que los tests estandarizados] acentúa exámenes singulares que evalúan el conocimiento de una sola mente en un momento singular” (Armstrong, 1999, Pág. 157).

### Los riesgos de enseñar para el test

Las evaluaciones estandarizadas han sido utilizadas para evaluar el nivel académico, ya sea de un estado, distrito o de una institución misma. Algunos autores, como Eisner, sostienen que una de las ventajas de estos *tests* es que plantean la posibilidad de acceder a resultados de una población extensa, lo cual sería imposible, o al menos más costoso, si se adoptase cualquier otro tipo de evaluación (Eisner, 1987). Del resultado de estas evaluaciones estandarizadas estatales surgen consecuencias importantes para la escuela, no sólo vinculadas a su financiamiento, sino también a su prestigio. Esto genera una clara tentación a hacer todo lo posible por elevar los resultados de las evaluaciones. Existen muchas estrategias que pueden ser implementadas, algunas de ellas son lícitas y otras ilícitas. De estas últimas me encargaré en el apartado que sigue; en esta parte describiré aquellas estrategias “lícitas” que pueden ser utilizadas por los profesores y directivos para mejorar los resultados de los *tests* (lo cual no implica, necesariamente, un incremento en el nivel educativo).

Koretz (2002) describe como caso ejemplar una anécdota referida a algo que sucedió en el estado de Kentucky: los educadores dedicaban 5,2 horas por semana a ciencias, materia que era evaluada ese año con una prueba estandarizada, mientras que los educadores de quinto año le asignaban a la misma materia solamente 3,5 horas por semana. Inversamente, los maestros de quinto grado le asignaban 6,4 horas a la enseñanza de las matemáticas, ya que esa era la materia que se evaluaba ese año, comparado con las 4,9 horas que le dedicaban los profesores de cuarto grado. Similarmente, autores como Linn (2000) y Popham (2001), sostienen que muchas veces el incremento en los resultados de este tipo de evaluaciones, no se debe a una profundización de los conocimientos, sino a una familiarización con la metodología del examen. En suma, como consecuencia de la presión de cumplir con los estándares propuestos, dentro de los docentes, surge el incentivo de sacrificar actividades que fomenten la creatividad y el pensamiento crítico, con el fin de alocar el tiempo en estrategias destinadas a mejorar el rendimiento de los alumnos en los exámenes estandarizados. Por ejemplo, Lipman (2004) realizó un estudio acerca de escuelas en Chicago. Uno de los datos alarmantes que arrojó su estudio es que en una escuela primaria, la cual contaba con una población estudiantil de origen mexicano-americano, no se pudo utilizar todo ese bagaje cultural para realizar actividades creativas que pudiesen despertar la noción de entendimiento internacional entre los alumnos, dado que ese tiempo implicaba una menor preparación para la evaluación estandarizada. Eisner (1987) quien criticó la postura de “volver a lo básico”, sugiere que las evaluaciones estandarizadas no contemplan muchas habilidades que los estudiantes poseen, lo cual es injusto para ellos. Es decir, si uno

posee una habilidad “x”, que puede ser muy útil tanto en la educación superior como también así en el campo laboral, pero las evaluaciones estandarizadas solo evalúan las habilidades “a” y “b”, este estudiante estará destinado a obtener un resultado más bajo del que podría obtener si las evaluaciones estandarizadas hiciesen hincapié en sus habilidades. Es decir, si bien la estandarización no implica *per se* “volver a lo básico”, lo que sucede es que indirectamente se termina volviendo a lo básico.

Desafortunadamente, al tener que obtener buenos resultados, los docentes tienen la tendencia de privilegiar los contenidos que serán evaluados, los cuales concuerdan con aquellas asignaturas defendidas por quienes bregaban por volver a lo básico, en desmedro de otras asignaturas como ser las artes e incluso, ciertas áreas de las ciencias experimentales. En otro trabajo, McNeil (2000) relata que a raíz de la implementación de las evaluaciones estandarizadas los alumnos ya no aprenden a escribir bien, sino que se los instruye en como escribir un ensayo de cinco párrafos, formados cada uno de ellos por cinco oraciones, que les permite los alumnos obtener una buena calificación en el examen estandarizado al cual son sometidos. El problema con la selección de contenidos curriculares en función de las evaluaciones estandarizadas no es algo nuevo. De hecho, treinta años atrás, Barickman (1978) explicó que algunos docentes del estado de Nueva York asignaban la importancia que le darían a los diferentes aspectos de la enseñanza del uso del lenguaje en función de las evaluaciones estandarizadas; es decir, si las posibilidades de obtener un puntaje más alto en las evaluaciones de vocabulario son mayores a las posibilidades de obtener puntajes altos en las evaluaciones de composición, entonces dedicarán más tiempo a la enseñanza de vocabulario, dejando rezagada la enseñanza de composición (Barickman, 1978). Sin embargo, a medida que crece el impacto de los resultados de esas evaluaciones, es de esperar que los docentes tengan el incentivo de dedicar más tiempo a esas áreas que serán evaluadas. De ello se desprende que el peligro de las evaluaciones aumenta por la importancia que tienen los resultados de esos exámenes, es decir, si los exámenes fuesen simplemente uno de muchos otros elementos –posiblemente cualitativos- que se utilizaran para evaluar el desempeño de los docentes, probablemente el incentivo de enseñar para el test, al menos disminuiría.

Algunos educadores insistimos en pedirles a nuestros estudiantes se preocupen menos por las notas y más por lo que aprenden. El sistema de evaluaciones estandarizadas, sin embargo, atenta contra este principio ya que los estudiantes terminan dándole más importante al “valor moneda” de lo aprendido que al valor intrínseco que tiene el conocimiento (ver Haladyna et al., 1991). No obstante ello, como indiqué anteriormente no son sólo los estudiantes los que se preocupan por “subir” los resultados de las evaluaciones: existe también una clara estructura de incentivos para profesores y directivos.

Como director de una escuela, George Wood (2004) admitió que en un caso particular de una alumna llamada Angélica, sintió la tentación de pedirle que abandone la institución, con el fin de lograr un mejor resultado a nivel colectivo. Es decir, lo que la presión de las evaluaciones estandarizadas forzó, en el caso de Wood, es tener que elegir entre una alumna en particular, la cual probablemente jamás alcanzaría los estándares que el estado prevé, y el bienestar de la institución en su conjunto. En su trabajo, Wood demuestra que, por ejemplo, en la el Sharpston High School, de Houston, en el año 2001-2002, 462 alumnos, de los 1700 que tenía la institución, abandonaron la escuela. No obstante ello, ninguno fue reportado como un caso de deserción escolar. En otro trabajo, Darling Hammond (2004) demostró que en el estado de Massachussets en el 2002 el porcentaje de alumnos afroamericanos que lograron graduarse, con respecto a la cantidad de alumnos que había en el noveno año, era del 71 por ciento; sin embargo, en el año 2003 el porcentaje de

alumnos graduados fue del 59,5 por ciento. Algo similar ocurrió con los alumnos de descendencia latina: por ejemplo, del 54 por ciento de los alumnos que se graduó en el 2002, en el 2003 solo lo logró el 45 por ciento (Darling Hammond, 2004, Pág. 21).

Similarmente Karp (2004) sostiene que esta encrucijada la viven los directivos de las escuelas muy frecuentemente. El autor, sostiene que las evaluaciones estandarizadas, en cuanto ponen mucho en juego, generan el incentivo de enviar a los alumnos que difícilmente aprueben las evaluaciones estandarizadas, a los grupos de educación especial, o incluso a aquellos de alumnos con dificultades en el aprendizaje del idioma inglés. En uno de sus trabajos Haas incluyó una reflexión de una profesora, mediante la cual explicaba que ella debía dedicar mucho tiempo a la enseñanza de técnicas para obtener mejores resultados en las evaluaciones, tiempo que podría ser utilizado para la enseñanza de otro tipo de conocimientos adicionales (ver Haas, 1989). Si bien no es probable que los directivos diseñen el plan institucional, tomando como único referente las evaluaciones estandarizadas, sí es probable afirmar, en función de los ejemplos expuestos en esta parte, que le dediquen mucho de su tiempo (recurso escaso) a la preparación de una prueba, en detrimento de otras aptitudes que pudiesen ser enseñadas; es decir, existe una clara estructura de incentivos para que los docentes enseñen para el *test* (ver Furst, 1963). Si bien los ejemplos presentados se corresponden a experiencias de los Estados Unidos, algo similar ocurre en países como Holanda o Nueva Zelanda, en los que el impacto de las evaluaciones estandarizadas al final de la escuela secundaria delimita las posibilidades que los alumnos tienen en el campo de la educación superior. Una vez más, es importante mencionar que la fuerza de esta crítica aumenta en los países en los que los resultados de estas evaluaciones estandarizadas —es decir, los datos puramente cuantitativos— son utilizados como los principales métodos de evaluación del rendimiento y funcionamiento del sistema educativo. Probablemente, la crítica tenga menos fuerza en lo que concierne a sistemas como el británico en el que además de los resultados de los llamados “A” y “O” Level Examinations, también se utilizan “visitas” o inspecciones a las escuelas, a través de las cuales se obtienen datos cualitativos. La confluencia de los datos cualitativos con los datos cuantitativos disminuye los riesgos de enseñar para el test, dado que el desempeño de los docentes no se evalúa únicamente a través del rendimiento de los alumnos en las evaluaciones estandarizadas.

Algunas de las estrategias que pueden ser utilizadas para mejorar los resultados de las evaluaciones son: enseñar estrategias útiles para resolver las preguntas de ese tipo de exámenes, presentar modelos de exámenes similares y basar el currículo y el proceso de aprendizaje en ellos, y/o utilizar libros de textos que contengan ejercicios similares a los de los exámenes. En principio uno podría decir que no hay nada de malo en esto. Es totalmente lícito que los profesores quieran que sus alumnos obtengan mejores resultados en las evaluaciones, ya que en definitiva ello no perjudica a nadie. Sin embargo, como bien el principio de la demanda y de la oferta los recursos son escasos por definición. El tiempo, uno de los tantos recursos sobre los cuales descansa el sistema educativo, no es una excepción a esta regla. Por lo tanto, el problema surge cuando se genera una clara estructura de incentivos para dedicar tiempo a la enseñanza de estrategias específicas para un tipo de evaluación también específico.

### **Incentivos para hacer trampas con los resultados**

En el apartado anterior evalué las posibilidades, mediante vías lícitas, que tienen tanto los docentes como los directivos para mejorar los resultados de las evaluaciones. Sin

embargo, existen muchas otras artimañas fraudulentas que pueden ser puestas en marcha para mejorar los resultados de las evaluaciones.

Existen muchos métodos mediante los cuales los directivos pueden fraguar los resultados de las evaluaciones estandarizadas, Haladyna, entre otros, sugiere que se pueden fraguar las hojas en las que los estudiantes vuelcan sus resultados, borrando las respuestas escogidas por los estudiantes y marcando la respuesta correcta con mayor vehemencia. En otras oportunidades se le permite a los alumnos con bajo rendimiento no asistir a clases los días que se tomaran los exámenes (ver Haladyna et al., 1991). En todos estos casos se percibe cómo la maquinaria institucional está puesta en función de la corrupción de los resultados que arrojan las evaluaciones.

Considero que la importancia y la publicidad que se le dio a este tipo de evaluaciones en los últimos años generaron mayores incentivos para mejorar los resultados, cueste lo que cueste. Con el paso del tiempo los resultados empezaron a ser publicados en diarios, tanto locales como nacionales y en varios sitios de Internet. A través de la publicación de los resultados se confeccionaron *rankings* que van modificándose todos los años en función de los resultados de las evaluaciones. Es harto evidente que en la medida que se le siga dando cada vez más importancia a las evaluaciones, los incentivos para corromper los resultados serán cada vez mayores. Por ejemplo, al describir el uso de las evaluaciones estandarizadas en China, Suen reflexiona que como hay mucho en juego cuando los alumnos deben rendir los gaokaos, es lógico concluir que existe una clara tentación a fraguar los resultados (ver Suen, 2006). Por lo tanto, siguiendo la lógica de Suen es posible concluir que cuanto mayor importancia se le de a los resultados de las evaluaciones estandarizadas, mayores serán los incentivos a corromper los resultados.

En los Estados Unidos, al menos veinte estados han implementado un sistema mediante el cual las escuelas o los profesores reciben sanciones o beneficios, en función del resultado de las evaluaciones estandarizadas de sus alumnos. En suma, algunos estados como Nueva York o Chicago han implementado evaluaciones denominadas “*promotional gates*”, las cuales son requisitos para que los alumnos puedan ser promovidos de grado en grado. En su trabajo, Koretz (2002), analiza como una determinación excesiva de la importancia de este tipo de evaluaciones puede ser responsable de la creación de un sistema de incentivos perversos, para “inflar” los resultados de los alumnos de las escuelas en los Estados Unidos. De hecho, el autor denunció que para el año 2002, al menos 71 escuelas en 22 distritos del estado de Michigan se encontraban bajo investigación por posible caso de corrupción durante las evaluaciones estatales.

Por otro lado, los estados pueden utilizar su discrecionalidad a la hora de establecer los estándares. Por ejemplo, Murnane (2007) sostiene que uno de las tentaciones que plantea NCLB es que los estados pongan estándares más bajos con el fin de lograr un mejor financiamiento. De hecho, Linn (2005) sostiene que, por ejemplo, en el 2003 el 77% de los alumnos de cuarto grado de las escuelas de Alabama obtuvieron resultados de aprobación en los exámenes estatales. No obstante ello, en el mismo año, solamente el 22% aprobó el examen administrado por NAEP (Linn, 2005). A este problema, Murnane le agrega otro: si bien es cierto que dentro de los lineamientos generales se prevé que las escuelas deben entregar no sólo los resultados obtenidos, sino también los niveles de deserción escolar, lo que sucede es que no deben revelar en cuales de los tantos grupos que conforman el alumnado se ha sentido con mayor vehemencia esa deserción. Es decir, si bien la totalidad de la deserción, en cuanto al alumnado tomado como un ente colectivo, puede ser baja, ello no implica que sea alta en algunos subgrupos en particular (Murnane, 2005, Pág. 165). De estos ejemplos citados se demuestra con claridad que los estados cuentan con herramientas para

adaptar los resultados y las evaluaciones a sus necesidades reales. Esa artimaña claramente desnaturaliza el espíritu de las evaluaciones estandarizadas, por ejemplo recogido en NCLB, dado que en la norma se sugiere la importancia que todos los estudiantes del país, mas allá de su contexto ya sea geográfico, como también socio económico, gocen de los mismos derechos a una educación de calidad.

### **Las evaluaciones estandarizadas y los contextos sociales**

Es común escuchar que el ser humano es un ser social, por lo tanto, lo que sucede a su alrededor, en diferentes grados, moldea y afecta tanto sus acciones como su personalidad. El estudiante no escapa a esta regla, dado que la educabilidad se encuentra, al menos en parte, definida por las condiciones en que tiene lugar la escolarización (Terigi, 2005; ver también Oxford, 1997). Sin lugar a duda el contexto en el que el estudiante vive deja sus huellas en su proceso de educación.

Los sistemas de evaluación estandarizada se utilizan con todos los estudiantes del sistema educativo. Ello, trae problemas en países en los que conviven integrantes de diferentes culturas, quienes no siempre tienen como lengua nativa al idioma oficial del país en el que residen. Los alumnos que deben, por ejemplo, rendir una evaluación de alcance nacional en inglés, pero tienen como lengua primaria el español, se encuentran claramente en una situación de desventaja (ver García y Pearson, 1994). Por lo tanto, en estos casos, en los que existen diferencias en cuanto al nivel de manejo del idioma del examen, se puede determinar erróneamente que el nivel educativo de una escuela es malo cuando en realidad existen otros factores que determinan que los resultados sean bajos. Haladyna (1991), entre otros, expresa con mucha claridad que los resultados de las evaluaciones están estrechamente vinculados a la influencia de las condiciones socio económicas en las que vive el estudiante, por lo tanto, el origen étnico y cultural de los estudiantes debe ser un factor a tener en consideración a la hora de evaluar el rendimiento de las evaluaciones (ver Haladyna et al., 1991). Por lo tanto, no sería descabellado afirmar que por su metodología, las evaluaciones estandarizadas hacen la vista gorda a todas estas particularidades que son elementos constitutivos del aprendiz.

Como ya he mencionado en otras partes de este trabajo, los resultados de las evaluaciones estandarizadas adoptaron mucha importancia con el correr del tiempo. Como consecuencia de ello se creó una “industria” que ofrece libros de texto específicos para que los alumnos se preparen para los *tests*. A su vez, es muy común en países como Corea y Japón, que los estudiantes contraten profesores particulares para perfeccionar sus técnicas de rendición de examen. Teniendo en cuenta la idea que cuanto mas se ponga en juego según los resultados obtenidos en las evaluaciones, mayor será la preocupación que los estudiantes tengan por su puntaje, es lógico afirmar que existe una gran demanda, por parte de los estudiantes de las clases mas pudientes, a contratar a todo aquel que pueda ayudar a obtener un mejor resultado en las evaluaciones estandarizadas. Suen sostiene que en Japón existe un sistema de educación paralelo, al que la autora ha denominado “*shadow education*” (ver Suen, 2006) mediante el cual los estudiantes acuden a profesionales especializados en la enseñanza de estrategias para obtener mejores resultados en las evaluaciones. En una sociedad tan heterogénea y con niveles de desigualdad tan grandes como la nuestra, esto plantea un inconveniente: la mala distribución de recursos pasa a ser un determinante a la hora de obtener un puntaje alto en estos exámenes.

Por otro lado, en los Estados Unidos, por ejemplo, uno de los valores que se persiguieron a traves de las evaluaciones estandarizadas es el de equidad: es decir, que todos

los estudiantes, más allá del contexto social y económico al que pertenecen puedan progresar. Hasta este punto, uno podría afirmar que no existe nada malo en ese fin, de hecho es un objetivo de valor. El problema surge cuando el incumplimiento trae consecuencias graves para los estudiantes. Por ejemplo, ¿cuál fue el impacto que tuvo NCLB en la brecha que existe entre los alumnos más favorecidos y los pertenecientes a grupos más marginados? Causey-Bush (2005) se refirió a los casos de California y de Texas, dado que el 34% de la población de color en los Estados Unidos, pertenece a estos dos estados. De acuerdo a los datos del Departamento de Políticas Educativas y división de evaluación de California, y de la Agencia de Educación de Texas, en el año 2003 solamente los alumnos “blancos” (no de origen hispanico) fueron los únicos de cumplir con los estándares propuestos, tanto en California como en Texas. Tal como sugiere Kohn (2004), uno de los daños colaterales de esta iniciativa fue que como resultado se penalizó —a través de un menor financiamiento— a aquellas instituciones que no lograron buenos resultados. (Ver Kohn, 2004, Pág. 91). Darling Hammond (2004) sugiere que es peligroso que el financiamiento educativo se vea “pegado” a los resultados de las evaluaciones. En el caso de NCLB, la autora sostiene que esa ley generó un sistema inadecuado de financiamiento del sistema educativo, mediante el cual se castiga a las escuelas que instruyen a alumnos que pertenecen a estos grupos desfavorecidos. Esto sucede, ya que para este tipo de instituciones, cumplir con los estándares federales, lo cual es necesario para recibir el mayor nivel de financiamiento posible, es prácticamente imposible. Es decir, la ley se basa en los resultados de las evaluaciones estandarizadas, a la hora de determinar la calidad educativa de la escuela, desconociendo el contexto socio cultural y económico al que pertenece el alumnado.

Por lo tanto, cuando del resultado de las evaluaciones estandarizadas se fija el grado de financiamiento que recibirán las escuelas se genera un perverso círculo vicioso, ya que aquellas escuelas que no obtienen buenos resultados no obtienen suficiente financiación para mejorar los servicios educativos; por lo tanto, ¿sin la financiación apropiada, cómo harán para obtener buenos resultados? La idea del círculo vicioso es también sugerida por autores como Dworkin (2005, Pág. 172), dado que de sus argumentos se desprende la idea que las escuelas que se encuentran por debajo de los estándares estatales, probablemente sigan estándolo, aunque mejoren (dado que año tras año los estándares deben mejorar para poder alcanzar el 100% de efectividad para el año 2014). Es decir, una vez que la institución es catalogada como ineficiente, debe, con menos recursos, mejorar para poder estar a la altura de estándares aun más altos. Financiar a través del cumplimiento de estos estándares pone en riesgo la continuidad de la instrucción para muchos alumnos, ya que tal como sugiere Darling-Hammond (2004) el sistema crea una estructura de incentivos para no aceptar a los alumnos que provengan de contextos socioeconómicos débiles. Paulatinamente los niveles de deserción escolar en varios estados de los Estados Unidos han crecido durante los últimos años. Tal como sugiere Darling Hammond los porcentajes de graduación de Nueva York, Florida y Massachusetts, entre otros, ha descendido a partir de la revalorización de los resultados de las evaluaciones estandarizadas, que se dio en los últimos años.

Son muchos los autores que han sugerido la importancia del contexto socio económico en el proceso de aprendizaje. Karp (2004) menciona que en el año 2003 *Educational Testing Services*, una de las empresas que confeccionan exámenes estandarizados, más importantes del mundo, concluyó que existen alrededor 14 factores responsables de la brecha entre los que logran buenos resultados y aquellos que fracasan. Entre aquellos factores se mencionaban el nivel de pobreza, la alimentación de los niños, la educación de los padres, la atención médica que recibieron durante la infancia. Similarmente, Karen

(2005) sostuvo que el acceso a un buen servicio médico, la estabilidad emocional en el entorno familiar del niño, la integración del niño en la comunidad y las condiciones socio económicas de la comunidad a la que pertenece el alumno, son solamente algunos de los tantos factores que afectan el nivel educativo del educando (Karen, D., 2005, Pág. 168). Con lo cual, creo que estos datos deberían ser suficientes para, al menos, ser cautelosos a la hora de asignarle importancia a los resultados de las evaluaciones estandarizadas. En definitiva, ¿es la escuela la institución facultada para confeccionar políticas que garanticen el libre acceso para todos al servicio médico público, que aseguren la asignación de recursos del presupuesto anual para la implementación de recursos modernos en el proceso de formación educativa, entre otros casos? La respuesta es claramente no.

Lógicamente existen políticas públicas que exceden a aquellos encargados de las políticas educativas. Es decir, no sería justo pedirle a quienes diseñan las políticas educativas que resuelvan el problema de que en algunas casas por el tipo de pintura con altos grados de plomo existan alumnos con deficiencias en su proceso de aprendizaje. Probablemente tampoco este dentro del campo de acción de los encargados de delinear las políticas educativas resolver los problemas de acceso a los servicios médicos o al tipo de nutrición que reciben los alumnos de los sectores menos pudientes. Sin embargo, ello no implica que no haya nada que se pueda hacer. Al menos lo que se puede hacer es no agravar la situación de los alumnos que asisten a este tipo de instituciones. Al atar el resultado de las evaluaciones estandarizadas a cuestiones como el financiamiento que reciben las escuelas (tal como sucede en los Estados Unidos) o a la posibilidad de acceder al circuito de educación superior (como sucede en países como Chile, Holanda o Nueva Zelanda), lo que se está haciendo es perpetuar esas inequidades. Probablemente una mejor estrategia sea reconocer el crecimiento de esos alumnos, en definitiva un estilo de evaluación teniendo en cuenta el valor agregado, más que el resultado final. Al menos, de esa manera se estaría dando la posibilidad a que estos alumnos asistan a escuelas con un mejor financiamiento, y que en última instancia, puedan crecer en la escala social a través de una mejor educación superior.

## **Predicciones acerca del éxito de los estudiantes en el mercado laboral**

La última de las críticas apunta contra la validez de las evaluaciones estandarizadas para determinar la aptitud y la capacidad de los estudiantes para completar sus estudios universitarios y, principalmente, para poder desenvolverse con facilidad en el terreno laboral. Si bien de alguna manera todas las críticas anteriores debilitan la validez de las evaluaciones en su conjunto, en este apartado me centraré en la poca utilidad de los resultados de estas evaluaciones como factor predictivo del éxito profesional y académico post- escuela de los estudiantes que son sometidos a ellas.

Sternberg, quién junto a Gardner ha sido uno de los más reconocidos exponentes del constructivismo norteamericano, sostuvo que los resultados de las mediciones de la inteligencia práctica para predecir el éxito en la escuela así como en el ámbito laboral no se correlacionan significativamente con la inteligencia medida psicométricamente (Sternberg, 1996). Si bien el tema de las inteligencias múltiples ya fue tratado, es importante notar que si las habilidades cognitivas requeridas en el campo laboral, no se encuentran representadas en estos *tests*, entonces los mismos no son buenos indicadores del éxito de los estudiantes a la hora de convertirse en “empleados”.

Las evaluaciones estandarizadas tienen, por lo general, un componente que es muy tomado en cuenta: tiempo. Es decir, en este tipo de evaluación se requiere que el alumno

responda las preguntas que le son formuladas en un tiempo determinado. Esta variable es necesaria para poder comparar resultados, sin ella los resultados no tendrían valor alguno (si es que en esencia lo tienen). Esta característica tampoco condice con la realidad que el estudiante va a vivir a la hora de desenvolverse en el mercado laboral. Los *tests* estandarizados miden la capacidad de un individuo en ejercicios breves, para lo cual el alumno debe entrenar sus habilidades para contestar con velocidad y cambio. Sin embargo, muchos de los problemas que se espera que los individuos resuelvan en la vida real requieren mucho tiempo, concentración e implicación. Con respecto a ello, Armstrong (1999) señaló que los tests estandarizados tienen “límites de tiempo que fuerzan los procesos de pensamiento de muchos alumnos [mientras que la evaluación auténtica] ofrece a los alumnos el tiempo que necesitan para trabajar en la resolución de un problema, proyecto o proceso” (Armstrong, 1999, Pág. 158). Es decir, no pareciera existir entre las evaluaciones estandarizadas y la realidad, en la que la gran mayoría de las tareas no se realiza en una cantidad de tiempo estandarizada, una correlación lógica (Gartner, 1999). En coincidencia con esa postura, Armstrong (1999) sostuvo que existen dos tipos principales de evaluación, por un lado se encuentran las evaluaciones estandarizadas, mientras que por el otro se encuentran las evaluaciones auténticas. Estas últimas se refieren “a criterios, a puntos de referencia que sirven para evaluar el aprendizaje y comparan el rendimiento del alumno con sus propios desempeños anteriores” (Armstrong, 1999, págs. 156 – 157). Según la postura de autor, y las ideas de Gardner (1993) lo importante es que los alumnos sean evaluados en un contexto similar “al medio donde se esperara de ellos que demuestren su aprendizaje en la vida real” (Armstrong, 1999, Pág. 156). A su vez, indica que las evaluaciones estandarizadas, reducen las vidas complejas y ricas de los niños a una colección de notas, porcentajes y calificaciones [mientras que las evaluaciones auténticas] dan a los maestros una sensación ‘vivencial’ del niño como estudiante” (Armstrong, 1999, Pág. 157).

En la misma línea de pensamiento, y teniendo en cuenta lo anterior, es preciso mencionar el trabajo de Causey-Bush (2005), en el que sostiene que cumplir con estándares rígidos no predice, necesariamente el éxito de los alumnos en sus estudios universitarios. La autora sostiene que el fin de la educación secundaria debería estar centrado en proveerles a los alumnos las herramientas de análisis y pensamiento crítico que puedan ser utilizados en cualquier contexto, no limitándose a la metodología de un tipo de examen en particular. El hecho de obtener buenos resultados en las evaluaciones, no es suficiente, por sí solo, para asegurar el éxito del estudiante en su desarrollo profesional (Causey Bush, 2005). George Wood (2004), similarmente, sostiene que no existe en la actualidad ningún estudio sólido que demuestre que como consecuencia del énfasis en la estandarización, los alumnos logran ser mejores empleados, estudiantes universitarios e incluso, mejores ciudadanos. Por otro lado, Deborah Meier (2004), sostiene que las legislaciones, independientemente del país del que provengan, que tienen como fin la estandarización genera un debilitamiento de los principios democráticos. La autora sostiene que es importante fomentar la participación de los constituyentes de los estados en la regulación del sistema educativo. Meier defiende la idea que la participación de los estudiantes en materia de regulación del bien educación, fomenta un aprendizaje acerca de cómo funciona el “mundo adulto”. Es decir, la posibilidad que las familias solían tener, de opinar en materia de decisiones locales con respecto a la educación de sus hijos, fortalecía la democracia en su totalidad. La implementación de las legislaciones con un acento en el poder estatal, lo cual es característico de la mayoría de las legislaciones en las que se utilizan estándares nacionales, implica un fortalecimiento de los poderes políticos y burocráticos más altos, en cuanto a la regulación de la educación. Como consecuencia de ellos, los directorios locales pierden poder; la estandarización logra alejar a

las familias y a los docentes de las decisiones fundamentales en cuanto a la regulación y la administración de la educación.

Pareciera existir, en suma, una relación entre esta crítica, y la presentada en el apartado anterior. En la mayoría de los casos, los alumnos de los sectores socio económicos más bajos, son aquellos que se encuentran en un mayor riesgo de no alcanzar los estándares requeridos. Teniendo en cuenta que una de las ventajas de las evaluaciones estandarizadas es que resultan menos costosas que otros métodos que requieren, por ejemplo datos cualitativos como ser inspecciones a las escuelas, no es ilógico afirmar que en la mayoría de los casos solamente se evalúan algunas áreas. Por ejemplo, en el caso de los Estados Unidos, si bien existen excepciones como ser el caso de Virginia, lo que la mayoría de las legislaciones requiere es que se evalúen las áreas de lengua y matemática. Como consecuencia de ello, son los alumnos de esos sectores menos pudientes (los cuales tradicionalmente obtuvieron bajos resultados), los que van a ser sometidos a un currículum aún más focalizado en las áreas que se serán evaluadas (obviamente, ello en detrimento de los contenidos de otras áreas como ser las artes, las ciencias sociales y la educación física). Por ende, a largo plazo, cuando estos alumnos que tuvieron un currículum basado, en mayor medida, en habilidades cognitivas inferiores (en contraposición a asignaturas que requieren habilidades cognitivas superiores como ser la educación en valores o la filosofía, entre tantas otras), se van a encontrar en una situación de desventaja con respecto a aquellos alumnos que, por no tener problemas en lengua y matemática, sí fueron expuestos a otras asignaturas, como ser la educación cívica, la filosofía, la biología, la historia y la economía, entre tantas otras (Rothstein, 2008). A la hora de competir por empleos que requieran conocimientos que excedan las habilidades instruidas a traves de la enseñanza de la matemática y de la lengua, probablemente los alumnos de los sectores menos privilegiados, no se encuentren tan bien preparados como aquellos que asistieron a instituciones que no tuvieron que hacer tantas acomodaciones para enseñar para el test. Es decir, si bien uno de los principios de las legislaciones que contemplan las evaluaciones estandarizadas es lograr la equidad, a largo plazo, pueden servir para perpetuar esas inequidades.

## **Conclusión**

En este trabajo he planteado seis críticas centrales a los modelos basados en evaluación estandarizada. En la primera de ellas he demostrado la evolución que ha experimentado el concepto de inteligencia en los últimos años. Es decir, la inteligencia ya no es vista como un factor estático de características únicas, sino que hoy en día gran parte de la literatura reconoce la multiplicidad de tipos de inteligencia que los seres humanos poseemos. Ante el surgimiento de esta idea, las evaluaciones estandarizadas han perdido su utilidad, ya que son funcionales para algunos tipos de inteligencias, perjudicando a quienes tienen otros tipos de inteligencia no contemplados. Estamos viviendo en una era en la que es necesario volcarse a un sistema de enseñanza centrado en el aprendizaje, a través del cual se tomen en cuenta las diferencias entre los educandos, con el fin de individualizar la educación (Gardner, 1998).

La segunda de las críticas consintió en determinar que existen diferencias entre el currículum real y el currículum prescripto. Por tratarse de evaluaciones estandarizadas es necesario que estén basadas en un currículum prescripto, dejando de lado el currículum real y los métodos de enseñanza que una escuela puede llegar a implementar.

La tercera y cuarta crítica están relacionadas entre sí. Ambas plantean la posibilidad de afectar los resultados de las evaluaciones estandarizadas. En la tercera de las partes

describí los métodos lícitos de hacerlo, mientras que en la cuarta parte expliqué los métodos fraudulentos que las instituciones educativas pueden implementar para mejorar sus resultados. Es importante recordar que a medida que se le siga dando tanta importancia a los resultados de las evaluaciones, por ejemplo para determinar salarios, mayor será la estructura de incentivos para corromper los resultados de las evaluaciones.

En la quinta parte critiqué el hecho que las evaluaciones estandarizadas no hacen reparos en los diferentes contextos en los que se llevan a cabo las mismas. El efecto homogenizador al que apunta este tipo de evaluaciones no condice con la heterogeneidad de nuestra sociedad.

Por último, en la sexta parte apunté contra el poder “predictivo” que tienen las evaluaciones estandarizadas para predecir el éxito universitario, y en el mercado laboral, de los estudiantes secundarios. A lo largo de esta sección he demostrado que no existe una correlación entre las condiciones en las que se llevan a cabo las evaluaciones estandarizadas y la educación superior, e incluso, en últimas instancias, las condiciones en las cuales se desarrollan la mayoría de las profesiones. Mientras las evaluaciones estandarizadas priorizan el trabajo rápido con límites preestablecidos, la vida adulta de los estudiantes requieren que ellos puedan utilizar su pensamiento crítico, lograr trabajar cooperativamente en equipos, entre otras cuestiones (Armstrong, 1999, Pág. 158).

A esta altura muchos podrán preguntarse si este trabajo sugiere la eliminación de las evaluaciones estandarizadas en los sistemas educativos nacionales. La respuesta es no, o al menos, no necesariamente. En definitiva, las evaluaciones son parte de nuestra vida, más allá de la experiencia escolar. A su vez, sería necio negar que las evaluaciones estandarizadas hayan sido útiles para proporcionar datos acerca de deficiencias y fortalezas de los sistemas educativos. El problema, sin embargo, es cuando son utilizadas como el único barómetro para determinar cuestiones trascendentales en las vidas de los diferentes actores del sistema educativo. Por ejemplo, es inaceptable, por las razones expuestas en este trabajo, que solamente a través de las evaluaciones estandarizadas se determinen las posibilidades que los alumnos tienen para acceder al sistema universitario (tal como sucede en China, Turquía e Inglaterra entre otros países). A su vez, también pareciera ser excesivo que se utilicen los resultados de las evaluaciones estandarizadas para determinar el grado de financiamiento que recibirán las instituciones educativas, tal como sucede en el caso de los Estados Unidos.

¿Cómo pueden entonces utilizarse estas evaluaciones estandarizadas? En el caso de aquellos países que actualmente las utilizan para determinar las posibilidades de los alumnos de acceder a la universidades, las evaluaciones estandarizadas se podrían utilizar en conjunto con otro tipo de evaluaciones que se lleven a cabo “in situ”. En definitiva, la propuesta consiste en complementar una evaluación externa (las evaluaciones estandarizadas) con evaluaciones internas (por ejemplo, portfolios). De esa manera se estarían sorteando varias de las críticas formuladas en este trabajo. En definitiva, el incentivo de enseñar para el test disminuiría, dado que los docentes le otorgarían más valor al trabajo diario. De esta manera también se podría atacar la falta de reconocimiento que las evaluaciones estandarizadas tienen con respecto a los diferentes tipos de aprendizaje. Es decir, a través de la evaluación interna, el alumno accedería a un tipo de evaluación más personalizado.

Como tantas otras cosas en la vida, la utilización de las evaluaciones estandarizadas en exceso, es peligroso. ¿Hasta qué punto es justo que se utilicen las evaluaciones estandarizadas como el único, o incluso el más importante, método para determinar posibilidades de los alumnos, en un contexto socio económico tan desigual? Incluso en muchas de las naciones industrializadas existen diferencias socio económicas (Rotberg, 2006); la brecha entre ricos y pobres no es exclusiva de los países pobres. Este tipo de

evaluación no hace demasiado reparo en esas diferencias, trata a todos los alumnos como iguales, sin embargo, no todos gozan de las mismas oportunidades (Linn, 1993). Una vez más, existen diferentes mecanismos a través de los cuales esta limitación de las evaluaciones estandarizadas puede ser minimizada. Pero por lo general, para que ello suceda, la respuesta está en utilizar mecanismos de evaluación que no se basen únicamente en un frío resultado.

## Bibliografía

- Aedo, C. y Larrañaga, O. (1994), Sistemas de entrega de los servicios sociales: La experiencia chilena, *www.iadb.org* consultado el 9 de febrero del 2008)
- Apple, M. (1996) *Cultural politics and education*. Buckingham, Open University Press.
- Armstrong, T. (1999). *Las inteligencias múltiples en el aula*, Manantial, Buenos Aires
- Barickman, J. (Oct., 1978). Evaluation Procedures: False and True, *The English Journal*, Vol. 67, No. 7. Págs. 24 - 29
- Beber, L. y Liston, D. (2002) *El currículo en conflicto. Perspectivas sociales, propuestas educativas y reforma escolar progresista*, Akal, Madrid
- Beck, R. (Mar., 1976). Equal Time: Standardized Test Results: Answering Our Critics--And Ourselves, *The English Journal*, Vol. 65, No. 3. Págs. 65 - 71
- Becker, W. (Spring, 1998). Standards and Testing: Another View, *The Journal of Economic Education*, Vol. 29, No. 2, Págs. 182-186
- Campbell, L. y Campbell, B. (1999). *Multiple intelligences and students achievement: Success stories from six schools*, ASCD, Virginia
- Capello, C. (2004) Blowing the whistle on the Texas miracle: an interview with Robert Kimball, Rethinking schools. Available online at: <http://www.rethinkingschools.org> (consultada el 10 de mayo del 2009).
- Causey Bush, T. (2005). Keep Your Eye on Texas and California: A Look at Testing, School Reform, No Child Left Behind, and Implications for Students of Color, *The Journal of Negro Education*, Vol. 74, No. 4, Págs. 332-343
- Cox, C. (1997). La Reforma de la Educación chilena, *PREAL* (consultado online el 10 de febrero del 2008)
- Cronbach, L. (1960). *Essentials of psychological testing*. New York: Harper
- Darling-Hammond, L. (Aug., 1994). National Standards and Assessments: Will They Improve Education? *American Journal of Education*, Vol. 102, No. 4, Educational Reform through National Standards and Assessment.
- Darling-Hammond, L., & Wise, A. E. (1989). Beyond standardization: State standards and school improvement. *Elementary School Journal*, 85, Págs. 315-336.
- Darling-Hammond, L. (2001). *El derecho de aprender*. Buenas escuelas para todos, Ariel, Barcelona
- Darling-Hammond, L. (2004) From 'Separate but Equal; to 'No Child Left behind': The Collision of New Standards and Old inequalities. En Meier, D. y Wood, G. (Eds.), *Many Children Left Behind*, Beacon Press, Boston.
- Dworkin, G. (2005). The No Child Left Behind Act: Accountability, High-Stakes Testing, and Roles for Sociologists, *Sociology of Education*, Vol. 78, No. 2, Págs. 170-174
- Elmore, R. F. (2003). A plea for strong practice. *Educational Leadership*, 61, Págs. 6-10.
- Eisner, E. (2001). *Procesos cognitivos y curriculum. Una base para decidir lo que hay que enseñar*, Martinez Roca, Barcelona
- Firestone, W. A., Bader, B. D., Massel, D., & Rosenblum, S. (1992). Recent trends in state educational reform: Assessment and Prospects. *Teachers College Record*, 94, 255-277.

- Fowler, F. C. (2004). *Policy studies for educational leaders: An introduction*. (2<sup>nd</sup> ed.) Columbus, OH: Merrill.
- Furst, E. (Oct., 1963). The Question of Abuses in the Use of Aptitude and Achievement Tests, *Theory into Practice*, Vol. 2, No. 4, Testing and the Schools.
- Gallart, Antonia. (2006). *La construcción de la escuela media*, Editorial Stella, Buenos Aires
- Gardner, E. (1953). Development and Applications of Tests of Educational Achievement in Schools and Colleges, *Review of Educational Research*, Vol. 23, No. 1, Educational and Psychological Testing.
- Gardner, H. (1998). A Reply to Perry D. Klein's "Multiplying the Problems of Intelligence by Eight", *Canadian Journal of Education / Revue canadienne de l'éducation*, Vol. 23, No. 1. Págs. 96 - 102
- Gardner, H. (Dec., 1995) Multiple Intelligences as a Catalyst, *The English Journal*, Vol. 84, No. 8, Multiple Intelligences. Págs. 16 -18
- Gardner, H. y Hatch, T. (1989). Multiple Intelligences Go to School: Educational Implications of the Theory of Multiple Intelligences, *Educational Researcher*, Vol. 18, No. 8. Págs. 4 - 9
- Gartner, A.; Geer, C. y Riessman, F. (comps.). (1999). *Nuevo ataque contra la igualdad de oportunidades*. Cociente intelectual y estratificación social, Morata, Madrid
- Goodson, I. (2000). *El cambio en el currículo*, Octaedro, Barcelona
- Earnest García, G. y Pearson, D. (1994) Assessment and Diversity, *Review of Research in Education*, Vol. 20. Págs. 337 - 391
- Hackney Gray, L. (2005). The 2004 Charles H. Thompson Lecture-Colloquium Presentation: "No Child Left Behind: Opportunities and Threats", *The Journal of Negro Education*, Vol. 74, No. 2, Págs. 95-111
- Haladyna, T.; Bobbit Nolen, S. y Haas, N. (Jun. - Jul., 1991). Rising Standardized Achievement Test Scores and the Origins of Test Score Pollution, *Educational Researcher*, Vol. 20, No. 5. Págs. 2 - 7
- Haney, C. y Hurtado, A. (Jun., 1994). The Jurisprudence of Race and Meritocracy: Standardized Testing and "Race-Neutral" Racism in the workplace, *Law and Human Behavior*, Vol. 18, No. 3 Págs. 223 - 248
- Hargreaves, A., Earl, L. y Schmidt, M (2002). Perspectives on Alternative Assessment Reform, *American Educational Research Journal*, Vol. 39, No.1 Págs. 69 - 95
- Hursh, D. (2005). The Growth of High-Stakes Testing in the USA: Accountability, Markets and the Decline in Educational Equality, *British Educational Research Journal*, Vol. 31, No. 5, Education Policy and Social Justice, Págs. 605-622
- Ingersoll, R. (2005). The Problem of Underqualified Teachers: A Sociological Perspective, *Sociology of Education*, Vol. 78, No. 2, Págs. 175-178
- Jacob, B. (Summer, 2001). Getting Tough? The Impact of High School Graduation Exams, *Educational Evaluation and Policy Analysis*, Vol. 23, No. 2. Págs. 99 – 102
- Karen, D. (2005). No Child Left Behind? Sociology Ignored! *Sociology of Education*, Vol. 78, No. 2, pp. 165-169
- Karp, S. (2001). Paige Leads Dubious Cast of Education Advisors, *Rethinking Schools*, Pág. 4
- Karp, S. (2004). NCLB's selective vision of equality: Some gaps count more than others. En D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (pp. 53-63). Boston: Beacon Press.

- Khon, A. (2004). NCLB and the Effort to Privatize Public Education. En D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (Págs. 79 - 97). Boston: Beacon Press
- Koretz, D. (Autumn 2002). Limitations in the Use of Achievement Tests as measures of Educators' Productivity, *The Journal of Human Resources*, Vol. 37, No. 4 Págs 1 - 42
- Laistch, D., Shaker, P. & Heilman, E. (2002) Teacher education, pro-market policy and advocacy research, *Teaching Education*, 13(3), 251-272.
- Langer, J., & Pradl, G. (1984). Standardized testing: A call for action. *Language Arts*, 61, 764-766.
- Lewin, K. (2001). *Financing Secondary Education in Developing Countries: Strategies for Sustainable Growth*, Paris: Institute for Educational Planning/UNESCO (consultado online el 10 de febrero del 2008)
- Lee, T. (2005). A Teacher like Me: Does Race, Ethnicity, or Gender Matter? *The American Economic Review*, Vol. 95, No. 2, Papers and Proceedings of the One Hundred Seventeenth Annual Meeting of the American Economic Association, Philadelphia, PA, January 7-9, 2005, Págs. 158-165
- Linn, R. (Spring, 1993). Educational Assessment: Expanded Expectations and Challenges, *Educational Evaluation and Policy Analysis*, Vol. 15, No. 1. Págs. 3 – 38.
- Linn, R.; Baker, E. y Dunbar, S. (Nov., 1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria, *Educational Researcher*, Vol. 20, No. 8. Págs. 1 – 16.
- Linn, R. L. (2003, September 1). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives*, 11(31). Consultado el 15 de mayo del 2009 de <http://epaa.asu.edu/epaa/v11n31/>.
- Lipman, P. (2004) *High-stakes education: inequality, globalization, and urban school reform* (New York, RoutledgeFalmer).
- Meier, D. (2004). NCLB and democracy. En D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (Págs. 66-78). Boston: Beacon Press.
- McNeil, L. (2000) *Contradictions of school reform: educational costs of standardized testing* (New York, Routledge). El artículo se encuentra publicado en [www.nochildleft.com](http://www.nochildleft.com), consultado el 15 de mayo del 2009).
- Muller, C. y Kerbow, D. (1993). Parent involvement in the home, school and community. En Schneider, B. y Coleman, J. (Eds.), *Parents, their children, and schools*. Boulder, CO: Westview Press.
- Murnane, R (2007). Improving the Education of Children Living in Poverty, *The Future of Children*, Vol. 17, No. 2, *The Next Generation of Antipoverty Policies*, Págs. 161-182.
- Murnane, R. y Steele, J. (2007). What Is the Problem? The Challenge of Providing Effective Teachers for All Children, *The Future of Children*, Vol. 17, No. 1, *Excellence in the Classroom*, Págs. 15-43.
- Narodowski, M. (1999). *Después de clase*, Ediciones Novedades Educativas, Buenos Aires.
- Narodowski, M. y Andrada, M. (2004). Cuasi monopolio estatal y descentralización educativa. Una exploración sobre América Latina” *Revista de Educación*, Nro. 333 (consultado online el 10 de febrero del 2008).

- Narodowski, M. y Nores, M. (2002). Socioeconomic segregation with (without) competitive education policies. A comparative analysis of Argentina and Chile, *Comparative Education*, 38(4) (consultado online el 12 de febrero del 2008).
- Nevo, D. (1997). *Evaluación basada en el centro*, Ediciones Mensajero, Bilbao
- New York State School Boards Association (2002) Title I accountability status updated for March 10, 2003. Disponible online at: <http://www.nyssba.org/adnews/misc/thenewaccountability-5.htm> (consultado el 5 de mayo el 2009).
- Newmann, F.; Marks, H. y Gamoran, A. (Aug., 1996). Authentic Pedagogy and Student Performance, *American Journal of Education*, Vol. 104, No. 4.
- Noble, K., Tottenham, N. y Casey, B. (Spring, 2005). Neuroscience Perspectives on Disparities in School Readiness and Cognitive Achievement, *The Future of Children* Vol. 15, No. 1 Págs. 71 – 89.
- Novak, J. y Fuller, B. (2003). Penalizing Diverse Schools? Similar Test Scores But Different Students Bring Federal Sanctions, Berkeley: Policy Analysis for California Education. En Meier, D. y Wood, G. (Eds.), *Many Children Left Behind*, Beacon Press, Boston. Págs. 3 – 32.
- Oxford, R. (1997). Constructivism: Shape-Shifting, Substance, and Teacher Education Applications, *Peabody Journal of Education*, Vol. 72, No. 1, Teachers and Teacher Education in the United States: Perspectives from Members of the Japanese-United States Teacher Education Consortium. Págs. 35 – 66.
- Paris, S. (1991). A Developmental Perspective on Standardized Achievement Testing, *Educational Researcher*, Vol. 20, No. 5. Págs. 19 – 26.
- Petrosko, J. (Mar. 1978). The Quality of Standardized High School Mathematics Tests, *Journal for Research in Mathematics Education*, Vol. No. 2. Págs. 137 - 148
- Pipho, C. (1985). Tracking the Reforms, *Education Week* 4(35):19.
- Popham, W. J. (2004). Ruminations regarding NCLBs most malignant provisions: Adequate yearly progress. Consultado el 12 de mayo del 2009 de [http://www.ctredpol.org/pubs/ Forum28July2004](http://www.ctredpol.org/pubs/Forum28July2004).
- Rotberg, I. (2006). Assessment Around The World, Educational Leadership, November 2006.
- Rothstein, R. (2008). *Grading education Getting accountability right*. Economic Policy Institute and Teachers College Press .
- Ruggieri, C. (Nov., 2002). Multigenre, Mutiple Intelligences and Trascendentalism, *The English Journal*, Vol 92, No. 2. Págs. 60 – 68.
- Schiller, K., Khmelkov, V. and Wang X. (Aug. 2002). Economic development and the Effects of Family Characteristics on Mathematics Achievement, *Journal of Marriage and the Family*, Vol. 64, No. 3. Págs. 730 – 742.
- Shepard, L. y Bliem, C. (Nov. 1995). Parents' Thinking about Standardized Tests and Performance Assessments, *Educational Researcher*, Vol. 24 No. 8. Págs. 25 - 32
- Sternberg, R. (Mar., 1996). Myths, Countermyths, and Truths about Intelligence, *Educational Researcher*, Vol. 25, No. 2. Págs. 11 – 16.
- Stufflebeam, D. y Webster:, W. (May - Jun., 1980). An Analysis of Alternative Approaches to Evaluation, *Educational Evaluation and Policy Analysis*, Vol. 2, No. 3. Págs. 5 - 20
- Suen, H. (2006). Psychometric paradox of very high-stakes assessment and solutions, Korean Educational Development Institute, Págs. 113-129.

- Terigi, Flavia: Sobre el aprendizaje y las condiciones para el aprendizaje. En AAVV (2005): *Desigualdad, fragmentación social y educación*. Buenos Aires, Instituto Internacional de Planificación de la Educación.
- US Department of Education, Office of Elementary and Secondary Education (2002) No Child Left Behind: a desk reference (Washington, DC, US Department of Education).
- US Department of Education, Office of the Secretary (2002b) What to know and where to go: a parents' guide to No Child Left Behind (Washington, DC, US Department of Education).
- Wexler-Sherman, C.; Gardner, H. y Feldman, D. (Winter, 1988). A Pluralistic View of Early Assessment: The Project Spectrum Approach, *Theory into Practice*, Vol. 27, No. 1. Págs. 77 – 83.
- Winerip, M. (2003) The 'zero dropout' miracle: Alas! Alack! A Texas tall tale, New York Times, B7.
- Wood, G. (2004). A view from the field: NCLB's effects on classrooms and schools. En D. Meier & G. Wood (Eds.), *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools* (Págs. 33-50). Boston: Beacon Press.

**Acerca del autor: Ignacio Barrenechea** es docente de la Escuela Escocesa San Andrés, en el Departamento de Ciencias Sociales. Su título de grado es de Abogado, de la Universidad Torcuato Di Tella. A su vez, es becario de la Comisión Fulbright y, actualmente, cursa sus estudios de posgrado en George Washington University, en los Estados Unidos. Sus investigaciones se enfocan en el área de políticas educativas, en especial en lo que se refiere a la evaluación en la escuela primaria y secundaria.

**Afiliación:** George Washington University (GWU).

**Email:** [ibarrenechea@sanandres.esc.edu.ar](mailto:ibarrenechea@sanandres.esc.edu.ar)

---

## archivos analíticos de políticas educativas

Volumen 18 Número 8

10 de abril, 2010

ISSN 1068-2341

---

Los/as lectores/as pueden copiar, mostrar, y distribuir este artículo, siempre y cuando se de crédito y atribución al autor/es y a Archivos Analíticos de Políticas Educativas, se distribuya con propósitos no-comerciales, no se altere o transforme el trabajo original. Más detalles de la licencia de Creative Commons se encuentran en <http://creativecommons.org/licenses/by-nc-sa/3.0> Cualquier otro uso debe ser aprobado en conjunto por el autor/es, o AAPE/EPAA.

AAPE/EPAA es publicada conjuntamente por el *Mary Lou Fulton Institute and Graduate School of Education, Arizona State University*. Los artículos que aparecen en AAPE son indexados en EBSCO Education Research Complete, [Directory of Open Access Journals](#), ERIC, H.W. WILSON & Co, QUALIS – A 2 (CAPES, Brazil), SCOPUS, Socolar-China.

Contribuya con comentarios y sugerencias en <http://epaa.info/wordpress/>

Por errores y sugerencias contacte a [Fischman@asu.edu](mailto:Fischman@asu.edu).

---

archivos analíticos de políticas educativas  
consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

**Armando Alcántara Santuario** Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

**Claudio Almonacid** Universidad Metropolitana de Ciencias de la Educación, Chile

**Pilar Arnaiz Sánchez** Universidad de Murcia, España

**Xavier Besalú Costa** Universitat de Girona, España

**Jose Joaquín Brunner** Universidad Diego Portales, Chile

**Damián Canales Sánchez** Instituto Nacional para la Evaluación de la Educación, México

**María Caridad García** Universidad Católica del Norte, Chile

**Raimundo Cuesta Fernández** IES Fray Luis de León, España

**Marco Antonio Delgado Fuentes** Universidad Iberoamericana, México

**Inés Dussel** FLACSO, Argentina

**Rafael Feito Alonso** Universidad Complutense de Madrid, España

**Pedro Flores Crespo** Universidad Iberoamericana, México

**Verónica García Martínez** Universidad Juárez Autónoma de Tabasco, México

**Francisco F. García Pérez** Universidad de Sevilla, España

**Edna Luna Serrano** Universidad Autónoma de Baja California, México

**Alma Maldonado** Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

**Alejandro Márquez Jiménez** Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

**José Felipe Martínez Fernández** University of California Los Angeles, USA

**Fanni Muñoz** Pontificia Universidad Católica de Perú

**Imanol Ordorika** Instituto de Investigaciones Económicas – UNAM, México

**Maria Cristina Parra Sandoval** Universidad de Zulia, Venezuela

**Miguel A. Pereyra** Universidad de Granada, España

**Monica Pini** Universidad Nacional de San Martín, Argentina

**Paula Razquin** UNESCO, Francia

**Ignacio Rivas Flores** Universidad de Málaga, España

**Daniel Schugurensky** Universidad de Toronto-Ontario Institute of Studies in Education, Canadá

**Orlando Pulido Chaves** Universidad Pedagógica Nacional, Colombia

**José Gregorio Rodríguez** Universidad Nacional de Colombia

**Miriam Rodríguez Vargas** Universidad Autónoma de Tamaulipas, México

**Mario Rueda Beltrán** Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

**José Luis San Fabián Maroto** Universidad de Oviedo, España

**Yengny Marisol Silva Laya** Universidad Iberoamericana, México

**Aida Terrón Bañuelos** Universidad de Oviedo, España

**Jurjo Torres Santomé** Universidad de la Coruña, España

**Antoni Verger Planells** University of Amsterdam, Holanda

**Mario Yapu** Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas  
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)  
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**  
(Universidade Federal do Rio Grande do Sul)

- |  |   |
|--|---|
| <b>Dalila Andrade de Oliveira</b> Universidade Federal de Minas Gerais, Brasil             | <b>Jefferson Mainardes</b> Universidade Estadual de Ponta Grossa, Brasil          |
| <b>Paulo Carrano</b> Universidade Federal Fluminense, Brasil                               | <b>Luciano Mendes de Faria Filho</b> Universidade Federal de Minas Gerais, Brasil |
| <b>Alicia Maria Catalano de Bonamino</b> Pontifícia Universidade Católica-Rio, Brasil      | <b>Lia Raquel Moreira Oliveira</b> Universidade do Minho, Portugal                |
| <b>Fabiana de Amorim Marcello</b> Universidade Luterana do Brasil, Canoas, Brasil          | <b>Belmira Oliveira Bueno</b> Universidade de São Paulo, Brasil                   |
| <b>Alexandre Fernandez Vaz</b> Universidade Federal de Santa Catarina, Brasil              | <b>Antônio Teodoro</b> Universidade Lusófona, Portugal                            |
| <b>Gaudêncio Frigotto</b> Universidade do Estado do Rio de Janeiro, Brasil                 | <b>Pia L. Wong</b> California State University Sacramento, U.S.A                  |
| <b>Alfredo M Gomes</b> Universidade Federal de Pernambuco, Brasil                          | <b>Sandra Regina Sales</b> Universidade Federal Rural do Rio de Janeiro, Brasil   |
| <b>Petronilha Beatriz Gonçalves e Silva</b> Universidade Federal de São Carlos, Brasil     | <b>Elba Siqueira Sá Barreto</b> <a href="#">Fundação Carlos Chagas</a> , Brasil   |
| <b>Nadja Herman</b> Pontifícia Universidade Católica – Rio Grande do Sul, Brasil           | <b>Manuela Terrasêca</b> Universidade do Porto, Portugal                          |
| <b>José Machado Pais</b> Instituto de Ciências Sociais da Universidade de Lisboa, Portugal | <b>Robert Verhine</b> Universidade Federal da Bahia, Brasil                       |
| <b>Wenceslao Machado de Oliveira Jr.</b> Universidade Estadual de Campinas, Brasil         | <b>Antônio A. S. Zuin</b> Universidade Federal de São Carlos, Brasil              |