

¿RESPUESTA SELECCIONADA O RESPUESTA CONSTRUIDA?: EL DILEMA ENTRE LA OBJETIVIDAD Y LA VALIDEZ EN LA CONSTRUCCIÓN DE PRUEBAS

Raúl Marín Conus¹

Fernando Soto Soto²

Macarena Casanueva Sanhueza³

Universidad Católica de la Santísima Concepción

Resumen

El estudio intenta poner de manifiesto las diferencias que se observarían en el rendimiento escolar de los alumnos al aplicar dos pruebas, referidas a los mismos objetivos, pero que difieren en la forma de responder: Selección de una respuesta correcta o construcción de la respuesta. Se trabaja con una muestra de cuartos años básicos de la Intercomuna de Concepción y con las asignaturas de Lenguaje y Comunicación y Matemáticas. Se toma como base una prueba SIMCE aplicada en años anteriores. Los resultados indican que los alumnos obtienen más bajos puntajes en las pruebas de respuesta construida aún cuando ambas pruebas están referidas a los mismos objetivos. Esto podría considerarse una señal en el sentido que la selección de respuesta obedece a un tipo de proceso psicológico (reconocimiento) en tanto que las de respuesta construida estarían midiendo el recuerdo como proceso psicológico, el cual resultaría más difícil para los alumnos, lo cual concuerda con la teoría psicológica al respecto. Esto puede ser importante a la hora de tomar decisiones basadas en los resultados del SIMCE. Los resultados muestran además que los colegios varían en cuanto a los resultados dependiendo de la dependencia del colegio.

Palabras claves: Evaluación, Educación, Rendimiento, Pruebas, Formas de responder.

Abstract

This study reveals the observable differences in academic performances of students when two different tests are given, which aim at the same objectives but differ in the way they are answered: Multiple Choice or Writing an answer. It is carried out with Fourth Class Primary school samples of Concepción town in Language and Communication and Mathematics areas. The SIMCE test, tried out the previous years, is used as a basis for it. The outcomes of this research indicate that students get lower scores in the test in which they have to write an answer than in the multiple choice ones, even though both tests have got similar target aims. This could be

¹ Profesor de Educación Básica, Magíster en Ciencias de la Educación.

² Profesor de Ciencias Naturales, Licenciado en Educación.

³ Educadora de Párvulos, Magíster en Ciencias de la Educación ©.

⁴ Este trabajo es el resultado de una investigación financiada por la Dirección de Investigación de la Universidad Católica de la Santísima Concepción. (Proyecto DIN 301-I.2001.)

considered as a significant result in the sense that Multiple choice tests are easier than writing an answer ones because both appeal to different psychological strategies. That is, the former leads to the recognition strategy whereas the latter measures the remembering psychological strategy, which match with the corresponding psychological theory. These outcomes can be quite meaningful to make decisions based on the SIMCE results.

Key words: Evaluation, Education, Yield, Test-ways to respond.

Introducción

Recientes estudios internacionales (TIMMS OCDE) han puesto en tela de juicio la calidad de la Educación chilena. Estudios nacionales, por su parte, basados en el SIMCE dan cuenta acerca de los niveles de logro en dichas pruebas son también preocupantes. Sin embargo, la realidad demostrada nacionalmente parece ser diferente de la que muestran los estudios internacionales, diferencia que sería de utilidad explicar de alguna manera. En este estudio se intenta verificar la capacidad de los instrumentos que se utilizan en el SIMCE para dar cuenta de la calidad de la Educación, sobre todo en la concepción de calidad en que el Ministerio se sitúa.

El estudio se centra en comparar dos pruebas para verificar el grado en que ambas dan cuenta de similares resultados. Por un lado, se aplica una prueba SIMCE (4º año, Lenguaje y Matemáticas) y por otra una versión diferente de la misma prueba en la cual se modifican solamente dos aspectos: Se solicita que el alumno construya una respuesta en vez de seleccionarla como lo hace el SIMCE y se demanda a lo menos tres de cuatro ejercicios de respuesta construida correctos para dar por logrado el objetivo, en tanto que el SIMCE sólo demanda un reconocimiento. Se aplican los resultados a una muestra y se comparan los resultados de ambas pruebas.

El problema

El Ministerio de Educación implementa un programa denominado SISTEMA DE MEDICION DE LA CALIDAD DE LA EDUCACION, cuyo objetivo es realizar mediciones periódicas respecto del grado en que el sistema educativo nacional consigue los logros educativos que se propone. Esta información sirve de base para orientar políticas públicas de focalización de recursos e incentivos, entre otros propósitos. En 1999 se aplicó la Prueba correspondiente a los cuartos años. Esta prueba introduce dos modificaciones importantes respecto de las mediciones de los años anteriores:

- a) Los objetivos medidos corresponden esta vez a los objetivos educativos que incorpora la Reforma Educacional y,
- b) Se introducen preguntas de respuesta elaborada, aunque en una baja proporción. Las pruebas miden contenidos referidos a las asignaturas de Lenguaje y Comunicación, Educación Matemáticas, y Comprensión del Medio Natural y Social.

Los resultados alcanzados se han prestado para cuestionar la eficacia de la reforma educacional en cuanto a ser un conjunto de estrategias para mejorar la calidad de la Educación. En este contexto, lo que se quiere investigar es si las Pruebas SIMCE, de acuerdo con la tecnología bajo la cual se construyen, son apropiadas para medir CALIDAD de la Educación entendida ésta como la eficacia de la Escuela para el logro de los objetivos educativos propuestos.

Diversos autores han cuestionado la capacidad de las pruebas elaboradas bajo una tecnología sicométrica para medir logro de objetivos, toda vez que su diseño obedece al propósito central de establecer diferencias entre los sujetos (diferencias interindividuales respecto de resultados alcanzados), y por lo tanto su construcción entrega información normativa respecto de cómo se distribuyen los alumnos respecto de la característica medida en torno a un promedio, pero carecen de la capacidad para captar las diferencias intra-individuales, que son las diferencias a las cuales debería apuntar un proceso educativo que persigue lograr ciertos objetivos en los alumnos.

La comprobación de la capacidad real de las Pruebas para medir de manera efectiva la calidad de la educación es una cuestión de máxima relevancia en el contexto actual, toda vez que se está cuestionando todo un proceso de reforma educativa a partir de ciertos resultados, y por otro lado, las políticas de incentivos que utiliza el gobierno se basan en la información que de allí proviene para establecer los colegios de bueno, regular o mal rendimiento.

Así planteado, esta investigación puede contribuir a hacer claridad respecto del concepto de Evaluación Educativa y la naturaleza de las pruebas que deben emplearse para medir logro de objetivos. Asimismo los resultados pueden contribuir, en el contexto nacional, a clarificar el concepto de calidad de la educación y de los modelos y tecnología apropiadas para medirla.

En síntesis lo que esta investigación se propone es llegar a establecer:

- a) Si el formato que adoptan las pruebas SIMCE es el más adecuado para medir objetivos educativos que son producto de la acción de la escuela y, en consecuencia,
- b) Si un formato diferente de prueba, proporcionaría información diferente, suponiendo que se miden los mismos objetivos, y que esta información reflejaría en mejor medida los resultados de la acción de la Escuela.

El problema se analiza en función del tipo de dependencia del colegio y se centra fundamentalmente en el análisis del comportamiento de las pruebas.

Antecedentes teóricos

En general los estudios sobre la calidad de la educación han evolucionado desde una comparación entre colegios en base a resultados con pruebas de naturaleza sicométrica a modelos más recientes que incorporan la noción de valor agregado como expresión de la eficacia de los colegios y centros educativos. A ello se agrega

el uso de pruebas que incorporan planteamientos tomados de la teoría evaluativa edumétrica.

Los últimos 20 años han sido testigos de un sostenido avance de los estudios referidos a la calidad de la educación. Toranzos (1999) analiza las razones que han llevado a la preocupación por la calidad de la educación y muestra como esta preocupación ha pasado por la utilización de distintos enfoques, para llegar, en la actualidad, a relacionar la calidad con un mejoramiento en la gestión educativa. Los trabajos que se discuten más adelante tienen que ver con las orientaciones específicas que asumen los estudios en algunas partes importantes del mundo occidental. Marchesi (1999), hace una revisión de las orientaciones teóricas que ha asumido la evaluación de los centros escolares en España durante los últimos años. Analiza las funciones que cumple la evaluación en este contexto y el papel complementario que tienen las evaluaciones internas y externas. Termina analizando los modelos aplicados a la evaluación, los cuáles clasifica en enfoques unidimensionales, cualitativa de niveles múltiples y cuantitativa de niveles múltiples. La evaluación cualitativa la reconoce adecuada como para una evaluación interna y la cuantitativa como un modelo apropiado para evaluar la eficacia de los centros educativos, proponiendo varios niveles de indicadores. Uno de los supuestos en que hace descansar este modelo consiste en que se necesita efectuar análisis a distintos niveles, de manera que se pueda ofrecer a cada centro el valor añadido a partir de la aplicación de pruebas comunes iniciales y finales, teniendo en cuenta las condiciones de cada centro y sus procesos educativos. Este concepto, el de valor añadido en Educación, es analizado por Munició (1999), quien lo sitúa originariamente en la industria y en el comercio, entendido como la diferencia de estado entre los momentos de iniciar y finalizar un proceso. Sitúa el concepto en el paradigma de la calidad. Sin embargo, aplicado a la educación implica que el alumno ha sido *transformado* o mejorado por la escuela y en consecuencia le ha dado mayor valor. Esto es lo que entiende por valor agregado o añadido. El autor analiza las distintas maneras que se han utilizado para medir valor añadido o agregado, todas las cuales implican una medición inicial y una medición final.

Thomas, (1998) por su parte, describe algunas experiencias realizadas en Inglaterra respecto de medir el valor añadido como estrategia para establecer la eficacia de la Escuela. Llama la atención, en primer lugar, sobre lo peligroso que resulta el empleo de un modelo de resultados para establecer la eficacia de la escuela, ya que este resultado puede ser influido por múltiples factores, entre los cuales los aprendizajes iniciales del alumno juegan un papel importante. Define valor añadido como la medida del progreso medio de los alumnos a lo largo de un periodo de tiempo. Esta definición implica que se debe medir en dos momentos al alumno a objeto de poder determinar su progreso. Plantea los esfuerzos que se hacen en el Reino Unido para aceptar un sistema de valor añadido. Respecto de las técnicas estadísticas y metodologías que se han utilizado para medir el valor agregado, plantea como técnicas apropiadas el análisis de regresión múltiple y la modelización a múltiples niveles. Thomas (1996) condujo una investigación de este tipo que abarcó 402 Centros Docentes de nivel secundario. Los resultados de comparar resultados brutos y valor añadido en la clasificación de las escuelas indicaron que 6 centros fueron clasificados en el 25 % superior en función de los resultados brutos. Sin embargo, en los resultados de valor añadido esas mismas 6 Escuelas fueron clasificadas en el 25 % inferior.

En una línea similar se inscribe el trabajo de Strand (1998), si bien éste enfatiza la comparación entre la clasificación de las escuelas en base a la simple medición de resultados y la clasificación de las mismas en un sistema de valor agregado, indicando que en las primeras se obtienen visiones engañosas respecto de la real influencia de la Escuela en los aprendizajes que los alumnos demuestran en pruebas finales cuyos resultados no han sido ajustados por otras variables como nivel inicial de aprendizaje y nivel socioeconómico. El autor destaca el uso de los modelos estadísticos de análisis de multinivel como técnica apropiada para comparar establecimientos en una evaluación sobre la base de valor agregado.

Saunders (1999) hace un análisis de las principales experiencias realizadas en Inglaterra en el marco de la Evaluación de los centros educativos con el enfoque del valor agregado. Examina tres momentos distintos que van desde la investigación sobre las escuelas eficaces, pasa por un análisis de los principales puntos de discusión en torno al concepto de valor añadido o agregado y describe algunos de los trabajos realizados por la National Fundación for Educational Research (NFER). El artículo es especialmente importante en tanto ayuda a clarificar los factores que es necesario controlar para establecer el valor añadido por la escuela. Estos se refieren al contexto sociocultural y al nivel inicial de aprendizaje. Norris (1998) contribuye al tema analizando la forma en que se ha introducido en educación el concepto de indicadores de rendimiento y el incremento que su utilización ha tenido en los últimos 10 años. Sitúa el origen del concepto en la industria y en la empresa. La utilización de los indicadores de rendimiento parte del supuesto que el fin de las escuelas es producir cambios en los alumnos y que estos cambios pueden ser medidos. La eficiencia de un sistema, según el autor, se mide por los cambios que experimenta el que aprende y que estos cambios pueden ser medidos por test o pruebas adecuadas.

Entre las experiencias internacionales se puede citar el TIMSS (Tercer Estudio Internacional de Matemáticas y Ciencias). Esta experiencia tiene como objetivo medir, explicar y comparar el aprendizaje en Ciencias y Matemáticas en 41 países. Se basa fundamentalmente en un modelo de resultados, cuyo objetivo principal es proporcionar elementos para que los países examinen si existe correspondencia entre los currículos, los textos escolares y las experiencias de aprendizajes en las salas de clases. Otra experiencia similar es la que desarrolla PISA (Escudero, 2000) PROYECTO INTERNACIONAL DE INDICADORES DE RENDIMIENTO DE LOS ALUMNOS. Esta experiencia es desarrollada por la Organización para la Cooperación y Desarrollo Económico (OCDE). Se trata de un proyecto experimental dirigido a establecer un sistema internacional de indicadores de la situación de la educación. Aún cuando el Proyecto apunta a cuatro diferentes tipos de productos, uno de ellos está relacionado con el rendimiento académico (en disciplinas de Matemáticas, lenguaje y Ciencias).

La bibliografía comentada permite establecer de que manera el concepto de valor agregado se ha ido desarrollando en las últimas décadas y como estas experiencias han adquirido un mayor desarrollo en Europa, en países como Inglaterra, Francia, España y Estados Unidos. Inicialmente el concepto de valor añadido es tomado de la industria y la empresa y adaptado a la realidad educacional, concibiendo la acción de la escuela como orientada a introducir cambios en los alumnos, cambios que se

pueden identificar, cuantificar y, en consecuencia, medir. Esta concepción de educación implica que los resultados de la acción educativa deben ser traducidos en metas cuantificables.

El desarrollo de las evaluaciones educacionales en América Latina es analizado por Wolf (1998). El autor sitúa el desarrollo de esta preocupación como resultado directo de la preocupación mundial al respecto, citando experiencias como el TIMSS e indicando sus principales resultados. Analiza asimismo otras experiencias norteamericanas y francesas, en donde también se han empleado técnicas de valor agregado y se detiene en lo que ocurre en América Latina, citando las experiencias chilenas y de otros cinco países. Comenta, asimismo, el programa de Evaluación Regional de la UNESCO/OREALC. El autor finaliza señalando las lecciones aprendidas en los países latinoamericanos y los desafíos que deben enfrentar en un futuro. Llama la atención que en ninguna de las experiencias latinoamericanas se esté utilizando el concepto de valor agregado, y que las pruebas que se utilizan están centradas en los ítem de alternativas múltiples, situación que está cambiando tanto en Estados Unidos como en Europa.

En Latinoamérica, en el contexto de la recomendación para la ejecución del Proyecto Principal para América Latina y el Caribe, para el período 1993-1996, se crea el Laboratorio de Evaluación de Calidad de la Educación, constituyendo un proyecto de cooperación regional. (Casassus, 1998) La orientación conceptual del laboratorio es la comprensión del significado de la calidad en estrecha relación con el nivel de logro de los objetivos educacionales, en el marco de los programas oficiales de estudio tomando en cuenta las variables de insumo y especialmente las de proceso.

Diversos países latinoamericanos han desarrollado Servicios de Medición de la Calidad de la Educación. Por ejemplo Argentina creó el Sistema Nacional de Evaluación de la Calidad (SINEC), cuyo objetivo es evaluar el funcionamiento del sistema educativo en todas las jurisdicciones, niveles, ciclos y regímenes a partir del diseño de sistema de evaluación y control periódico de la calidad. (Toranso, 1999). En nuestro país funciona un Sistema de Medición de la Calidad de la Educación desde 1988 y su objetivo es diseñar y administrar un sistema de evaluación periódica del cumplimiento de los Objetivos Fundamentales y Contenidos Mínimos, tanto en la Enseñanza Básica como en la Media (MINEDUC Chile).

Estos estudios, se basan fundamentalmente en las Pruebas SIMCE, que son pruebas que se aplican a los alumnos de Cuarto y Octavo Año Básico y Segundo Año Medio. En general, aunque no exclusivamente estas pruebas miden el rendimiento de los alumnos en las asignaturas de Castellano y Matemáticas y son pruebas que se construyen bajo una concepción sicométrica de la medición.

En estas condiciones se considera la calidad de la educación en función de los puntajes que los alumnos obtienen en dichas pruebas. Son considerados colegios de mayor calidad los que alcanzan altos puntajes, y colegios de menor calidad cuando sus alumnos logran bajos puntajes. Los responsables de las adopciones de política y los investigadores han puesto de relieve lo peligroso que es utilizar tablas de clasificación basadas exclusivamente en los resultados brutos ya que las escuelas que

tienen un elevado nivel inicial de conocimientos tenderán a obtener mejores resultados tan sólo por este motivo.

Uno de los problemas más serios que se enfrenta al momento de evaluar la calidad de la educación es la definición restrictiva del PRODUCTO ESCOLAR. Normalmente se ha tenido en cuenta sólo el rendimiento del alumno en test referidos a NORMA. No se encuentran estudios que consideren otros aspectos, tales como satisfacción de los padres y estudiantes, cumplimiento de objetivos, adquisición de valores, etc. (Fernández 1997).

En general, las pruebas que se utilizan en los estudios sobre calidad de la educación son construidas en base a una tecnología sicométrica y en consecuencia, son instrumentos útiles para aportar información acerca de las diferencias interindividuales. Sin embargo no aportan al conocimiento de la variación intraindividual, variación que se supone es el efecto principal de la acción del colegio y sobre cuya base se debería hablar de calidad y eficacia. Toda la tecnología puesta en práctica en la elaboración de estos instrumentos tiende a producir una dispersión de puntajes que permita establecer diferencias entre los alumnos antes que a detectar avances en los procesos de aprendizaje previstos con deliberación. Por otra parte las variaciones interindividuales pueden ser consecuencias de muchos factores, algunos de los cuales no están relacionados con la calidad de la educación de un colegio.

Frente a este modelo de construcción de pruebas se presenta el modelo edumétrico, cuya característica más importante es justamente el desarrollo de una tecnología que permite medir efectivamente logros educativos (ganancias intraindividuales) (Popham 1989). Este modelo también ha sido denominado como instrumentos referidos a criterios (Casassus y otros, 1998). En este enfoque los instrumentos se diseñan para verificar si un individuo ha logrado un nivel predeterminado en una variable dada. La construcción de tales instrumentos requiere de una descripción precisa de qué es lo que se quiere medir y luego elaborar el diseño de situaciones de prueba que permitan que el alumno pueda demostrar que ha logrado el objetivo que se ha descrito.

Los indicadores de calidad de la educación sirven de base a las políticas de incentivos que utilizan los gobiernos. Algunas de éstas tienen que ver con la asignación de fondos, estímulos a la excelencia académica y otros beneficios de naturaleza similar. En tal sentido se hace necesario contar con información válida y confiable acerca de la calidad de la educación que imparten los colegios, para asegurar una correcta política de incentivos. Dicha información debe provenir de instrumentos que efectivamente sean capaces de poner en evidencia los progresos que los alumnos experimentan en la consecución de los objetivos educacionales propuestos y que son resultados específicos de la acción del colegio propiamente tal. (valor agregado por el colegio).

Específicamente, lo que se intenta probar es que pruebas de naturaleza edumétrica deberían dar resultados diferentes a pruebas sicométricas que midan los mismos objetivos cuando son aplicadas a un mismo grupo de alumnos y que dadas algunas características de las pruebas sicométricas (selección de una entre varias alternativas), los resultados deberían ser más altos que en las pruebas edumétricas.

Esta hipótesis se fundamenta en el supuesto de que una prueba de selección de alternativas es una prueba que implica como proceso psicológico el reconocimiento, en tanto que pruebas que demanden que el alumno elabore o construya la respuesta implica, como proceso psicológico, el recuerdo, proceso este último que es reconocido como de mayor complejidad por los psicólogos.

Por otra parte, el hecho de que se tomen decisiones acerca de logro o no logro de objetivos a partir de una actuación única, implica decisiones con alta probabilidad de error, probabilidad que disminuye, de acuerdo a principios estadísticos, en la medida en que la decisión se toma sobre la base de un conjunto mayor de observaciones.

Para los efectos de este estudio se consideró como prueba sicométrica las pruebas elaboradas y aplicadas como parte del SIMCE y como prueba edumétrica una prueba que reúne dos características que la diferencian de la anterior:

- a) El alumno debe elaborar la respuesta (lo que implica como proceso psicológico el recuerdo).
- b) Se demandan a los menos tres de cuatro o cuatro de cinco ejercicios correctos para dar por logrado el objetivo (lo que disminuye a un 20 o 25 % la probabilidad de error en la decisión).

Metodología

Se trabajó sobre la base de un universo constituido por las Escuelas Básicas de la Intercomuna Concepción-Talcahuano. De este universo se seleccionó, en una primera etapa, una muestra intencionada (se utilizó como criterio la facilidad de acceso a los investigadores y la aceptación voluntaria, por parte de los colegios, a la experiencia), formada por dos establecimientos de cada tipo de dependencia (municipal-subvencionado y particular pagado). Durante el proceso de confirmación de la muestra se agregó voluntariamente un colegio particular subvencionado, por lo cual la muestra final quedó constituida por siete establecimientos según la siguiente distribución:

- Dos establecimientos municipalizados.
- Tres establecimientos particulares subvencionados.
- Dos establecimientos particulares pagados.

Tamaño de muestra por colegio

Colegio	Alumnos
1	17
2	36
3	31
4	22
5	22
6	30
7	27
Total	185

Todos los establecimientos seleccionados pertenecían a la Intercomuna de Concepción—Talcahuano (salvo uno de la Comuna de Coronel que solicitó incorporarse voluntariamente).

El procedimiento seguido para la construcción de las pruebas implicó los siguientes pasos:

Se seleccionó una prueba SIMCE en Castellano y otra en Matemáticas para Cuarto Año Básico, de las que fueron aplicadas el año 1999. Por cada ítem se identificó el objetivo que se medía y sobre la base de dichos datos se elaboró la tabla de especificaciones correspondiente (Objetivos-items). Cada prueba fue dividida en dos mitades tratando de que cada mitad fuera equivalente a la otra desde el punto de vista de los objetivos medidos, y que en cada una de ellas se reflejara la tabla de especificaciones elaborada.

La partición fue sometida al criterio de jueces para obtener un juicio externo, y se ajustó hasta lograr consenso entre los jueces.

Se seleccionó una de las dos mitades y para ello se elaboró un reactivo de prueba alternativo para cada uno de los ítem que contenía. Los reactivos construidos debían responder a los siguientes criterios.

- Debían medir el mismo objetivo que el ítem de alternativa que se reemplazaría.
- Deberían exigir una respuesta construida (en oposición a la seleccionada de la prueba que se estudiaba).
- Debían plantearse tres o cuatro ejercicios del mismo tipo en cada reactivo.
- Cada reactivo se daría por aprobado cuando dos de los tres o tres de los cuatro ejercicios, a lo menos, estuvieran correctos.

La versión construida fue sometida a consulta de jueces, en el sentido de que el ejercicio elaborado midiera el mismo objetivo que la prueba de alternativas, y con el mismo nivel de complejidad. Los jueces fueron tres especialistas en cada disciplina, los cuales eran docentes de la Facultad en las especialidades medidas (Lenguaje y Matemáticas). Los reactivos construidos fueron ajustados de acuerdo con los elementos de juicio que los jueces aportaron. En algunos casos este ajuste se hizo en presencia de los mismos jueces. De esta manera se consiguieron dos pruebas que, supuestamente, medían los mismos objetivos, pero que variaban en cuanto a la forma de responder y a la cantidad de ejercicios requeridos para dar por logrado el objetivo.

Procedimientos de administración. La prueba fue administrada a todos los alumnos del Cuarto Año seleccionado que estuvieran asistiendo a clases el día de la administración. El día y hora de administración fue acordado con los respectivos Directores y/o Jefes Técnicos. La actividad de aplicación de la prueba en el colegio fue presentada como formando parte de un trabajo práctico asignado a alumnos de la carrera de Pedagogía en Educación General Básica que se encontraban realizando su práctica. Cada prueba, en sus dos versiones, fue administradas en un mismo día para asegurar que ambas fueran respondidas por los mismos alumnos. En caso de inasistencia a una de las dos pruebas, se eliminaba al alumno también de la otra, de

manera de disponer de la misma cantidad de alumnos en ambas pruebas. La prueba de alternativas tenía una duración de 45 minutos y la de respuesta construida de 60 minutos, con 20 minutos de descanso entre ambas.

Procedimientos de asignación de puntajes. Para cada prueba (de alternativa o de respuesta construida) se elaboró una pauta de respuestas correctas, la cual fue consensuada por criterio de jueces. En la prueba de alternativas se asignó un punto por cada ítem que tuviera marcada la respuesta correcta. En cambio, en la prueba de alternativas se asignó un punto por cada reactivo que cumpliera con el nivel de exigencia preestablecido: 75 o 80 % de respuestas correctas, dependiendo si el reactivo tenía cuatro o cinco ejercicios. Luego se contabilizaron los puntos obteniendo así el puntaje correspondiente a cada alumno.

Instrucciones de administración de las pruebas.

La administración de las pruebas se hizo conforme al Manual de Administración de las pruebas SIMCE, variando las instrucciones en lo relativo a la forma de responder, para el caso de la prueba de respuesta construida.

Resultados

El estudio se proponía verificar si las pruebas SIMCE darían los mismos resultados que una prueba con características edumétricas. Para este efecto se consideró como prueba edumétrica una prueba que reuniera tres características principales:

- a) Que midiera lo que el objetivo proponía como conducta a lograr.
- b) En ella el alumno debía construir la respuesta (en oposición a seleccionarla)
- c) Y donde el alumno debía responder a lo menos un 75 % o el 80 % de los ejercicios de manera correcta para dar por logrado el objetivo. Sólo en este caso se asignaba un punto al alumno.

Los resultados arrojados por esta prueba fueron comparados con una prueba SIMCE ya aplicada (pero no a los alumnos con quienes se hace el ensayo). La prueba SIMCE aplicada es puntuada de acuerdo con el procedimiento utilizado por el SIMCE, es decir una respuesta correcta igual un punto sin corrección por adivinación. La prueba SIMCE contenía aproximadamente un 50 % de ítem referido a objetivos propios de Segundo Año Básico y el otro 50 % correspondiente al Cuarto Año Básico propiamente tal, lo cual se igualó en la prueba edumétrica elaborada. Los resultados indican que:

Las pruebas son equivalentes entre sí, es decir, están midiendo los mismos objetivos, lo cual es demostrado por la alta correlación entre los puntajes de ambas pruebas. La correlación obtenida por el coeficiente de Spearman alcanza a 0.85.

Los puntajes que alcanzan los alumnos en la versión alternativa (respuesta construida), son notoriamente inferiores y las diferencias son estadísticamente significativas, al 0,01 nivel de significación. Estas diferencias son más acentuadas en la parte de la prueba en que se miden objetivos de Cuarto Año Básico.

Cuando se comparan establecimientos por distinto nivel de dependencia, éstos confirman los resultados tradicionales del SIMCE, es decir, los colegios particulares pagados son los que alcanzan más altos rendimientos, seguidos por los subvencionados y en tercer lugar los municipalizados, diferencias que también son significativas al 0,01 nivel de significación (Prueba T de Student). Las diferencias se observan por igual en Lenguaje y Comunicación como en Matemáticas.

Ambas mediciones están correlacionadas, pero esta correlación es menor en la parte de la prueba que mide objetivos correspondientes a Cuarto Año Básico. (0.85 de Spearman)

La siguiente secuencia de gráficos muestra la forma en que se distribuyen los porcentajes de respuestas correctas en las pruebas SIMCE y construida por cada ítem de las pruebas

SECCION A. Prueba de matemáticas

Gráfico1. Comparación de porcentaje de respuestas correctas entre pruebas SIMCE y construida en un Colegio Municipal

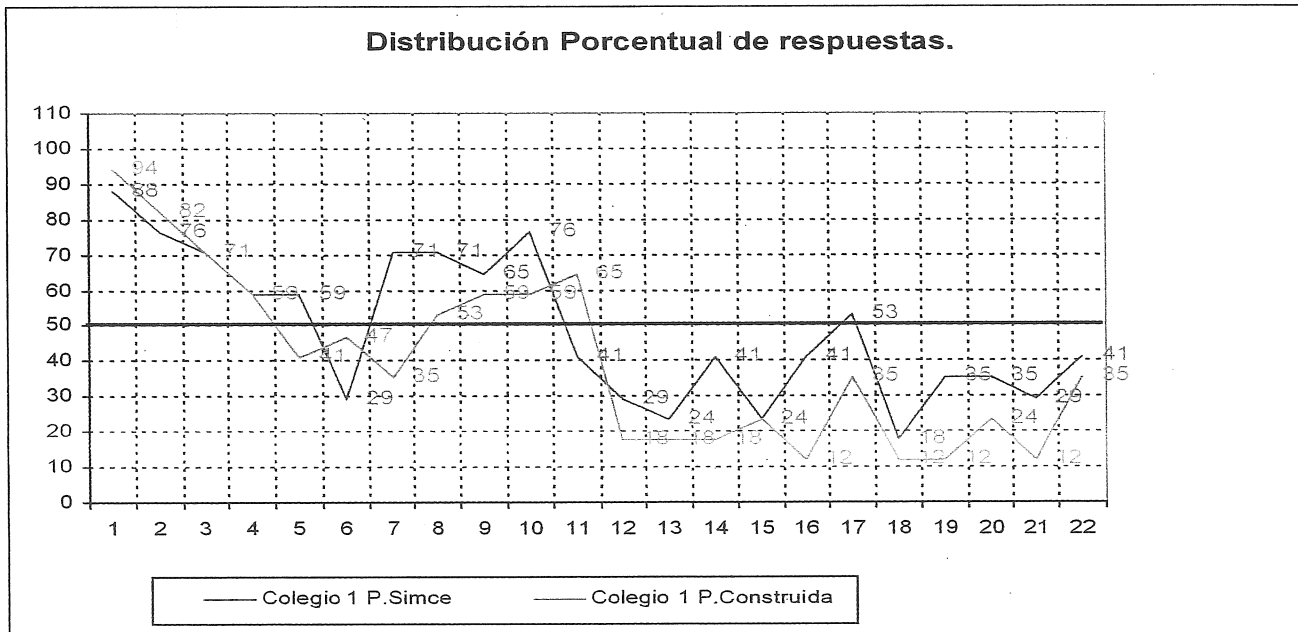


Gráfico 2. Comparación de porcentaje de respuestas correctas entre pruebas SIMCE y construida en un Colegio Particular Pagado

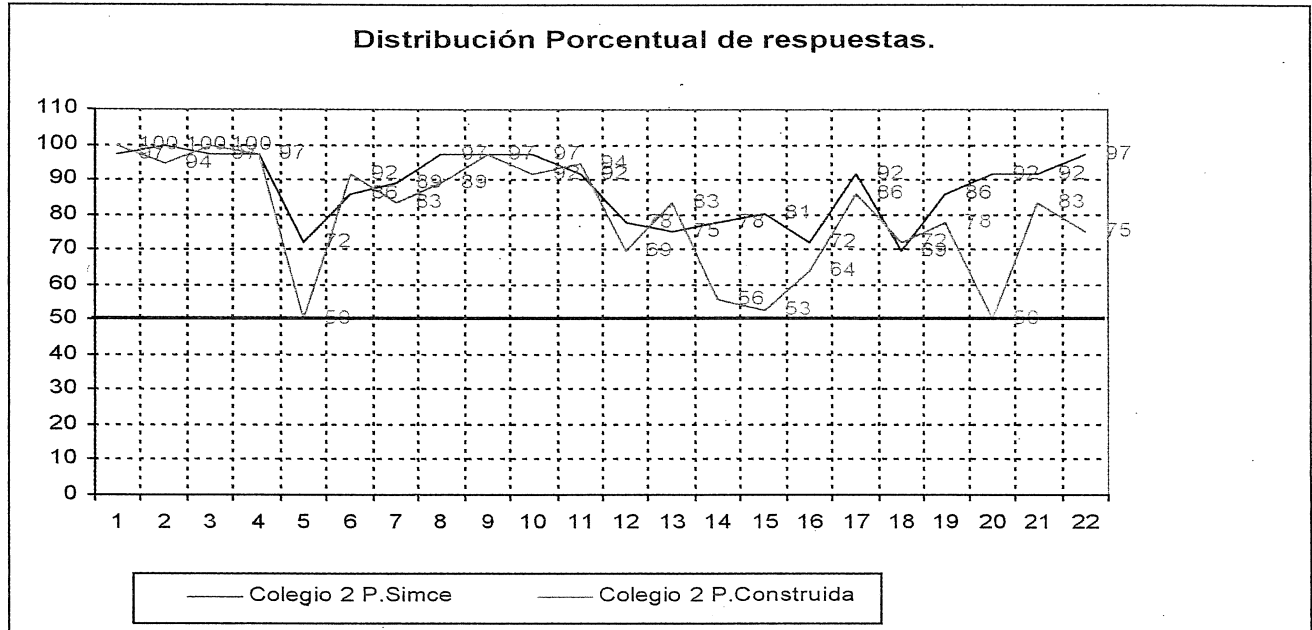
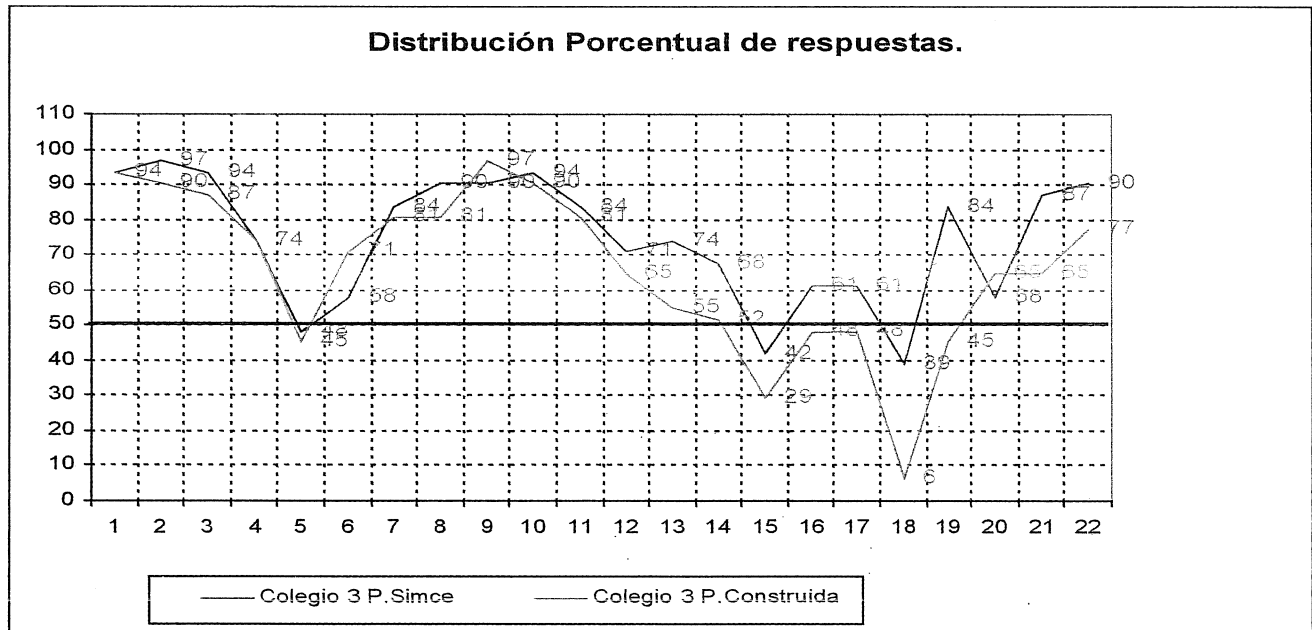


Gráfico 3. Comparación de porcentaje de respuestas correctas entre pruebas SIMCE y construida en un colegio Particular Subvencionado.



Sección B. Lenguaje.

Gráfico 4. Comparación de porcentaje de respuestas correctas entre pruebas SIMCE y construida en un colegio Municipal.

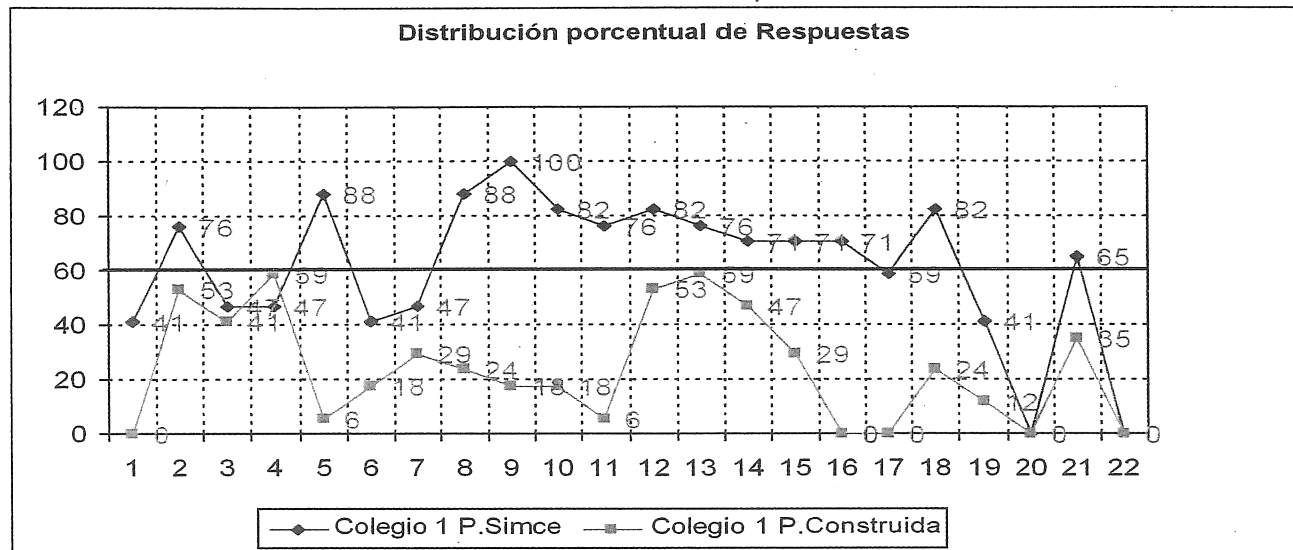


Gráfico 5. Comparación de porcentaje de respuestas correctas entre pruebas SIMCE y construida en un Colegio Particular Pagado.

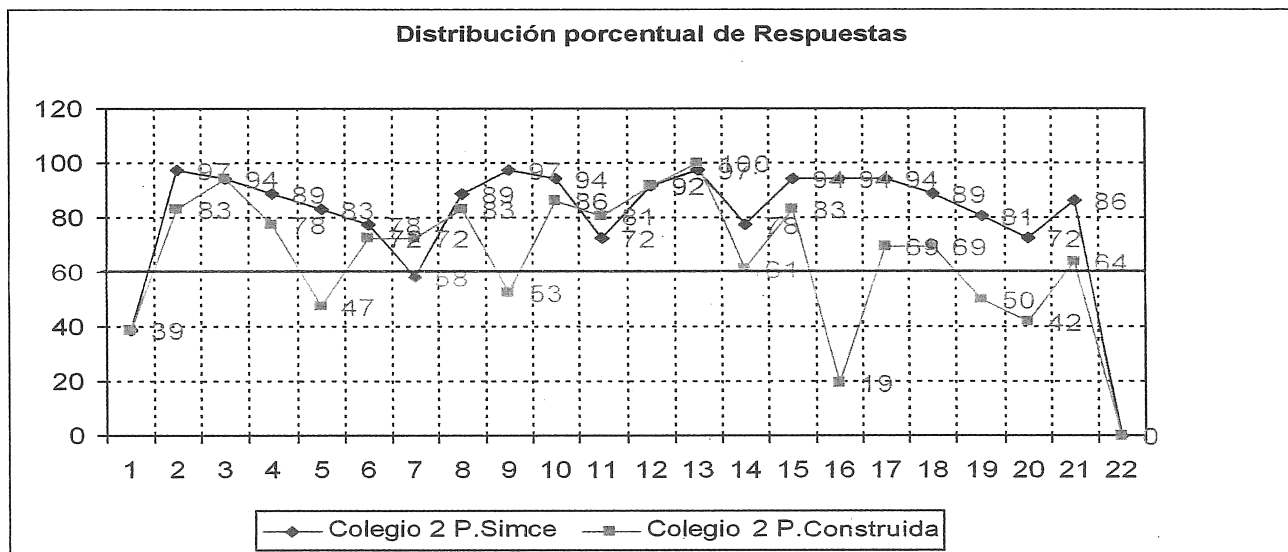
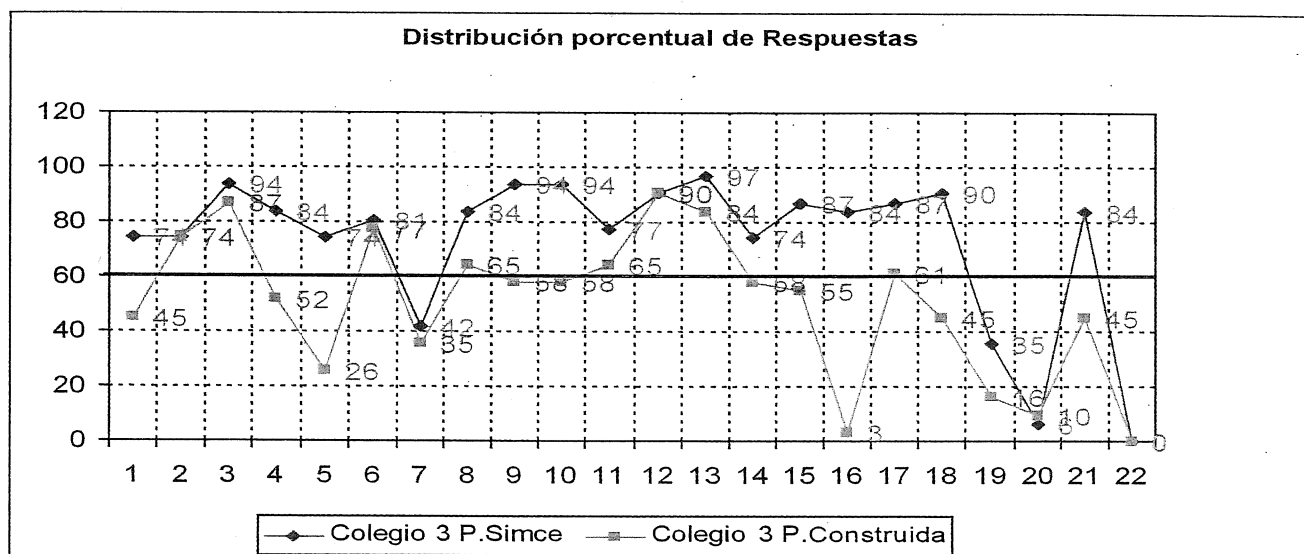


Gráfico 6. Comparación de porcentaje de respuestas correctas entre pruebas SIMCE y construida en un Colegio Particular Subvencionado.



TABLAS DE RESULTADOS

TABLA 1. Promedio de respuestas correctas por colegio según prueba SIMCE y prueba construida. Prueba de Matemáticas.

Matemática

Colegio	SIMCE	Construida
1	11	9
2	19	18
3	16	14
4	14	16
5	14	12
6	19	16
7	13	7

TABLA 2. Promedio de respuestas correctas por colegio según prueba SIMCE y prueba construida. Prueba de Lenguaje.

Lenguaje

Colegio	SIMCE	Construida
1	14	5
2	18	14
3	16	11
4	14	14
5	16	12
6	16	13
7	10	4

TABLA 3. Ordenamiento por rangos (ranking) según pruebas SIMCE y construida en matemática y lenguaje.

Ranking Matemática			Ranking Lenguaje		
Colegio	SIMCE	Construida	Colegio	SIMCE	Construida
1	7	6	1	5	6
2	1	1	2	1	1
3	3	4	3	2	5
4	4	2	4	5	1
5	4	5	5	2	4
6	2	2	6	2	3
7	6	7	7	7	7

Conclusiones

La investigación se proponía establecer en qué medida los resultados de una prueba con características edumétricas entregaría distintos resultados que una prueba SIMCE que es construida bajo una tecnología sicométrica. Los resultados muestran que son diferentes, y notoriamente más bajos en la prueba edumétrica. Esto confirma que están midiendo procesos psicológicos diferentes y que la prueba de respuesta construida es efectivamente más difícil. Si se considera además la disminución en la probabilidad de error en la toma de decisiones, esta prueba parece dar cuenta de mejor manera de los aprendizajes de los alumnos. A partir de ello se puede concluir qué decisiones que se tomen considerando unos resultados serían diferentes a si se toman considerando otros. De hecho, los ranking elaborados con ambos tipos de puntajes son diferentes, de manera que las decisiones afectarían de distinta manera a los establecimientos considerando uno u otro puntaje. Esto afectaría las políticas de equidad. Por otra parte debe considerarse que las pruebas, al ser aplicadas imprevistamente y en un periodo de cierre de las actividades escolares, están midiendo residuos de aprendizaje, es decir, aquello que los alumnos realmente recuerdan.

Se observó una correlación más alta en la primera parte de la prueba y un mayor nivel de facilidad en todos los casos analizados. Ello puede estar significando que los alumnos están alcanzando al término de cuatro años de escolaridad, aproximadamente el 75 % de los aprendizajes previstos para ser conseguidos al término del segundo año. En el caso de las escuelas municipalizadas, este nivel de logro es aún más reducido. Esto nuevamente pone de relieve el tema de la equidad.

Por otra parte se pensaba que los resultados de ambas pruebas no debían correlacionar ya que ambas se suponía medían habilidades diferentes. En el caso de la prueba SIMCE medía una habilidad de carácter general y en la prueba alternativa se medía logro del objetivo, definido el logro como la demostración reiterada de una habilidad. Esto resultó ser parcialmente correcto ya que ambas pruebas correlacionaron significativamente. Probablemente esto indique simplemente que el rendimiento está ligado a la habilidad general cuando no se atienden las diferencias individuales.

Estos antecedentes pueden aportar algunas explicaciones respecto a porqué en las pruebas internacionales de medición de calidad (como el TIMMS y OCDE, PISA y OREALC) las comparaciones nos dejan en lugares relativamente bajos.

Bibliografía

- CASASSUS, J. *et.al* (1998) "Laboratorio Latinoamericano de evaluación de la calidad de la educación". *Revista Iberoamericana de Educación* Nº 10. OEI.
- CASTEJON J. L. y otros (1993) "Investigación sobre la eficacia de centros de enseñanza secundaria. Un modelo de identificación y funcionamiento". *Revista de Educación*. España (301).
- CLARK, D. y otros (1984) "Effective School and school improvement: A comparative analysis two lines inquiry". *Educational Administration Quaterly*, Vol 20 Nº 3.
- DE LA ORDEN H. A.(1997) "Desarrollo y validación de un modelo de calidad Universitaria como base para su evaluación". *Revista Electrónica de Investigación y Evaluación educativa*. Volumen 3 Nº 1-2.
- DELGADO, KENNETH (1996) *Evaluación y Calidad de la Educación*. Editorial Magisterio, Colombia
- ESCUADERO, E. T. "Enfoques modélicos y estrategias en la Evaluación de Centros educativos". *Revista Electrónica de Investigación y Evaluación Educativa*. 1997. Vol. 3 Nº 1-1
- FERNANDEZ D. JOSE *et.al* (1997) "Desarrollo y situación actual de los estudios de eficacia escolar". *Revista Electrónica de Investigación y Evaluación Educativa*. 1997.Vol. 3 Nº 11
- HILL, P.W. y ROWE, H.J.(1996) Multilevel modelling in school effectiveness research. *School Effectiveness and school improvement*. Vol 7 Nº 1.
- HOUSE, E. (1994). *Evaluación, ética y poder*. Ed. Morata, Madrid. Infancia y Aprendizaje. 1999, 85, 19-32.
- LAFOURCADE P.(1982) *La evaluación de organizaciones educativas centradas en logros*. Edit. Trillas México.
- LANZA, HILDA M. (1996) "La Evaluación de los sistemas educativos latinoamericanos. Algunas reflexiones en torno a su especificidad". Ponencia presentada al Congreso Internacional de evaluación de la calidad. Madrid. España. Febrero.
- LeCOMPTE, MARGARET(1995) "Un Matrimonio conveniente: Diseños de investigación cualitativa y estándares para la evaluación de programas". Volumen 1 Nº 1.
- MADAUS, G. (1983). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Kluwer Nijhoff Publishing, Boston.
- MINEDUC. (Chile) "Sistema de medición de la calidad de la educación". www.mineduc.cl./media/currículum/sistema.htm
- MORTIMORE, P. *et.al* (1994) "School effectiveness and value-added measure. Assessment in educación: Principles, policy & practice". Vol. 1 Nº 3. Sage Publications Inc. USA.
- Popham, W (1987) *Problemas y técnicas en evaluación educativa*. Editorial Anaya. España
- SANTOS G. MIGUEL A.(1996) *Evaluación Educativa* (1 y 2). Editorial Magisterio del Río de la Plata. Buenos Aires
- STRAND, STEVE (1998) "A value added analysis of the 1996 primary School performance table". En *Educational Research*, Volume 40, Number 2, Summer.

- THOMAS, S; MORTIMORE, P. (1996) "Comparison of value-added models for secondary school effectiveness". *Research papers in education*. Vol 11 N° 1
- THOMAS, SALLY (1998) "Medir el valor añadido de la eficacia de las escuelas en el Reino Unido". *Perspectivas*. Vol XXVIII, N° 1.
- TORANZOZ, LILIA (1999) "El problema de la calidad en el primer plano de la agenda educativa". *Revista Iberoamericana de Educación*. N° 10 OEI. Madrid.
- WEBSTER, W. J. y MENDRO, R.L. (1994) "Effectiveness indices: A value Added approach to measuring school effect". *Estudies in Educational Evaluation* N° 20.