

TEST ESTANDARIZADOS, ANÁLISIS DEL RUIDO DEL INSTRUMENTO USADO PARA MEDIRLO Y SU EFECTO EN POLÍTICAS PÚBLICAS. CASO CHILENO

ISSN 2219-6722

ISSNE 2222-2707

ALEJANDRO SIERRA, Departamento Ingeniería Comercial
Universidad Técnica Federico Santa María Chile, Chile
LIONEL VALENZUELA, Departamento Ingeniería Comercial
Universidad Técnica Federico Santa María Chile, Chile
JUAN TAPIAS GERTOSIO, Departamento Ingeniería Comercial
Universidad Técnica Federico Santa María Chile, Chile
email: juan.tapia@usm.cl

RESUMEN

El puntaje de test estandarizados es frecuentemente uno de los principales factores a tomar en cuenta en decisiones de políticas públicas en educación. Además, que en aquellos países con un sistema de voucher para el financiamiento de la educación, influyen en las decisiones de matrículas de los estudiantes. No obstante ellos, hay dudas permanentes sobre qué tan amplio es el efecto del error (ruido) sobre la señal entregada por el test. Previos estudios han mostrado que dicho error tiene un importante efecto en el puntaje final del test. Esta investigación examina un extenso set de datos, usando los puntajes de test estandarizados en Chile desde el años 2005 al 2008, para analizar la presencia de ruido en dichos resultados y ampliar la discusión a un país que posee un amplio historial de una política de libre elección de escuelas en qué matricularse. Este estudio es una contribución, primero, porque hay un escaso número de estudios previos, y a su vez usa una set de datos no usados anteriormente para esta metodología y mucho más grandes que los previamente utilizados. Se encontró que una vez la variable ruido es aislada, la componente que es verdaderamente una señal del puntaje del test se mueve entre 45% y 47% del puntaje total para los test de matemáticas. Para los test de lenguaje, obtuvimos menores valores donde la señal se mueve entre 37.3% y 44.2%.

Palabras clave: Test estandarizados, políticas públicas de educación, señal.

STANDARDIZED TESTS, NOISE ANALYSIS OF THE TOOL USED TO MEASURE IT AND ITS EFFECT ON PUBLIC POLICIES. CHILEAN CASE

ISSN 2219-6722

ISSNE 2222-2707

ALEJANDRO SIERRA, Departamento Ingeniería Comercial
Universidad Técnica Federico Santa María Chile, Chile

LIONEL VALENZUELA, Departamento Ingeniería Comercial
Universidad Técnica Federico Santa María Chile, Chile

JUAN TAPIAS GERTOSIO, Departamento Ingeniería Comercial
Universidad Técnica Federico Santa María Chile, Chile
email: juan.tapia@usm.cl

ABSTRACT

Standardized Tests scores are often a main factor in public policy decisions in education. In addition, those countries with voucher systems for the financing, have an influence in the enrollment decisions of the students. However, there are doubts about how large is the effect of the error (noise) considering the signal given by the test. Previous studies have shown that this error has an important effect on the final test scores. This research examines a large data set, using test scores from students in Chile from 2005 to 2008 to understand the presence of noise in testing environments to analyze the presence of error in the results and increase the test-noise discussion to a country with a long history of free school choice policies. This study serves as a contribution, first, because the previous research has been sparse, and because we incorporated previously unused information and a larger, previously unused database. We found that the average signal moves between 45% and 47% for the mathematical test. For the language test, we found smaller values with an average signal between 37.3% and 44.2%.

Keywords: Standardized Tests, public policy education, signal

1. INTRODUCCIÓN

Los resultados de los test estandarizados en educación son uno de los factores más importantes para las decisiones en educación y en las políticas vinculadas a ella, además son una información de gran importancia para los padres cuando deciden dónde van a estudiar sus hijos. El estudio busca ser un aporte en el análisis de la calidad de dicha información, separando el ruido del instrumento empleado para obtener un indicador que muestra el real efecto de la escuela en los resultados de los estudiantes. No obstante, no sólo es importante cuantificar cuánto mide el aporte de la escuela durante un año específico, sino establecer cuánto de dicho aporte permanece en el tiempo y cuánto se debió sólo a un efecto transitorio. La investigación puede ser un aporte fundamental en dicha discusión al analizar además el efecto sobre el ordenamiento de las escuelas que tendría el utilizar un indicador que refleje su verdadero desempeño permanente. Para ello se aplican modelos econométricos que permitan analizar dichos efectos en los resultados del SIMCE, en el periodo 2005 al 2008 y ver su efecto en las decisiones de los padres y los otros agentes del sector educacional.

Se generará una base de datos con los resultados del SIMCE de 4° básico entre los años 2005 al 2008 por estudiante que permitirá mediante diversas herramientas econométricas y elementos de series de tiempo aislar las distintas componentes (ruido y señal). Para ello primero se desarrollarán modelos a nivel de estudiante, buscando controlar por sus características y la de sus padres, para lograr aislar un primer estimador del aporte de la escuela, a continuación considerando otras experiencias como las de Kane y Staiger (2001) se usará para estimar aislar los ruidos del instrumento y las componentes no persistentes de los indicadores. Se encontró que una vez aislada la variable ruido, la componente que es verdaderamente una señal del puntaje del test se mueve entre 37.3% y 47% del puntaje total, dependiendo del tipo de test.

2. MARCO TEÓRICO

En varios países y regiones hay diferentes test estandarizados que se aplican a los estudiantes para estimar el desempeño de ellos mismos, de sus escuelas y profesores. Estos test frecuentemente son usados para tomar decisiones de políticas públicas en materia educacional (asignación de incentivos, presupuestos, escuelas a intervenir con determinados proyectos, corregir programas de estudio, etc.). A su vez los indicadores de bienestar cada vez van integrando poco a poco el factor dentro de los mismos.

Cada vez más países toman decisiones guiándose por estos test, por lo cual cada vez el estudio de los puntajes de los mismos ha sido de interés para los estudios científicos en ciencias de la educación y políticas públicas a nivel internacional, así como economía

de sectores públicos. Por ejemplo, en Estados Unidos, No Child Left Behind Act (2001) requiere que cada estado de dicho país establezca test estandarizados para medir el desempeño de sus estudiantes, siendo obligatorias estas pruebas estandarizadas para todas aquellas escuelas que reciban fondos federales. En base a dichos test varios estados pueden imponer sanciones a escuelas cuyos estudiantes logren bajo puntaje. Por ejemplo, en Florida The Student Success Act (2011) requiere que al menos un 50% de la evaluación de desempeño del personal y administradores de los establecimientos educacionales sea fundamentado en el resultado de sus alumnos en dichos test.

Como ya se mencionó también, estos puntajes son frecuentemente uno de los recursos de información que poseen los padres para tomar decisiones respecto a dónde sus hijos van a estudiar y por extensión también afectan estos test otras decisiones como es la residencia o lugar de trabajo de la familia. En Estados Unidos se han hecho varios estudios comparando el precio de casas cercanas a distintos distritos escolares, utilizando el hecho que en dicho país el estudiante debe ir a la escuela del distrito que le corresponde por dónde vive. Estos estudios muestran una relación positiva entre el precio de las casas y el puntaje de las escuelas de los distritos escolares correspondientes, comparando casas similares en vecindarios colindantes, Black (1999), Kane et al. (2005) y Dougherty et al. (2009). Black (1999) muestra un incremento en el precio de 2.1% cuando el puntaje del test se incrementa en una desviación standard; Kane et al. (2005) encontraron un incremento del 2% y Dougherty et al. (2009) del 1.9%.

Otros países también han desarrollado test estandarizados para la toma de decisiones en políticas públicas. Por ejemplo, Chile ha administrado test nacionales estandarizados (SIMCE) desde 1988. Estos test se aplican a todas las escuelas y sus resultados son de conocimiento público. Tanto el Ministerio de Educación como las Municipalidades basan muchas de sus decisiones con respecto a presupuesto y políticas de educación en los puntajes de dichos test. Por ejemplo, cambios en las políticas locales, ranking de escuelas y profesores. Además, por un lado, proyectos dirigidos a incentivar a aquellas escuelas con mejores resultados, y por otro, motivar a aquellas de puntajes promedios o bajos a subir sus puntajes.

Además, el sistema nacional de evaluación del desempeño de profesores está fuertemente asociado a los resultados de sus alumnos en estos test nacionales. Mizala y Romaguer (2000), discute la importancia de los sistemas de evaluación e incentivos aplicados a la educación, en particular analiza el diseño e implementación del Sistema Nacional de Evaluación de los Establecimientos Educacionales Subvencionados (SNED), éste introduce dos elementos adicionales muy importantes dentro del conjunto de reformas implementadas con anterioridad, que refuerzan los incentivos para mejorar la calidad de educación: posibilita una mejor información de los resultados del establecimiento

escolar y establece incentivo a los docentes para mejorar la calidad de la enseñanza, por lo que el SNED tiene un componente principal los resultados académicos de los alumnos, a través de la prueba SIMCE.

Finalmente, la importancia de estos test también se ve reflejada en el hecho de que el Ministerio de Educación cada vez más incrementa su frecuencia y aumenta el número de asignaturas evaluadas. En los establecimientos educacionales privados también el resultado del test es importante, ya que un buen resultado incide directamente en un aumento de la demanda de matrícula y si el resultado es deficiente aplican rápidamente acciones remediales para mantener la cuota de matrícula. De este modo los directivos de los colegios suelen analizar series temporales interanuales de resultados para ver el comportamiento en plazos medios.

En Chile, donde existe libre elección a la escuela a la que postular, los test estandarizados son frecuentemente uno de los recursos más prácticos para tomar decisiones respecto a las escuelas donde matricular a sus hijos. Gallego y Hernando (2010), muestra como las familias valoran algunos atributos de las escuelas, siendo las dos dimensiones más importantes los resultados de las pruebas y la distancia a la escuela. Curiosamente, al mismo tiempo, los resultados de su estudio, sugieren que existe una gran cantidad de heterogeneidad en las preferencias, ya que la valoración de la mayoría de los atributos de la escuela depende de las características del hogar.

Independientemente de la forma en que se calcula la diferencia de fines de lucro, las escuelas tienden a obtener mejores resultados que las escuelas municipales. Según Gallego et al. (2008), la elección de escuelas, aumenta el bienestar general de los estudiantes, pero existe una gran cantidad de heterogeneidad en el tamaño e incluso en el signo del cambio del bienestar. La alta segregación socioeconómica del sistema escolar en Chile parece estar impulsada principalmente por factores del lado de la demanda, como lo demuestra el hecho de que una potencial abolición de las tasas y el uso de las loterías en el contexto de la aplicación gratuita para todas las escuelas no parecen disminuir la segregación de manera significativa.

Para el debate de las temáticas en ciencias de educación y de bienestar es un problema fundamental determinar la calidad de la información, especialmente del puntaje de los test escolares para analizar la calidad de la información que están usando para los estudios y comparaciones involucradas, así para una serie de decisiones en el ámbito educacional y político. Un primer cuestionamiento es si esos test en verdad son una buena aproximación para medir la calidad de la escuela, de un determinado profesor, del programa, o del alumno. Pero aunque el objeto central del test fuera lo antes expresado, permanece el problema de cuánto de su valor reflejan en verdad la calidad del colegio y cuánto es el ruido del instrumento.

Kane y Staiger (2002) muestran cómo el ruido del instrumento y otros efectos de corto plazo podían afectar la medición de la calidad de la escuela al considerar el resultado del test en un año de sus alumnos. Ellos analizaron 1.163 escuelas básicas en Carolina del Norte entre 1994 y 1998 encontrando en promedio que el 38% de los puntajes en test de comprensión lectora de los alumnos de 5° grado estaban asociados a problemas de ruido debido a problemas del tamaño de la muestra. Una escuela con pocos estudiantes es más proclive a mostrar mayores cambios en el puntaje de los test de año a año, a causa de que un cambio en unos pocos estudiantes tiene un mayor impacto en el promedio del establecimiento al contrario de aquellos casos donde la escuela posee un mayor número de estudiantes. También Kane y Staiger (2001) encontraron que en aquellas escuelas más pequeñas el ruido representaba un 27% del puntaje en los test de matemáticas y un 44% en los test de lectura. Mientras, que en aquellas escuelas más grandes, éste representaba sólo un 8.5% del test de matemáticas y 16.1% de los test de lectura. Ellos demostraron que el problema era aún mayor cuando en vez de analizar el valor del test se analizaba como indicador el cambio del mismo de un año a otro, estableciendo un 80% ruido en el caso de las escuelas más pequeñas.

Esto como se precisó anteriormente, puede generar problemas en el tema de asignación de recursos, por ejemplo, en el caso de Chile, Chay et al. (2005) analizaron el efecto de incrementar recursos para 900 escuelas “promedios” que fueron asignados considerando el puntaje de un año del en el test estandarizado y los problemas en su medición derivados de no considerar los aspectos antes enunciados. Dejando de manifiesto, la necesidad de incorporar una metodología que los considere para la distribución de recursos.

Por otro lado Mizala et al. (2007) también se analizaron el caso Chileno. Comparando diferentes maneras de crear rankings de puntajes de escuelas, demostrando que las diversas formas para llevarlo a cabo presentan diversos problemas en cada una de ellas. Algunos métodos usados para “rankear” escuelas hacen que los resultados de los mismos estén altamente correlacionados con las características de los estudiantes (siendo más un ranking del nivel socioeconómico de los estudiantes que de la escuela en sí), mientras que otros métodos que buscan eliminar esto generan medidas altamente volátiles a través del tiempo, variando considerablemente año a año. Los autores de hecho indican en la discusión final del artículo que “Our results also suggest that it might be desirable for Chile to further explore filtering schemes that would more explicitly exploit the time-series dimension in school performance measures, trying to extract whatever signal they contain. Unfortunately, this is unlikely to be easy and there is not much accumulated experience on how to implement these techniques.” (Mizala et al., 2007). La presente investigación justamente en buena parte busca seguir el acercamiento al problema, planteado por Mizala et al. (2007), como desafío, al final de su paper, y

aplicar y desarrollar ratings basados en las metodologías por ellos indicados y que son de gran importancia para el desarrollo científico del área.

Dicha metodología dada su complejidad, que fue desarrollada y aplicada a esta área de estudio por Kane y Staiger (2001), no ha sido replicada por completo más allá de su estudio, teniendo como único punto de referencia de los resultados obtenidos para las 1.163 escuelas básicas en Carolina del Norte. Pero dichos valores pueden ser particulares para ese caso y no necesariamente tienen que ser los mismos en el caso de otros países, con otras realidades y otros test. Siendo un gran aporte para la investigación tener otros puntos de referencia de estudio.

El aporte del estudio de la calidad de la información de los puntajes y en base al cuál se comparan las distintas escuelas, no es menor, tanto nacional como internacionalmente, ya que en aquellos países en que existe un sistema de distritos escolares, uno de los argumentos usados contra el sistema de libre elección, entre otros argumentos, es el hecho que los padres no poseen la suficiente información para tomar la información. Luego es un aporte real el analizar la calidad de la información con que los padres están tomando dichas decisiones. Además dicha calidad es básica para toma de decisiones de políticas públicas, dado el número importante de decisiones país que se fundamentan en ello, y además reorienta parte de las estrategias de los colegios privados.

3. METODOLOGÍA

En el caso Chileno, se tiene un test anual estandarizado denominado SIMCE (Sistema de medición de calidad de la educación). Este test es tomado en todas las escuelas desde el año 1988. El evalúa los logros de los estudiantes en los principales objetivos de estudio del currículum escolar de las asignaturas que evalúa. Actualmente el SIMCE se aplica a alumnos de 2°, 4°, 6° y 8° año básico y a estudiantes de 2° y 3° año medio, en áreas tan diversas como son: Lenguaje y Comunicación (Comprensión de Lectura y Escritura); Matemática; Ciencias Naturales; Historia, Geografía y Ciencias Sociales; Inglés y Educación Física. La cobertura que le da la prensa a dicha información cada año cuando los puntajes son publicados por el Ministerio de Educación, también hay acceso a dicha información tanto en las páginas web del Ministerio de Educación y de la Agencia para la Calidad de la Educación. Además, está presente, cuando esta es positiva, en la publicidad y en la información que las escuelas presentan a los padres durante el proceso de matrícula o selección de escuela.

La propuesta de estudio busca aprovechar el hecho que los test del SIMCE en Chile han sido aplicados anualmente para cuarto año básico en lenguaje y matemáticas desde el 2005 a la actualidad. Buscando en primer lugar aplicar la metodología de filtered esti-

mator desarrollada por Kane y Staiger (2001) a un conjunto de datos mucho mayor y en un periodo mayor de tiempo, considerando datos de aproximadamente 5.000 escuelas entre los años 2005 y 2008, es una base de datos mucho mayor a la empleada por Kane y Staiger que la aplicó a 1.163 escuelas. La investigación busca estimar cuánto anualmente es ruido y cuánto en verdad refleja el aporte de la escuela (señal). Siendo importante para las decisiones de políticas públicas tener claridad sobre las magnitudes de cada una de ellos dado la importancia que se da al puntaje de los test en las decisiones de políticas públicas como se mencionó anteriormente.

También este estudio tiene una gran importancia práctica a nivel internacional, en el sentido que como antes se indicó generaría otro punto de referencia después de la investigación Kane y Staiger (2001) sobre el tema. Con la diferencia de que se analiza una base de datos mucho mayor. Al desarrollar esta metodología aplicada al caso Chileno esta será replicable para la toma de otras decisiones de los actores económicos en las decisiones de políticas públicas, es decir es un aporte para el nivel de aplicabilidad de dichas metodologías.

Se aplica las metodologías planteadas por Kane y Staiger (2002) y, Kane y Staiger (2001). Estas a su vez se encuentran basadas en las metodologías planteadas por McClellan et al. (1999) y, McClellan y Staiger (1999). Dichas metodologías fueron aplicadas a indicadores del sector salud en Estados Unidos. McClellan y Staiger (2001) proponen separar la varianza en una componente de ruido y una se señal. En esta investigación se utiliza dicha metodología para separar las componentes de ruido y de señal en el test estandarizado. Pueden haber diversas fuentes de variación en el resultado, por lo cual, es necesario cuantificar la influencia del efecto real de la escuela y el efecto del ruido. Con este fin primero se usa una regresión a nivel de estudiantes, para determinar el efecto de la escuela sobre su rendimiento en el test, además de controlar por distintas características del estudiante:

$$(1) Y_{irt} = \delta_{rt} + X_{irt}'\Omega + \mu_{irt}$$

Donde: μ_{irt} es el error del modelo. Y_{irt} es el puntaje normalizado por la desviación estándar del SIMCE del estudiante i de la escuela r en el año t . X_{irt} representa el vector de características del estudiante i (por ejemplo, sexo, ingreso de los padres, nivel de educación de los padres, etc.). Ω es un vector $M \times 1$, donde M es el número total de características de los estudiantes. δ_{rt} muestra del efecto de la escuela r en el año t sobre el puntaje. Este es independiente de las características de los estudiantes ese año. δ_r será un vector que muestre el efecto de la escuela los diversos años. Una vez realizada la regresión no es posible obtener directamente δ_r sino una estimación del efecto de la escuela. Este estimador tiene cierto nivel de ruido, ya que la precisión del mismo se incrementa con

el número de estudiantes que haya en cada escuela. Luego se tiene la ecuación (2):

$$(2) E(\delta r' = \delta r + \varepsilon),$$

Donde ε representa una estimación de la componente de ruido y δr es el efecto real de las escuelas (señal). Para estimar el efecto, primero aplicando mínimos cuadrados ordinarios obtenemos de la ecuación (1). Este es un vector $N \times 1$ que contiene los estimadores del efecto de la escuela a través de la dicha regresión. Segundo, se estima la matriz varianza covarianza de los mismos de la componente de ruido: $E(\varepsilon' \varepsilon)$. Las varianzas de $E(\varepsilon' \varepsilon)$ estará dada por la varianza de los estimadores de la regresión (1). Se calcula $E(\delta r' \delta r)$ basado en la matriz varianza covarianza de entre escuelas cada año. Luego, ya obtenidos $E(\varepsilon' \varepsilon)$ y $E(\delta r' \delta r)$, se puede obtener $E(\delta r)$, aplicando:

$$(3) E(\delta r' \delta r) = E(\delta r) - E(\varepsilon' \varepsilon)$$

Finalmente para filtrar el ruido de los que es señal una alternativa es usar la siguiente relación:

$$(4) E(\delta r) = \beta^*$$

Donde se obtiene β de la forma:

$$(5) \beta = (E(\delta r' \delta r) + E(\varepsilon' \varepsilon))^{-1} E(\delta r' \delta r)$$

El estimador β nos permitiría filtrar lo que es señal de lo que es ruido en los diversos años.

4. ANÁLISIS DE DATOS

Con el fin de cuantificar los efectos antes descritos en la metodología, se utilizarán los puntajes de la prueba nacional estandarizada (SIMCE) correspondientes a los estudiantes de cuarto grado de educación elemental en las áreas de matemáticas y lenguaje desde el año 2005 al 2008.

Estas bases de datos se encuentran disponibles, algunas de ellas públicamente (la información a nivel de escuela) y otras están en forma privada disponibles para ser solicitadas por investigadores de las universidades (a nivel de estudiante) en la página web de la Agencia para la Calidad de la Educación.

Actualmente se poseen los datos para el periodo 2005 al 2008, considerando un número cercano a 5.700 escuelas por años.

Como ejemplo, en la figura 1, podemos ver los puntajes medios en los test de matemáticas como función de los números de estudiantes matriculados en la escuela correspondientes al año 2008. Donde se puede apreciar como a menor tamaño de escuela mayor variabilidad en ellos tal como se esperaba. Esta diferencia es aún más clara cuando graficamos el cambio entre 2007 y 2008 versus el número de alumnos matriculados en la escuela como se muestra en la figura 2.

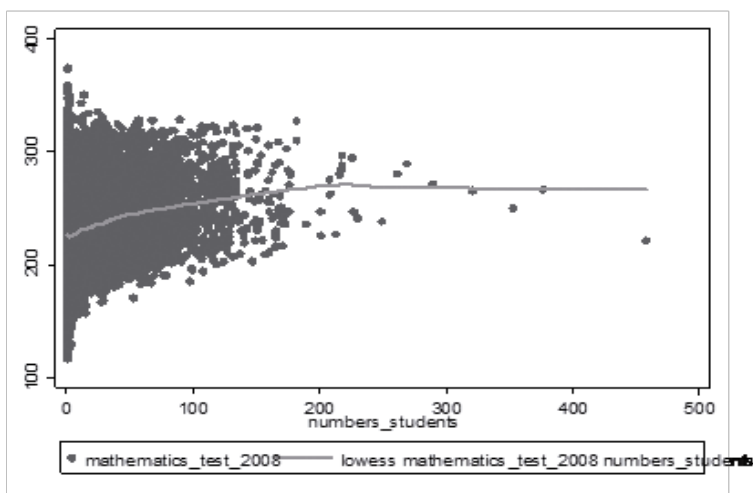


Figura No. 1: Puntaje promedio del SIMCE de matemáticas 2008 versus la matrícula de la escuela

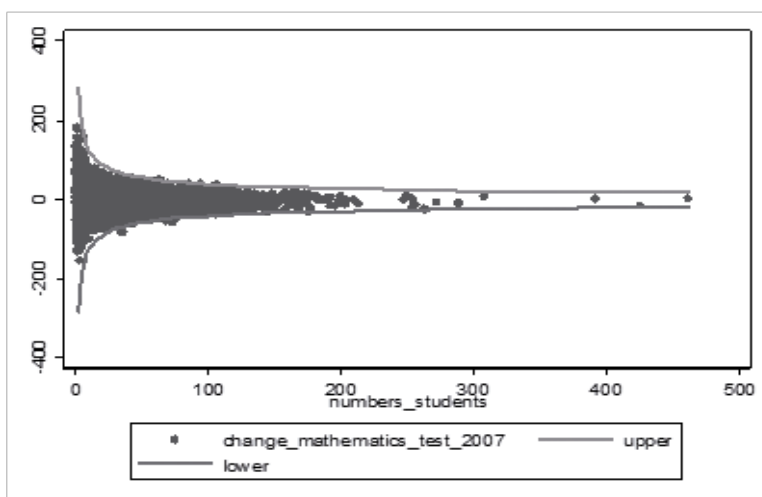


Figura No. 2: Cambio en el puntaje promedio del SIMCE de matemáticas entre 2007 y 2008 versus la matrícula de la escuela (2008)

5. RESULTADOS Y DISCUSIÓN

Se calculó la regresión (1) mostrada en metodología. Basados en los supuestos indicados en dicho punto se obtuvieron las siguientes matrices varianza y covarianza para la componente de ruido ($V\varepsilon$) y la componente de señal ($V\delta$) para los test de matemáticas y lenguaje:

Test de Matemáticas:

$V\varepsilon$:

| | | | |
|---------|---------|---------|---------|
| 0.03538 | 0 | 0 | 0 |
| 0 | 0.03154 | 0 | 0 |
| 0 | 0 | 0.03504 | 0 |
| 0 | 0 | 0 | 0.03233 |

$V\delta$:

| | | | |
|---------|---------|---------|---------|
| 0.02956 | 0.0077 | 0.00678 | 0.00672 |
| 0.0077 | 0.02819 | 0.00877 | 0.00843 |
| 0.00678 | 0.00877 | 0.02897 | 0.00721 |
| 0.00672 | 0.00843 | 0.00721 | 0.02644 |

Test de Lenguaje:

$V\varepsilon$:

| | | | |
|---------|---------|---------|---------|
| 0.03528 | 0 | 0 | 0 |
| 0 | 0.03128 | 0 | 0 |
| 0 | 0 | 0.03487 | 0 |
| 0 | 0 | 0 | 0.03190 |

$V\delta$:

| | | | |
|---------|---------|---------|---------|
| 0.02792 | 0.00743 | 0.00579 | 0.00524 |
| 0.00743 | 0.02242 | 0.00535 | 0.00704 |
| 0.00579 | 0.00535 | 0.02368 | 0.00715 |
| 0.00524 | 0.00704 | 0.00715 | 0.01895 |

Los resultados muestran una varianza del ruido mayor que los de la señal. En la mayoría de los casos la diferencia es sobre el 20%. Por otro lado, si comparamos la varianza del ruido de ambos test, la varianza del ruido es similar. Pero, el test de lenguaje muestra una menor varianza de la señal.

Basado en estos resultados se aplicó la metodología de filtro mostrada en la ecuación (5), de lo cual se obtuvo que en promedio el componente de señal del puntaje del test iba entre 45% y 47% para el test de matemáticas (45.5% en 2005, 47.2% en 2006, 45.2% en 2007, 45% en 2008). Para el test de lenguaje se encontraron porcentajes menores, que van entre 37.3% y 44.2% (44.2% en 2005, 41.8% en 2006, 40.4% en 2007, 37.3% en 2008).

Se encontraron porcentajes de ruido mayores a los encontrados por Kane y Staiger (2001) en su investigación en Carolina del Norte. Pero, debemos considerar que en este caso estamos hablando de otro test, con otras características y con una muestra más grande. Por otro lado, Kane and Staiger (2001) también encontraron, al igual que esta investigación, valores distintos para el test de matemáticas y el de lectura.

Los resultados obtenidos confirman la importancia de considerar las componentes de ruido y señal al momento de tomar los resultados de los test para decisiones de políticas públicas, matrícula o de ser incorporados a índices de bienestar. A su vez los resultados, muestran como diferentes test pueden tener diferentes proporciones entre ambas componentes.

6. CONCLUSIONES E IMPLICACIONES

En Estados Unidos, Chile y varios otros países, decisiones de políticas pública y matrícula son frecuentemente basadas en los resultados de los test estandarizados. Este estudio, corroborado por investigaciones previas, muestra que el ruido presente en dichos test no es menor, y esto podría llevar a los distintos agentes involucrados en las decisiones a conclusiones erradas.

Este estudio utiliza una muestra mayor para obtener dichos componentes. Hasta el momento los únicos porcentajes que se tenían eran los obtenidos por Kane y Staiger (2001), utilizándose estos como puntos de referencia. Pero como muestra esta investigación, hay que ser cuidadosos en su aplicación, ya que no obstante se coincide en la importancia de la componente ruido, se muestra que diferentes test pueden variar a su vez en la magnitud que esta representa.

Finalmente es importante indicar para futuras investigaciones que este es un primer

paso. Ajustar ruido y señal no es suficiente. Otros factores también pueden estar afectando la componente de señal: un profesor que estuvo un año en la escuela pero no duró, un problema de clima puntual, etc. Hay componentes transitorios que, no obstante ser parte de la señal, no permanecen en el tiempo y sería importante en un siguiente estudio aislar estas una vez se haya aislado el ruido. Para esto sería importante imponer nuevos supuestos a los modelos. Por otro lado, ver cómo estas consideraciones afectan los estudios de rating de las escuelas o posibles futuros indicadores de bienestar. Pudiendo ampliarse esta metodología a otras áreas de estudio.

REFERENCIAS BIBLIOGRÁFICAS

Black, S. (1999). “Do better schools matter? Parental valuation of elementary education”, *The Quarterly Journal of Economics*, Vol. 114 No.2, pp. 577– 599.

Chay, K.; McEwan, P.; Urquiola, M. (2005). “The central role of noise in evaluating interventions that use test scores to rank schools”. *American Economic Review*, Vol. 95 No.4, pp. 1237–1258.

Dougherty; Maloney; Smith; Snow. (2009). “School choice in suburbia: test scores, race, and housing markets”. *American Journal of Education*, 115(4), pp. 523–548.

Gallego, F.; Hernando, A. ; Flabbi, L.; Tartari, M (2008). “On the Determinants and Implications of School Choice: Semi-Structural Simulations for Chile”. *Economía*, pp. 197-244.

Gallego, F.; Hernando, A (2010). “School choice in Chile: Looking at the demand side.” *Pontificia Universidad Católica de Chile Documento de Trabajo* 35.

Kane, T.; Staiger, D. (2001). “Improving school accountability measures”. Retrieved from *National Bureau of Economic Research website: <http://www.nber.org/papers/w8156>*.

Kane, T.; Staiger, D. (2002). “The promise and pitfalls of using imprecise school accountability measures”. *Journal of Economic Perspectives*, Vol.16 No.4, pp. 91–114.

Kane, T.; Staiger, D.; Riegg, S (2005). “School quality, neighborhoods and housing prices: the impacts of school desegregation”. Retrieved from *National Bureau of Economic Research website: <http://www.nber.org/papers/w11347>*.

McClellan, M; Staiger, D. (1999). “The Quality of Health Care Providers”. *National Bureau of Economic Research website. Working paper* 7327.

McClellan, M; Staiger, D. (2001). “Comparing the Quality of Health Care Providers”. *Forum for Health Economics & Policy*. Vol. 3 No.1.

Mizala A.; Romaguera P (2000). “Sistemas de incentivos en educación y la experiencia del SNED en Chile”. No. 82. *Centro de Economía Aplicada, Universidad de Chile*.

Mizala, A.; Romaguera P; Urquiola M. (2007). “Socioeconomic status or noise? Tradeoffs in the generation of school quality information”. *Journal of Development Economics* 84, pp. 61-75.