

Estimando modelos de valor agregado: Evidencias sobre la eficacia de los maestros de las escuelas municipales de São Paulo

Estimating Value-Added Models: Evidence on Teacher Effectiveness from São Paulo's Municipal Schools

Gabriela Miranda Moriconi

Fundación Carlos Chagas, Brasil

Resumen

Dada la importancia de los docentes en el proceso educativo y la disponibilidad de datos para la estimación de modelos de valor agregado, este trabajo es parte de los esfuerzos para estimar los efectos docentes y su variación y obtener evidencias para explicarla. Este trabajo pretende estimar medidas de valor agregado individuales de una muestra de docentes de 4° grado de las escuelas municipales de São Paulo, a partir del crecimiento en las puntuaciones de sus alumnos en pruebas de lectura y matemáticas en 2010. Se encontró una variación de 0.062 a 0.45 en la desviación estándar de los efectos docentes medidos en términos de desviaciones estándares de las notas de los alumnos en lectura y de 0.059 a 0.53 en matemáticas. Hubo asociaciones positivas entre las siguientes variables y los efectos docentes: la frecuencia de tarea y la frecuencia de uso de los libros de apoyo. El análisis de fiabilidad y estabilidad de los efectos estimados mostró una capacidad limitada para apoyar las decisiones de personal, pero permitió identificar un 13% de maestros con efectos diferentes del promedio. Así, son el público ideal para la investigación futura sobre sus prácticas docentes, sobre todo las positivamente asociadas con los efectos docentes estimados.

Palabras clave: eficacia, efectos docentes, modelos de valor agregado

Correspondencia a:

Gabriela Miranda Moriconi
Departamento de Investigación en Educación de la Fundación Carlos Chagas,
Brasil.

Rua Quitanduba, 363 – CEP: 05516-030 – São Paulo/SP, Brazil.

Correo electrónico: gmoriconi@fcc.org.br

Este informe es un resultado de la tesis doctoral de la autora en el área de Administración Pública y Gobierno de la Fundación Getulio Vargas (EAESP-FGV). La autora agradece a GVpesquisa y a la Coordinación de Perfeccionamiento del Personal Académico (CAPES), por haber financiado los estudios doctorales. Agradece al profesor guía, Profesor Nelson Marconi y a los profesores Reynaldo Fernandes, Paula Louzano, José Francisco Soares y André Portela, quienes contribuyeron de diversas maneras al desarrollo de esta tesis.

© 2014 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

ISSN: 0719-0409 DDI: 203.262, Santiago, Chile
doi:10.7764/PEL.51.1.2014.21

Abstract

Given the importance of teachers to the educational process and availability of data for estimating a value-added model, this paper is part of the current efforts to estimate teacher effects and their variation, and to obtain evidences on what can explain it. This paper intends to estimate individual value-added measures of a sample of 4th grade teachers from *São Paulo's* municipal schools, based on student achievement gains in tests on reading and math in 2010. The results indicated a variation from 0.062 to 0.45 standard deviation of teacher effects measured in terms of standard deviations of student achievement in reading and of 0.059 to 0.43 in math. There were positive associations between the following variables and teacher effects: frequency of homework assignment and frequency of use of Support Workbooks. The analysis of reliability and stability of estimated teacher effect measurements has shown a limited capacity to support recommendations regarding personnel policies, but permitted to identify 13% of the teachers with different effects from the estimated average. Due to this distinction, these teachers are the ideal target to future research on their teaching practices, especially on those observed as positively associated with the estimated teacher effects.

Keywords: effectiveness, teacher effects, value-added models

Los docentes cumplen una función central en el proceso educativo. Son quienes interactúan cada día con los estudiantes, y el trabajo que hacen se enfoca en torno al aprendizaje que logran los estudiantes. Los docentes son responsables de tomar varias decisiones cruciales, como: qué métodos utilizarán para presentar ideas y desarrollar destrezas, qué nivel y tipo de participación exigirán de los estudiantes, qué tipo de procedimientos utilizarán para motivar a sus alumnos y cómo evaluarán que sus alumnos hayan adquirido conocimientos. Por lo tanto, es fácil entender por qué los investigadores y encargados de las políticas han puesto tanta atención en instaurar políticas que mejoren la enseñanza.

En las últimas décadas, se han presentado varias iniciativas a nivel nacional e internacional para evaluar el aprendizaje de los estudiantes por medio de exámenes estandarizados, principalmente de comprensión lectora y conocimientos matemáticos. Dado que las destrezas de lectura y matemáticas se consideran como muy importantes para la vida de los estudiantes en las sociedades contemporáneas, y dado que los docentes tienen la gran responsabilidad de promover el desarrollo de estas destrezas, resulta lógico intentar crear métodos para descubrir cuánto contribuyen los docentes al logro académico de los estudiantes.

Recientemente, ha surgido un debate intenso sobre la eficacia docente; es decir, la capacidad que tiene el docente para mejorar los logros académicos de sus estudiantes en exámenes a gran escala. Existen dos enfoques principales relacionados con el uso de estimaciones respecto a la eficacia docente. El primero es tratar de identificar los atributos que puedan explicar la eficacia docente y luego utilizar esta información para subsidiar políticas que mejoren la calidad de los docentes. Algunos atributos analizados son los conocimientos, prácticas y características propias del profesor o profesora. El segundo enfoque es considerar las estimaciones del valor agregado de parte del profesor o profesora al logro académico del estudiante como una representación aceptable de la eficacia de cada profesor o profesora y utilizar estos datos, ya sea por sí mismos o junto a otros criterios, para tomar decisiones de personal, tales como la opción a un cargo permanente, un ascenso laboral o compensaciones.

El énfasis que pone el gobierno federal de los Estados Unidos en las políticas basadas en los logros académicos de los estudiantes y en las mediciones de la contribución de los profesores a dicho logro, impulsó la generación de más estudios y artículos sobre los desafíos y dificultades metodológicas al estimar modelos de valor agregado para generar aquellas mediciones.

En Brasil no se cuenta con información sobre algún estado o municipalidad que haya implementado políticas de alto impacto, basadas en mediciones del efecto docente en el resultado que logran los estudiantes en los exámenes. Solo se cuenta con algunas experiencias sobre la evaluación de estudiantes a lo largo de varios años consecutivos, las que nos podrían ayudar a estimar las mediciones individuales del valor agregado del profesor.

Mientras que las preocupaciones de los investigadores norteamericanos suelen centrarse en torno a los efectos directos de las políticas basadas en las estimaciones del valor agregado docente, en este estudio la motivación es mejorar los conocimientos que ya se tienen sobre la eficacia docente en Brasil. A pesar de que la docencia es una profesión compleja y la enseñanza de calidad es un concepto que puede abarcar muchos otros aspectos, la capacidad de los profesores de mejorar el aprendizaje de sus estudiantes en comprensión lectora y matemáticas es un tema muy importante que se debería explorar.

En este estudio se pretende obtener una estimación de las medidas de valor agregado individual de una muestra de docentes de escuelas municipales en *São Paulo*, basadas en los logros y resultados académicos que obtuvieron los estudiantes en *Prova São Paulo*, un sistema de evaluación que mide la comprensión lectora y conocimientos en matemáticas de los estudiantes. El propósito es analizar si es posible utilizar aquellas mediciones para tomar decisiones sobre profesores de manera individual y brindar evidencias de prácticas y características de docentes eficaces en general. Este es un tema que aun no se explora en Brasil, dado que solo se cuenta con pocos datos longitudinales recientes que permitan realizar estudios como esos.

Método

Marco teórico

Los modelos de valor agregado para estimar los efectos de los docentes y las escuelas son de gran interés, tanto para los investigadores como para los encargados de las políticas. McCaffrey, Lockwood, Koretz y Hamilton (2003) indican que las mediciones de los efectos docentes son interesantes como un modo de responder al menos dos preguntas amplias: (a) ¿Influyen los profesores de diferentes maneras en los resultados de los estudiantes? (b) ¿Qué tan eficaz es cada profesor respecto a la capacidad de mejorar los logros académicos de los estudiantes? y ¿Cuáles profesores son más o menos eficaces? A partir de la respuesta a la segunda pregunta, podemos llegar a otra pregunta de gran interés: (c) ¿Qué sustenta y cómo se explica la eficacia docente?

La primera pregunta requiere llegar a estimaciones sobre la variabilidad entre los efectos docentes. Si los datos y los modelos estadísticos pueden describir de manera precisa las contribuciones de los profesores al logro académico de los estudiantes, los modelos pueden brindar estimaciones de la variabilidad entre los efectos docentes y determinar la proporción de la variabilidad en los logros académicos o el crecimiento que se atribuye a los docentes (McCaffrey et al., 2003).

Para lograr esto, los estudios más recientes han utilizado una variable ficticia para identificar a cada profesor que hizo clases a los estudiantes en la muestra y su coeficiente estimado representa el efecto del profesor. La estrategia más común es calcular la desviación estándar de los efectos docentes estimados, corregir los errores de muestreo y vincularla a los puntajes de los estudiantes, que suelen estandarizarse a una media de cero y una varianza de uno (Aaronson, Barrow, & Sander, 2007; Goldhaber & Hansen, 2010; Koedel & Betts, 2011; Rockoff, 2004; Rothstein, 2009).

Este tipo de estudios fue creciendo, principalmente en los Estados Unidos, desde que en el año 2002 la *Ley No Child Left Behind (Que Ningún Niño se Quede Atrás)* exigió a todos los estados evaluar cada año a cada estudiante desde 3^o a 8^o año de educación escolar en comprensión lectora y matemáticas. Entonces, se cuenta con suficientes datos longitudinales de los Estados Unidos para obtener una estimación de los modelos de valor agregado. Por ejemplo, en la revisión de la literatura que realizaron Hanushek y Rivkin (2010), estos autores encontraron 11 estudios con este enfoque, basados en datos de ciudades o estados de EE.UU. publicados desde 2004 hasta 2010. Estos estudios indican variaciones de la desviación estándar de 0,08 a 0,36 de los efectos docentes medidos en referencia a las desviaciones estándar del logro académico de los estudiantes, con una desviación estándar promedio de 0,11 para comprensión lectora y 0,15 para matemáticas. Según lo explican los autores, de los resultados de matemáticas se puede deducir que tener un profesor o profesora en el percentil 25, en comparación con el percentil 75 de la distribución de calidad, significaría una diferencia en la adquisición de aprendizaje de aproximadamente 0,2 desviación estándar en un único año. Este es un impacto mayor si se considera que se dispone de solo un factor escolar en la función de la producción educativa: el profesor. Según Wright, Horn y Sanders (1997), en algunos estudios basados en los datos del Sistema de Evaluación del Valor Agregado en Tennessee (*Tennessee Value-Added Assessment*

System, TVAAS) se mostró que la eficacia docente fue el factor determinante más importante en el logro académico de los estudiantes, el que incluso provocaba un efecto mayor que el trasfondo de los estudiantes.

Como indica Ravitch (2010), al concluir que el profesor es el factor más importante que afecta el aprendizaje de los estudiantes y que existe una gran variación en la eficacia docente, las implicancias inmediatas que pueden derivarse de estos estudios es que debería ser posible hacer más para mejorar la educación si nos enfocamos en mejorar la eficacia docente, más que cualquier otro factor. Los hallazgos de aquellos estudios apoyaron la creación del programa federal estadounidense llamado “*Race to the Top*” (“Carrera hacia la cima”). La idea de este programa es generar una carrera para obtener fondos federales entre los estados y en la cual los estados deben adoptar algunas políticas de reforma. Una de esas políticas es utilizar las mediciones del valor agregado de los docentes al logro académico estudiantil, a fin de guiar las decisiones respecto al equipo docente, como compensaciones, ascensos laborales y la opción a un cargo permanente (Department of Education, 2009). Tanto la Ley *No Child Left Behind* como el programa *Race to the Top* han llevado a un aumento de las iniciativas que utilizan estimaciones de los efectos docentes para guiar las decisiones que afectan al equipo docente, pero también ha llevado a un aumento de los estudios que abordan los desafíos y problemas que conlleva utilizar estimaciones en propuestas de alto impacto (Braun, 2005; Koedel & Betts, 2011; McCaffrey et al., 2003; Rothstein, 2009).

De hecho, la posibilidad de utilizar un tipo de estimaciones para recompensar o sancionar a profesores es lo que recalca el interés en la segunda pregunta propuesta por McCaffrey et al. (2003), la que requiere tener una estimación de los efectos de cada profesor.

En este caso, los estudios también hacen estimaciones de los efectos docentes al utilizar variables ficticias para identificar a cada profesor. No obstante, suelen enfocarse en el análisis de la calidad de estas mediciones, especialmente respecto a temas como precisión y estabilidad. La idea es que las consecuencias posibles unidas a los efectos estimados, determinarán los niveles aceptables para la precisión y estabilidad de la estimación.

Una de las muchas preocupaciones relacionadas con la precisión de las estimaciones de la eficacia docente es la influencia de errores de muestreo (McCaffrey et al., 2003). Un buen ejemplo de factores que no guardan relación con la calidad de los docentes pero que sí pueden afectar el logro académico de los estudiantes en un examen específico, es cuando los alumnos tienen un problema familiar. Estos factores no relacionados llevan a errores de muestreo en las estimaciones de valor agregado docente.

Al igual que con cualquier otra medición, cuanto mayor es el tamaño de la muestra, menor es la influencia del error sobre los efectos estimados. Dado que los docentes imparten clases a un número moderadamente reducido de estudiantes, la variabilidad en el error de muestreo en los efectos estimados será probablemente grande en relación a la varianza verdadera de los efectos docentes (McCaffrey et al., 2003).

Lockwood, Louis y McCaffrey (2002) descubrieron que, a menos que la razón de la varianza del error de muestreo respecto a la varianza de los efectos docentes sea menor a 0,1, las estimaciones de clasificación no serán lo suficientemente precisas como para permitir que se distinga entre todos los docentes, sino que solo se podrán identificar los casos más extremos.

Este parece ser un objetivo difícil de alcanzar. En otro estudio, McCaffrey, Lockwood, Koretz, Louis y Hamilton (2004) descubrieron que los errores de muestreo en el efecto docente estimado alcanzaba un rango aproximado de entre un 20% a un 40% de la varianza de los efectos docentes. A pesar de que las clasificaciones estimadas no eran lo suficientemente precisas como para distinguir a los diferentes profesores, los autores fueron capaces de identificar entre un tercio y un cuarto de los docentes como diferentes a la media. Esta información puede ser muy útil para propósitos de diagnóstico, como lo que indica McCaffrey et al. (2003): identificar a profesores que tienen un desempeño bueno o malo es un punto de inicio para que los administradores (como los directores y superintendentes) identifiquen a los profesores que deben someterse a una revisión más detallada. Desde el punto de vista del investigador, estos profesores con efectos docentes altos o bajos también pueden ser el objetivo para llevar a cabo investigaciones cualitativas, a fin de encontrar patrones de enseñanza eficaz.

Dado que muchos encargados de las políticas se muestran interesados en utilizar las estimaciones de los efectos docentes para propósitos de alto impacto, una preocupación principal es también determinar

un estándar aceptable de estabilidad. Una evaluación en donde las estimaciones ubican al profesor en el décimo percentil menor en un año y en el décimo percentil mayor al año siguiente, suele dar suficientes razones para dudar de su calidad. De todos modos, no es posible esperar una estabilidad completa. Si la correlación entre las estimaciones de los efectos de un profesor en dos años consecutivos fuese perfecta, significaría que el profesor más eficaz en un año también debería ser el profesor más eficaz al año siguiente (Goldhaber & Hansen, 2010). Si esto ocurriese, indicaría que la eficacia docente es algo permanente sin variación de año a año.

En algunos estudios, también basados en datos de los Estados Unidos, el foco se centra en el análisis de la estabilidad de las estimaciones de los efectos docentes. McCaffrey, Sass, Lockwood y Mihaly (2009) han obtenido correlaciones de las estimaciones de los efectos docentes desde 0,22 hasta 0,67. Según ellos, estas correlaciones implican clasificaciones de estabilidad moderada: cerca de un tercio de los docentes del quintil superior en un año siguen estando en el quintil superior al año siguiente, mientras que un décimo desciende hasta el quintil inferior. Goldhaber y Hansen (2010) han trabajado con estimaciones de efectos docentes acumuladas durante 10 años y han obtenido correlaciones de 0,32 a 0,59 para los pares de años. Según ellos, el nivel de variación observada es coherente con las mediciones de trabajadores de otras ocupaciones de gran complejidad como la enseñanza, en donde las correlaciones van de 0,33 a 0,40. Estas evidencias parecen confirmar que hay un componente permanente de los efectos docentes, pero ese componente no representa el efecto docente total.

Goldhaber y Hansen (2010) indican que, además del componente constante y los errores de muestreo, entra en juego un componente inconstante en los efectos docentes. El componente inconstante puede captar la variación en el desempeño del profesor a lo largo de los años, pero también puede captar una “buena relación” profesor-alumno en un curso en particular o que un profesor cualquiera estuvo enfermo por una semana durante la temporada de gripe, lo que afectará el efecto docente estimado en un mismo año.

McCaffrey et al. (2009) también comenzaron a analizar las estimaciones de los efectos docentes como la suma de tres variables independientes: el efecto docente constante, los efectos docentes inconstantes y los errores de muestreo. A diferencia de los errores de muestreo y los efectos constantes, en ambos estudios se muestra que no es posible estimar de manera directa la variabilidad de los efectos inconstantes. Más bien, es necesario estimar de manera indirecta la varianza de los efectos docentes inconstantes, como la porción que queda de la varianza de las estimaciones de los efectos docentes, después de restar la porción que se debe a efectos constantes y a errores de muestreo.

Luego de seguir al mismo profesor durante 10 años, Goldhaber y Hansen (2010) encontraron evidencias que los errores de muestreo tienen un rol importante cuando se determina la estabilidad de las estimaciones de los efectos docentes con el paso del tiempo. Los errores de muestreo corresponden a alrededor de un tercio de la variación de los efectos docentes. De la varianza que queda, entre uno y dos tercios se deben a la variación de los efectos constantes, y entre uno y dos tercios se debe a la variación en los efectos docentes inconstantes.

Como ya se indicó, a partir de la segunda pregunta presentada es posible llegar a otra pregunta de interés constante en esta literatura: ¿Cómo se explica la variación en la eficacia docente? En otras palabras, ¿qué características, prácticas, creencias y actitudes del profesor se pueden vincular a efectos docentes mayores?

Se han encontrado varios estudios que apuntan a responder esta pregunta. En otra revisión de la literatura basada en datos de las ciudades o estados de EE.UU., realizada por Hanushek y Rivkin (2004), el nivel de estudios académicos del profesor fue la variable con la evidencia más débil de vínculo con los logros académicos de los estudiantes. También apareció en otros estudios publicados después de la revisión de la literatura que hicieron ellos, como en el caso de Rivkin, Hanushek y Kain (2005), Clotfelter, Ladd y Vigdor (2007) y Croninger, Rice, Rathbun y Nishio (2007).

En estos estudios basados en datos de los Estados Unidos, los investigadores suelen comparar a docentes que tienen un magíster o un doctorado con docentes que solo tienen una licenciatura. De esta manera, en la mayoría de los estudios se ha mostrado que tener un grado de magíster o doctorado no está asociado a efectos docentes mayores.

Por otro lado, existen muchos estudios que presentan evidencia de que la experiencia como profesor tiene efectos positivos y significativos en los logros académicos de los estudiantes (Clotfelter et al., 2007; Croninger et al., 2007; Goldhaber & Hansen, 2010; Rivkin et al., 2005; Rockoff, 2004). A partir de algunas estimaciones de Rockoff (2004) y Rivkin et al. (2005), se ha deducido que el aumento en la experiencia como profesor produce efectos solo en los primeros años de enseñanza, pero no después del tercer año de enseñanza.

Pero quedan muchas dudas sobre las evidencias respecto a la asociación entre la experiencia y la eficacia docente. La principal preocupación es que los profesores con más experiencia por lo general pueden elegir la escuela que quieren y tienden a elegir aquellas con los estudiantes que tienen los mejores resultados (Hanushek, 1986; Hanushek, Kain, O'Brien, & Rivkin, 2005). Esto genera la posibilidad de una relación opuesta: los estudiantes con mejores resultados atraen a profesores con más experiencia profesional, o esta relación de causalidad opera en ambas direcciones. Este tipo de problema solo se puede reducir si es que en más estudios, como el realizado por Rockoff (2004), se encuentran efectos positivos de los años de experiencia docente, mediante la comparación de los docentes dentro de las escuelas o un estudio de los mismos docentes en diferentes escuelas durante un periodo de tiempo.

Otras características, como los puntajes de profesores en pruebas de rendimiento y el tipo de certificación, han recibido bastante atención. Hanushek y Rivkin (2004) señalan que los puntajes de profesores suelen guardar una relación más cercana con los resultados de los estudiantes, más que la experiencia y la educación de los docentes, pero esta evidencia no es muy cuantiosa. Por el momento, en la literatura se identifica una evidencia mixta de los efectos de la certificación en la calidad de los docentes (Ballou & Podgursky, 2000; Clotfelter et al., 2007; Croninger et al., 2007; Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Hanushek et al., 2005; Kane, Rockoff, & Staiger, 2006; Rockoff, Jacob, Kane, & Staiger, 2008).

A pesar de no ser muchos, también hay algunos estudios basados en datos brasileños que se enfocan en el análisis de la relación entre las características del profesor y el logro académico del estudiante. Por ejemplo, aquellos realizados por Barros et al. (2001), Albernaz et al. (2002), Soares, J. F. (2005) y Soares, T. M. (2003).

A excepción de Soares, T. M. (2003), estos estudios se centran específicamente en el grado académico del profesor. Tanto Barros, Mendonça, Santos y Quintaes (2001) como Albernaz, Ferreira y Franco (2002), encontraron evidencia empírica de la relación entre el nivel de estudios académicos de los profesores de educación primaria y el logro académico de los estudiantes. Soares, J. F. (2005) obtuvo evidencia de que los alumnos de profesores con una licenciatura en matemáticas tenían mejores resultados en esta materia escolar. En Brasil, en los estudios se suele evaluar la diferencia entre profesores con un certificado de educación secundaria, con una licenciatura, una especialización (*especialização*), un título de magíster académico y un doctorado, pues en el país todavía hay profesores que no tienen su certificado de educación secundaria, especialmente en algunas áreas de escasos recursos.

La *especialização*, que es bastante común en Brasil y que en este artículo denominamos especialización para diferenciarla del título de magíster académico, (*mestrado*), es un curso de capacitación profesional para graduados universitarios que incluye 360 horas de clases como mínimo. A diferencia de un magíster académico, al final del curso de *especialização* los estudiantes no presentan una tesis sino una monografía.

Cuando se utiliza la función de producción educativa para estudiar el caso de Brasil, el mayor desafío es la falta de datos longitudinales en el país. Solo hace poco comenzaron las municipalidades y estados a recopilar datos que permiten hacer estimaciones sobre los modelos de valor agregado. Por esta misma razón, en ninguno de los estudios basados en los datos brasileños se pudo hacer estimaciones de las mejoras en los logros académicos de los estudiantes, y el análisis se limitó a la relación entre las características de profesores y los puntajes obtenidos por los estudiantes al final del año escolar.

Al ser el primer estudio que entrega estimaciones de los efectos de los docentes según datos de una municipalidad brasileña, en esta investigación se pretende contribuir a los debates de la eficacia docente al ofrecer evidencia de la calidad de tales estimaciones grupales e individuales y su utilidad para subsidiar políticas públicas, a fin de mejorar la enseñanza.

Modelo

Para estimar los efectos docentes de cada profesor, se aplicó el modelo de ajuste de covariables que se expresa en la siguiente ecuación:

$$A_{ijst} = \theta A_{it-1} + \beta \gamma_i + \alpha L_{it} + \tau_j T_{ijt} + \delta_s S_{ist} + \eta_i \quad (1)$$

Esta ecuación especifica el puntaje actual del estudiante i bajo la tutela del profesor j en la escuela s el año t como función de:

A_{it-1} , que representa el puntaje del estudiante i en el año $t-1$;

γ_i , que representa un vector de características fijas del estudiante i , lo que incluye su género y si completó su educación preescolar;

L_{it} , que representa un vector de variadas características del estudiante i en el año t , lo que incluye ingreso familiar, nivel de educación de los padres, edad adecuada según el año escolar que cursa, y participación en un programa para niños que todavía se consideraban analfabetos en la evaluación anterior (llamado PIC – *Programa Intensivo de Ciclo*);

T_{ijt} , que representa un vector de variables ficticias que identifican al profesor j que hizo clases al estudiante i en el año t ;

S_{ist} , que representa un vector de variables ficticias que identifican a la escuela s a la que el estudiante i asistió el año t ;

η_i , que representa el error aleatorio.

El efecto estimado para cada profesor j en el año t se evidencia en el coeficiente de variable ficticia de este profesor en el modelo. Todos los puntajes de los estudiantes están estandarizados en la media cero y varianza uno. Este procedimiento permite analizar la desviación estándar estimada de la eficacia docente, medida en referencia a la desviación estándar de los logros académicos del estudiante.

Los efectos de los docentes y las escuelas se especifican como efectos fijos. El efecto docente luego se mide según el promedio de todos los docentes en una escuela dada. Como indicaban McCaffrey et al. (2003), los efectos fijos y aleatorios tienden a arrojar conclusiones similares sobre la variabilidad de los docentes, pero brindan estimaciones diferentes de los efectos de cada profesor por separado. Dado que el método de efectos fijos utiliza solo a los alumnos de un profesor para estimar su efecto, las estimaciones pueden verse altamente sensibles a los errores de muestreo, pues los profesores suelen enseñar a un número reducido de estudiantes. Por lo tanto, se debe incluir procedimientos para estimar y ajustar los errores de muestreo.

Como ya se indicó, sabemos muy bien que la varianza residual de los puntajes de los estudiantes, la que se debe a errores de medición y a las fuentes de variabilidad en la eficacia aparte de los profesores, escuelas y otros componentes del modelo, provoca errores de muestreo en las estimaciones de los efectos docentes. Por esta razón, parte de la desviación estándar de las estimaciones de los efectos docentes se debe a errores de muestreo.

En este análisis se adoptó el mismo procedimiento que utilizan Aaronson et al. (2007), Goldhaber y Hansen (2010) y Koedel y Betts (2011) para ajustar los errores de muestreo. En este, se presume que el efecto docente estimado en la ecuación (1) es la suma del efecto real de un profesor más un margen de error, que no está directamente relacionado con el efecto real del profesor:

$$\hat{\tau}_j = \tau_j + \lambda_j \quad (2)$$

Luego, la varianza de los efectos docentes estimados puede descomponerse en dos partes: una que corresponde a la varianza real de los efectos docentes y otra que corresponde a los errores de muestreo:

$$Var(\hat{\tau}) = Var(\tau) + Var(\lambda) \quad (3)$$

Al igual que estos autores, se utilizó la media del cuadrado de las estimaciones de error estándar de $\hat{\tau}_j$ como una estimación de la varianza del error de muestreo. Luego se restó esto a la varianza observada de $\hat{\tau}_j$ para obtener una varianza ajustada y, después, la desviación estándar ajustada.

Con el objeto de entender mejor la variabilidad de los efectos docentes, se descompusieron los efectos docentes estimados y la variabilidad de estas estimaciones, tomando en consideración la estabilidad de estas mediciones, según lo proponen McCaffrey et al. (2009).

De esta manera, los efectos docentes estimados en un año en particular pueden descomponerse en:

$$\hat{\tau}_j = \tau_j pers + \tau_j npers + \lambda_j \quad (4)$$

La diferencia entre las ecuaciones (2) y (4) es que, en la segunda, el efecto real del profesor se divide en un componente constante y otro inconstante. El componente constante se refiere a la porción de los efectos que se evidencia en las estimaciones de todos los años y que, por lo tanto, se puede entender como una parte intrínseca de la eficacia docente. El componente inconstante se refiere a la porción de los efectos que es específica a un año.

Este componente transitorio, además de capturar la variación del desempeño de los docentes en las diferentes etapas de su carrera profesional en educación, evidencia otros factores del curso en ese año en particular. En la muestra de este estudio, cada profesor se observa con solo un curso por año. Por lo tanto, no es posible adoptar procedimientos para separar los efectos docentes de los efectos de cada curso. De esta manera, el grupo de estudiantes que conforma el curso es probablemente responsable de una gran parte de este componente transitorio de los efectos docentes estimados. Entonces, la porción de los efectos estimados, que no se debe a errores de muestreo no se llamará “efectos reales del profesor” porque seguramente se ve influenciado por otros factores, aparte del desempeño del profesor, en ese año en particular.

La variabilidad de los efectos docentes estimados se puede descomponer en:

$$Var(\hat{\tau}) = Var(\tau pers) + Var(\tau npers) + Var(\lambda) \quad (5)$$

Dado que se utilizan datos del año 2010 para el análisis principal, y la varianza de los efectos docentes constantes solo está disponible para un grupo limitado de profesores que impartieron clases a un curso de cuarto año de educación primaria en 2009 y 2010, se descompuso la varianza de los efectos docentes estimados en 2010 para este grupo pequeño.

De los resultados de la estimación en la ecuación (1), se calculó la varianza de los efectos docentes estimados para este grupo de profesores. Dado que la correlación entre los efectos docentes de dos años consecutivos se considera como una medición del coeficiente de estabilidad de la eficacia docente, y este coeficiente puede considerarse como la razón entre la varianza de los efectos docentes constantes y la varianza de los efectos docentes estimados, como lo indica McCaffrey et al. (2009), es posible obtener una estimación de la varianza de los efectos docentes constantes.

Este mismo procedimiento utilizado para calcular la varianza de los errores de muestreo para todo el grupo de docentes, puede aplicarse en este caso para este grupo más pequeño de profesores. Si se hace esto, es posible descomponer la varianza de los efectos docentes estimados, según se describe en la ecuación (5).

Una de las principales razones para hacer una estimación de los efectos docentes por medio del uso de los modelos de valor agregado es tratar de averiguar si es que hay características y prácticas de los profesores que guardan relación con estos efectos. Encontrar vínculos entre algunas variables y los efectos estimados puede ser muy útil para guiar las políticas de calidad en la docencia.

Con esto en mente, la idea era realizar una regresión de los efectos docentes estimados para cada profesor por separado en la primera etapa $\hat{\tau}_j$ en un grupo de variables representadas por Z :

$$\hat{\tau}_j = \phi Z_j + u_j \quad (6)$$

Sin embargo, como demostraron Goldhaber y Hansen (2010), las estimaciones de mínimos cuadrados comunes incluyen errores estándares que son muy pequeños porque ignoran el error en la variable dependiente, que es el efecto docente, el cual se obtuvo en la estimación previa. En su lugar, se adoptó un enfoque de mínimos cuadrados generalizado. Este enfoque toma en cuenta la incertidumbre en la variable dependiente al ponderar las observaciones en proporción con la confiabilidad de los efectos estimados de cada profesor (Aaronson et al., 2007; Koedel & Betts, 2007; Goldhaber & Hansen, 2010).

Luego de esto, el paso siguiente fue un análisis de la relación entre los efectos docentes estimados y un grupo de variables basadas en los datos administrativos disponibles y en un cuestionario que respondieron los profesores cuando sus alumnos eran evaluados por *Prova São Paulo*.

Otros estudios con análisis similares, todos con datos de Estados Unidos, suelen incluir variables como el nivel de estudios académicos del profesor y la materia que enseña, los años de experiencia, las certificaciones con las que cuentan y los puntajes que obtuvieron, entre otras. En esta base de datos, solo se cuenta con información sobre el nivel de estudios académicos de los profesores y sus años de experiencia, lo que se incluyó en el modelo, como también las variables para controlar el género y la ascendencia étnica de los profesores.

En el cuestionario para docentes hay casi 200 ítems, con preguntas sobre sus características socio-económicas, su opinión acerca del entorno escolar, su propia práctica en las clases, entre otros temas.

Para componer el modelo de análisis de los efectos docentes, se eligieron las preguntas que estaban directamente relacionadas con los docentes y su trabajo. El modelo incluyó solo las preguntas que pueden considerarse como objetivas, excluyendo aquellas sobre las opiniones y percepciones de los docentes.

Datos

Para este estudio se utilizaron datos obtenidos de los registros del Sistema de evaluación de estudiantes de las escuelas municipales de São Paulo, denominado *Prova São Paulo*. Otra información se obtuvo de los registros administrativos y de cuestionarios entregados a los estudiantes y profesores, al mismo tiempo que estos exámenes se realizaban.

Prova São Paulo utiliza exámenes estandarizados para medir el logro académico de los estudiantes de 3° a 8° año de educación en comprensión lectora y matemáticas. A diferencia de las evaluaciones de los efectos docentes encontradas en la literatura, *Prova São Paulo* se administra en base a un censo en el caso de estudiantes de un curso par y en base a una muestra para estudiantes en cursos impares. Esto significa que existen puntajes para casi todos los estudiantes de 4°, 6° y 8° año de educación escolar, pero que solo se examinó de un 25% a un 40% de estos estudiantes el año anterior. Debido a lo anterior, se eligió a los estudiantes de 4° año de educación primaria para el análisis, ya que este es el curso evaluado con el mayor número de estudiantes que cuentan con el puntaje actual y el anterior: cerca del 40% de todos los estudiantes de 4° año.

Desafortunadamente, esta limitación dejó a cada docente con pocos estudiantes con el puntaje actual y el anterior: cerca de la mitad de todos los docentes de la muestra tiene solo 10 estudiantes. Como un intento de garantizar que exista una cantidad mínima de estudiantes por docente, la muestra de docentes analizada contenía solo profesores con al menos 6 estudiantes que contaran con ambos puntajes. Con el fin de presentar la distinción entre los efectos de la escuela y los efectos docentes, la muestra de docentes estaba compuesta solo por profesores de escuelas que contaban con al menos dos profesores de 4° año de educación primaria en el 2010.

Las principales estimaciones se basan en los datos del 2010 de estudiantes y docentes, donde los estudiantes contaban con los datos de sus puntajes del 2009. Para el análisis de estabilidad, agregamos los datos del 2009 de estudiantes y docentes, además de los datos de los puntajes de los estudiantes del 2008.

En las Tablas 1 y 2 se muestran estadísticas descriptivas de los estudiantes y docentes de 4° año de educación primaria de las escuelas municipales de São Paulo en el 2010. En las primeras dos columnas de estas tablas se muestran las características de la población, a saber, todos los estudiantes y docentes de 4° año en 2010. En la tercera y cuarta columna de estas tablas se muestran las características de los estudiantes y docentes de la muestra, a saber, los estudiantes de 4° año en 2010 que fueron seleccionados para el examen el año anterior, cuando cursaban su 3° año de educación, y sus respectivos docentes. Debido a que los estudiantes del PIC (*Programa Intensivo de Ciclo*) se examinan en base a un censo cuando cursan su 3° año de escolaridad y los otros estudiantes de acuerdo con una muestra, los estudiantes del PIC están sobre-representados en la muestra, en comparación con la población. Sin contar este hecho, no existen otras diferencias importantes entre los estudiantes de la muestra y la población, ni entre los docentes de la muestra y la población. Esta es una evidencia que refuerza la idea de que la muestra no es representativa de un grupo de estudiantes y docentes distintos de la población original. A pesar de que parece no haber diferencias importantes entre la muestra y la población, las inferencias obtenidas con este estudio se restringen a la muestra analizada.

Tabla 1
Estadísticas descriptivas para estudiantes que cursaban 4° año en 2010

	Población		Muestra	
	% de estudiantes	% de respuestas	% de estudiantes	% de respuestas
Mujeres	47%	100%	44%	100%
<u>Edad</u>				
Menor de 10 años	1%	100%	1%	100%
10 años	68%	100%	57%	100%
11 años	22%	100%	29%	100%
12 años	7%	100%	10%	100%
Mayor de 12 años	2%	100%	3%	100%
Participación en PIC	7%	100%	14%	100%
Educación preescolar	57%	69%	56%	72%
<u>Nivel de educación del padre</u>				
Inferior a una educación primaria	22%	71%	24%	72%
Educación primaria	28%	71%	29%	72%
Educación secundaria inicial	21%	71%	21%	72%
Educación secundaria terminal	26%	71%	23%	72%
Educación superior	3%	71%	3%	72%
<u>Nivel de educación de la madre</u>				
Inferior a una educación primaria	16%	75%	18,40%	77%
Educación primaria	29%	75%	30,20%	77%
Educación secundaria inicial	22%	75%	21,50%	77%
Educación secundaria terminal	30%	75%	27,70%	77%
Educación superior	3%	75%	2,30%	77%
<u>Ingreso familiar</u>				
Inferior a R\$ 850,00	43%	56%	45%	57%
R\$ 851,00 a R\$ 1.275,00	31%	56%	30%	57%
R\$ 1.276,00 a R\$ 2.125,00	18%	56%	18%	57%
R\$ 2.126,00 a R\$ 4.250,00	7%	56%	6%	57%
Mayor a R\$ 4.250,00	1%	56%	1%	57%
Número de estudiantes		70.367		25.777
Número de docentes		535		519

Nota: Basado en datos de 2009 y 2010 de *Prova São Paulo*.

Tabla 2
Estadísticas descriptivas para docentes de cursos de 4° año en 2010

	Población		Muestra	
	% de docentes	% de respuestas	% de docentes	% de respuestas
Mujeres	96%	100,00%	96%	100,00%
Blanco o asiático ^a	70%	87%	70%	87%
<u>Edad</u>				
Menor de 26 años	2%	100%	2%	100%
26 a 35 años	17%	100%	18%	100%
36 a 45 años	39%	100%	39%	100%
46 a 55 años	29%	100%	28%	100%
Mayor de 55 años	13%	100%	13%	100%
<u>Grado de educación superior</u>				
Licenciatura	76%	88%	75%	88%
Especialización (<i>Especialização</i>)	23%	88%	24%	88%
Magíster académico (<i>Mestrado</i>)	1%	88%	1%	88%
Doctorado	0%	88%	0%	88%
<u>Experiencia docente</u>				
Hasta 5 años	7%	89%	7%	89%
6 a 10 años	12%	89%	12%	89%
11 a 15 años	17%	89%	17%	89%
16 a 20 años	22%	89%	22%	89%
Más de 20 años	42%	89%	42%	89%
<u>Ingreso familiar</u>				
R\$ 1.000,00 a R\$ 2.000,00	2%	86%	2%	86%
R\$ 2.001,00 a R\$ 3.000,00	16%	86%	15%	86%
R\$ 3.001,00 a R\$ 4.000,00	28%	86%	29%	86%
R\$ 4.001,00 a R\$ 5.000,00	25%	86%	25%	86%
Más que R\$ 5.000,00	29%	86%	29%	86%
<u>Nivel de educación del padre</u>				
Inferior a una educación primaria	30%	87%	30%	87%
Educación primaria	37%	87%	37%	87%
Educación secundaria inicial	13%	87%	13%	87%
Educación secundaria terminal	13%	87%	13%	87%
Universitaria	7%	87%	7%	87%
Número de docentes		2.209		2.055

Nota: Basado en datos de 2009 y 2010 de *Prova São Paulo*.

Resultados

En la Tabla 3 se muestran los resultados de la estimación del modelo de ajuste de covarianza expresado en la ecuación (1). Según se discutió anteriormente, las variables ficticias incluidas en la ecuación permitieron estimar cada efecto docente de forma individual, según los coeficientes de $\hat{\tau}_i$, y también para analizar cuánto varían estos efectos entre los docentes de la muestra. Esto también permitió estimar el impacto de los cambios en la eficacia docente en los puntajes de los estudiantes al analizar la desviación estándar estimada de la eficacia docente, medida en términos de las desviaciones estándar de los logros académicos de los estudiantes.

^a En São Paulo existe una gran parte de la población que son descendientes de japoneses, chinos y coreanos, los que denominamos asiáticos.

Tabla 3
Resultados obtenidos de la estimación de la ecuación (1)

	Comprensión lectora	Matemáticas
Puntajes anteriores	0,645*** [0,009]	0,640*** [0,008]
Mujeres	0,100*** [0,015]	-0,016 [0,015]
Edad apropiada según el curso	0,082*** [0,027]	0,115*** [0,026]
Participación en PIC	-0,044 [0,288]	0,228 [0,256]
Educación preescolar	-0,006 [0,015]	-0,010 [0,015]
<u>Nivel de educación de la madre</u>		
Educación primaria	-0,016 [0,026]	0,052** [0,025]
Ed. secundaria inicial	-0,044 [0,028]	0,039 [0,027]
Ed. secundaria terminal	0,071*** [0,028]	0,117*** [0,027]
Educación superior	0,158*** [0,053]	0,132*** [0,051]
<u>Nivel de educación del padre</u>		
Educación primaria	0,038* [0,023]	0,037* [0,022]
Ed. secundaria inicial	0,002 [0,025]	0,074*** [0,024]
Ed. secundaria terminal	0,049* [0,026]	0,084*** [0,025]
Educación superior	0,126** [0,050]	0,168*** [0,048]
<u>Ingreso familiar</u>		
R\$ 851,00 a R\$ 1.275,00	0,050*** [0,018]	0,038** [0,018]
R\$ 1.276,00 a R\$ 2.125,00	0,085*** [0,022]	0,067*** [0,022]
R\$ 2.126,00 a R\$ 4.250,00	0,146*** [0,034]	0,119*** [0,033]
Mayor a R\$ 4.250,00	0,067 [0,067]	0,058 [0,064]
<u>Desviación estándar</u>		
De las estimaciones de los efectos docentes	0,622	0,621
De las estimaciones de los efectos escolares	0,500	0,483
R ²	0,665	0,677
Número de estudiantes	9.425	9.567
Número de profesores	2.055	2.055

Nota: *** denota un 1% en el nivel de significación; ** denota un 5% en el nivel de significación; * denota un 10% en el nivel de significación.

Como se muestra en la Tabla 3, la desviación estándar calculada de los efectos docentes estimados fue de aproximadamente 0,62, tanto en comprensión lectora como matemáticas. No obstante, se espera que esta medición esté inflada, ya que el margen de error es especialmente problemático cuando las estimaciones del efecto profesor se basan en muestras pequeñas como esta. En estos casos, las variaciones de la muestra podrían sobreestimar o subestimar los efectos docentes, ya que las pocas observaciones con resultados cercanos a los resultados extremos de la distribución podrían influir en los efectos docentes estimados (Aaronson et al., 2007).

Luego de corregir los errores de muestreo, la desviación estándar ajustada de los efectos docentes estimados fue de 0,53 en comprensión lectora y en matemáticas.

Esto aun puede considerarse como una gran variación de los efectos docentes, ya que en 10 estudios con análisis similares, revisados por Hanushek y Rivkin (2010) se obtuvo una desviación estándar ajustada para los efectos docentes entre 0,08 y 0,26 en comprensión lectora y entre 0,11 y 0,36 en matemáticas.

Luego se obtuvo un coeficiente de confiabilidad de estas estimaciones, el que se calculó como la razón entre la varianza de los efectos docentes ajustados y el total de la varianza de las estimaciones, tal como la de McCaffrey et al. (2009).

El coeficiente de confiabilidad que se obtuvo fue de 0,73, similar al obtenido por McCaffrey et al. (2009), el que fue entre 0,6 y 0,8. Todos estos resultados empíricos proporcionan una clasificación estimada que no es lo suficientemente precisa como para distinguir la eficacia entre dos profesores. Según Lockwood et al. (2002), para lograr esta distinción, la razón entre la variabilidad de los errores de muestreo y la variabilidad de las estimaciones debe ser menor a 0,1, lo que significa que el coeficiente de confiabilidad debe ser por lo menos de 0,9.

Tal como lo hicieron McCaffrey et al. (2004), los coeficientes estimados para cada docente se pusieron a prueba para verificar si podían ser distintos de 0, lo que significa que la eficacia docente de los mismos podría ser significativamente diferente a la media.

A pesar de que el análisis de estos autores demostró que entre un tercio y un cuarto de los docentes de la muestra podían considerarse significativamente diferentes de la media, los resultados que se obtuvieron de *Prova São Paulo* indican que solo un 12% de los docentes tuvo efectos estimados significativamente diferentes de la media en comprensión lectora y un 15% de los docentes tuvo efectos estimados distintos a la media en matemáticas.

Para analizar la estabilidad de los efectos docentes individuales, se hizo una correlación entre los efectos estimados para el mismo docente en dos años seguidos (McCaffrey et al., 2009). Esta es la manera más simple de calcular el coeficiente de estabilidad, y fue el único procedimiento disponible dada la limitación de nuestros datos. Si se contara con 12 años de datos en la muestra de docentes, como en el análisis de Goldhaber y Hansen (2010), se podría haber realizado un análisis más sofisticado.

Solo se estimaron los efectos docentes en dos profesores de 4° año, tanto en 2010 como 2009, para así contar con dos mediciones para cada profesor. De los 2.055 docentes que impartieron clases a alumnos de 4° año en 2010, solo 714 habían sido docentes de 4° año en el 2009. Otros docentes estaban ya sea en su primer año de enseñanza, o lo más probable, enseñando en otros cursos.

Debido a que la nueva muestra fue diferente a la original, valió la pena verificar si la desviación estándar de los efectos docentes estimados para este grupo en 2010 no era considerablemente diferente a la de la estimación inicial. La desviación estándar de los efectos docentes estimados para este grupo fue de 0,64 en comprensión lectora y 0,61 en matemáticas, cercano al 0,62 obtenido del grupo original en ambas áreas. La desviación estándar ajustada de los efectos estimados fue de 0,54 en comprensión lectora y 0,51 en matemáticas, también cercana al 0,53 obtenido del grupo original en ambas áreas. Estos resultados refuerzan la idea de que la variabilidad de los efectos docentes de esta nueva muestra podría representar la varianza de la muestra original de docentes.

La correlación entre los dos años de efectos docentes estimados fue de 0,097 en comprensión lectora y 0,081 en matemáticas. Estas correlaciones demuestran una estabilidad mucho menor de los efectos

docentes que aquellas obtenidas por McCaffrey et al. (2009), a saber entre 0,22 y 0,67, o Goldhaber y Hansen (2010) que obtuvieron entre 0,32 y 0,59, cuando estos autores también trabajaron con correlaciones entre las estimaciones de pares de años seguidos.

Con el fin de intentar comprender mejor la variabilidad de los efectos docentes estimados en este estudio, se proporciona una descomposición de esta variabilidad para el grupo de docentes de 4° año, tanto en 2009 como en 2010. La descomposición se exhibe en la Tabla 4.

Tabla 4
Descomposición de la varianza de los efectos docentes estimados

	Comprensión lectora		Matemáticas	
	Varianza	%	Varianza	%
$Var(\hat{\tau})$	0,403	100%	0,368	100%
Descomposición:				
$Var(\tau_{pers})$	0,039	9,7%	0,030	8,1%
$Var(\tau_{pers})$	0,250	61,9%	0,230	62,7%
$Var(\lambda)$	0,114	28,4%	0,108	29,2%

Nota: Basado en datos de 2008, 2009 y 2010 de *Prova São Paulo*.

Solo un 9,7% de la varianza de los efectos docentes estimados en comprensión lectora y un 8,1% de la varianza de los efectos docentes estimados en matemáticas se deben a efectos constantes. Esto significa que solo es posible asegurar que cerca del 9% de la varianza de los efectos docentes estimados se debe a la parte de efectos docentes que no cambió en el transcurso de los años. Estos valores pueden considerarse como límites inferiores para la variación de los efectos estimados que se deben realmente a profesores individuales, y pueden verse como efectos docentes verdaderos.

De estas evidencias empíricas se deduce que si se mueve la distribución de los efectos docentes a una desviación estándar más alta, esto aumentaría la distribución de los puntajes de los estudiantes al menos en una desviación estándar de 0,062 en comprensión lectora y en una desviación estándar de 0,059 en matemáticas.

Estos datos no permiten distinguir entre la parte de la varianza de los efectos docentes, que son inconstantes debido a los cambios en el desempeño docente con el tiempo, y la parte que se debe a la calidad de la clase específica, o la parte que se debe a la interacción o “la química” entre el profesor y sus alumnos. Es posible que la totalidad de los efectos inconstantes se deban a la eficacia docente. Por lo tanto, el límite superior para la varianza de efectos docentes verdaderos es de 71,6% en comprensión lectora y 70,8% en matemáticas. Esto significa que el aumento de una desviación estándar aumentaría la distribución de los puntajes de los estudiantes en una desviación estándar en la distribución de efectos docentes máxima de 0,45 en comprensión lectora y 0,43 en matemáticas.

Aunque posible, es poco probable que toda la variación de los efectos docentes inconstantes se deba solo a la eficacia docente. Por lo que puede ser que se esté sobrestimando el límite superior de la variación de los efectos docentes verdaderos.

En un intento por simular una distribución más probable de la variación del efecto docente inconstante, se puede asumir que los efectos docentes inconstantes se deben igualmente a: (a) docentes, (b) clases y (c) la interacción entre docentes y clases. Luego, cada uno de estos tres factores correspondería a un tercio de la variación debido a componentes inconstantes. En esta simulación, la parte de la varianza de los efectos docentes verdaderos sería de 30,3% en comprensión lectora y 29% en matemáticas. En esta situación hipotética, un aumento en la desviación estándar de los efectos docentes significaría un aumento de 0,19 en la desviación estándar de los puntajes de los estudiantes en comprensión lectora y de 0,18 en

matemáticas. Esta estimación parece ser más realista de los diferentes efectos que tienen los docentes en el logro académico de sus estudiantes.

A continuación, se intentó estimar la relación que hay entre las características y prácticas docentes y los efectos docentes estimados de la primera etapa. Los resultados se muestran en la Tabla 5.

Tabla 5
Resultados de la estimación para la relación entre las características y prácticas docentes y los efectos docentes

	Comprensión lectora	Matemáticas
Mujeres	0,196** [0,088]	[0,064] [0,083]
Blanco o asiático	-0,075* [0,039]	0,026 [0,039]
Especialización (<i>Especialização</i>)	-0,103*** [0,040]	-0,072* [0,041]
Magíster académico (<i>Mestrado</i>)	-0,254 [0,186]	0,030 [0,178]
Desde 1 año a 5 años de experiencia docente	0,219 [0,220]	0,067 [0,224]
Desde 6 año a 10 años de experiencia docente	0,201 [0,216]	0,154 [0,221]
Desde 11 años a 15 años de experiencia docente	0,153 [0,216]	0,097 [0,220]
Desde 16 años a 20 años de experiencia docente	0,190 [0,217]	0,154 [0,221]
Más de 20 años de experiencia docente	0,146 [0,216]	0,118 [0,220]
Desde 1 año a 5 años de experiencia docente en esta escuela	-0,019 [0,051]	0,040 [0,051]
Desde 6 años a 10 años de experiencia docente en esta escuela	0,039 [0,061]	0,053 [0,060]
Desde 11 años a 15 años de experiencia docente en esta escuela	-0,000 [0,066]	0,002 [0,065]
Desde 16 años a 20 años de experiencia docente en esta escuela	0,074 [0,088]	-0,005 [0,086]
Más de 20 años de experiencia docente en esta escuela	[0,026] [0,106]	-0,026 [0,100]
Enseña en más de una escuela	-0,031 [0,037]	-0,008 [0,037]
Tiene una actividad profesional diferente a la de enseñar en educación primaria	0,045 [0,110]	-0,109 [0,105]
Dedica desde 4 a 8 horas por semana al trabajo pedagógico fuera de la escuela	0,055 [0,047]	0,025 [0,046]
Dedica más de 8 horas por semana al trabajo pedagógico fuera de la escuela	0,071 [0,058]	-0,025 [0,057]

Utiliza computador e Internet en clases	-0,013 [0,038]	0,020 [0,038]
Convoca a reuniones a los apoderados para hablar sobre sus hijos	0,101 [0,144]	0,145 [0,144]
Asigna tareas para la casa frecuentemente	0,039 [0,046]	0,106** [0,045]
Utiliza Libros de Apoyo una o dos veces a la semana	0,219* [0,125]	-0,027 [0,122]
Utiliza Libros de Apoyo tres o cuatro veces a la semana	0,176 [0,126]	-0,016 [0,122]
Utiliza Libros de Apoyo diariamente	0,284** [0,140]	0,016 [0,137]
Cubrió desde un 50% a un 70% del currículo	0,119 [0,247]	0,245 [0,215]
Cubrió más del 70% del currículo	0,120 [0,246]	0,255 [0,212]
Dedica desde 20 a 30 minutos a la organización de los estudiantes	0,060 [0,081]	-0,109 [0,081]
Dedica desde 10 a 20 minutos a la organización de los estudiantes	0,049 [0,070]	-0,086 [0,072]
Dedica menos de 10 minutos a la organización de los estudiantes	0,109 [0,075]	-0,040 [0,077]
Constante	-0,784** [0,375]	-0,436 [0,368]
R ²	0,031	0,024
Número de docentes	1.161	1.155

Nota: *** denota un 1% en el nivel de significación; ** denota un 5% en el nivel de significación; * denota un 10% en el nivel de significación.

Los resultados de estas estimaciones deben analizarse con cuidado. En primer lugar, el número de docentes en la muestra se redujo a casi la mitad desde la primera a la segunda etapa, debido a la falta de respuestas de los docentes en el cuestionario. En segundo lugar, el análisis se basa en datos transversales donde se realiza un análisis de regresión entre los efectos docentes de un año y las características y prácticas docentes recogidas en ese mismo año. Esto significa que los resultados proporcionan evidencias de la asociación entre las variables dependientes e independientes, pero no permiten establecer relaciones causales entre ellas.

Asimismo, el R² del valor más alto de las estimaciones fue de 0,11, lo que demuestra que la mayor parte de la variación de los efectos docentes no se explica por las características y prácticas docentes disponibles. Esta evidencia es similar a la encontrada por Aaronson et al. (2007), y Koedel y Betts (2007). Sin embargo, según señalan estos autores, el R² es una subestimación del poder explicativo de esta relación, pues una parte considerable, quizás un tercio de la variación en $\hat{\tau}_j$, se debe a errores de muestreo.

De acuerdo con lo que Aaronson et al. (2007) hicieron, si se multiplica el total de la suma de los cuadrados por una tasa del 50% que representa los errores de muestreo, lo que los autores consideran

que es moderado, el valor del R^2 sería el doble. Aun así, sería como máximo 0,22, lo que aun puede considerarse como un valor bajo para el poder explicativo de una regresión.

Los resultados no indicaron ninguna relación entre la experiencia de los docentes y la eficacia docente, una de las variables tradicionales que se utilizan en estudios de este tema. Por otro lado, se encontró una asociación negativa y estadísticamente relevante entre poseer una especialización (*especialização*) y los efectos docentes, tanto en comprensión lectora como en matemáticas.

Se esperaba una relación positiva o ninguna relación entre estas variables, pero no una negativa y estadísticamente relevante. Este resultado muestra que los docentes con una especialización (*especialização*) tienen, en promedio, menos efecto en el rendimiento de los estudiantes. Los docentes con grados académicos de magíster, por otra parte, tiene un efecto similar al de los docentes que solo poseen una licenciatura.

Debido a que los docentes pueden realizar diferentes tipos de especializaciones (*especialização*) con diferentes temas y enfoques, es difícil encontrar una explicación para estos resultados que sea válida para todos los tipos de estudios. Ya que se trata de la asociación más fuerte que se encontró en el análisis, vale la pena intentar recolectar información que permita explicar estos resultados. Esta iniciativa debería incluir la recolección de información de las características comunes de los docentes que participaron en un programa de especialización (*especialização*). Por ejemplo, se puede partir con el análisis de los criterios de selección de docentes para los cursos ofrecidos o proporcionados por el gobierno. Estos criterios pueden ser la causa de que se elija a docentes menos eficaces para participar en los cursos, y es probable que estos cursos no ayuden a mejorar la eficacia de dichos docentes.

Los docentes que asignan tareas en comparación con otros presentan con frecuencia un mayor efecto docente en matemáticas. Además, los docentes que utilizan Libros de Apoyo con más frecuencia presentan un mayor efecto docente en comprensión lectora. Los Libros de Apoyo se utilizan como un material especial para guiar la práctica docente en aquellos temas en que los estudiantes necesitaron más ayuda según pruebas anteriores. De acuerdo con los resultados, también sería interesante realizar más estudios y en más detalle acerca del uso de Libros de Apoyo en las clases de comprensión lectora, para así entender por qué los docentes que los utilizan de manera más frecuente son más eficaces. Existen varias explicaciones posibles, por ejemplo, la calidad de este tipo de recurso pedagógico en comparación con otros recursos que utilizan los docentes, la participación de docentes en cursos de capacitación para el uso de Libros de Apoyo, o el hecho de que este recurso tiene un carácter estructural que ayuda a los docentes a organizar sus clases, lo que deriva en un mejor uso del tiempo para enseñar a los estudiantes a leer.

Discusión

El objetivo de este artículo fue el de usar un modelo de valor agregado para estimar y analizar la eficacia de una muestra de docentes de 4° año de las escuelas municipales de São Paulo.

A partir de estas estimaciones, se obtuvo algunas evidencias con el fin de contribuir al entendimiento de los efectos docentes en los logros académicos de los estudiantes de *Prova São Paulo* y los desafíos de usar dichas mediciones.

La estimación de los efectos docentes individuales dio como resultado una desviación estándar de los efectos docentes estimados de aproximadamente 0,62, tanto en comprensión lectora como en matemáticas. Incluso, luego de corregir los errores de muestreo, la desviación estándar ajustada de los efectos docentes estimados fue de 0,53 en comprensión lectora y en matemáticas.

Esto aun puede considerarse como una gran variación de los efectos docentes, en comparación con otros análisis similares que se revisaron en este artículo. Además, representa efectos que son demasiado grandes para un solo factor escolar, incluso si se trata de docentes. Si se considera lo anterior, se intentó descomponer la varianza de los efectos estimados, con el fin de separar la parte que se debe a cada profesor y no a otros factores que pueden influir en las estimaciones.

Al descomponer la varianza de efectos docentes, se observó que aproximadamente un 29% de la varianza de los efectos docentes estimados se debía a errores de muestreo. Solo cerca de un 9% de la varianza de efectos docentes estimados se debió a la parte de efectos docentes constantes, y se deben realmente a los efectos de profesores individuales. Por otra parte, cerca de un 62% de la varianza de efectos docentes estimados se debe a la parte de efectos docentes inconstantes. Como estos datos no permitieron distinguir la parte de la varianza de efectos docentes inconstantes que se debe a cada profesor o curso y a la interacción entre el docente y su curso, la varianza de los efectos docentes verdaderos está entre un 9% y un 62% de la varianza del efecto docente estimado. Esto significa que el aumento de una desviación estándar en los efectos docentes derivaría en un aumento de la desviación estándar de 0,062 a 0,45 en los puntajes de los estudiantes en comprensión lectora y de 0,059 a 0,43 en matemáticas.

Según los estudios de Lockwood et al. (2002), McCaffrey et al. (2009) y Goldhaber y Hansen (2010), se puede concluir que estos datos no proporcionan información suficiente para entregar recomendaciones sobre políticas para el equipo docente. Según Lockwood et al. (2002), la razón entre la varianza de los errores de muestreo y la varianza de los efectos docentes estimados debería ser menor a 0.1 para que la clasificación estimada pueda ser lo suficientemente precisa como para que se pueda dilucidar la eficacia docente. Este no fue el caso de este estudio, ya que la razón fue de 0,27.

A pesar de estas limitaciones, del análisis se obtiene que un 12% de los docentes en la muestra presentara efectos que pueden considerarse significativamente distintos de la media en comprensión lectora y un 14% de los docentes en la muestra presentó efectos que pueden considerarse significativamente distintos de la media en matemáticas.

En el análisis de los factores que pueden asociarse con los efectos docentes, se encontró que completar una especialización (*especialização*) era un factor negativo y asociado de manera significativa con los efectos docentes en comprensión lectora y matemáticas. Por otro lado, la frecuencia de asignar tareas para la casa y el uso de Libros de Apoyo mostró ser un efecto positivo y significativo en los efectos docentes.

Según estos resultados, se puede recomendar una investigación más detallada que se base en los docentes que producen efectos distintos a la media. El punto de partida deberían ser sus prácticas docentes, con especial atención a la asignación de tareas para la casa y el uso de recursos pedagógicos.

Este es un ejemplo interesante en el uso de la investigación cuantitativa para generar algunas hipótesis que puedan proporcionar ideas para la investigación cualitativa, lo que dará como resultado una mejor comprensión de la eficacia docente. En conjunto, ambos tipos de investigación tienen un gran potencial para ayudar a definir políticas orientadas a mejorar la eficacia docente.

Dado que los modelos de valor agregado docente pueden contribuir al desarrollo de políticas docentes, se recomienda esforzarse por obtener muestras más grandes de estudiantes, docentes y escuelas, para así mejorar la estimación de los efectos docentes y su utilidad al momento de definir políticas docentes.

El artículo original fue recibido el 2 de mayo de 2013
El artículo revisado fue recibido el 18 de agosto de 2013
El artículo fue aceptado el 22 de noviembre de 2013

Referencias

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teacher and student achievement in Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135. Recuperado de <http://www.press.uchicago.edu/ucp/journals/journal/jole.html>
- Albernaz, A., Ferreira, F., & Franco, C. (2002). Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, 32(3), 453-476. Recuperado de <http://www.memoria.nemesis.org.br/index.php/ppe>
- Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record*, 102(1), 5-27. Recuperado de <http://www.tcrecord.org/>
- Barros, R. P., Mendonça, R., Santos, D. D., & Quintaes, G. (2001). Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, 31(1), 1-42. Recuperado de <http://www.memoria.nemesis.org.br/index.php/ppe>
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models. A Policy Information Center's Report*. Princeton, NJ: Educational Testing Service (ETS).
- Clotfelter, C., Ladd, H., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673-682. doi:10.1016/j.econedurev.2007.10.002
- Croninger, R., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, 26(3), 312-324. doi:10.1016/j.econedurev.2005.05.008
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, 13(42), 1-51. Recuperado de <http://epaa.asu.edu/ojs/>
- Goldhaber, D. D., & Hansen, M. (2010). *Is it just a bad class? Assessing the stability of measured teacher performance*. (CEDR Documento de trabajo 2010-3). Seattle, WA: University of Washington.
- Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141-1177. Recuperado de <http://www.aeaweb.org/jel/>
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. (Documento de trabajo 11154). Massachusetts: NBER.
- Hanushek, E. A., & Rivkin, S. G. (2004). How to improve the supply of high-quality teachers. *Brookings Papers on Education Policy*, 7, 7-25. doi: 10.1353/pep.2004.0001
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267-271. doi:10.1257/aer.100.2.267
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City*. (Documento de trabajo 12155). Massachusetts: NBER.
- Koedel, C., & Betts, J. (2007). *Re-examining the role of teacher quality in the educational production function*. (Documento de trabajo). San Diego, CA: University of Missouri.
- Koedel, C., & Betts, J. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, 6(1), 18-42. doi: 10.1162/EDFP_a_00027
- Lockwood, J. R., Louis, T. A., & McCaffrey, D. F. (2002). Uncertainty in rank estimation: Implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics*, 27(3), 255-270. doi: 10.3102/10769986027003255
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Mónica, CA: RAND Corporation.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A., & Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101. doi: 10.3102/10769986029001067
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The inter-temporal variability of teacher effects estimates. *Education Finance and Policy*, 4(4), 572-606. doi: 10.1162/edfp.2009.4.4.572
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. Nueva York: Basic Books.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458. Recuperado de <http://www.econometricsociety.org/aims.asp>
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252. Recuperado de <http://www.aeaweb.org/aer/index.php>

- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008). *Can you recognize an effective teacher when you recruit one?* (Documento de trabajo 14485). Massachusetts: NBER
- Rothstein, J. (2009). Student sorting bias in value added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537-571. doi:10.1162/edfp.2009.4.4.537
- Soares, J. F. (2005). Qualidade e equidade na educação básica brasileira: Fatos e possibilidades. En C. Brock, & S. Schwartzman (Eds.), *Os Desafios da Educação no Brasil* (pp. 91-117). Rio de Janeiro, Brasil: Nova Fronteira.
- Soares, T. M. (2003). Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados pelo Simave 2002. *Estudos em Avaliação Educacional*, 28, 103-123. Recuperado de <http://www.fcc.org.br/institucional/2012/08/31/estudos-em-avaliacao-educacional-4/>
- United States of America, Department of Education (2009). *Race to the Top Program: Executive Summary*. Recuperado de <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67. Recuperado de <http://old.library.georgetown.edu/newjour/j/msg03680.html>