

DESEMPEÑO EDUCATIVO EN MÉXICO: LA PRUEBA ENLACE*

Raymundo M. Campos Vázquez
El Colegio de México, A.C.

Freddy D. Urbina Romero
Coneval

Resumen: Este artículo tiene como objetivo medir la relación entre los resultados de una prueba estandarizada a nivel nacional, Enlace, con pruebas bimestrales administradas a nivel aula. Los resultados para la asignatura de matemáticas implican que un aumento de una desviación estándar en el aprendizaje acumulado medido en las pruebas bimestrales aumentan en 0.84 desviaciones estándar el resultado de la prueba Enlace. El resultado es robusto a diferentes métodos de estimación por regresión cuantil, efectos fijos a nivel escuela y modelos de efectos mixtos. Por tanto, concluimos que lo enseñado en el aula se refleja en la prueba Enlace.

Abstract: This article measures the relationship between a standardized test applied at the national level (Enlace) with bimonthly test results at the classroom level. The results for mathematics imply that a one standard deviation increase in cumulative learning measured with the bimonthly tests increase the Enlace tests scores by 0.84 standard deviations. The result is robust to different estimation methods like quantile regression, fixed effects at the school level and mixed models. Therefore, we conclude that what it is taught in the classroom is reflected in the Enlace test.

Clasificación JEL/JEL Classification: H40; I20; I21, O10; O54

Palabras clave/keywords: México, educación, pruebas estandarizadas, Enlace, profesores, education, standardized tests, teachers.

Fecha de recepción: 17 XII 2010

Fecha de aceptación: 21 IX 2011

* Agradecemos los comentarios de dos dictaminadores anónimos que ayudaron a mejorar sustancialmente nuestro trabajo. rmcampos@colmex.mx, <http://raycampos.googlepages.com>, furbina@coneval.gob.mx.

Estudios Económicos, vol. 26, núm. 2, julio-diciembre 2011, páginas 249-292

1. Introducción

Desde 2006 México ha instrumentado un sistema de evaluación del logro educativo basado en un puntaje estandarizado que se mide a través de la *Evaluación nacional de logro académico en centros escolares* (Enlace). Esta prueba ha sido desde su creación el estandarte del desempeño educativo en México y de la evolución del mismo. Sin embargo, las pruebas estandarizadas presentan serias dificultades asociadas a la interpretación de los resultados. Por ejemplo, las pruebas estandarizadas pueden estar sujetas a errores, inflación de resultados o bien a mejorar habilidades para elevar el resultado de la prueba sin que afecte el conocimiento permanente.¹

La prueba Enlace está sujeta a dichas dificultades y poco se ha hecho para tratar de darle un sentido a los resultados, y verificar lo que en verdad mide la prueba. El presente trabajo tiene como objetivo medir el grado en que la prueba Enlace está relacionada con las calificaciones bimestrales que obtienen los alumnos a lo largo del ciclo escolar (tanto en matemáticas como en español). El análisis comprende a todos los alumnos en educación primaria del Distrito Federal, DF, para el periodo 2006-2009. Los resultados de Enlace han generado varios cuestionamientos sobre el desempeño de los alumnos y de las escuelas, sobre si se debería o no aplicar la prueba o si los resultados deberían estar sujetos a compensaciones económicas para los docentes (véase Hiriart, 2010; Vargas, 2010). Sin embargo, son pocos los trabajos que analizan los resultados de Enlace que consideren su carácter intertemporal y la comparen con otra medida del desempeño, por ejemplo las calificaciones bimestrales de los alumnos.

Por medio de una solicitud de información se construyó una base de datos que permite dar seguimiento al puntaje obtenido por estudiantes de primaria en la prueba Enlace junto con sus calificaciones bimestrales en un mismo año. Es importante mencionar que relacionamos la prueba Enlace de cada área, ya sea matemáticas o español, con su correspondiente prueba bimestral. Enlace tiene un identificador único para cada individuo, lo que permitió anexarles sus calificaciones bimestrales, cuya información es confidencial, y no se conoce ningún otro estudio que cuente con ella. Con la información es posible analizar las calificaciones bimestrales y de Enlace de un mismo individuo entre 2006 y 2009 para el nivel de primaria.

La prueba Enlace se ha focalizado en evaluar el desempeño educativo de los alumnos. La medición del desempeño académico es la

¹ Este efecto es conocido como *teaching to the test* y ha sido ampliamente investigado en Figlio y Getzler (2002) y Figlio (2006).

piedra angular para la instrumentación de políticas educativas debido a que no sólo permite evaluar al alumno sino también analizar qué factores afectan el desempeño de las escuelas. Por tanto, el presente trabajo representa una importante contribución para entender la relación entre pruebas estandarizadas, como Enlace, y conocimiento en el salón de clase, como las pruebas bimestrales. En la literatura no existen estudios similares por lo que el análisis nos permitirá entender el grado de relación entre pruebas estandarizadas y lo enseñado en el salón de clases.

El Distrito Federal cuenta con más de 3 400 escuelas de educación primaria básica, donde casi dos terceras partes de ellas son de sostenimiento público. Alrededor de 600 mil alumnos cursan del tercer al sexto grado, segmento al que anualmente se le aplica la prueba Enlace. En 2009 se aplicó la prueba a 573 713 alumnos en el Distrito Federal, en la asignatura de español 8% de los alumnos obtuvo un logro excelente, 34.8 alcanzó un logro bueno, 43.9% un logro elemental y el resto tuvo un logro insuficiente. Este comportamiento no ha cambiado radicalmente desde 2006, año en el que 56.6% de los alumnos alcanzó un logro elemental, 27.6% un logro bueno y 3.3% un logro excelente. Enlace tiene como objeto evaluar los conocimientos y las habilidades de los alumnos, no obstante, estos son evaluados cinco veces durante el ciclo escolar por medio de pruebas bimestrales administradas a nivel escuela y grado. En términos de aprendizaje de los alumnos no queda claro cual de las pruebas refleja mejor el desempeño de ellos. Por eso nuestro objetivo es determinar el grado de relación entre Enlace y las evaluaciones bimestrales, en particular, se estudia si las calificaciones bimestrales antes de tomar la prueba Enlace determina, de alguna forma, su puntaje. Lo que indicaría que lo enseñado en el salón de clases es relevante para otras pruebas con diferente dominio.

Para evitar la comparación de diferentes puntajes de las pruebas todos los resultados se estandarizan a nivel grupo y, además, se compara cada asignatura por separado. Después de estandarizar los resultados se estiman modelos donde la variable dependiente es la calificación de Enlace y la variable independiente principal es un promedio de las calificaciones bimestrales antes de tomar la prueba Enlace. Si la transmisión de aprendizaje entre pruebas es alta, entonces esperamos que también lo sea la relación econométrica. Es decir, si ambas pruebas miden el mismo conocimiento y el ordenamiento de los estudiantes en las pruebas fuera el mismo, entonces esperaríamos que la relación entre las pruebas bimestrales y Enlace fuera igual a uno o cercana. Sin embargo, esta hipótesis nula puede resultar inapropiada si

la calificación en la prueba Enlace es determinada asimismo por otros factores. Por tanto, también estimamos modelos donde la variable independiente es la prueba Enlace anterior. Si el aprendizaje en el aula, medido con la calificación bimestral, se transmite a la prueba Enlace entonces esperaríamos que el coeficiente de la calificación bimestral sea cercano al coeficiente de la prueba Enlace anterior. Es importante mencionar que este artículo no tiene como finalidad la causalidad de una prueba sobre otra, sino la relación que mantienen ambas pruebas entre sí.

En lugar de seguir una metodología en particular estimamos la relación de calificaciones de diferentes formas. En primer lugar, estimamos modelos de regresión lineal simple. No obstante, tales estimaciones no toman en cuenta que las escuelas son diferentes entre sí, por lo tanto realizamos estimaciones con efectos fijos a nivel escuela. Por otro lado, puede existir heterogeneidad en los datos para diferentes cuantiles por ello también presentamos regresiones cuantiles. Finalmente, estimamos un modelo de efectos mixtos donde seguimos el trabajo de McCaffrey et al. (2004), en el cual se trata de aislar las diferentes contribuciones de las escuelas y de los docentes utilizando la mayor cantidad de información sobre los puntajes y calificaciones pasadas de los alumnos.

Los resultados encontrados nos llevan a concluir que la prueba Enlace y los exámenes bimestrales mantienen una alta relación entre sí. Esto se debe a que la relación entre la prueba Enlace actual y la prueba Enlace anterior es muy similar a la relación de la prueba Enlace y prueba bimestral inmediata anterior. En general, los resultados de la prueba Enlace actual (o bimestral) y Enlace anterior (o bimestral) para el mismo estudiante es alta. Es decir, las pruebas por sí solas mantienen un patrón intertemporal consistente. En la asignatura de matemáticas, un incremento de una desviación estándar en la prueba Enlace del año anterior explica un aumento de 0.76 desviaciones estándar en la prueba Enlace actual. Asimismo, un incremento de una desviación estándar en la prueba bimestral explica un aumento en la prueba Enlace actual de 0.67 desviaciones estándar. Si tomamos en cuenta los exámenes de 2006 a 2009 los efectos aumentan a 0.88 y 0.84 desviaciones estándar, respectivamente. Los resultados nos llevan a concluir que la relación entre pruebas es relativamente alta y que el ordenamiento en las calificaciones bimestrales se mantiene, en general, en la prueba Enlace.

El resultado es robusto al método de estimación. Estimamos regresiones lineales simples, modelos de efectos fijos y modelos de efectos mixtos. Los resultados entre modelos son sumamente similares entre

sí. También estimamos modelos diferenciados por sexo del estudiante y por tipo de escuela (privada y pública), encontramos pocas diferencias económicamente significativas entre ellas. Asimismo, estimamos regresiones cuantiles para analizar si el efecto de las pruebas es heterogénea a través de la distribución de puntajes; aspecto en el que encontramos poca heterogeneidad y evidencia de reversión a la media en las colas de la distribución.

El resto del artículo está organizado como sigue: en la siguiente sección se presentan una serie de estadísticas que nos permiten tener una idea general de la heterogeneidad de las escuelas, en la tercera sección se da una introducción sobre la prueba Enlace y de las dificultades asociadas a las pruebas estandarizadas. El problema principal es la interpretación de los resultados y los problemas de interpretación de los resultados generados por el tamaño de las escuelas y de los grupos, se observa que, en general, los grupos pequeños alcanzan un puntaje medio mayor que los grupos grandes. En la cuarta, se explica la estrategia de investigación del trabajo, discutimos regresión simple y regresión con efectos fijos, regresión cuantil y modelos de efectos mixtos. En la quinta y sexta sección se presentan los resultados y las conclusiones sobre el desempeño educativo en México, respectivamente.

2. Las pruebas estandarizadas y el desempeño educativo

Koretz (2002) ofrece una visión de las debilidades de los sistemas de evaluación basados en puntajes estandarizados y concluye que dicho tipo de pruebas son una medida incompleta, ya que están sujetas a errores de medición y son susceptibles a inflación (Shepard, 1990). La evaluación del desempeño educativo es reciente en México, varios de sus resultados han sido considerados definitivos sin tomar en consideración los problemas asociados a este tipo de pruebas. La mayoría de las mediciones que provienen del campo de los estudios sociales se encuentran expuestas a errores de medición, en el sentido de que la repetición de un mismo proceso de medición no produce un resultado idéntico (Woodhouse *et al.*, 1996).

Las dificultades de las pruebas estandarizadas no son inherentes únicamente a la estructura de la prueba. El puntaje de logro académico es sumamente volátil y muy susceptible a la inflación. Una vez que el sistema de evaluación se lleva a cabo, las escuelas empiezan a jugar con el sistema para obtener resultados que los beneficien indebidamente (Rubin, Stuart y Zanutto, 2004). Por su parte, Kane

y Staiger (2001 y 2008) señalan que existe la posibilidad de que las autoridades escolares puedan manipular que los estudiantes tomen o no los exámenes de evaluación, con la finalidad de sesgar de manera positiva los resultados.²

Las acciones que llevan a cabo tanto los directivos como los docentes para incrementar el puntaje promedio de los grupos y de las escuelas se conoce como comportamiento estratégico. Sin embargo, tal comportamiento puede ser controlado o restringido. Por ejemplo, en Massachusetts y Colorado las evaluaciones de los alumnos ausentes se consideran como reprobados directamente, mientras que en Florida y Michigan las escuelas son descartadas del certamen cuando hay una gran ausencia de alumnos por grupo. Figlio y Getzler (2002) señalan que las escuelas reaccionan a los incentivos derivados de la evaluación con un incremento en el tiempo destinado a las asignaturas que se presentan en el examen, lo cual reduce el tiempo destinado a las asignaturas ajenas a la evaluación (es decir, hay sustitución entre pruebas). Jacob (2005) encuentra que el puntaje de los alumnos en ciencias decae luego de la introducción del sistema de rendición de cuentas en Chicago, dado que hay una mayor focalización de las escuelas a otras asignaturas. Figlio (2006) ofrece evidencia de un comportamiento inusual en la suspensión de alumnos previo a la aplicación de exámenes. Un resultado de la preparación de los alumnos es que los docentes dedican más tiempo a algunas materias o alumnos. Neal y Schanzenbach (2007) encuentran evidencia sobre este comportamiento entre estudiantes del quinto grado en Chicago, concluyen que los sistemas de rendición de cuentas determinan la cantidad de tiempo que los profesores dedican a los estudiantes con distintas habilidades.³

En principio, en México, con la prueba Enlace no es posible identificar todos estos comportamientos. No obstante, para tratar dichos problemas se hacen tres supuestos sobre las calificaciones y so-

² Al respecto Figlio y Getzler (2002) analizan la probabilidad de que los alumnos modifiquen su estatus para no ser elegibles para la aplicación de la prueba. Encuentran que los alumnos con bajo desempeño tienen una alta probabilidad de cambiar de un estatus de aceptable a no aceptable, previo a la aplicación de la prueba.

³ Otro ejemplo es el de Glewwe, Ilias y Kremer (2010), donde se analiza un programa aleatorio que otorga un beneficio económico al profesor que logre una mejora en los exámenes de los estudiantes. El estudio encuentra que el programa mejora los puntajes de los estudiantes mediante un incremento en el tiempo de preparación para la prueba (no hay efectos para el número de tareas o ausencia del profesor).

bre el comportamiento de las escuelas. Primero, el comportamiento estratégico tiene como único fin incrementar el puntaje medio de la escuela/grupo (inflación). El supuesto se entiende debido a que en varios países las evaluaciones de desempeño educativo están condicionadas a recompensas económicas a los docentes y/o escuelas. Segundo, el comportamiento estratégico en la prueba Enlace no es sistemático e independiente del comportamiento estratégico en las calificaciones bimestrales.⁴ Tercero, el comportamiento estratégico en las calificaciones bimestrales es mínimo dado que los resultados de Enlace son, a diferencia de otros países, únicamente de divulgación. Para prevenir tales dificultades se decidió utilizar al total de escuelas del Distrito Federal en los cuatro años de aplicaciones de la prueba y todas las calificaciones bimestrales de los alumnos en el mismo periodo.

Bajo esos supuestos esperamos que las calificaciones bimestrales sean las que menos ruido contengan. Por lo tanto, el utilizarlas como regresores del puntaje de la prueba Enlace nos dice el grado de transmisión de la prueba bimestral a la prueba Enlace. Más aún, utilizamos las calificaciones bimestrales anteriores a la aplicación de la prueba Enlace. Mientras la estrategia de inflación de las instituciones educativas no sea sistemática, el utilizar toda la información del desempeño educativo de los alumnos brinda mejores estimaciones sobre la evolución del mismo. Sin embargo, como se mencionó anteriormente, no existen incentivos para manipular las calificaciones bimestrales ya que no son públicas y, además, las pruebas Enlace y bimestrales son calificadas por diferentes instituciones. Por ello, nuestro objetivo es estimar la relación de la calificación en la prueba bimestral, antes de tomar la prueba Enlace, con respecto al resultado en esa prueba.

El conocimiento medido en la prueba Enlace es limitado, sólo evalúa (todos los años) las asignaturas de español y matemáticas, la medición incompleta se manifiesta desde la construcción de la prueba, en la definición del dominio y la estructura.⁵ La prueba Enlace ha evaluado en todos sus años de aplicación las asignaturas de español y matemáticas, más una tercera materia rotativa. Aunque las asignaturas de español y matemáticas se consideran las más generales, la restricción del dominio limita, necesariamente, la medición del desem-

⁴ Es decir, el error de medición en la prueba Enlace y en las calificaciones bimestrales no está correlacionado. El supuesto tiene sentido y es factible dado que el profesor que evalúa el desempeño escolar no es el mismo a quien califica, ni influye en el proceso de calificación.

⁵ El dominio se refiere al contenido que debería medirse para cierto grupo, la estructura hace referencia a la delimitación del contenido específico.

peño y evalúa sólo una parte del aprendizaje escolar. La estructura delimita el subespacio generado por el dominio, lo que define la estandarización del examen. Esta delimitación es el grado de dificultad que hace posible la estandarización de las pruebas, dependerá del objetivo de la evaluación si sólo se desea medir lo que se ha enseñado. Aunque lo anterior queda fuera del análisis, debe tomarse en cuenta que la definición de dominio y estructura generan sesgos importantes de medición que son ajenos al desempeño de los alumnos. Ya que el dominio es diferente entre pruebas, puede ser difícil de interpretar las relaciones entre pruebas distintas.

Por otro lado, existen más tipos de retos al analizar pruebas estandarizadas, Koretz (2002) y Kane *et al.* (2002) mencionan tres dificultades principales sobre la medición e interpretación de las pruebas. La primera, es la variación del tamaño de la población en las escuelas. Las escuelas y grupos están conformados por distinto número de estudiantes, por lo que no es pertinente comparar la media del puntaje obtenido, ya que, el número de alumnos, varía entre las escuelas.

El segundo problema es que, el desempeño de los alumnos, se ve afectado tanto por factores persistentes como no persistentes. Estos últimos, sólo afectan el puntaje del alumno al momento de la aplicación de la prueba (ruido), sin influir de manera definitiva en su desempeño. Los factores persistentes tienen un efecto más duradero sobre el desempeño. Ambos factores tienen incidencia directa sobre el aprendizaje y el desarrollo de las habilidades (Abreu y Martín del Campo, 2007); sin embargo, de las pruebas escolares no es posible determinar si las variaciones en los puntajes son resultado de factores persistentes o transitorios.

La tercer, y mayor, dificultad es la misma definición de desempeño educativo. En general se consideran las variaciones en los puntajes como una medida simple del desempeño educativo, no obstante, la variación en un punto sobre el examen de un individuo no proporciona mucha información (Kane *et al.*, 2002). Es decir, la medida del desempeño depende del intervalo de la puntuación y de su número de puntos. Por ejemplo, los puntos de los exámenes bimestrales son números de uno a dos decimales en un intervalo de cinco a diez, mientras que la prueba Enlace maneja puntos hasta con ocho decimales sobre el intervalo 200-800. El uso de demasiados puntos conduce a atribuir un significado inadecuado a las diferencias de puntuación, mientras el uso de muy pocos puntos conduce a una pérdida de información útil (Kolen, 1988).

No se puede tratar en conjunto los problemas anteriores de manera simultánea y satisfactoria. En nuestro caso, cada problema se

trata, en lo que cabe, de acuerdo con nuestras posibilidades. Para tratar la variación del tamaño de la población en las escuelas se utilizan modelos de efectos fijos a nivel escuela y un modelo jerárquico que captura las correlaciones entre e intra escuelas (modelo de efectos mixtos). Asimismo, para las diferencias del número de individuos en los grupos no se utiliza sólo información actual de los alumnos, sino toda la historia posible de cada uno de ellos. Para tratar los factores persistentes y no persistentes es conveniente utilizar información específica sobre los alumnos. Sin embargo, dicha información está restringida como confidencial y protegida bajo la ley. Las características que se tienen por individuo son únicamente aquellas a las que se pueden acceder mediante el identificador único.⁶ Para el problema de la definición y la medición del desempeño educativo, dado que el puntaje de las calificaciones bimestrales está definido bajo un número diferente de puntos, respecto a la prueba Enlace, se estandarizaron ambos puntajes por año/grupo, para hacerlos lo más comparables posible.

3. La prueba Enlace

Al ser Enlace una prueba estandarizada ha sido definida como una prueba objetiva para medir el grado de conocimiento y habilidades, de los alumnos, que se establecen en los programas oficiales de estudio (SEP, 2011). En México es la única prueba de aplicación masiva y censal que desde 2006 y hasta 2009 ha contado con la participación de más de 100 000 escuelas. Como en muchos otros países la prueba evalúa las asignaturas de español y matemáticas, desde 2008 se incluyó una asignatura rotativa.⁷ La aplicación de la prueba se realiza

⁶ Se ha mostrado que los factores socioeconómicos son determinantes importantes del puntaje. Por ejemplo, los artículos de De Hoyos, Espino y García (2010) y Mizala, Romaguera y Urquiola (2007) muestran la importancia de esos factores. Lamentablemente el presente estudio no cuenta con esa información. De Hoyos, Espino y García (2010) utilizan los cuestionarios de contexto, pero no tienen información sobre las calificaciones bimestrales. Estudios futuros deben analizar la robustez de los resultados presentados aquí al incluir variables socioeconómicas a nivel hogar o más características a nivel individual. De cualquier manera, los cuestionarios de contexto se aplican a una sola muestra y no al universo, esta muestra no es seguida a través del tiempo, por lo que no es posible utilizar los cuestionarios de contexto en nuestro análisis.

⁷ En 2008 se agregó Ciencias Naturales, en 2009 Formación Cívica y Ética, Historia se evaluó en 2010 y Geografía en 2011.

a los alumnos que cursan del tercer al sexto grado de educación primaria.

La aplicación de la prueba Enlace es coordinada por la Dirección General de Evaluación de Políticas (DGEP) de la Secretaría de Educación Pública (SEP). En cada escuela participa uno de los 133 mil coordinadores externos comisionados por la SEP para la aplicación de la prueba. Adicionalmente, el director de la escuela invita a padres de familia para participar como supervisores y a personas externas que fungen como observadores. Como medida de seguridad los docentes no pueden supervisar a sus propios alumnos, así como los padres de familia no pueden supervisar a sus hijos. Sin embargo, no hay un requerimiento de asistencia mínima de alumnos por grupo. Para garantizar la transparencia se utilizan los métodos *K-index* y *Scrutiny* que determinan la probabilidad de copia basados en los trenes de respuesta.⁸ En caso de que detecten alguna irregularidad ésta se registra en los reportes individuales del alumno y se omite de los cálculos.

La prueba Enlace se aplica de manera regular entre el tercer y cuarto bimestre del ciclo escolar, los resultados son publicados en el mes de octubre y no están sujetos a compensaciones monetarias o en especie, ni para los docentes, ni para las escuelas, pero se utilizan como indicadores de desempeño.

3.1. Características de Enlace

La base de Enlace para los cuatro años de aplicación cuenta con 2 333 304 individuos para el Distrito Federal distribuidos de manera uniforme en cada año. La base de las características de las escuelas cuenta con alrededor de 3 128 escuelas del Distrito Federal, con información sobre el turno (matutino, vespertino y nocturno) y tipo de sostenimiento (públicas y privadas). La base de calificaciones bimestrales no se tiene de manera física, sin embargo, mediante la clave del ciclo escolar, de la escuela (CCT), del grado y del identificador único de los alumnos, se juntan los resultados de la prueba Enlace y las calificaciones bimestrales. Toda la información fue solicitada a través del Instituto Federal de Acceso a la Información Pública y Protección de Datos, IFAI, aunque la respuesta fue relativamente rápida, la información del identificador único fue un proceso de tres

⁸ Estos paquetes de *software* de *detección de probabilidad de copia* tienen como base patrones de respuestas incorrectas similares.

meses, desde su solicitud, debido a que está clasificada como personal y, por lo tanto, protegida bajo la ley de acceso a la información. No obstante, la Dirección General de Planeación y Programación de la SEP accedió a proporcionar esa información y, finalmente, con asistencia de la Administración Federal de Servicios Educativos del Distrito Federal se juntaron los resultados de Enlace con las calificaciones bimestrales de la mayoría de los alumnos del Distrito Federal. Sin embargo, un número significativo de observaciones fueron eliminadas al encontrarse incongruencias en el identificador⁹ o por que las escuelas no presentaron información en los cuatro años seguidos.¹⁰ Como se ha mencionado, las escuelas eliminadas de nuestra base final no tienen registros en, al menos, uno de los cuatro años de aplicación de la prueba Enlace. No obstante, no tenemos registros de que alguna escuela no haya presentado al menos un examen en los cuatro años. Por lo cual, no se considera que haya problemas de autoselección en nuestros datos. El número de escuelas que deciden no participar en al menos uno de los cuatro años de la aplicación de la prueba Enlace es de alrededor de 150, de más de 3 000 escuelas registradas, por ello no se contemplan posibles sesgos por problemas de autoselección.

A esta base se agregan una base de características de las escuelas que incluye variables como: número de alumnos inscritos en cada grado escolar, total de aulas asignadas a cada grado, número y tipo de docentes, personal administrativo, número de alumnos que reprobaron, que desertaron, etc. Esta base, conocida como el formato 911, fue proporcionada por la SEP para todas las escuelas públicas y privadas desde el año 2000. Además contamos con información sobre si la escuela pertenecía al Programa Escuelas de Calidad (PEC).

Después de juntar las tres bases y eliminar todas las anomalías posibles, la base de datos final queda conformada por 3 076 escuelas. Para dar seguimiento a los alumnos se construye un panel balanceado para cada una de las tres generaciones sobre las que se realiza el análisis. El primer panel (panel 1) tiene información de los alumnos que cursaban en 2006 el tercer grado hasta el sexto grado en 2009, el segundo (panel 2) los alumnos que cursaban el tercer grado en 2007 hasta el quinto grado en 2009 y el tercero (panel 3) los alumnos que cursaron del cuarto grado en 2006 hasta el sexto grado en 2008. No

⁹ Se realizó un análisis sobre la congruencia en los dígitos de la clave, en particular se buscó que hubiera congruencia sobre el estado de nacimiento, el género y la fecha de nacimiento.

¹⁰ Por parsimonia se eliminaron los alumnos que: repitieron algún año escolar, se cambiaron de escuela, no completaron el ciclo escolar.

conocemos ningún otro artículo académico que haya reunido tal cantidad de información, por lo que consideramos que el presente estudio es una contribución significativa a la literatura sobre la educación en México.

Como ya se mencionó, el objetivo es relacionar el efecto de las calificaciones bimestrales sobre la prueba Enlace. Para disminuir el error de medición en las pruebas bimestrales utilizamos el promedio de las calificaciones bimestrales en los tres primeros bimestres, justo antes de aplicarse la prueba Enlace. Asimismo, cada calificación y el puntaje de Enlace fueron estandarizados a nivel grupo, para que ambas medidas estuvieran en la misma escala de medición, y facilitar así su comparación. Es importante señalar que la información con la que contamos sobre las calificaciones de los alumnos es bastante completa, aunque no tenemos información particular sobre las características de los alumnos, salvo género, año y lugar de nacimiento.

3.2. *Estadística descriptiva*

Se juntan las cuatro bases de datos de la prueba Enlace (2006-2009) y sólo se mantienen las escuelas que presentan resultados los cuatro años consecutivos. En esta subsección presentamos resultados generales sobre la estructura de los datos finales. En el cuadro 1 se muestra la distribución de las escuelas por delegación,¹¹ se observa que Iztapalapa abarca casi 20% de las escuelas del Distrito Federal, seguida por la Gustavo A. Madero con 15 por ciento. Las delegaciones Milpa Alta, Cuajimilpa y Magdalena Contreras son las que tienen un menor número de escuelas.

La distribución de los alumnos por delegación es relativamente similar a la distribución de las escuelas. Iztapalapa tiene el mayor número de observaciones, mientras que Milpa Alta tiene el menor. De las 3 076 escuelas con las que contamos con suficiente información, 33.2% son particulares. En el turno matutino se concentra 76% de las escuelas. En el turno vespertino y nocturno el servicio educativo lo prestan, en su mayoría, las escuelas públicas. La delegación Benito Juárez concentra un mayor número de escuelas privadas que públicas, en tanto que en la de Milpa Alta es a la inversa.

¹¹ AZC: Azcapotzalco, ALV: Álvaro Obregón, MAG: Magdalena Contreras, COY: Coyoacán, TLH: Tláhuac, MIG: Miguel Hidalgo, CAJ: Cuajimilpa, TLP: Tlalpan, MIL: Milpa Alta, GAM: Gustavo A. Madero, XOC: Xochimilco, VEC: Venustiano Carranza, IZTC: Iztacalco, BEN: Benito Juárez, IZTP: Iztapalapa y CUA: Cuauhtémoc.

Cuadro 1
Composición de las escuelas. Distribución por sostenimiento, turno y composición de alumnos

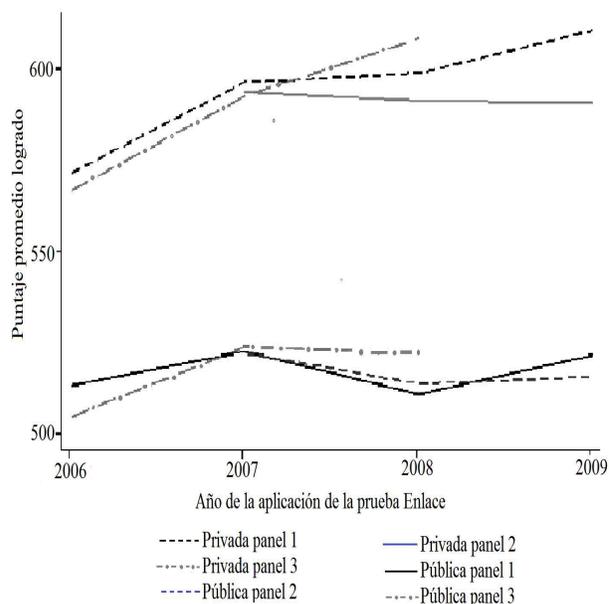
	<i>Composición de las escuelas</i>			<i>Composición de los alumnos</i>			
	<i>Núm. de escuelas escuelas</i>	<i>Escuelas privadas %</i>	<i>Escuelas matutinas %</i>	<i>Núm. de alumnos 2006</i>	<i>Asistencia escuelas privadas</i>	<i>Núm. de alumnos 2009</i>	<i>Asistencia escuelas privadas</i>
AZC	156	21.8	76.9	26 947	19.5	26 294	18.6
COY	204	42.6	85.8	35 538	24.5	35 551	24.2
CAJ	71	45.1	74.6	14 458	35	15 208	33.8
GAM	480	32.7	74.4	85 996	21.6	83 587	19.6
IZTC	146	25.3	74	26 960	12.9	25 854	11.3
IZTP	589	25.8	67.6	125 819	11.6	125 633	10.4
MAG	72	27.8	73.6	15 240	13.2	15 778	11.9
MIL	41	22	65.9	9 503	3.3	9 747	2.8
ALV	228	32.9	73.7	45 193	23.3	45 465	22.8
TLH	110	28.2	65.5	26 752	8.5	26 968	8.1
TLP	216	46.8	78.2	42 206	25.4	43 935	24.9
XOC	118	33.9	69.5	24 330	11.9	27 868	11.3
BEN	142	64.1	97.9	19 723	52	20 762	50.4
CUA	199	28.6	83.9	25 283	26	30 083	22.9
MIG	131	38.9	87	20 692	31.9	20 546	29.4
VEC	173	27.2	79.2	28 478	14.6	28 475	13.7
DF	3076	33.2	76	573 118	19.5	581 754	18.4

Notas: Elaboración propia a partir de los resultados de Enlace. La información de los alumnos se refiere a los que cursaban del tercer al sexto grado de educación primaria. Como se observa la distribución de las escuelas privadas no es homogénea espacialmente, mientras la delegación Milpa Alta concentra el menor número de escuelas privadas, la delegación Benito Juárez cuenta con el mayor número de escuelas privadas. De la misma manera, la distribución de los alumnos depende del número de habitantes en cada delegación, Iztapalapa es la que concentra el mayor número de alumnos.

La gráfica 1 ofrece el puntaje promedio alcanzado por toda una generación de alumnos, separados por sostenimiento (pública y privada), en la asignatura de matemáticas; incluye resultados para cada

panel (panel 1 es de tercer grado en 2006 a sexto grado en 2009, panel 2 es de tercer grado en 2007 a quinto grado en 2009 y panel 3 es de cuarto grado en 2006 a sexto grado en 2008). Como se puede observar el puntaje logrado a lo largo de los años es creciente dentro de las escuelas privadas, en tanto que en las escuelas públicas parece que se regresa a la media de aprovechamiento. También se aprecia como la brecha en el puntaje logrado se hace más grande al pasar los años. La gráfica muestra que el aprendizaje medido por la prueba Enlace no crece para las escuelas públicas.

Gráfica 1
Evolución del puntaje logrado en Enlace por sostenimiento
Asignatura: Matemáticas



Nota: Elaboración propia a partir de los resultados de Enlace.

El comportamiento del puntaje en la prueba estandarizada y de las calificaciones bimestrales se muestran en los cuadros 2 y 3, los cuales incluyen calificaciones por asignatura y grado escolar para el

periodo escolar 2006-2009. Para la prueba Enlace se observa como las medias se mantienen relativamente estables con el paso del tiempo, tanto para la asignatura de español como de matemáticas. Sin embargo, el puntaje del percentil 90 se ha incrementado con el paso de los años, mientras el puntaje obtenido por el percentil 10 ha permanecido más o menos estable en ambas asignaturas. Este comportamiento se debe, principalmente, al incremento del puntaje logrado de las escuelas privadas, tal como se muestra en la gráfica 1.

Cuadro 2
*Puntaje promedio logrado en la prueba Enlace
 por año y grado escolar*

	<i>Español</i>				<i>Matemáticas</i>			
	<i>Promedio</i>	<i>p10</i>	<i>p50</i>	<i>p90</i>	<i>Promedio</i>	<i>p10</i>	<i>p50</i>	<i>p90</i>
<i>2006</i>								
3°	536.7	401.9	553.6	660.5	532.9	399.6	556.4	655.4
4°	537.7	410.0	556.4	655.8	525.3	407.5	544.6	651.7
5°	541.7	408.0	553.9	659.4	526.2	395.2	548.0	655.1
6°	535.1	396.7	553.0	658.5	525.7	394.9	537.2	669.6
<i>2007</i>								
3°	553.2	416.2	559.5	679.0	545.8	409.5	557.1	663.1
4°	556.4	409.5	559.1	686.7	547.0	416.0	555.3	654.8
5°	558.6	419.3	560.1	676.7	546.5	412.2	555.9	657.9
6°	575.7	443.9	564.1	704.5	569.6	426.4	565.9	698.4
<i>2008</i>								
3°	557.7	410.5	565.3	686.4	549.7	399.6	557.2	684.5
4°	549.2	391.7	557.9	692.0	539.1	394.6	540.8	675.7
5°	565.2	414.3	565.0	711.3	539.8	384.1	543.2	688.2
6°	561.5	410.2	564.8	705.9	550.3	390.0	546.9	712.7
<i>2009</i>								
3°	571.5	411.5	582.7	708.4	559.4	412.4	566.2	697.8
4°	550.9	399.3	552.0	700.5	542.7	388.4	546.0	693.3
5°	552.7	401.3	554.2	703.9	540.3	387.6	535.7	694.4
6°	551.1	391.9	557.3	700.3	550.7	397.0	548.5	708.0

Nota: Elaboración propia a partir de los resultados de Enlace.

El cuadro 3 muestra el promedio de las calificaciones bimestrales al tercer bimestre. Las calificaciones presentan un patrón un poco más estable, en cuanto a la cota inferior y superior, en tanto están previamente limitadas. Una particularidad de las calificaciones es que la desviación con respecto a la media es relativamente constante, esto se debe a los pocos puntos que se utilizan para la evaluación bimestral. Como señala Kolen (1988), el bajo número de puntos puede resultar en una pérdida considerable de información, que se observa en la poca variabilidad de las calificaciones bimestrales.

Cuadro 3
*Calificación promedio obtenida en el tercer bimestre
por año y grado escolar*

	<i>Español</i>				<i>Matemáticas</i>			
	<i>Promedio</i>	<i>p10</i>	<i>p50</i>	<i>p90</i>	<i>Promedio</i>	<i>p10</i>	<i>p50</i>	<i>p90</i>
<i>2006</i>								
3°	8.39	7.79	8.37	9.04	8.21	7.64	8.18	8.87
4°	8.32	7.72	8.3	8.99	8.06	7.48	8.02	8.72
5°	8.31	7.73	8.28	8.95	7.95	7.36	7.91	8.60
6°	8.4	7.84	8.39	8.99	8.07	7.50	8.04	8.68
<i>2007</i>								
3°	8.35	7.72	8.36	9.03	8.16	7.53	8.16	8.85
4°	8.31	7.7	8.3	8.97	8.06	7.48	8.03	8.71
5°	8.25	7.66	8.25	8.92	7.91	7.32	7.89	8.56
6°	8.31	7.71	8.31	8.96	8.01	7.39	7.99	8.67
<i>2008</i>								
3°	8.07	7.21	8.25	8.94	7.90	7.04	8.08	8.74
4°	8.07	7.35	8.2	8.87	7.83	7.17	7.94	8.62
5°	8.05	7.3	8.17	8.85	7.73	7.02	7.82	8.53
6°	8.23	7.66	8.27	8.90	7.93	7.35	7.96	8.60
<i>2009</i>								
3°	8.25	7.37	8.32	9.14	8.09	7.22	8.15	8.99
4°	8.29	7.71	8.29	8.93	8.05	7.48	8.03	8.69
5°	8.23	7.62	8.25	8.88	7.92	7.31	7.92	8.58
6°	8.32	7.76	8.31	8.92	8.02	7.46	8.01	8.69

Nota: Elaboración propia a partir de los resultados de Enlace.

Parte de corregir esta pérdida de información es estandarizar los puntajes de Enlace y las calificaciones bimestrales por grupos, se sintetizan en una sola medida ambos puntajes. La estandarización resuelve el problema de que el puntaje de la prueba Enlace se incrementa en el tiempo, debido a que la media del puntaje es igual a cero. Por tanto, si las calificaciones bimestrales miden el mismo aprendizaje que la prueba Enlace sus distribuciones deberían ser idénticas (con igual ordenamiento, por ejemplo). La gráfica 2 muestra la dispersión de la calificación promedio en el tercer bimestre, junto con el puntaje de Enlace, ambas pruebas son para la asignatura de matemáticas, una vez estandarizadas. A la dispersión de los datos se agrega un ajuste no paramétrico (*lowess*) y una línea de 45 grados.¹² Si la relación entre ambos puntajes fuera perfecta, los datos y el ajuste no paramétrico (*lowess*) estarían muy cercanos a la línea de 45 grados.

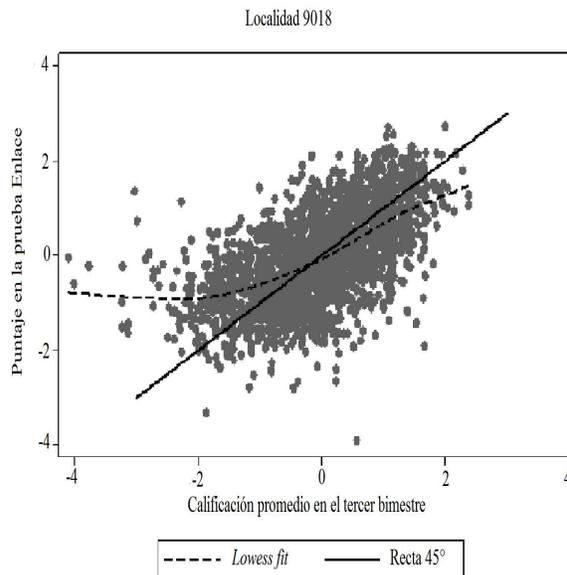
El ajuste no paramétrico muestra en general que: 1) los alumnos con una calificación promedio bimestral por debajo del promedio (o más lejana al promedio en la parte baja de la distribución) alcanzaron en su prueba Enlace un puntaje más cercano al promedio general y 2) los alumnos que tuvieron un buen desempeño en su evaluación bimestral también lo obtuvieron en Enlace. Esto implica que existe una relación positiva entre las calificaciones bimestrales y el puntaje de Enlace para el segmento de alumnos que obtuvieron buenas notas y que hay una baja relación para los alumnos que estuvieron por debajo del promedio. Es decir, si el alumno obtiene una calificación bimestral baja no obtiene una calificación tan baja, relativamente, en la prueba Enlace. Analizaremos este punto con más detalle y con una mayor robustez cuando analicemos regresiones cuantiles.

Como ya se había mencionado, una de las dificultades en la medición es el hecho que el puntaje medio de los grupos depende del número de alumnos que integran un grupo. Por esta razón es importante incluir efectos fijos por nivel escuela o bien modelos de efectos mixtos. La gráfica 3 incluye distribuciones de densidad en el año 2009 por tipo de escuela, donde el eje de las x es el número de alumnos o bien el puntaje en la prueba de matemáticas de Enlace (los resultados para la prueba de español son muy similares). La gráfica muestra que las escuelas privadas concentran grupos con un menor número

¹² La gráfica corresponde a una localidad del Distrito Federal seleccionada aleatoriamente, se utiliza en este ejemplo en lugar de una delegación ya que el tamaño de muestra (número de alumnos) no permite observar el patrón de comportamiento por cuestiones de memoria en la computadora. Las gráficas son muy similares si utilizamos una muestra aleatoria pequeña del Distrito Federal en su totalidad.

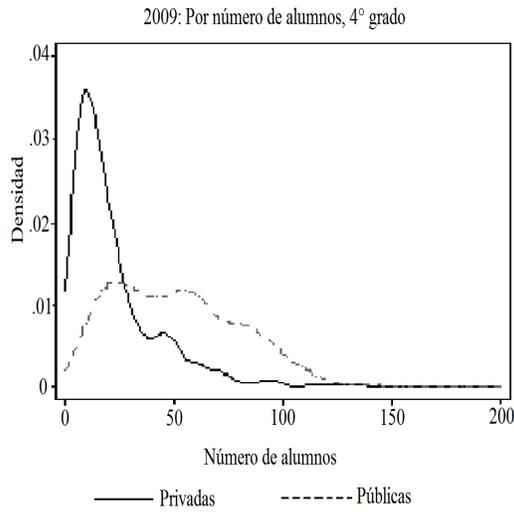
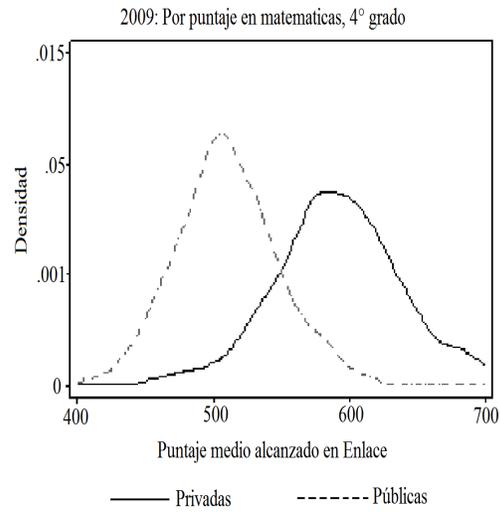
de alumnos que las escuelas públicas, además, el puntaje logrado por las mismas es mayor. Como lo señalamos antes, este patrón de comportamiento es observado por Kane *et al.* (2002) y Kane y Staiger (2002), donde los grupos pequeños tienden a lograr puntajes relativamente más altos que los grupos grandes. Es importante señalar que el tamaño de los grupos está fuertemente influenciado por el tipo de escuela, privada o pública. Las escuelas privadas son las que tienden a tener grupos con menor número de alumnos, por ello es importante diferenciar entre el tipo de escuelas.

Gráfica 2
Enlace y calificaciones bimestrales estandarizadas
Relación básica

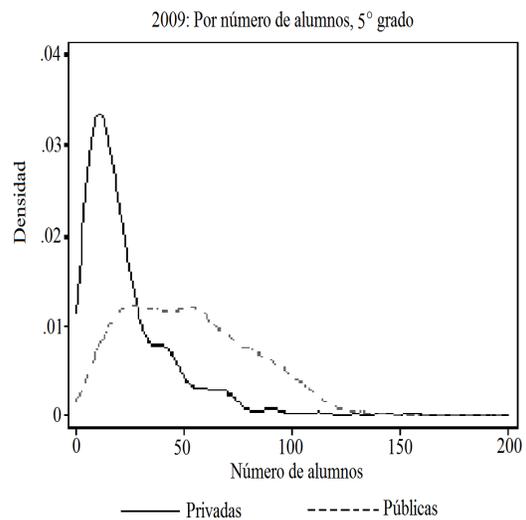
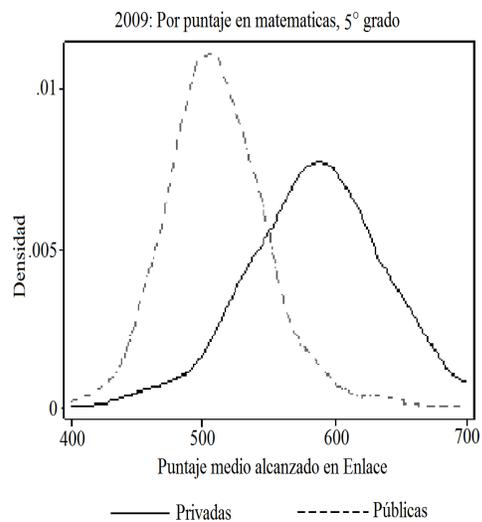


Nota: Se selecciona una localidad aleatoriamente. Por espacio no se presenta información para todos los grados en todos los años, el comportamiento es similar para otras muestras. Elaboración propia a partir de los resultados de Enlace y las calificaciones bimestrales.

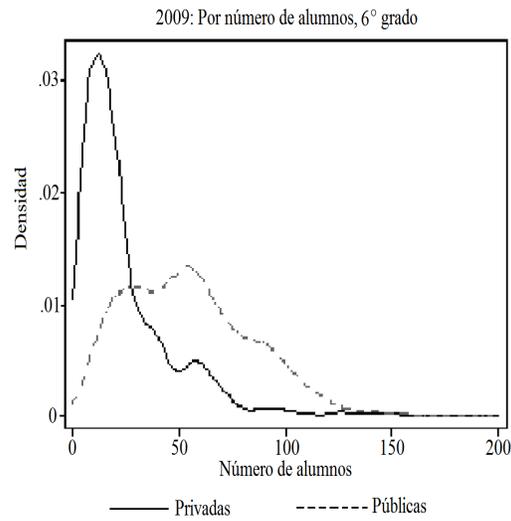
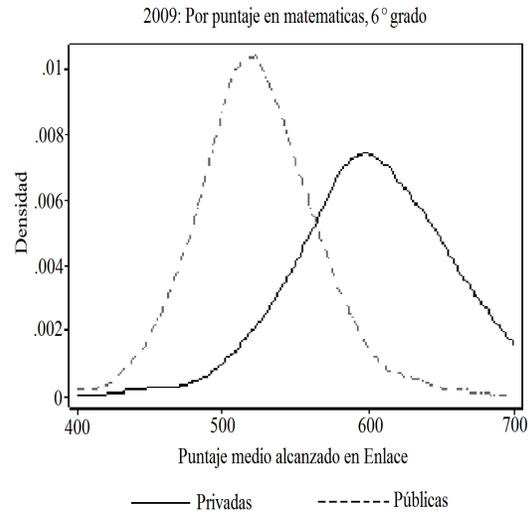
Gráfica 3
Distribución de las escuelas por sostenimiento



Gráfica 3
(continuación)



Gráfica 3
(continuación)



Nota: Elaboración propia a partir de los resultados de Enlace.

4. Estimación empírica

El objetivo es analizar el grado en que el desempeño académico de los alumnos en las pruebas bimestrales se ve reflejado sobre el puntaje de la prueba Enlace. La estrategia a seguir utiliza el hecho de que Enlace se aplica de manera conjunta a todos los individuos entre el tercer y cuarto bimestre del ciclo escolar. En total, tenemos cuatro grados escolares para evaluar las asignaturas de español y matemáticas cada año. Además, contamos con la información de la prueba bimestral en cada bimestre para esos mismos alumnos. Por tanto, utilizamos la información de la prueba bimestral (el promedio en los primeros tres bimestres) antes de que el alumno tome la prueba Enlace para las asignaturas de español y matemáticas.

Como hemos comentado, en lugar de seguir una metodología en particular, presentamos diferentes estrategias de estimación para hacer los resultados más robustos y transparentes. En primer lugar, se sigue una estrategia de regresión lineal simple. Después, estimamos modelos con efectos fijos a nivel escuela con variables de control a nivel individual. Para analizar si la estimación varía para diferentes cuantiles estimamos modelos de regresión cuantil. Finalmente, seguimos la literatura especializada en educación y estimamos modelos jerárquicos o de efectos mixtos.

4.1. Regresión lineal y efectos fijos

¿Cómo sería el comportamiento del aprendizaje si pudiéramos evaluarlo continuamente? Supongamos que tenemos evaluaciones para los periodos $\{t, t-1, t-2, \dots\}$. Se esperaría que el puntaje en t estuviera más relacionado con el puntaje en $t-1$ que en $t-2$ debido a que el puntaje en $t-1$ abarca el aprendizaje en $t-2$. Una manera sencilla de evaluar este comportamiento es estimar el puntaje actual en función de sus rezagos.¹³ Tenemos la siguiente especificación:

$$Enlace09_i = \beta_0 + \beta_1 Enlace08_i + \beta_2 Enlace07_i + \beta_3 Enlace06_i + \varepsilon_i \quad (1)$$

La ecuación (1) considera al mismo individuo en el tiempo y no se agrega ninguna otra variable explicativa (por ejemplo, no incluye

¹³ Hay que recordar que todos los puntajes han sido estandarizados por escuela y grupo.

ninguna variable a nivel escuela). La regresión resume el efecto de los resultados del estudiante en la prueba Enlace anterior a la del año 2009. Si creemos que el aprendizaje en la prueba Enlace es permanente, entonces la suma de los coeficientes debería ser cercano a 1, es decir,

$$\sum_{i=1}^3 \beta_i = 1.$$

Sin embargo, en la mayoría de los estudios no se observa una relación de 1 a 1 entre pruebas, ver, por ejemplo, los resultados en Andrabi *et al.* (2009), Jacob, Lefgren y Sims (2008), Kane *et al.* (2002) y Rothstein (2010). Asimismo, se pueden sustituir las variables de la prueba Enlace por las calificaciones de la prueba bimestral; por lo que, los coeficientes de esa regresión, nos dirían qué tan persistentes son los resultados en las pruebas bimestrales. Lo que hacemos ahora es estimar dos regresiones con la misma variable dependiente (Enlace), pero con diferentes variables explicativas. Así, al estimar una regresión con la misma variable dependiente, lo que se busca es comparar el coeficiente de la prueba Enlace rezagada con el coeficiente de la prueba bimestral inmediata anterior a la prueba Enlace. Como se espera que esta última tenga el mismo dominio a lo largo de los años, las pruebas Enlace deben de tener una relación alta entre sí. Queremos comparar qué tan cercana es la relación de las pruebas bimestrales con la prueba Enlace; si la relación es la misma, quiere decir que la prueba bimestral tiene el mismo dominio que la prueba Enlace y, por lo tanto, el aprendizaje en las aulas sí se ve reflejado en la calificación bimestral y en la prueba Enlace. En términos estadísticos queremos comprobar que

$$\sum_{i=1}^3 \beta_i^{Bimestral} = \sum_{i=1}^3 \beta_i^{Enlace}$$

o bien que el aprendizaje mostrado en Enlace es el mismo que se refleja en las pruebas bimestrales.

Por otro lado, la regresión (1) no toma en cuenta posibles diferencias entre escuelas (tamaño o tipo de escuela), por lo que un posible análisis para estimar coeficientes más robustos incluye efectos fijos en la ecuación (1), así como variables de control. Sin embargo, es importante mencionar que no se tienen variables de control a nivel individual, con excepción de sexo, indicadora de copia en el examen y estado de nacimiento. Variables de control a nivel escuela incluyen

número de profesores, intendentes, profesores en educación física, tecnológica o idiomas, turno y tipo de sostenimiento de la escuela, total de alumnos y si la escuela pertenece al Programa de escuelas de calidad.

De igual forma es importante comprobar si la estimación correcta es mediante efectos fijos, y no efectos aleatorios. Dado que, si la estimación correcta es por medio de efectos fijos, entonces el procedimiento de efectos mixtos que se señala más adelante es incorrecto. La modelación por medio de efectos fijos siempre es consistente, mientras que los parámetros obtenidos por efectos aleatorios no lo son, si la modelación de efectos fijos es correcta. Por otro lado, si la modelación correcta es por medio de efectos aleatorios, los parámetros estimados por efectos fijos son ineficientes. Por tanto, es importante mencionar el resultado de la prueba de Hausman para discriminar entre modelos.

4.2. Regresión cuantil

Estimamos regresiones cuantiles de la ecuación (1) y sus variantes (pruebas bimestrales) para estimar cómo afecta el puntaje de la prueba bimestral o Enlace rezagada, en cierto cuantil q del puntaje de la prueba Enlace actual. A diferencia de mínimos cuadrados ordinarios, regresiones cuantiles sirven para observar la heterogeneidad de los datos. Con mínimos cuadrados ordinarios se obtiene el efecto de la variable independiente sobre la media de la distribución de la variable dependiente. En nuestro caso, las regresiones cuantiles son de utilidad, porque queremos analizar si la transmisión de aprendizaje entre pruebas es homogéneo a través de toda la distribución de puntajes, y no únicamente en la media.¹⁴

La hipótesis nula ahora se refiere a que el aprendizaje entre Enlace y prueba bimestral es el mismo para todos los cuantiles. Es posible que en mínimos cuadrados ordinarios se cumpla, pero no se así para ciertos cuantiles. Por ejemplo, es de esperarse que en los cuantiles bajos (0 al 10) y altos (90 al 100) la relación de aprendizaje sea menor entre pruebas, si consideramos que existe reversión a la media. Por otro lado, si pensamos que existe fuerte persistencia en los estudiantes de altos puntajes, esperaríamos una relación más estrecha; de la misma forma, para estudiantes más bajos, si es que no

¹⁴ Si se desea conocer más sobre regresiones cuantiles se puede consultar el libro de Cameron y Trivedi (2005), capítulo 4.

pueden mejorar sus puntajes. En general, este método sirve para analizar posibles efectos heterogéneos y como método de robustez de los resultados principales.

4.3. Modelos jerárquicos: metodología de McCaffrey et al. (2004)

Debido a la estructura de las escuelas estos modelos son apropiados para capturar la varianza entre e intra escuelas, se asume que los efectos aleatorios no están correlacionados con las variables explicativas. Sean y_{igk} y x_{igk} los puntajes estandarizados de dos pruebas: Enlace y el promedio de las calificaciones bimestrales al tercer bimestre, del alumno i , en el grado g de la escuela k . Dado que x_{igk} es temporalmente anterior a y_{igk} tenemos que la relación entre ambos es (por simplicidad omitimos el término de tiempo):

$$y_{igk} = \pi_{gk} + \beta_{gk}x_{igk} + \varepsilon_{igk} \quad (2)$$

En particular x_{igk} es la calificación promedio de los primeros tres bimestres del año, donde β_{gk} expresa como el desempeño de ésta se ve reflejado en y_{igk} . Si β_{gk} es cercano a cero entonces el aprendizaje en un periodo no influye en el siguiente. Cuando β_{gk} es cercano a uno el aprendizaje en un periodo se mantiene en el siguiente. En el caso de modelos de efectos mixtos a β_{gk} se le puede denominar el efecto alumno-docente debido a que representa la variación en el puntaje de la prueba Enlace atribuible a las calificaciones bimestrales, las cuales dependen del desempeño tanto del alumno como del docente. Es importante señalar que este efecto comprende lo relacionado con las características socioeconómicas del alumno, pero lamentablemente no se cuenta con este tipo de información, razón por la cual dicho efecto (socioeconómico) no es tratado de manera separada. El coeficiente π_{gk} es el puntaje medio del grado g en la escuela k , en tanto que ε_{igk} son todos los efectos no persistentes asociados a cada alumno.

Una de las ventajas de los modelos jerárquicos o de efectos mixtos es que, por su estructura, se pueden modelar los diferentes efectos aleatorios.¹⁵ Este tipo de modelos nos brinda información sobre la dispersión del efecto, lo que ayuda a distinguir mejor entre los efectos de cada una de las escuelas y apreciar las diferencias entre un

¹⁵ $y = X\beta + Zb + \varepsilon$, $b : N_q(0, \Psi)$, $\varepsilon : N_n(0, \Omega)$, donde b y ε son ortogonales.

conjunto de escuelas y otras, por ejemplo, entre las escuelas públicas y privadas. A diferencia de la modelación de los efectos fijos, los efectos aleatorios son más apropiados para representar un *clúster* con estructura jerárquica. Si agrupamos a todos los alumnos de un mismo grado, g , tenemos que el aprendizaje medio del grupo, π_{gk} , puede expresarse como:

$$\pi_{gk} = \delta_k + \mu_{gk} \quad (3)$$

Donde δ_k es el puntaje medio de todos los grados de la escuelas k , μ_{gk} es un efecto aleatorio atribuible a la estructura de cada grado. La ecuación (3) nos dice que el desempeño del grado g de la escuela k es igual al desempeño medio de la escuela, más un efecto único para cada grado y escuela.

Finalmente, el puntaje medio del sistema escolar en el grado δ_k , puede escribirse como:

$$\delta_k = \varphi_0 + \eta_{0k} \quad (4)$$

Donde φ_0 es la media del sistema escolar, mientras η_{0k} es un efecto atribuible a cada escuela. De (2), (3), (4) tenemos:

$$y_{igk} = \varphi_0 + \eta_{0k} + \mu_{gk} + \beta_{gk}x_{igk} + \varepsilon_{igk}$$

Por simplicidad se generan tres paneles balanceados, conformados por tres generaciones diferentes de alumnos (panel 1, 2 y 3). Por ejemplo, un panel considera la información únicamente sobre los alumnos que cursaron del tercer al sexto grado escolar. Lo que implica que el efecto asociado a la estructura del grado se desvanece al no haber otros grados, mientras el efecto de las escuelas se mantiene. El tratar ambos efectos de manera separada complica seriamente el proceso de estimación, la convergencia del modelo puede no darse. La ecuación anterior se reescribe como:

$$y_{igk} = \alpha_0 + \eta_{0k} + \beta_{gk}x_{igk} + \varepsilon_{igk}$$

Al abusar de la notación y que, por construcción, el efecto de los grados ya no está presente, se puede cambiar la notación de η_{0k} por η_{gk} sin cambiar su interpretación como efecto escuela.

$$y_{igk} = \alpha_0 + \eta_{gk} + \beta_{gk}x_{igk} + \varepsilon_{igk} \quad (5)$$

Donde η_{gk} es denominada como el efecto de las escuelas sobre el aprovechamiento. Este, junto con el efecto de los alumnos-docentes, β_{gk} , son modelados como efectos aleatorios donde $\beta_{gk} : N(0, \sigma_\beta^2)$ y $\eta_{gk} : N(0, \sigma_\eta^2)$. La ecuación (5) resume toda la información del puntaje del alumno i en una ecuación jerárquica que, por construcción, mantiene agrupaciones por grupos y escuela, el desempeño educativo es una combinación de efectos por grado y escuela. Los coeficientes η_{gk} y β_{gk} nos permiten identificar los efectos asociados a la escuela, al grupo y a los alumnos-docentes. En resumen, η_{gk} es el efecto aleatorio a nivel escuela que nos indica cómo se ve modificado el desempeño del alumno en la prueba estandarizada cuando éste pertenece a la escuela k . Es importante mencionar que, al seguir a los mismos individuos en el tiempo y después de estandarizar los puntajes, el efecto que tiene el grado o grupo sobre el desempeño de los alumnos queda absorbido en el efecto de las escuelas.¹⁶ Finalmente, β_{gk} es el efecto aleatorio que es atribuible a los alumnos-docentes, y queda reflejado por el desempeño del alumno en la prueba x sobre la prueba y .

Similar a la ecuación (1), la ecuación (5) puede ser aumentada para incluir los rezagos de las pruebas anteriores. Por tanto, nuestros parámetros de interés son las sumas

$$\sum_{j=3}^g \beta_j = \beta_0,$$

que son las contribuciones de los alumnos-docentes y donde g es el número de rezagos. Si $\beta_0 \equiv 1$, entonces el aprendizaje es perpetuo, es decir, todo el aprendizaje depende del efecto alumno-docente. Si

¹⁶ Observemos que el efecto de grado o grupo se interpretaría como el efecto que tiene el grado g de la escuela k en el desempeño del alumno (número de alumnos, proporción de varones en el grupo, etc.). Debido a que se han construido paneles de tal manera que se sigue al mismo individuo en el tiempo, el efecto de pertenecer a un determinado grado g queda representado, únicamente, por el efecto que tiene la escuela k sobre el desempeño.

$\beta_0 \equiv 0$, la contribución al aprendizaje es nulo por parte del alumno-docente. Como nuestro objetivo es determinar la magnitud de la relación entre la prueba bimestral y Enlace, el parámetro β_0 es clave.¹⁷

Todd y Wolpin (2003) señalan que los *Added Value Models* o modelos jerárquicos, en su forma más básica presentan sesgos importantes en la estimación de los parámetros debido a la omisión de variables (crítica similar a Raudenbush, 2004). Esto se debe a que la mayoría de las veces no se cuenta con información disponible sobre la trayectoria educativa de los alumnos. No obstante, Todd y Wolpin (2003) también señalan que incluir una variable adicional del desempeño de los alumnos permite corregir tales problemas. El modelo estimado incluye una variable de desempeño anterior, así como varias. Por ejemplo, el panel 1 incluye hasta cuatro variables de desempeño anterior, por lo que consideramos que el modelo estimado aminora los sesgos indicados. De cualquier manera, estimamos el modelo bajo diferentes métodos y comparamos sus resultados como prueba de robustez.

5. Resultados

5.1. Enlace y las evaluaciones bimestrales

El cuadro 4, columna 1 y 2, presenta resultados de la ecuación (1) para una regresión con un sólo regresor y con varios regresores para la prueba de español panel 1, mientras que el cuadro 5 muestra los resultados para la asignatura de matemáticas. Los resultados para los paneles 2 y 3 son muy similares, y se omiten sin pérdida de información. La columna 1 muestra que la prueba Enlace de 2008 explica 0.7 desviaciones estándar de la prueba Enlace 2009. Cuando se incluyen todas las pruebas Enlace anteriores, estas explican cerca de 0.9 desviaciones estándar, de una desviación estándar en la prueba 2009. Los resultados sugieren que el aprendizaje es persistente. El cuadro 5 muestra resultados similares para la asignatura de matemáticas, aunque para ésta el efecto es ligeramente más persistente. La R^2 indica que cerca de la mitad de la variación en la variable dependiente no es explicada por las variación en las variables explicativas. Esto

¹⁷ Recordemos que $\eta \sim N(0, \sigma_\eta^2)$ y $\beta \sim N(0, \sigma_\beta^2)$. Para una explicación más a fondo sobre los modelos de efectos mixtos o modelos multiniveles, así como del método de estimación, se puede consultar el capítulo 22 de Cameron y Trivedi (2005).

es consistente con el hecho de que hay otras variables que afectan también el puntaje fuera de las pruebas anteriores.

Cuadro 4
Relación entre el puntaje de Enlace y las calificaciones bimestrales para el panel 1. Asignatura: español

	<i>Enlace09</i>	<i>Enlace09</i>	<i>Calf. Prom09</i>	<i>Calf. Prom09</i>	<i>Enlace09</i>	<i>Enlace09</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Enlace08	0.699*** [0.003]	0.367*** [0.004]				
Enlace07		0.305*** [0.004]				
Enlace06		0.202*** [0.004]				
Calf. Prom09					0.623*** [0.003]	0.259*** [0.005]
Calf. Final08			0.744*** [0.003]	0.417*** [0.005]		0.182*** [0.005]
Calf. Final07				0.251*** [0.005]		0.189*** [0.005]
Calf. Final06				0.200*** [0.004]		0.177*** [0.005]
N	76 877					
R ²	0.491	0.588	0.508	0.584	0.371	0.462

Notas: Estimación por mínimos cuadrados ordinarios, sin variables control para las calificaciones de español. *Calf. Prom*: calificación promedio de los tres primeros bimestres del año señalado, *Calf. Final*: calificación promedio final del ciclo escolar señalada. Las variables *Enlace*, *Calf. Prom* y *Calf. Final* son puntajes normalizados por escuela/año. Errores estándar robustos entre paréntesis. ***significativo a 1%.

Cuadro 5
Relación entre el puntaje de Enlace y las calificaciones bimestrales para el panel 1. Asignatura: matemáticas

	<i>Enlace09</i> (1)	<i>Enlace09</i> (2)	<i>Calf. Prom09</i> (3)	<i>Calf. Prom09</i> (4)	<i>Enlace09</i> (5)	<i>Enlace09</i> (6)
Enlace08	0.759*** [0.003]	0.477*** [0.004]				
Enlace07		0.241*** [0.004]				
Enlace06		0.165*** [0.004]				
Calf. Prom09					0.663*** [0.003]	0.295*** [0.005]
Calf. Final08			0.757*** [0.003]	0.439*** [0.005]		0.204*** [0.005]
Calf. Final07				0.248*** [0.005]		0.181*** [0.005]
Calf. Final06				0.188*** [0.004]		0.162*** [0.004]
N	76 877	76 877	76 877	76 877	76 877	76 877
R ²	0.569	0.622	0.529	0.598	0.371	0.511

Notas: Estimación por mínimos cuadrados ordinarios, sin variables control para las calificaciones de matemáticas. *Calf. Prom*: calificación promedio de los tres primeros bimestres del año señalado, *Calf. Final*: calificación promedio final del ciclo escolar señalada. Las variables *Enlace*, *Calf. Prom* y *Calf. Final* son puntajes normalizados por escuela/año. Errores estándar robustos entre paréntesis. ***significativo a 1%.

Los resultados pueden ser comparados de manera similar para las calificaciones bimestrales. Esto se logra al estimar una regresión similar a la ecuación (1) pero se sustituyen los valores de la prueba Enlace por la prueba bimestral. Los resultados se muestran en los cuadros 4 y 5 en las columnas (3) y (4). Los resultados son sorprendentemente similares entre la prueba Enlace y las calificaciones bimestrales. Es decir, las calificaciones bimestrales anteriores predicen, en buena medida, el desempeño actual de la prueba bimestral. En general, se encuentra que ambas pruebas son intertemporalmente consistentes entre sí.

Sin embargo, dado nuestro objetivo, en particular queremos analizar si lo que se está enseñando a los alumnos en las aulas se ve reflejado en la prueba Enlace. Por tanto, las columnas (5) y (6) de los cuadros 4 y 5 incluyen los resultados de una regresión de la prueba Enlace sobre las calificaciones bimestrales por asignatura, mismos que son muy similares a los obtenidos en las columnas anteriores. La prueba bimestral tomada antes de la prueba Enlace explica cerca de 0.65 desviaciones estándar de una desviación estándar de la prueba Enlace. Es decir, con base en los resultados podríamos concluir que, las calificaciones bimestrales, son importantes y miden, efectivamente, el conocimiento permanente de los estudiantes al transmitirse éste en la calificación de la prueba Enlace. La suma de coeficientes estimados de las calificaciones bimestrales son cercanos a 0.83 desviaciones estándar, lo que implicaría, en principio, que los resultados de la prueba Enlace reflejan el aprovechamiento de los alumnos en clase, al menos en lo referente a la asignatura de español y de matemáticas. Si las calificaciones bimestrales fueran una medida completa del desempeño educativo, entonces la prueba Enlace refleja gran parte del desempeño educativo.¹⁸

No obstante, el modelo especificado no controla por las características del alumno, ni de la escuela. Es decir, los resultados no consideran la heterogeneidad del estudiante, ni de la escuela, ni tampoco la estructura jerárquica de las escuelas. Por ejemplo, el número de alumnos en las aulas puede influir de manera importante sobre el aprovechamiento, mientras más amplio sea el número de alumnos menor es el tiempo de atención que brinda el docente a cada alumno. Asimismo, el modelo anterior implicaría que todas las escuelas son iguales y no afectan el proceso de aprendizaje del estudiante.

Por tanto, es importante considerar ciertas características dentro de la estructura del modelo. Para mejorar la estimación de los parámetros de aprendizaje se instrumentan modelos de efectos fijos, regresión cuantil y un modelo jerárquico o multinivel donde la unidad básica de análisis es el desempeño de los alumnos.

5.2. Resultados para el modelo con efectos fijos

La sección anterior estima regresiones simples de la relación entre la prueba Enlace y las pruebas bimestrales, pero no analiza cómo cambian los resultados al incluir variables de control. El cuadro 6 incluye

¹⁸ Los resultados son muy similares entre sí para diferentes paneles. Por cuestiones de espacio se omite esta información.

estimaciones por efectos fijos a nivel escuela, con algunas variables de control para los diferentes paneles en la asignatura de matemáticas (los resultados a nivel español se omiten, aunque son muy similares entre sí). Esta regresión absorbe cualquier variable omitida a nivel escuela y, además, incluye variables a nivel individual como sexo, año de nacimiento y si el alumno copió en la prueba. Para facilitar la lectura del cuadro se han eliminado los asteriscos y no se presentan los estimadores para cada rezago, sólo la suma de los estimadores,

$$\sum_{j=3}^g \beta_j = \beta_0.$$

Cada renglón y columna representan una regresión diferente. La variable dependiente para los paneles 1 y 2 siempre es la prueba Enlace 2009 y para el panel 3 la 2008. La columna *Enlace* se refiere a que la o las variables explicativas es (son) únicamente la(s) prueba(s) anterior(es), es similar para la prueba bimestral. El renglón “1 coeficiente” indica que sólo se incluye en la regresión la prueba inmediata anterior, en tanto que el renglón “Suma coeficientes” considera la suma de los coeficientes de las pruebas rezagadas. Por ejemplo, en Suma coeficientes para el panel 1, en la columna Enlace se incluye la prueba del 2006 al 2008 (esta es la ecuación (1)) y en la columna Bimestral se expresa el promedio de la prueba en el 3er bimestre de 2009 y el promedio total de 2006 a 2008. Todas las estimaciones incluyen efectos fijos a nivel escuela.

Las estimaciones por efectos fijos son sumamente similares al cuadro 5, lo que indica que, aun cuando se tomara únicamente la variación dentro de cada escuela y se eliminara cualquier variable omitida, los resultados son los mismos. Esto sugiere que no existen variables omitidas a nivel escuela, de otra forma los resultados cambiarían, entre paneles son consistentes entre sí, por lo que, más adelante, se presentan sólo los resultados para el panel 1. Si tomamos en cuenta nada más el aprendizaje corriente (promedio de las pruebas bimestrales en el 3er bimestre) la prueba bimestral explica cerca de 0.67 desviaciones estándar el resultado en la prueba Enlace en matemáticas. Por otro lado, la prueba Enlace anterior explica alrededor de 0.76 desviaciones estándar el resultado actual. Es decir, las pruebas bimestrales explican cerca de 88 por ciento de la variación posible. Si bien se puede comparar contra un coeficiente de 1, sabemos que la misma prueba Enlace no consigue ese nivel, por lo que la prueba bimestral debe compararse contra el coeficiente de la prueba Enlace. En este sentido, los resultados indican que las pruebas bimestrales tienen un ordenamiento similar a la prueba Enlace.

La razón que se explica por las pruebas bimestrales aumenta cuando se incluyen más rezagos. Para el panel 1, la suma de coeficientes resulta en una explicación de 0.85 desviaciones estándar, mientras que la prueba Enlace lo hace en 0.88 desviaciones estándar. Es decir, la prueba bimestral consigue explicar cerca del 96% de lo que explica la prueba Enlace. El resultado es similar para los otros paneles.

Cuadro 6

Estimación de efectos fijos entre el puntaje de Enlace y las calificaciones bimestrales para paneles 1, 2 y 3
Asignatura: matemáticas

	<i>Panel 1</i>		<i>Panel 2</i>		<i>Panel 3</i>	
	<i>Enlace</i> <i>(1)</i>	<i>Bimestral</i> <i>(2)</i>	<i>Enlace</i> <i>(3)</i>	<i>Bimestral</i> <i>(4)</i>	<i>Enlace</i> <i>(5)</i>	<i>Bimestral</i> <i>(6)</i>
1 coeficiente	0.758 [0.003]	0.665 [0.004]	0.754 [0.003]	0.657 [0.003]	0.743 [0.003]	0.676 [0.003]
Suma coeficientes	0.881 [0.003]	0.847 [0.003]	0.843 [0.003]	0.813 [0.003]	0.839 [0.002]	0.830 [0.003]

Notas: Estimación por mínimos cuadrados con efectos fijos a nivel escuela para la asignatura de matemáticas. La variable dependiente siempre es la prueba Enlace: 2009 para el panel 1 y 2, y 2008 para el panel 3. Cada renglón y columna representan una regresión diferente, la columna de Enlace se refiere a que las variables explicativas incluyen pruebas Enlace anteriores, de manera similar es para la columna Bimestral. El renglón 1 coeficiente significa que la variable explicativa se refiere al examen inmediato anterior, el renglón Suma coeficientes se refiere a la suma de coeficientes de los exámenes anteriores a la prueba. El panel 1 para la prueba Enlace incluye rezagos de la prueba en 2008, 2007 y 2006, y para las pruebas bimestrales considera la suma de coeficientes de la prueba al 3er bimestre de 2009 (promedio) y total de 2006 a 2008; de la misma manera se obtiene para los paneles 2 y 3. Errores estándar robustos entre paréntesis.

De Hoyos, Espino y García (2010) muestran que factores socioeconómicos tienen una influencia positiva sobre el desempeño de los alumnos. Las escuelas privadas representan un costo monetario adicional a la educación de los alumnos y, en general, las personas que envían a sus hijos a estas escuelas presentan un estatus socioeconómico

por arriba de la media. Es decir, la asistencia a una escuela privada mantiene cierta correlación con el estatus socioeconómico de los padres de los alumnos, lo cual puede afectar la relación entre pruebas.

Por tanto, es importante también analizar cómo el resultado cambia para grupos de interés. El cuadro 7 incluye estimaciones por efectos fijos a nivel escuela, para niños y niñas por separado, así como para estudiantes en escuelas públicas y privadas. El cuadro comprende regresiones donde la variable explicativa cambia. El renglón de Enlace indica que se contempla únicamente la prueba Enlace, es similar para la prueba bimestral. Asimismo, se analiza el efecto de los rezagos, el cuadro indica que los resultados son muy similares a lo encontrado en el cuadro 6. No existen diferencias sustanciales entre niños y niñas o por tipo de escuela. Sin embargo, es importante resaltar que la prueba inmediata anterior pesa ligeramente más para estudiantes en escuelas privadas, que para estudiantes en escuelas públicas, aunque ya cuando se toma en cuenta todo el efecto de los rezagos el resultado es similar.

Cuadro 7

Estimación de efectos fijos entre el puntaje de Enlace y las calificaciones bimestrales para el panel 1, asignatura matemáticas, para diferentes grupos

	<i>Niños</i> (1)	<i>Niñas</i> (2)	<i>Públicas</i> (3)	<i>Privadas</i> (4)
<i>1 coeficiente</i>				
Enlace	0.760 [0.004]	0.755 [0.004]	0.756 [0.003]	0.766 [0.006]
Bimestral	0.673 [0.005]	0.655 [0.005]	0.659 [0.004]	0.687 [0.007]
<i>Suma de coeficientes</i>				
Enlace	0.884 [0.003]	0.877 [0.004]	0.880 [0.003]	0.885 [0.005]
Bimestral	0.858 [0.004]	0.834 [0.005]	0.849 [0.003]	0.836 [0.006]

Notas: Estimación por mínimos cuadrados con efectos fijos a nivel escuela para el panel 1, por sexo y tipo de escuela, para la asignatura de matemáticas. La variable dependiente siempre es la prueba Enlace 2009. Cada renglón y columna representan una regresión diferente, el renglón de Enlace se refiere a que las variables explicativas incluyen pruebas Enlace anteriores, de forma similar es para

el renglón Bimestral. El renglón 1 coeficiente significa que la variable explicativa se refiere al examen inmediato anterior, el renglón Suma coeficientes es la suma de coeficientes de los exámenes anteriores a la prueba. El panel 1 para la prueba Enlace incluye rezagos de la prueba en 2008, 2007 y 2006, para las pruebas bimestrales se incluye la suma de coeficientes de la prueba al 3er bimestre de 2009 (promedio) y total de 2006 a 2008. Errores estándar robustos entre paréntesis.

Los cuadros 6 y 7 indican que las pruebas bimestrales tienen un efecto muy similar sobre la prueba Enlace actual, como con las pruebas Enlace anteriores. Este resultado se mantiene para estudiantes en escuelas privadas o públicas, así como para niños o niñas. Es decir, los resultados implican que el conocimiento medido en las pruebas bimestrales es similar al conocimiento medido en la prueba Enlace. Sin embargo, es importante recalcar que aún la prueba Enlace inmediata anterior no puede pronosticar, completamente, el resultado de la prueba Enlace actual. Aun cuando tomamos los resultados de las pruebas rezagadas existe un porcentaje relativamente menor por explicar (0.15 desviaciones estándar aproximadamente, de acuerdo con la regresión) en la calificación de Enlace.

También realizamos pruebas de especificación de Hausman para analizar si la forma correcta de modelar los datos es con efectos fijos. La hipótesis nula en la prueba de Hausman es que la modelación con efectos fijos es la correcta. La modelación con efectos fijos provee de estimadores consistentes bajo cualquier hipótesis, pero si se rechaza la hipótesis nula los estimadores son ineficientes, y debe estimarse con efectos aleatorios para lograr eficiencia.

Cuadro 8
*Pruebas de Hausman. Comparación de efectos fijos
vs. efectos aleatorios. Panel 1*

	<i>Español</i>			<i>Matemáticas</i>		
	<i>Fijos</i> (1)	<i>Aleatorios</i> (2)	<i>Diferencia</i> (3)	<i>Fijos</i> (4)	<i>Aleatorios</i> (5)	<i>Diferencia</i> (6)
1 coeficiente	0.6119	0.6122	-0.0004 [0.0006]	0.6649	0.6649	0.0000 [0.0005]
Suma coeficientes	0.802	0.8107	-0.009 [0.006]	0.8467	0.8452	0.002 [0.002]

Notas: Estimación por efectos fijos o aleatorios con el panel 1, para cada asignatura. La variable dependiente es el resultado de la prueba Enlace 2009,

el renglón 1 incluye sólo la prueba bimestral al 3er bimestre antes de tomar la prueba Enlace. El renglón de suma de coeficientes considera la prueba bimestral al 3er bimestre y, además, las pruebas bimestrales totales de 2006 a 2008. Las columnas (3) y (6) representan la diferencia entre el estimador de efectos fijos y aleatorios, así como su error estándar calculado por la prueba de Hausman.

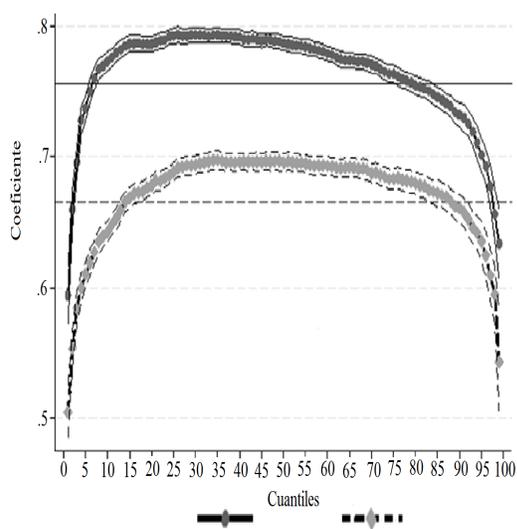
El cuadro 8 incluye las pruebas de Hausman correspondientes al panel 1 para ambas asignaturas. La prueba indica que ambos modelos implican resultados muy similares, casi idénticos. Por lo que la prueba de Hausman rechaza la hipótesis nula que el modelo de efectos fijos sea eficiente. Esto es consistente con el resultado que mínimos cuadrados ordinarios obtenga un efecto muy similar a los estimadores de efectos fijos o aleatorios.

5.3. Regresión cuantil

En esta sección incluimos regresiones cuantiles de la relación entre la prueba Enlace y la prueba bimestral. Como se ha mencionado en las secciones anteriores, las estimaciones si se utilizan efectos fijos o no dan resultados muy similares, la prueba de Hausman indica que la modelación con efectos fijos no es eficiente. Por tal razón, en esta sección se presentan resultados de regresiones cuantiles sin incluir efectos fijos. Como se documenta en la literatura de regresión cuantil, cuando se tiene una gran cantidad de efectos fijos se complica el proceso de estimación, además de que la interpretación de los efectos fijos no es inmediata. Dado que tenemos más de 2 500 efectos fijos para todas las escuelas y no es posible incluirlas manualmente (como variables dicotómicas) en la estimación, se estiman regresiones cuantiles únicamente con variables de control.

Estimamos 100 regresiones cuantiles donde la variable dependiente es la prueba Enlace 2009 para el panel 1 y la variable explicativa es la prueba Enlace anterior, 2008, o bien la prueba bimestral al 3er bimestre en la asignatura de matemáticas. Es decir, se sigue la ecuación (1), pero con estimaciones para cada cuantil. Por simplicidad estimamos únicamente el efecto cuando incluimos una sola variable explicativa: la prueba Enlace o la prueba bimestral antes de tomar la prueba Enlace 2009. Se quiere analizar si el efecto es el mismo, independiente del cuantil. Después de estimar cada regresión cuantil, graficamos cada coeficiente en la gráfica 4, que incluye el estimador por mínimos cuadrados ordinarios (líneas horizontales) y el intervalo de confianza para la estimación cuantil.

Gráfica 4
Estimación por regresión cuantil,
panel 1, asignatura: matemáticas



Notas: Estimaciones de regresión cuantil. Se estiman 100 regresiones para cada cuantil del 1 al 99 y se grafica su coeficiente. La línea sólida con círculos es el coeficiente de una regresión donde la variable dependiente es la prueba Enlace 2009 y la variable explicativa la prueba Enlace 2008, la línea punteada con diamantes se refiere al coeficiente donde la variable explicativa es la prueba bimestral promedio al 3er bimestre. Las líneas horizontales es el efecto en la media estimado por mínimos cuadrados ordinarios, las líneas alrededor de los cuantiles son intervalos de confianza a 95%.

La gráfica nos indica que los resultados son homogéneos, con excepción de los encontrados en ambas colas de la distribución. Aproximadamente del cuantil 20 al 80 los resultados son muy similares, pero en las colas el efecto disminuye. Se puede decir que existe una reversión a la media en los resultados de las pruebas, de lo contrario esperaríamos que los resultados en las colas fueran aún más persistentes que el efecto generalizado. Es decir, los resultados implican que, si un estudiante está en la cola inferior o en la superior en la prueba bimestral o en la de Enlace antes de tomar la prueba Enlace

2009, no necesariamente ese estudiante seguirá en la misma posición relativa. De hecho, es probable que cambie su posición más que cuando está en la mitad de la distribución. En resumen, la gráfica 4 indica que el efecto de ambas pruebas (bimestral y Enlace) anterior a la prueba Enlace 2009 es similar, independiente del cuantil, por lo que, el efecto estimado, es un efecto homogéneo para toda la distribución. Es decir, las pruebas bimestrales miden, aproximadamente, lo mismo que la prueba Enlace independiente del cuantil (por encima de 85% del coeficiente en la prueba Enlace).

5.4. Resultados para McCaffrey et al. (2004)

Los resultados anteriores nos dan una idea general de la relación entre las calificaciones bimestrales y Enlace, sin embargo, puede ser que el procedimiento no sea correcto, en tanto no se reconoce la estructura jerárquica de los datos.¹⁹ Además, al utilizar una prueba de Hausman se comprobó que la modelación por efectos fijos es ineficiente; por lo que se prosigue a estimar la especificación (5) para los tres paneles y el total de escuelas con el modelo de efectos mixtos.²⁰ El método de estimación fue máxima verosimilitud, los resultados se presentan en el cuadro 9 para la asignatura de matemáticas, y es comparable con el cuadro 6. Los resultados de la asignatura español son similares.

¹⁹ Las variables control utilizadas a nivel escuela fueron: *sostenimiento*: uno si la escuela es privada, *turno*: uno si el turno de la escuela es matutino, *idioma*: número de docentes en educación de idiomas diferentes al español, *tecnología*: número de docentes en educación tecnológica, *artística*: número de docentes en educación artística, *física*: número de docentes en educación física, *conserje*: uno si la escuela tiene, al menos, un conserje, *intendente*: uno si la escuela tiene, al menos, un intendente, *PEC*: uno si la escuela participó en el Programa Escuelas de Calidad en el año indicado, *tot. alumnos*: total de alumnos inscritos en la escuela en el año indicado, *aula múltiple*: uno si la escuela tenía, al menos, una aula de usos múltiples, *tot. admistr*: número total de personal administrativo reportado por la escuela en el año indicado. Para el caso de los alumnos tenemos *sexo*: género del alumno, *nacimiento*: año en que nació el alumno, *al corriente*: uno si su edad corresponde con el grado escolar que cursa y *copia*: si el alumno obtuvo una alta probabilidad de haber copiado. Esta última es información que fue proporcionada con la información de los resultados de la prueba Enlace.

²⁰ En un principio se pensaba separar la muestra por sostenimiento, sin embargo no hubo convergencia de la mayoría de las especificaciones. Se encontraron regiones discontinuas que impidieron el proceso de convergencia, aun cuando se modificó el método de maximización los resultados fueron los mismos.

Es importante recordar que el modelo estimado incluye dos efectos aleatorios (en el intercepto y en la pendiente). El primero está asociado a las escuelas (η_{gk}) y el segundo a los alumnos-docentes (Calf. Prom, β_g). Los efectos de transmisión de aprendizaje están representados por la suma del parámetro actual y de los años anteriores,

$$\sum_{j=3}^g \beta_j = \beta_0.$$

La estimación de cada uno de los efectos depende de las variables que se utilizan como regresores.

Cuadro 9
Resultados con efectos mixtos
Asignatura: matemáticas

	<i>Panel 1</i>		<i>Panel 2</i>		<i>Panel 3</i>	
	<i>Enlace</i> (1)	<i>Bimestral</i> (2)	<i>Enlace</i> (3)	<i>Bimestral</i> (4)	<i>Enlace</i> (5)	<i>Bimestral</i> (6)
1 coeficiente	0.755 [0.003]	0.668 [0.003]	0.751 [0.003]	0.658 [0.003]	0.742 [0.003]	0.678 [0.003]
Varianza pendiente	0.046 [0.004]	0.077 [0.004]	0.0470 [0.004]	0.0715 [0.004]	0.052 [0.004]	0.065 [0.004]
Prueba LR		136.2*		131.1*		109.6*
Suma Coeficientes	0.879 [0.003]	0.846 [0.003]	0.840 [0.002]	0.814 [0.003]	0.838 [0.002]	0.831 [0.003]
Varianza pendiente	0.028 [0.005]	0.057 [0.004]	0.0365 [0.005]	0.0627 [0.004]	0.034 [0.004]	0.045 [0.004]
Prueba LR		61.9*		102.9*		43.4*

Estimación por el método de efectos mixtos, donde se permiten efectos aleatorios a nivel escuela tanto en la constante como en la pendiente del efecto principal, para la asignatura de matemáticas. La variable dependiente siempre es la prueba Enlace: 2009 para el panel 1 y 2 y 2008 para el panel 3. Cada renglón y columna representan una regresión diferente, la columna de *Enlace* se refiere a que las variables explicativas incluyen pruebas Enlace anteriores, de manera similar es para la columna *Bimestral*. El renglón “1 coeficiente” significa que la variable explicativa se refiere al examen inmediato anterior. El renglón “Suma coeficientes” se refiere a la suma de coeficientes de los exámenes anteriores a la prueba. El panel 1 para la prueba Enlace incluye rezagos de la prueba en 2008, 2007 y 2006; para las pruebas bimestrales incluye la suma de coeficientes de la

prueba al 3er bimestre de 2009 (promedio) y total de 2006 a 2008; de la misma manera se obtiene para los paneles 2 y 3. Se incluyen, también, resultados para la varianza del efecto aleatorio, así como su error estándar. LR prueba indica el valor de la prueba de verosimilitud entre un modelo con efectos aleatorios en la constante únicamente, contra un modelo de efectos aleatorios en la constante y en la pendiente. Errores estándar robustos entre paréntesis. *significancia a 1% en la prueba LR.

En el cuadro 9 se muestra la estimación obtenida de todos los paneles, los alumnos que cursaban el tercer grado en 2006 hasta el sexto grado en 2009. Para facilitar comparaciones entre métodos no reportamos los coeficientes obtenidos para todas las variables explicativas sino, únicamente, las de interés. Asimismo, reportamos una tabla similar a la explicada en el cuadro 6, donde incluimos el efecto del coeficiente de la prueba rezagada y la suma de coeficientes de las pruebas rezagadas. No obstante, al estimar un modelo de efectos mixtos, también reportamos la varianza estimada de los efectos aleatorios (la pendiente).²¹ En particular, la diferencia del modelo con efectos mixtos es que permitimos un efecto aleatorio en la pendiente, de otra forma la estimación sería idéntica a efectos aleatorios (estimada anteriormente). El cuadro incluye la varianza del efecto aleatorio de la pendiente y, además, el valor de la prueba de verosimilitud, donde se contrasta la hipótesis nula que el modelo correcto es el modelo con efectos aleatorios únicamente en la pendiente, contra la alternativa que el modelo correcto es el modelo con efectos aleatorios en la constante y en la pendiente (renglón Prueba LR).

El cuadro nos ofrece resultados sumamente similares a los encontrados en el cuadro 6. Es decir, para el panel 1 los resultados significan que un incremento de 1 desviación estándar en la prueba Enlace anterior implica un aumento de 0.76 desviaciones estándar en la prueba Enlace actual (2009). Asimismo, un aumento de 1 desviación estándar en la prueba bimestral anterior (promedio al 3er bimestre) se traduce en un aumento de 0.67 desviaciones estándar en la prueba Enlace actual. Por tanto, el efecto de la prueba bimestral es alto, si consideramos todas las pruebas anteriores, el efecto se incrementa aún más, al aumentar a 0.88 y 0.85 desviaciones estándar, respectivamente. Esto implica que el conocimiento medido en las pruebas bimestrales sí se traduce a la prueba Enlace y, por tanto, lo enseñado

²¹ La varianza de la constante era muy baja y cercana a cero por lo que se omitió.

en el aula de clase sí es relevante para la prueba Enlace. Los resultados para la prueba de español son similares en magnitud.

El cuadro también incluye el parámetro del efecto aleatorio de la pendiente principal, se muestra que el coeficiente es estadísticamente significativo. Por lo tanto, la prueba de verosimilitud rechaza la hipótesis nula de efectos aleatorios únicamente en la constante, a favor de un modelo de efectos aleatorios en la constante y en la pendiente principal. Asimismo, los resultados indican una mayor varianza del efecto aleatorio en las calificaciones bimestrales, lo que significa que existe una mayor variabilidad en el efecto de las pruebas bimestrales, que en el efecto de la prueba Enlace. En general, se concluye que los resultados son consistentes entre modelos y que lo enseñado en el salón de clases sí es relevante para la prueba Enlace.

6. Conclusiones

En este estudio se utilizaron datos confidenciales para unir las calificaciones bimestrales de estudiantes del Distrito Federal con las calificaciones de la prueba Enlace. En particular, utilizamos una base de datos con cerca de 300 000 estudiantes, para los cuales observamos longitudinalmente sus calificaciones de Enlace y sus calificaciones en el aula para diferentes grados de primaria (en las asignaturas de matemáticas y español). El objetivo del trabajo era determinar la relación que existe entre las calificaciones bimestrales y los resultados logrados en Enlace. Estimamos diferentes modelos de regresión para analizar la relación entre pruebas. Estimamos modelos de regresión lineal simple, efectos fijos, regresión cuantil y, por último, basados en McCaffrey *et al.* (2004) utilizamos un modelo jerárquico o de efectos mixtos. En particular, utilizamos una medida de calificación bimestral antes de aplicarse la prueba Enlace y estandarizamos ambas pruebas para poder compararlas.

Los resultados principales muestran que ambas pruebas sí están relacionadas. Para lograr esto comparamos el efecto de la prueba bimestral con el efecto de la prueba Enlace anterior a tomar la prueba Enlace actual. Es decir, las pruebas Enlace tienen el mismo dominio, por lo que una prueba diferente y con distinto dominio puede tener diferentes resultados. Entonces, comparar pruebas Enlace a través del tiempo implica comparar el efecto de pruebas con el mismo dominio, por lo tanto, si el efecto de la prueba bimestral implica un estimador cercano al de la prueba Enlace anterior, significa que la prueba bimestral mide aspectos similares que la prueba Enlace.

Realizamos estimaciones de regresión lineal simple, efectos fijos, regresión cuantil y modelos de efectos mixtos. Todos los resultados son similares y robustos e indican que el efecto de la prueba bimestral sobre la prueba Enlace es similar al efecto de la prueba Enlace anterior sobre la prueba Enlace actual. Los resultados encontrados nos llevan a concluir que la prueba Enlace y los exámenes bimestrales mantienen una alta relación entre sí. Lo cual se debe a que la relación entre la prueba Enlace actual y la prueba Enlace anterior es muy similar a la relación de la prueba Enlace y prueba bimestral inmediata anterior. En la asignatura de matemáticas un incremento de una desviación estándar en la prueba Enlace del año anterior explica un incremento de 0.76 desviaciones estándar en la prueba Enlace actual. Asimismo, un incremento de una desviación estándar en la prueba bimestral explica un incremento en la prueba Enlace actual de 0.67 desviaciones estándar. Si tomamos en cuenta los exámenes de 2006 a 2009 los efectos aumentan a 0.88 y 0.84 desviaciones estándar, respectivamente. Los resultados nos llevan a concluir que la relación entre pruebas es relativamente alta y que el ordenamiento en las calificaciones bimestrales se mantiene, en general, en la prueba Enlace. También estimamos modelos diferenciados por sexo del estudiante y por tipo de escuela (privada y pública), encontramos pocas diferencias económicamente significativas entre ellas.

Finalmente, futuras investigaciones deben profundizar y discutir sobre lo que se les enseña a los alumnos dentro de las aulas, así como analizar el rol de variables individuales en el puntaje. Debido a la confidencialidad de los datos, no podemos obtener variables a nivel del hogar del estudiante o demográficas al estudiante (a excepción de sexo y año de nacimiento). Asimismo, es importante se busque obtener dichos datos para analizar la robustez de los resultados, así como también analizar detalladamente los métodos de enseñanza utilizados en las escuelas, ya que nuestros resultados están estandarizados y no pueden comparar puntajes en valor absoluto a través de escuelas. De igual manera es necesario considerar el papel de los factores culturales, como factores que influyen en el aprendizaje de los alumnos. En suma, futuras investigaciones pueden analizar el rol de la escuela, en su conjunto, con el rol de la composición de los estudiantes (*peer effects* o efectos de vecindad).

Referencias

- Abreu, R. y D. C. Martín del Campo. 2007. *Índice compuesto de eficacia de los sistemas escolares, 2007*, Mexicanos Primero Visión 2030, A.C./Fundación IDEA, A.C.
- Andrabi, T., J. Das, A. I. Khwaja y T. Zajonc. 2009. Do Value-Added Estimates Add Value? Accounting for Learning Dynamics, The World Bank, Development Research Group, Policy Research Working Paper, núm. 5066.
- Cameron, A. C. y P. K. Trivedi. 2005. *Microeconometrics. Methods and Applications*, Cambridge University Press.
- De Hoyos, R., J. M. Espino y V. García. 2010. Determinantes del logro escolar en México: primeros resultados utilizando la prueba Enlace media superior, Subsecretaría de Educación Media Superior, México (inédito).
- Figlio, D. N. 2006. Testing, Crime and Punishment, *Journal of Public Economics*, 90(4-5): 837-851.
- y L. S. Getzler. 2002. Accountability, Ability and Disability: Gaming the System, NBER Working Paper Series, núm. 9307.
- Glewwe, P., N. Ilias y M. Kremer. 2010. Teacher Incentives, *American Economic Journal: Applied Economics*, 2(3): 205-227.
- Jacob, B. A. 2005. Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools, *Journal of Public Economics*, 89(5-6): 761-796.
- , L. J. Lefgren y D. Sims. 2008. The persistence of Teacher-Induced Learning Gains, NBER Working Paper Series, núm. 14065.
- Kane, T. J., D. Grissmer, D. O. Staiger y H. F. Ladd. 2002. Volatility in School Test Scores: Implications for Test-Based Accountability Systems, *Brookings Papers on Education Policy*, núm. 5, The Brookings Institution, pp. 235-283.
- Kane, T. J. y D. O. Staiger. 2001. Improving School Accountability Measures, NBER Working Paper Series, núm. 8156.
- . 2002. The Promise and Pitfalls of Using Imprecise School Accountability Measures, *The Journal of Economic Perspectives*, 16(4): 91-114.
- . 2008. Estimating Teacher Impacts on Student Achievement: An experimental Evaluation, NBER Working Paper Series, núm. 14607.
- Kolen, M. J. 1988. Defining Score Scales in Relation to Measurement Error, *Journal of Educational Measurement*, 25(2): 97-110.
- Koretz, D. M. 2002. Limitations in the Use of Achievement Tests as Measures of Educators' Productivity, *Journal of Human Resources*, 37(4): 752-777.
- McCaffrey, D. F., J.R. Lockwood, D. Koretz, T. A. Louis, y L. Hamilton. 2004. Models for Value Added Modeling of Teacher Effects, *Journal of Educational and Behavioral Statistics*, Valued-Added Assessment Special Issue, 29(1): 67-101.
- Mizala, A., P. Romaguera y M. Urquiola. 2007. Socioeconomic Status or Noise? Tradeoffs in the generation of school quality information, *Journal of Development Economics*, 84(1): 61-75.
- Neal, D. y D. W. Schanzenbach. 2007. Left Behind by Design: Proficiency Counts and Test-Based Accountability, NBER Working Paper Series, núm. 13293.

- Raudenbush, S. W. 2004. What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice? *Journal of Educational and Behavioral Statistics*, Value-Added Assessment Special Issue, 29(1): 121-129.
- Rothstein, J. 2010. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement, *Quarterly Journal of Economics*, 125(1): 175-214.
- Rubin, D. B., E. A. Stuart y E. L. Zanutto. 2004. A Potential Outcomes View of Value-Added Assessment in Education, *Journal of Educational and Behavioral Statistics*, Value-Added Assessment Special Issue, 29(1): 103-116.
- SEP. 2011. *Evaluación nacional del logro académico en centros escolares*, ENLACE, en: <http://enlace.sep.gob.mx/>
- Shepard, L. A. 1990. Inflated Test Score Gains: Is the Problem Old Norms or Teaching the Test? *Educational Measurement: Issues and Practice*, 9(3): 15-22.
- Todd, P. y K. I. Wolpin. 2003. On the Specification and Estimation of the Production Function for Cognitive Achievement, *The Economic Journal*, 113(485): F3-F33.
- Woodhouse, G., M. Yang, H. Goldstein y J. Rasbash. 1996. Adjusting for Measurement Error in Multilevel Analysis, *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(2): 201-212.