



Documento de trabajo del BID # IDB-WP-438

Teoría y evidencia sobre las políticas docentes en países desarrollados y en desarrollo

Emiliana Vegas
Alejandro Ganimian

Agosto 2013

Banco Interamericano de Desarrollo
División de Educación

Teoría y evidencia sobre las políticas docentes en países desarrollados y en desarrollo

Emiliana Vegas
Alejandro Ganimian



Banco Interamericano de Desarrollo

2013

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo

Vegas, Emiliana.

Teoría y evidencia sobre las políticas docentes en países desarrollados y en desarrollo / Emiliana Vegas,
Alejandro Ganimian.

p. cm. (IDB working paper series ; 438)

Incluye referencias bibliográficas.

1. Teacher effectiveness. 2. Teachers—Training of. 3. Teachers—Rating of. 4. Teachers—Recruiting. 5.
Teachers—Selection and appointment. I. Ganimian, Alejandro. II. Banco Interamericano
de Desarrollo. División de Educación. III. Título. IV. Serie.

IDB-WP-438

<http://www.iadb.org>

Las opiniones expresadas en esta publicación son exclusivamente de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.

Se prohíbe el uso comercial no autorizado de los documentos del Banco, y tal podría castigarse de conformidad con las políticas del Banco y/o las legislaciones aplicables.

Copyright © 2013 Banco Interamericano de Desarrollo. Este documento de trabajo puede reproducirse para fines no comerciales. Puede también reproducirse en cualquier revista académica indizada en el EconLit de la Asociación Americana de Economía, con el consentimiento previo del Banco Interamericano de Desarrollo (BID), siempre y cuando se reconozca la autoría del Banco y el autor o autores del documento no hayan percibido remuneración alguna derivada de la publicación.

Teoría y evidencia sobre las políticas docentes en países desarrollados y en desarrollo

Emiliana Vegas

Alejandro J. Ganimian*

Resumen

La década pasada ha sido testigo del surgimiento de un gran número de rigurosas evaluaciones de impacto de políticas docentes. En este trabajo se revisa la teoría económica y la evidencia empírica sobre ocho metas de las políticas docentes: 1) fijar expectativas claras para los docentes; 2) atraer a los mejores a la docencia; 3) preparar a los docentes con formación y experiencia útiles; 4) asignar a los docentes donde más se los necesita; 5) liderar a los docentes con buenos directores; 6) evaluar el aprendizaje y la enseñanza; 7) apoyar a los docentes para que mejoren la enseñanza y 8) motivar a los docentes en su desempeño. También se discuten los conceptos y métodos econométricos claves para comprender los estudios realizados hasta la fecha y ofrecer algunos lineamientos para futuras investigaciones.

Códigos JEL: I250 Educación y Desarrollo, J450 Mercados Laborales del Sector Público, I210 Análisis de Educación.

* Los autores desean agradecer a Susanna Loeb, Pilar Romaguera, Agustina Paglayán, Analía Jaimovich, Andrew Trembley y Nicole Goldstein, quienes fueron integrantes claves del equipo del proyecto SABER-Docentes del Banco Mundial, en el que nos basamos en buena medida para este trabajo. Emiliana Vegas es jefa de la División de Educación del Banco Interamericano de Desarrollo (evegas@iadb.org). Alejandro J. Ganimian es doctorando en Análisis Cuantitativo de Políticas Educativas en la Escuela de Posgrado de Educación de la Universidad de Harvard e investigador asociado en el Programa Multidisciplinario de Inequidad y Política Social de la Escuela de Gobierno John F. Kennedy de la Universidad de Harvard (alejandro_ganimian@mail.harvard.edu).

Introducción

La década pasada ha sido testigo del surgimiento de un gran número de rigurosas evaluaciones de impacto de políticas docentes. Estos estudios difieren de las investigaciones anteriores en tres aspectos que los hacen útiles a la hora de orientar las decisiones de política. En primer lugar, en vez de confiar en “proxies” (aproximaciones) del aprendizaje de los alumnos y de la eficacia docente, tales como la matrícula escolar o los índices de certificación docente, se valen de los datos sobre el aprendizaje estudiantil que resultan de evaluaciones nacionales e internacionales. En segundo lugar, en vez de identificar asociaciones entre políticas y resultados, emplean métodos que permiten distinguir los efectos de las intervenciones de otros factores que pueden confundir tales efectos. Por último, aunque evalúan el impacto de reformas específicas, también analizan cómo las reformas interactúan entre sí.

En este trabajo se condensan las principales conclusiones de algunos de los estudios sobre políticas docentes realizados más recientemente. La primera sección presenta el marco analítico mediante el cual se estructuró la revisión de la evidencia. Sobre la base de ese marco, la segunda sección analiza la teoría económica que impulsa los trabajos empíricos sobre políticas docentes así como las principales constataciones de los estudios más rigurosos en esta área. La última sección presenta las conclusiones interpretando la teoría y la evidencia acerca del impacto de diversas políticas docentes sobre el desempeño estudiantil y señalando cuáles son los principales hallazgos e interrogantes que abren las puertas a nuevas investigaciones.

Marco analítico: las políticas docentes como sistema

Una manera de abordar los temas que han motivado las investigaciones sobre políticas docentes es organizarlos de acuerdo con los desafíos que enfrentan los sistemas educativos en diferentes etapas de la carrera docente, que van desde el momento en que una persona decide ingresar a la docencia hasta el momento en que se incorpora a una escuela como integrante del plantel docente y comienza a dar clases. Las principales cuestiones de política que enfrentan los sistemas educativos para mejorar la eficacia de los docentes son: 1) fijar expectativas claras para los docentes, 2) atraer a los mejores a la docencia, 3) preparar a los docentes con formación y experiencia útiles, 4) asignar a los docentes donde más se los necesita, 5) liderar a los docentes con buenos directores, 6) evaluar el aprendizaje y la enseñanza, 7) apoyar a los docentes para que mejoren la enseñanza y 8) motivar a los docentes en su desempeño. Si bien hay otras maneras de organizar la evidencia, este marco nos recuerda que las políticas tales como aquellas sobre los requisitos para la certificación o sobre las evaluaciones se conciben con el objeto de alcanzar metas y, por consiguiente, deberían evaluarse en función de en qué medida logran alcanzarlas.

La eficacia con que un sistema educativo aborda los desafíos presentes en una etapa de la carrera docente puede facilitar o dificultar la respuesta del sistema a los desafíos de las etapas subsiguientes. Asimismo, la eficacia con que un sistema educativo afronta estos desafíos para una determinada proporción de docentes en un momento dado facilitará o dificultará la respuesta del sistema a los desafíos que se le presentarán a quienes se incorporen a la docencia en el futuro. Tal como expusimos en otro trabajo, las políticas docentes configuran un sistema dinámico: los cambios que se introducen en un área repercuten en todas las demás (Vegas *et al.*, 2011).

Teoría y evidencia: el mercado laboral docente

Esta sección se basa en la teoría y la práctica económicas para plantear la importancia de las metas de política docente antes enumeradas y revisar la evidencia más rigurosa disponible sobre las intervenciones realizadas para alcanzar cada una de estas metas.

La economía ha desempeñado un papel importante en investigaciones rigurosas recientes sobre políticas docentes, no sólo en definir los problemas sino también sus soluciones. Por lo tanto, si bien existen muchas maneras de examinar la evidencia sobre las políticas docentes, revisar la teoría económica que ha impulsado las investigaciones empíricas puede ayudarnos a entender por qué la evidencia se ha desarrollado del modo en que lo ha hecho.

Hay diversos aspectos que diferencian las recientes investigaciones sobre políticas docentes de los estudios empíricos previos. Tal vez la característica más distintiva de los estudios recientes es su capacidad para calcular los efectos de las intervenciones en los resultados de alumnos y docentes. A primera vista, el hecho de que esta no haya sido la norma puede resultar sorprendente. Sin embargo, hasta hace poco tiempo, para la mayor parte de las investigaciones empíricas en el campo de la educación no era posible separar los efectos de una intervención (conocidos como “efectos del tratamiento”) de los efectos asociados a las características de quienes participaban en el programa (“efectos de selección”). Este problema, conocido como “sesgo de selección”, ha limitado el impacto de las investigaciones sobre políticas en la esfera de la educación (Murnane y Willett, 2011).

No obstante, durante la década pasada, los investigadores del campo de la educación han hecho notables progresos para abordar el sesgo de selección. En particular, han mejorado en cuanto a dar respuestas confiables a dos preguntas de primordial interés para los diseñadores de políticas: 1) ¿cómo les habría ido a los participantes de un programa en ausencia del programa? y 2) ¿cómo les habría ido a los no participantes en presencia del programa? Responder a estas preguntas de manera concluyente es imposible porque, en

un momento determinado, un individuo está participando en un programa o no lo está. Intentar una respuesta a partir de comparar a quienes participan en un programa antes y después de su participación no es una buena opción, porque esos individuos habrán atravesado circunstancias diversas mientras participaban en el programa que pudieron haber influido en los resultados que deben medirse. Esto es lo que se conoce como el “problema de la evaluación de programas” (Duflo, Glennerster y Kremer, 2008). La innovación de los trabajos de investigación recientes ha consistido en usar métodos que permiten encontrar un grupo de control adecuado para ofrecer una estimación confiable para responder a las dos preguntas “contrafácticas” recién mencionadas.

La generación más reciente y más rigurosa de investigaciones sobre políticas docentes tanto en países desarrollados como en países en desarrollo recurre a métodos econométricos como los estudios de asignación aleatoria, el análisis de diferencias en diferencias, la estimación con variables instrumentales y los diseños de regresión discontinua para ofrecer cálculos sin sesgo de los efectos de las políticas docentes.

Este trabajo de revisión se concentra en estudios rigurosos que evalúan el impacto de las políticas docentes sobre el aprendizaje de los alumnos medido a través de pruebas estandarizadas. El objetivo no es transmitir que el único resultado de un sistema educativo que funciona bien es la mejora en el aprendizaje de los alumnos.¹ En cambio, esta revisión se ha visto motivada por el creciente interés de los gobiernos de los países tanto desarrollados como en desarrollo en las políticas educativas que aumentan el rendimiento estudiantil y por la evidencia que vincula el aprendizaje con el desarrollo económico (Hanushek y Woessmann, 2007). Se hará referencia a estudios menos rigurosos (que usan la correspondencia del puntaje de propensión o efectos fijos) en dos casos: 1) cuando no existen estudios rigurosos sobre una política docente y/o 2) cuando el objetivo es contrastar la evidencia de los países desarrollados y en desarrollo.

En los recuadros se resumen los métodos empleados en las investigaciones así como algunos conceptos estadísticos de uso frecuente. El objetivo es incorporar a los responsables de las políticas y demás partes interesadas al debate sobre la evidencia disponible e instarlos a contribuir activamente.

Los efectos de los programas se informan principalmente en unidades de desviación estándar. En estadística, la desviación estándar suele usarse para medir el nivel de dispersión de los valores (por ejemplo, cuánto varía el puntaje que los alumnos de una clase obtienen en una prueba). La desviación estándar se calcula a través de un proceso

¹ Investigaciones rigurosas han documentado el impacto de la educación sobre resultados importantes de corto plazo, como la satisfacción estudiantil, y de largo plazo, como los resultados en el mercado laboral, los logros educativos y el delito. Estos resultados no siempre se traducen en los puntajes obtenidos en las pruebas (véanse Chetty *et al.*, 2011; Deming, 2011; Kane y Staiger, 2010, 2012; Kemple y Willner, 2008).

que puede resultar complicado y arbitrario, pero se usa porque ofrece una métrica común para comparar los efectos de los diferentes programas sobre resultados tales como los puntajes de los alumnos. En las ciencias sociales en general, se considera que un efecto de 0,80 desviaciones estándar es grande, uno de 0,50 es moderado y uno de 0,20 es pequeño (Cohen, 1988). Sin embargo, tal como fue observado por Murnane y Willett (2011), en el campo de la educación hasta las intervenciones más exitosas tienen efectos pequeños conforme a estos estándares.

Se puede interpretar las desviaciones estándares en cinco maneras. Una es pensar qué dicen respecto de cuánto mejora el aprendizaje de los alumnos, normalmente en el caso de las pruebas de rendimiento estudiantil en matemática y comprensión lectora. Si se parte de la premisa de que las habilidades de los alumnos siguen una distribución “normal” de modo tal que la mayoría tiene un nivel promedio de habilidad y unos pocos se ubican en los niveles más bajos y más altos —los psicometristas que diseñan las pruebas se aseguran de que así sea—, entonces el tamaño del efecto en unidades de desviación estándar puede traducirse fácilmente en un cambio en el percentil en que se ubica un alumno (Recuadro 1). Por ejemplo, una mejora de 0,25 desviaciones estándar significa que un estudiante cuyo desempeño estaba en el percentil 50 se ubica ahora en el 60.

Recuadro 1. ¿Qué son los percentiles?

En estadística, los percentiles suelen usarse para comprender dónde se ubica un individuo en una distribución. El cálculo de los percentiles es sencillo, pero no es necesario entender cómo se hace para comprender lo que significan. Un percentil es simplemente el porcentaje de personas que están por debajo de ese individuo. Por ejemplo, si en una prueba un alumno obtiene un puntaje que lo coloca en el percentil 10, el desempeño del 10% de sus pares estuvo por debajo del de ese alumno o, lo que es equivalente, el desempeño del 90% de sus compañeros fue mejor. De manera similar, una mejora del percentil 10 al percentil 15 significa que el estudiante que antes había superado al 10% de sus pares ahora supera al 15% de ellos.

Otra manera de pensar los tamaños de los efectos es compararlos con puntos de referencia. Muchos estudios realizados en Estados Unidos comparan los efectos de un programa con la brecha que existe entre el rendimiento de los blancos y el de los negros, que se ubica entre 0,65 y 0,75 desviaciones estándar (Reardon, 2011). Otros comparan los efectos con las “equivalencias de grado”, es decir, cuánto aprenden los alumnos en un año escolar. Kane y Staiger (2012) equiparan 0,25 desviaciones estándar a nueve meses de escolaridad en la franja de 4° a 8° grado. Sin embargo, otros entienden estos efectos en términos de los beneficios tangibles que les reportan a los niños. Según los cálculos de Krueger (2003), un aumento de una desviación estándar en los puntajes de matemática de la escuela primaria equivale a un incremento del 8% en sus ingresos de adultos. También es posible comparar los efectos de nuevas políticas con los de intervenciones muy

conocidas. En Estados Unidos, muchos comparan los efectos de las nuevas intervenciones con aquellos de una famosa política de reducción de la cantidad de alumnos por clase —descrita más adelante—, cuyo efecto fue de 0,20 desviaciones estándar (Hanushek, 2003).

Fijar expectativas claras para los docentes

Según la teoría económica, fijar expectativas claras es importante tanto para contratar como para dirigir a docentes eficaces. Las expectativas influyen en cómo perciben la profesión quienes potencialmente opten por ingresar a la docencia. Tal como plantearon Roy (1951) y Borjas (1987), los individuos “se autoseleccionan” en los mercados de trabajo eligiendo aquel del que esperan obtener mayores ingresos después de evaluar sus propias habilidades y las habilidades exigidas por el puesto de trabajo. Los estándares profesionales pueden brindar a los empleadores un mecanismo para “filtrar” a aquellos con más probabilidades de tener éxito (Stiglitz, 1975). Una profesión con estándares bajos o poco claros crea un desequilibrio en la información que los empleadores y los potenciales empleados tienen sobre las habilidades de estos últimos. Esto se conoce en economía como “asimetría de la información” y es problemática porque alienta a los individuos poco calificados a intentar ingresar a profesiones para las cuales posiblemente no estén preparados al mismo tiempo que desalienta el ingreso de las personas más calificadas—un problema que los economistas denominan “selección adversa” (Akerlof, 1970; Greenwald, 1986).

En segundo lugar, las expectativas pueden orientar el trabajo docente. Tal como lo indicaron Baker (1992) y Prendergast (2002), si los estándares profesionales son claros, se puede minimizar el “problema del agente y el principal”, en el que un principal (por ejemplo, un sistema educativo) contrata a un agente (por ejemplo, un maestro) para realizar tareas que son de interés del principal pero que son costosas para el agente y difíciles de cumplir (por ejemplo, mejorar el aprendizaje de los alumnos). Los estándares profesionales pueden mitigar estos problemas ya que pueden sentar las bases para un “contrato” en el que empleador y empleado decidan por consenso las tareas que han de realizarse y la forma en que se medirá el desempeño para determinar la retribución del empleado.

Existen investigaciones cuyo propósito ha sido definir en qué medida las expectativas sobre el trabajo docente son suficientemente claras, justas y/o ambiciosas para producir los niveles esperados de aprendizaje estudiantil. Estas expectativas incluyen la existencia de directrices explícitas sobre el trabajo docente en el aula, tales como los estándares relativos a la planificación de clases o el aprendizaje, y sobre las características del sistema, por ejemplo: las horas de clase o la cantidad de alumnos por aula. Estas

directrices establecen implícitamente las condiciones en las que se espera que los docentes se desempeñen eficazmente.

Evaluar rigurosamente las expectativas sobre el trabajo docente es un desafío porque las evaluaciones, en general, se aplican a todos los alumnos de un sistema educativo nacional o subnacional. Por consiguiente, al igual que con otras políticas de carácter universal, a los investigadores les resulta difícil encontrar un grupo de control adecuado. No obstante, han hecho incursiones importantes experimentando con las expectativas existentes y asignando aleatoriamente grupos de escuelas de un sistema educativo a las condiciones habituales y a nuevas condiciones, un método que se detalla más adelante.

En el Cuadro 1 se resumen tres tipos de intervenciones evaluadas rigurosamente: 1) brindar andamiaje al trabajo docente, 2) aumentar las horas de clase y 3) reducir la cantidad de alumnos por clase. Los últimos dos tipos de intervenciones pueden no parecer tan relacionados con la meta de fijar expectativas claras como el primero, pero se incluyen aquí porque contribuyen a nuestro entendimiento sobre la adecuación de las actuales expectativas sobre la labor docente.

Brindar andamiaje al trabajo docente

Una manera en que los sistemas educativos han buscado mejorar el desempeño de docentes y alumnos consiste en ofrecer a los docentes directrices más estructuradas para su trabajo en el aula.

Una intervención realizada en Estados Unidos que brinda andamiaje o apoyo la labor de los docentes es el programa de lectura Success for All (SFA), que tiene un currículo muy estructurado que afecta a toda la escuela, utiliza novelas y libros de texto de lectura, reagrupa periódicamente a alumnos de distintas edades y grados y exige que los alumnos se comprometan a leer en sus hogares. En el año 2000, Borman *et al.* (2007) evaluaron el programa SFA mediante un estudio de asignación aleatoria (Recuadro 2). Los autores asignaron en forma aleatoria 41 escuelas muy pobres a SFA para los alumnos de Kindergarten a 2º grado o a los alumnos de los grados 3º a 5º y compararon a los alumnos de Kindergarten y de 1º grado de estos dos grupos. La conclusión fue que el programa tenía efectos positivos sobre tres resultados de alfabetización. El tamaño de los efectos fue de 0,21 desviaciones estándar para la comprensión de textos a 0,33 desviaciones estándar para la habilidad de reconocer palabras (“ataque de palabras”). Estos resultados indican que modificar las expectativas en las escuelas a las que asisten niños socialmente desfavorecidos podría incidir en su rendimiento al hacer que la labor docente resulte más manejable.

Recuadro 2. ¿Qué es un estudio de asignación aleatoria?

Los estudios de asignación aleatoria son cada vez más utilizados. Su principal beneficio es que permiten que los investigadores obtengan una estimación “limpia” del efecto promedio de un programa comparando los resultados de los individuos que participaron en él (el “grupo de tratamiento”) con los de los individuos que no participaron (el “grupo de control”). Cuando se asigna un programa aleatoriamente, el único factor que distingue a las personas del grupo de tratamiento de las personas del grupo de control es el azar; ningún grupo tiene características que puedan incidir en los resultados de interés. Por consiguiente, el riesgo del sesgo de selección desaparece y el efecto puede calcularse como la diferencia entre los resultados promedio del grupo de tratamiento menos los del grupo de control.

Hay evaluaciones rigurosas de iniciativas que aumentan el andamiaje o apoyo brindado a los docentes de los países en desarrollo, donde la capacidad docente tiende a ser más baja. De hecho, dos estudios recientes realizados en la India indican que estas intervenciones pueden tener un impacto considerable en la pedagogía del docente y en el aprendizaje del alumno. En Maharashtra, He, Linden y MacLeod (2007) usaron un estudio de asignación aleatoria para evaluar un programa destinado a enseñar inglés que podía implementarse a través de una máquina especialmente diseñada o de tarjetas didácticas (*flashcards*). Constataron que ambas versiones de este programa reportaban beneficios para el aprendizaje del idioma inglés de alrededor de 0,30 desviaciones estándar y que resultaban especialmente efectivos para los alumnos de mayor edad y menor rendimiento. También observaron que la versión del programa enseñada por los docentes no sólo mejoraba el puntaje de los alumnos en inglés, sino también en matemática. Esto indica que invertir en mejorar la pedagogía de los docentes (en lugar de reemplazarlos) tuvo efectos de derrame positivos.

Sin embargo, según el otro estudio realizado en la India, aunque el andamiaje puede ser eficaz para producir cambios sencillos a nivel pedagógico, la baja capacidad de la fuerza docente de algunas escuelas puede limitar las posibilidades de lograr cambios más complejos. En 2004, en Mumbai, He, Linden y MacLeod (2009) asignaron en forma aleatoria a alumnos de escuelas de nivel preescolar y primario y de clases independientes de lectura a un programa que modificaba las clases de comprensión lectora y que proponía nuevas actividades para los docentes del grupo de tratamiento. El programa produjo mejoras en el puntaje de los alumnos en comprensión lectora de 0,26-0,70 desviaciones estándar y demostró ser más efectivo con los alumnos de nivel preescolar y de bajo rendimiento. Sin embargo, la versión implementada fuera de la jornada escolar fue más efectiva que aquella llevada a cabo durante el horario de clase, produciendo un efecto adicional de 0,24 desviaciones estándar. Este mayor impacto del programa implementado fuera de la jornada escolar indica que tal vez no basta con el andamiaje

para producir en los docentes los cambios pedagógicos que son necesarios para mejorar sustancialmente el aprendizaje de los alumnos.

Aún cambios relativamente sencillos en las actividades del aula han demostrado mejorar el aprendizaje en los países en desarrollo. En el año 2009, en Tarlac, una provincia de las Filipinas, se asignó aleatoriamente 5.510 alumnos de 4º grado a un grupo de tratamiento con docentes que recibieron capacitación en servicio, con material de lectura y una maratón de lectura, o a un grupo de control que no recibió ninguna intervención (Abeberese, Kumler y Linden, 2011). Luego de un mes, la cantidad de libros leídos por los alumnos aumentó de 2,3 a 9,5 y el puntaje de los alumnos en comprensión lectora aumentó 0,13 desviaciones estándar. Más aún, los efectos persistieron con el paso del tiempo. Tres meses después de la maratón de lectura, los estudiantes del grupo de tratamiento aún leían 3,1 libros más que los del grupo de control y sus puntajes en lectura eran 0,06 desviaciones estándar más altos.

Sin embargo, dos estudios realizados en Madagascar indican que los problemas sistémicos pueden limitar la magnitud de las mejoras pedagógicas. En 2005, el país entregó a los docentes herramientas que especificaban sus deberes en detalle (por ejemplo, un manual instructivo). Esta iniciativa fue evaluada con un estudio de asignación aleatoria: los investigadores seleccionaron 30 distritos para la evaluación y eligieron aleatoriamente 15 de ellos para que recibieran esta intervención, mientras que los otros continuaron administrando sus escuelas de la manera en que lo hacían habitualmente.

En un primer estudio de esta intervención, Lassibille *et al.* (2010) constataron que, al cabo de dos años, las escuelas donde se había implementado el programa tenían tasas de asistencia mayores y tasas de repetición menores que las escuelas del grupo de control, pero sólo cuando el programa se combinaba con intervenciones a nivel distrital y subdistrital destinadas a mejorar el flujo de trabajo. Sin embargo, aun con estas intervenciones, los puntajes obtenidos en las pruebas de las escuelas del grupo de tratamiento no eran significativamente más altos desde el punto de vista estadístico que el obtenido en las escuelas del grupo de control. (Véase en el Recuadro 3 una definición de las diferencias estadísticamente significativas). Estos resultados indican que mejorar la gestión docente tal vez no sea suficiente cuando hay otros obstáculos sistémicos.

Recuadro 3. ¿Qué son las diferencias estadísticamente significativas?

Para expresarlo en términos sencillos, cuando una diferencia entre los resultados de dos grupos es “estadísticamente significativa” quiere decir que es improbable que haya ocurrido por azar. La frase tiene poco que ver con el tamaño de la diferencia entre ambos grupos. En una evaluación, los resultados del grupo de control pueden ser significativamente más altos que los del grupo de tratamiento desde el punto de vista estadístico, aún cuando la magnitud de la diferencia entre ambos resultados es pequeña. Para determinar si la diferencia entre dos resultados es estadísticamente significativa, se han desarrollado pruebas que calculan si la diferencia podría haberse dado por mera casualidad. Si las probabilidades de que así haya sido son demasiado altas, los investigadores son reacios a sostener que las diferencias entre los resultados son estables. Sin embargo, definir qué significa que las probabilidades sean “demasiado altas” es un proceso un tanto arbitrario. De hecho, los investigadores del campo de la educación suelen informar tres niveles de significancia estadística: 10%, 5% y 1%. Un nivel de significancia del 10% indica que hay una probabilidad de 0,10 de encontrar por azar una diferencia como la observada; es decir, hay una probabilidad de 1 en 10 de que la diferencia observada entre los dos grupos sea una observación espuria de la muestra seleccionada para el estudio.

En un segundo estudio realizado a partir de la misma intervención, Glewwe y Maïga (2011) concluyeron que el efecto promedio del programa fue el mismo para los docentes nombrados o de planta permanente que para los docentes por contrato. La primera diferencia entre ambos grupos fue que los primeros eran titulares, mientras que los segundos tenían contratos anuales renovables. Se trata de una cuestión interesante porque en Madagascar se incorporó a numerosos docentes por contrato, quienes normalmente están menos capacitados pero suelen tener más incentivos para desempeñarse bien a fin de asegurarse la renovación de sus contratos. Si se considera que el tipo de contrato es un buen indicador representativo de la experiencia o aptitud del docente, estos resultados implican que mejorar la eficacia de los docentes que tienen poca experiencia (y que lógicamente son los que más lo necesitan) tal vez no sea suficiente para mejorar el rendimiento estudiantil.

Estas investigaciones ilustran el potencial que tienen las intervenciones que fijan con claridad las expectativas sobre la labor docente a través de la introducción de mejoras pedagógicas. Pero también sirven como recordatorio de que dichas iniciativas pueden ser difíciles de implementar a gran escala o resultar ineficaces en presencia de obstáculos sistémicos.

Aumentar las horas de clase

En Estados Unidos se han experimentado con políticas que amplían la jornada escolar y prolongan el año lectivo, y los resultados de estas intervenciones son alentadores. Linden, Herrera y Grossman (2001) evaluaron un programa en Washington D. C. en el 2006 que

ofrecía actividades académicas, de enriquecimiento y de apoyo después de la escuela y durante el verano a alumnos de escuelas de nivel medio (6° a 8° grado). Para el segundo año del programa, se elevó el puntaje de los alumnos 0,09 desviaciones estándar en las pruebas de comprensión lectora y 0,12 en las de matemática. También observaron que los estudiantes que participaban en el programa eran más propensos a inscribirse en una escuela secundaria, conseguir información sobre estas escuelas, visitarlas y conversar con pares y adultos sobre a qué escuelas postularse.

En Chicago, el gobierno local adoptó en 1996 un plan obligatorio por el cual todos los alumnos de 3°, 6° y 8° grado que no aprobaban los exámenes de fin de año debían asistir a un curso remedial de seis semanas durante el verano, después del cual debían volver a rendir los exámenes. Quienes no aprobaban repetían de grado. Jacob y Lefgren (2004a) evaluaron los efectos de este programa de verano. Dado que los alumnos no fueron asignados aleatoriamente al programa, los investigadores compararon a los que no habían aprobado los exámenes por sólo unos pocos puntos con aquellos que los habían aprobado por apenas unos pocos puntos. Como se puede esperar que los alumnos ubicados cerca del puntaje de corte tengan un nivel similar —y dado que todos aquellos que no aprobaban debían asistir al campamento de verano mientras que los que habían aprobado no debían hacerlo—, comparar estos dos grupos de alumnos podía arrojar un cálculo confiable del efecto del programa en el caso de los alumnos ubicados cerca del puntaje de corte. Esta estrategia empírica se conoce como “diseño de regresión discontinua” (Recuadro 4).

Recuadro 4. ¿Qué es el diseño de regresión discontinua?

Los diseños de regresión discontinua se usan comúnmente cuando se asignan individuos a un programa dependiendo de si alcanzan un punto de corte en algún tipo de calificación, como las pruebas de rendimiento estudiantil o los exámenes de certificación docente. En estos casos, los investigadores pueden llegar a calcular los efectos del programa comparando los resultados de los individuos que no alcanzaron el umbral por poco margen con quienes lo alcanzaron por poco margen. Si una parte de la premisa razonable de que muy probablemente las personas que están apenas por debajo y por arriba del punto de corte tendrán resultados comparables (por ejemplo, si se presupone que los alumnos que aprobaron o reprobaron un examen con lo justo tienen habilidades similares), entonces la comparación de los resultados de estos dos grupos después de instrumentar el programa permitirá calcular el efecto promedio. Sin embargo, es importante tener presente que el cálculo sólo se aplicará a individuos ubicados cerca del punto de corte, razón por la cual esto comúnmente se conoce como efecto de tratamiento promedio “local”.

Los alumnos de 3° grado que asistieron al campamento de verano aprendieron cerca de un 20% de lo que aprenden en un año durante el primer año del programa, con un retroceso gradual del 25%-40% en el segundo año del programa. Los alumnos de 3° grado que no pasaron de grado también mejoraron su desempeño más que quienes

pasaron de grado. Sin embargo, Jacob y Lefgren (2004a) no detectaron ningún efecto en los alumnos de 6° grado. Estos resultados son curiosos, ya que no ofrecen una buena explicación sobre por qué el programa tuvo efectos sólo en algunos niveles o grados.

Estudios como el realizado por Jacob y Lefgren indican que aumentar las horas de clase puede ser una estrategia efectiva para mejorar los resultados de aprendizaje de los alumnos menos favorecidos. Sin embargo, es importante advertir que dichos estudios analizan programas focalizados en escuelas o estudiantes desfavorecidos y no necesariamente expresan la efectividad que tendría el aumento de las horas de clase si dicha medida se instrumentara con carácter universal.

Reducir la cantidad de alumnos por clase

Las intervenciones que reducen la cantidad de alumnos por clase han sido muy controversiales. Los partidarios de esta medida alegan que reducir la proporción alumno-docente puede hacer más manejables las clases y ayudar a los alumnos desfavorecidos porque pueden recibir la atención personalizada que necesitan. Sus detractores sostienen que la mera reducción de la cantidad de alumnos por aula no lleva a mejoras en el aprendizaje a menos que genere cambios en la pedagogía docente, lo que es improbable, y que la contratación de más docentes por alumno eleva considerablemente los costos.

Estudiar los efectos de reducir la cantidad de alumnos por clase es difícil porque los alumnos rara vez son asignados aleatoriamente a clases con diferente cantidad de estudiantes. Una manera de estimar el impacto de estas intervenciones es valerse del límite máximo de alumnos por aula para poder implementar un diseño de regresión discontinua y comparar el desempeño de alumnos de clases que estén a uno y otro lado de ese límite. Angrist y Lavy (1999) aprovecharon una norma con muchos años de vigencia en Israel que fijaba en 40 el límite máximo de alumnos por aula. Si en una clase con 40 alumnos se inscribía un alumno más, la norma exigía dividir la clase en dos grupos de 20,5 alumnos (en promedio) cada uno. Cuando compararon el desempeño de los alumnos pertenecientes a clases a uno y otro del punto de corte, los autores hallaron que el hecho de estar en una clase con menos alumnos tenía efectos positivos sobre el desempeño, pero este hallazgo luego fue refutado por Urquiola y Verhoogen (2009).²

² Urquiola y Verhoogen estudiaron el caso de Chile, donde rige una norma que establece que el límite máximo de alumnos por clase es 45. Hallaron que los datos seguían un patrón muy parecido al de Israel. Sin embargo, también observaron que la mayor parte de las escuelas privadas intentaban evitar que se inscribiera la cantidad de alumnos que habría de obligarlas a dividir la clase, ya sea ajustando sus aranceles o no admitiendo al nuevo alumno. (La única excepción eran aquellas escuelas en las que los padres valoraban las aulas de pocos alumnos y cuya población estudiantil provenía, en general, de familias más acomodadas). Por lo tanto, los autores mostraron que la estrategia usada por Angrist y Lavy (1999) podía dar lugar a cálculos sesgados sobre el impacto de reducir la cantidad de alumnos por clase, ya que los alumnos asignados a estas clases más pequeñas no son comparables con los alumnos de clases de tamaño regular.

Afortunadamente, los gobiernos de los países desarrollados y en desarrollo han venido mostrando cada vez más interés por saber si reducir la cantidad de alumnos por clase mejora el aprendizaje de los alumnos y han invertido en estudios de asignación aleatoria para obtener una respuesta definitiva a esta cuestión.

Sin lugar a dudas, la evaluación rigurosa más conocida sobre una política destinada a reducir la cantidad de alumnos por clase en un país desarrollado fue realizada en el estado de Tennessee, Estados Unidos. Allí, en 1985, el gobierno local adoptó un programa de reducción del número de alumnos por aula denominado Student/Teacher Achievement Ratio (rendimiento según la proporción alumno-docente), que pasó a ser conocido como Proyecto STAR, por sus siglas en inglés. El programa asignó aleatoriamente a 11.600 alumnos y a sus docentes a una de las siguientes opciones de clases: 1) clase reducida (13–17 alumnos), 2) clase de tamaño regular (22–25 alumnos) y 3) clase de tamaño regular con un docente auxiliar de tiempo completo. Krueger (1999) evaluó los resultados y constató que, en promedio, los alumnos que asistían a clases reducidas tuvieron un rendimiento 4 puntos porcentuales más alto en el primer año del programa (alrededor de 0,20 desviaciones estándar) y que esta ventaja aumentó un punto porcentual por año a lo largo de los años subsiguientes. Estos efectos fueron más marcados entre los alumnos pobres o pertenecientes a minorías. Sin embargo, se observó que la presencia de docentes auxiliares tuvo poca incidencia en el rendimiento de los alumnos. Si bien la magnitud de estos efectos ha sido objeto de gran debate (Hanushek, 2003), el efecto de la reducción del tamaño de las clases observado en este estudio siguió siendo un punto de referencia para otras evaluaciones de impacto realizadas en el campo de la educación. Se considera que este estudio demuestra el potencial que tiene reducir considerablemente la cantidad de alumnos por clase en el contexto de un país desarrollado.

La principal contribución del estudio de Krueger fue que usó un estudio de asignación aleatoria para evaluar la reducción de la cantidad de alumnos por aula. Sin embargo, los beneficios en términos de validez interna que reportan experimentos como el proyecto STAR pueden perderse en validez externa (Recuadro 5). Muchos han argumentado que si bien Krueger ofreció una respuesta rigurosa a la pregunta sobre si el proyecto STAR mejoró el aprendizaje, resulta menos claro determinar si acaso sus conclusiones pueden hacerse extensivas a otros contextos fuera de Tennessee.

Recuadro 5. ¿Qué son la validez interna y la validez externa?

La validez interna se refiere a si un estudio obtuvo un cálculo no sesgado del efecto de una intervención. La validez externa se refiere a si el efecto puede hacerse extensivo a una población de interés. Tomemos el ejemplo de un programa de capacitación implementado antes de que el docente comience a ejercer su profesión. La validez interna estaría relacionada a asuntos tales como si los docentes fueron asignados aleatoriamente al programa (sesgo de selección), si se trata de docentes principiantes que habrían mejorado aun en ausencia del programa (efectos de maduración) o si los docentes que abandonaron el programa antes de que se midieran los resultados fueron los que menos se beneficiaron (desgaste diferencial). Por su parte, la validez externa se ocuparía de asuntos tales como si el programa o sus principales beneficiarios son similares a otros grupos al punto que, al evaluar a esos otros grupos, los resultados obtenidos puedan considerarse comparables.

Hoxby (2000) contribuyó al debate sobre la validez externa del Proyecto STAR al evaluar el impacto del tamaño de las clases en Connecticut usando dos métodos rigurosos. En primer lugar, mediante un diseño de regresión discontinua comparó a los alumnos asignados a clases reducidas o de tamaño regular en función de las normas sobre tope mínimo/máximo de alumnos por clase. Luego, aprovechó los cambios en las tendencias de la población. Dado que estos cambios afectan la proporción alumno-docente aunque difícilmente influyen en el aprendizaje de los alumnos, Hoxby utilizó dichos cambios para hacer una estimación mediante el método de variables instrumentales (Recuadro 6). Con estos dos métodos, descubrió que el tamaño del grupo no producía un efecto estadísticamente significativo en el rendimiento de los alumnos. De hecho, hasta pudo descartar efectos modestos (0,02–0,04 desviaciones estándar en el caso de una reducción del 10% en la cantidad de alumnos por clase). Si bien Hoxby no pudo beneficiarse de la asignación aleatoria en este caso de Connecticut, su estudio es considerado un aporte importante para las implicaciones de los hallazgos sobre el programa STAR.

Recuadro 6. ¿Qué es la estimación con variables instrumentales?

La estimación con variables instrumentales suele usarse cuando no hay una asignación aleatoria de individuos a un grupo de tratamiento y no hay una discontinuidad que determine quién participó en un grupo de tratamiento y quién no lo hizo. Se basa en la observación de que la variación en los resultados de los participantes de un programa es, en parte, endógena (es decir, se relaciona con las características de los participantes) y, en parte, exógena (es decir, se relaciona con las características del programa). Así, una manera de evaluar un programa sin asignación aleatoria es obtener datos sobre una variable que esté relacionada con la variación exógena pero no con la variación endógena en los resultados. Luego, se puede calcular el efecto del programa usando esta variable. No es fácil identificar tal variable (el “instrumento”) y obtener datos sobre ella. Sin embargo, hay una serie de instrumentos que se usan comúnmente, como las características geográficas que inciden en el tratamiento (por ejemplo, la distancia entre los participantes y el programa), las instituciones jurídicas que determinan la intensidad del tratamiento (por ejemplo, leyes diferentes en estados vecinos similares) y los datos de fenómenos naturales (por ejemplo, las fechas de nacimiento que determinan si una persona tiene acceso a un grupo de tratamiento o no). Por último, la asignación aleatoria a un grupo de tratamiento también puede usarse como instrumento, ya que no está correlacionada con las características de los individuos pero influye en el nivel de aceptación a participar en el programa.

La otra cuestión es si los resultados del Proyecto STAR pueden aplicarse a las naciones en desarrollo, donde la cantidad promedio de alumnos por clase es mucho mayor que en Estados Unidos. Duflo, Dupas y Kremer (2007) evaluaron un programa en Busia y Teso, Kenia, en el que se asignaron en forma aleatoria 210 escuelas primarias a un programa que 1) reducía la proporción alumno-docente, en promedio, de 80 a 46; 2) combinaba esta reducción con mayores incentivos para los docentes (mediante la incorporación de docentes locales con contratos de corto plazo o el aumento de la supervisión por parte de los padres), y 3) combinaba la reducción con una división de alumnos por clase en base a su rendimiento inicial (*tracking*). La reducción de la proporción alumno-docente, en ausencia de otras reformas, llevó a una disminución en el esfuerzo de los docentes y a un aumento estadísticamente insignificante de los puntajes obtenidos por los alumnos en las pruebas. Sin embargo, al combinar la reducción de la cantidad de alumnos por clase con los mayores incentivos para los docentes, se obtuvieron mejoras significativamente más altas en los puntajes (0,19 desviaciones estándar) y, al combinar la reducción en la cantidad de alumnos por clase con el *tracking*, se obtuvieron mejoras aún mayores (0,25-0,31 desviaciones estándar). Este estudio pone en evidencia la importancia de las interacciones de las políticas al indicar que la reducción en la cantidad de alumnos por clase difícilmente mejore el rendimiento de los alumnos si esa medida no se acompaña con incentivos a los docentes para que se modifique su desempeño en el aula.

Estos estudios ofrecen una respuesta con ligeras variaciones a la pregunta sobre si la reducción del tamaño de las clases mejora el aprendizaje de los alumnos. Al parecer, la

respuesta es que la disminución de la proporción alumno-docente puede mejorar el desempeño de los alumnos, pero sólo si esa medida impulsa al docente a cambiar su pedagogía y, con ello, la experiencia educativa de los niños en el aula. Este resultado no se da automáticamente y puede depender de en qué medida pueden introducirse mejoras en la gestión del aula (es decir, el tamaño inicial de la clase) y de la combinación de incentivos para gestar cambios en la conducta docente.

Atraer a los mejores a la docencia

La teoría económica parece indicar que las personas con logros académicos y/o profesionales destacados tienen más probabilidades de ser docentes eficaces. Por una parte, la “teoría del capital humano” planteada por Becker (1964), Schultz (1963) y Mincer (1962) propone que los empleadores deberían contratar a personas con más formación porque cuanto más educación hayan recibido, mayor será su productividad dado que pueden entender y utilizar nueva información. Por otra parte, la “teoría de la señalización del mercado de trabajo” desarrollada por Spence (1974) y Arrow (1963), entre otros, postula que, aun cuando la educación adquirida por los potenciales empleados no aumente su productividad, los más calificados seguirán formándose para enviar una señal a sus empleadores de que tienen habilidades que otros carecen y, por consiguiente, merecen una mejor retribución. Si bien estas teorías difieren respecto de las razones por las cuales los empleadores deberían contratar a los más calificados, ambas suponen que atraer a los docentes mejor formados generará una fuerza laboral más calificada.

Hay estudios del campo de la economía que postulan que atraer a los más talentosos a la docencia también puede tener un efecto multiplicador. Becker y Murphy (2000) plantearon que las acciones de las personas suelen influir en la información y los incentivos de sus vecinos, y Montgomery (1991) mostró que esto ocurre de diversas maneras en los mercados de trabajo. A estas interacciones se las ha llamado “efectos multiplicadores sociales”. Estos modelos ofrecen razones para creer que si la docencia es capaz de atraer a los más calificados, los postulantes competitivos que no habían considerado ingresar a la docencia pueden sentirse atraídos a hacerlo.

Un mecanismo que los sistemas educativos pueden utilizar para atraer a individuos talentosos consiste en ofrecer salarios competitivos. Como sería de esperar, los economistas tienen mucho para decir sobre el papel que desempeña el salario. Como lo señalaron Akerlof (1982) y Shapiro y Stiglitz (1984), el salario cumple dos funciones: contratar trabajadores y brindar incentivos al esfuerzo. Alegan que esta es la razón por la que las empresas que buscan atraer a empleados sobresalientes estarían dispuestas a pagar un “salario de eficiencia”, es decir, más de lo que ofrecen sus competidores. No obstante, Lazear y Rosen (1981) plantean que las empresas pueden evitar algunos de los problemas inherentes a los salarios de eficiencia pagando a sus empleados en función de su

desempeño. En los últimos tiempos, esta idea ha vuelto a ocupar un lugar de interés en la literatura sobre educación. Específicamente, Barlevy y Neal (2011) mostraron que el criterio de “pago por percentil”, por el cual se remunera a los docentes de acuerdo con la mejora en el desempeño de sus alumnos en comparación con los pares con niveles de rendimiento similar al principio del año lectivo, incentiva al docente a distribuir sus esfuerzos en niveles socialmente óptimos entre todos los alumnos.

Numerosos estudios han procurado entender hasta qué punto los requisitos de ingreso a la docencia motivan a los más talentosos a incorporarse a esa profesión. También han intentado saber cómo identificar a aquellos que tendrán más probabilidades de tener un buen desempeño como docentes.

En el Cuadro 2 se resumen cuatro tipos de intervenciones que han sido evaluadas rigurosamente: 1) fijar requisitos de ingreso a la docencia, 2) flexibilizar los requisitos de ingreso a la docencia para aspirantes sobresalientes, 3) recompensar a los docentes más calificados y con mayor experiencia y 4) aumentar el salario docente.

Los estudios rigurosos sobre estas intervenciones son escasos porque los docentes no son asignados aleatoriamente a las características (por ejemplo, de raza y género) ni a las acreditaciones (títulos o experiencia) y los alumnos rara vez son asignados aleatoriamente a sus docentes. Los estudios incluidos en esta revisión bibliográfica se ubican entre los más rigurosos en esta línea de investigación, pero aún resta mucho por hacer para obtener respuestas más concluyentes a los interrogantes planteados. Para minimizar el riesgo de confundir los efectos de selección y tratamiento, la mayoría de las investigaciones sobre los requisitos impuestos al docente usan una estrategia conocida en econometría como “efectos fijos” (Recuadro 7).

Recuadro 7. ¿Qué son los efectos fijos?

Los efectos fijos se usan comúnmente en las investigaciones que identifican correlaciones entre dos factores. Su propósito principal es dar cuenta de los factores que no pueden ser observados o para los cuales no hay datos. Esto se hace desagregando la muestra en subgrupos (por ejemplo, subgrupos de alumnos, de escuelas y de docentes) y comparando las unidades de cada subgrupo entre sí a lo largo del tiempo. El modelo de efectos fijos permite controlar las características que se mantienen “fijas” o invariables (es decir, que no cambian) a lo largo del tiempo, pero no tiene manera de dar cuenta de los factores que varían en el tiempo. A diferencia de otras estrategias empíricas analizadas en este trabajo, los efectos fijos, independientemente de su sofisticación, por lo general no permiten establecer en forma concluyente relaciones de causa y efecto. En cambio, se usan para aumentar la confiabilidad de las correlaciones.

Fijar requisitos de ingreso a la docencia

Prácticamente todos los sistemas educativos exigen que para ser docente, el aspirante reúna ciertas calificaciones. Estos requisitos aspiran a fijar estándares mínimos para el ejercicio de la profesión, pero en muchos sistemas educativos se han cuestionado por considerarse que no son buenos predictores de la eficacia del docente y que hacen costoso el ingreso a la docencia, sobre todo para los mejores aspirantes que ya enfrentan un alto costo de oportunidad para incorporarse a la profesión docente.

La mayoría de las investigaciones sobre este tema se han centrado en determinar si los alumnos de los docentes certificados se desempeñan, en promedio, mejor que los alumnos a cargo de docentes no certificados. De acuerdo con los datos obtenidos, el efecto predictivo de la certificación suele ser pequeño.

Kane, Rockoff y Staiger (2006) usaron datos de los años lectivos 1998–1999 a 2004-2005 para estudiar la relación entre el rendimiento estudiantil y la certificación de los docentes de comprensión lectora y matemática de 4° a 8° grado de la ciudad de Nueva York. Estos “datos de panel” permitieron a los autores observar la eficacia de un docente en la misma escuela, el mismo grado y el mismo año a lo largo del tiempo y usar combinaciones de efectos fijos para tener en cuenta los factores que confunden el efecto de la certificación. Los efectos de la certificación docente fueron, en el mejor de los casos, pequeños. De hecho, en matemática, los alumnos asignados a docentes sin certificación tuvieron, en general, un desempeño igual al de los alumnos asignados a docentes que cumplían con los requisitos de certificación tradicionales. Sin embargo, estos resultados fueron un tanto distintos según el grupo de docentes sin certificación con el que se comparó a los docentes certificados por la vía tradicional.

Una posible explicación de la falta de efectos de la certificación docente en el estudio realizado en Nueva York es que la certificación sirva como indicador representativo de la calidad docente en ciertas materias y en ciertos niveles más que en otros o bien que la certificación en lugares como Nueva York no tenga el rigor suficiente para predecir la eficacia del docente. Esto implicaría que la certificación docente debería ser más estricta en vez de eliminarla por completo.

Eso es precisamente lo que parece indicar un estudio realizado por Clotfelter, Ladd y Vigdor (2007) en Carolina del Norte. Los autores midieron el efecto predictivo de una serie de acreditaciones docentes (incluida la certificación) sobre el rendimiento estudiantil en el nivel secundario durante el período 1999–2002, usando permutaciones a nivel de estudiante, escuela, materia y año como efectos fijos. En consonancia con la hipótesis de que la certificación podría desempeñar un papel más importante en la eficacia del docente de escuela secundaria, Clotfelter y sus colaboradores observaron que

tener un docente con una certificación alternativa *reduce* el rendimiento estudiantil 0,06 desviaciones estándar y que tener un docente sin certificación tradicional ni alternativa afecta negativamente el desempeño de los alumnos 0,05 desviaciones estándar. Los autores también encontraron fundamentos para respaldar el argumento de fijar estándares más altos de certificación al registrar efectos pequeños pero positivos cuando el docente tenía una certificación de la NBPTS (National Board of Professional Teaching Standards), un programa de certificación de alcance nacional considerado por muchos como más riguroso que los programas tradicionales de certificación de Estados Unidos. NBPTS exige que los docentes presenten videos de su trabajo en el aula y escriban trabajos o disertaciones, y se trata de un programa muy exigente en cuanto al esfuerzo que deben realizar los docentes y al tiempo que deben invertir.

Los datos diversos sobre las certificaciones llevaron a Estados Unidos a un estudio de asignación aleatoria para determinar si la certificación valía la inversión. Cantrell *et al.* (2008) compararon en 2003 el desempeño de alumnos de nivel primario de Los Ángeles asignados aleatoriamente a docentes postulantes y no postulantes a obtener la certificación de la NBPTS. Los alumnos de los docentes certificados por la NBPTS obtuvieron un puntaje 0,22 desviaciones estándar por encima de los alumnos cuyos docentes se habían postulado para recibir la certificación pero no habían sido aprobados. Los alumnos de los docentes que fueron certificados por la NBPTS también tuvieron un mayor aprendizaje que los alumnos de los docentes que se postularon y no fueron aprobados. Sin embargo, los docentes certificados por la NBPTS no resultaron más eficaces que los docentes que no se postularon para obtener la certificación. Esto parecería demostrar que si bien la NBPTS puede ser muy buena para identificar a los docentes más eficaces entre sus postulantes, tal vez no esté logrando atraer como postulantes a los más eficaces.

Otro requisito de ingreso al que recurren cada vez más los sistemas educativos es el examen, con el cual normalmente se evalúa el conocimiento de la/s materia/s o el conocimiento pedagógico específico del área del docente. La evidencia sobre la utilidad de estos exámenes está en plena evolución, pero parece alentadora a la hora de discernir entre los docentes a aquellos que tendrán un mejor desempeño en el aula.

En Perú, Meltzer y Woessmann (2010) aprovecharon que se evaluó en el mismo año a docentes y alumnos sobre dos asignaturas para determinar si un alumno a cargo del mismo docente en dos materias diferentes se desempeñaba mejor en una de ellas si el docente tenía un conocimiento relativamente superior en esa área temática. Los autores usaron el modelo de efectos fijos para analizar las variaciones intra-docentes e intra-alumnos en los resultados de los alumnos, controlando las características fijas de los alumnos, docentes y asignaturas. Los datos obtenidos sugieren que el conocimiento de la materia por parte de un docente afecta el rendimiento estudiantil: un aumento de una

desviación estándar en el puntaje del docente elevaba el puntaje del alumno en 0,10 unidades de desviación estándar. Esto significa que si un alumno pasaba de tener un docente en el percentil 5 de la distribución del conocimiento sobre la materia a un docente en el percentil 50, hacia fin de año su desempeño habría de mejorar en 0,17 desviaciones estándar.

Los resultados de un estudio reciente realizado en Estados Unidos fueron muy diferentes de los obtenidos en el estudio de Perú. Cantrell y Kane (2013) se valieron de correlaciones para analizar si los exámenes sobre conocimiento de contenidos para la enseñanza (Content Knowledge for Teaching o CKT) —que evalúan cuánto saben los docentes acerca de cómo los alumnos adquieren y comprenden las habilidades específicas para matemática y comprensión lectora— se correlacionaban con las mediciones de los puntajes de valor agregado de docentes de escuelas de nivel primario y medio de seis distritos estadounidenses: Charlotte-Mecklenburg (Carolina del Norte); Dallas (Texas); Denver (Colorado); condado de Hillsborough (Florida); ciudad de Nueva York y Memphis (Tennessee). Con el objeto de dar cuenta del efecto de la asignación de los docentes a los alumnos, los autores usaron los puntajes obtenidos en los exámenes CKT por los docentes para predecir su desempeño promedio en una serie de indicadores en dos secciones diferentes. Cantrell y Kane constataron que los exámenes de CKT se correlacionaban con los protocolos de observación de las clases específicas de cada materia. Sin embargo, no se correlacionaban con los puntajes de valor agregado de los docentes, independientemente de si fueron medidos usando exámenes de los estados o exámenes que miden habilidades más complejas.

En definitiva, los estudios sobre requisitos de ingreso a la docencia, al parecer, hallan poca evidencia que respalde la efectividad de los requisitos tradicionales (por ejemplo, la certificación) en los países desarrollados y plantean resultados alentadores en el caso de los requisitos alternativos (por ejemplo, exámenes de ingreso) usados en los países en desarrollo. Sin embargo, quedan por responder preguntas importantes sobre su validez interna y externa.

Flexibilizar los requisitos de ingreso a la docencia para aspirantes sobresalientes

Tal vez sea la falta de claridad sobre cuáles son los requisitos ideales de ingreso a la docencia lo que ha motivado a algunos sistemas educativos a flexibilizar las exigencias tradicionales en el caso de individuos talentosos que aspiran a incorporarse a la profesión con la esperanza de convertir la docencia en una opción atractiva para ellos.

Uno de los programas que buscan atraer a jóvenes talentosos a la docencia en Estados Unidos es Teach for America (TFA), que selecciona graduados terciarios o universitarios sobresalientes de una amplia variedad de carreras para enseñar en escuelas con muchas

necesidades durante dos años. La doble misión de TFA es elevar la calidad de la enseñanza y, al mismo tiempo, servir como experiencia transformadora para los propios participantes de manera tal que, después de dos años, lideren la reforma educativa en la disciplina o el sector donde elijan trabajar. Desde su institución en 1989, el programa TFA ha servido de inspiración para 28 países que han instaurado programas similares.

El estudio más riguroso sobre el impacto de TFA, que se realizó en el año 2001, analizó una muestra que abarcó de 1° a 5° grado en seis de las 15 regiones donde se desempeñan sus participantes (Baltimore, Chicago, Los Ángeles, Houston, Nueva Orleans y la región del delta del Mississippi). Decker, Mayer y Glazerman (2004) asignaron aleatoriamente a alumnos de la misma escuela y del mismo grado a docentes de TFA, por un lado, y a docentes del grupo de control, por el otro, y compararon su desempeño al cabo de un año. Los docentes del grupo de control eran todos aquellos a los que los alumnos habrían sido asignados en ausencia de docentes de TFA (que no necesariamente estaban certificados ni tenían experiencia). De hecho, era más probable que los docentes de TFA se hubieran formado en instituciones prestigiosas y hubieran obtenido un título para ejercer la docencia, pero era menos probable que tuvieran formación específica en educación y experiencia en práctica docente en comparación con los docentes del grupo de control. Aun así, en general, los alumnos de los docentes de TFA tuvieron un desempeño 0,15 desviaciones estándar más alto que los de los docentes del grupo de control y 0,26 más alto que los asignados a docentes principiantes (con uno a tres años de experiencia) del grupo de control. No obstante, no se registraron diferencias en comprensión lectora.

Datos recientes indican que el efecto de este tipo de programas podría ser mayor en países en desarrollo, donde la efectividad del docente promedio es más baja. Alfonso, Santiago y Bassi (2011) evaluaron en Chile el impacto de una adaptación de TFA denominada Enseña Chile (eCh) durante los años lectivos 2009–2010 y 2010–2011. Alfonso y sus colaboradores no pudieron asignar aleatoriamente docentes del programa eCh a los alumnos, de modo que recurrieron a la correspondencia del puntaje de propensión (Recuadro 8). Al igual que el modelo de efectos fijos, esta técnica es mejor que las correlaciones que tratan de explicar las medidas observables que podrían sesgar los cálculos del impacto de una política en particular, pero no puede distinguir los efectos de selección de los efectos de tratamiento. Al cabo de un año, los alumnos de las escuelas que recibieron a un docente del programa eCh tuvieron un puntaje 0,22–0,51 desviaciones estándar más alto en español y 0,17–0,43 en matemática. Transcurridos dos años, los alumnos obtuvieron un puntaje 0,75 desviaciones estándar más alto en español y 0,33 desviaciones estándar más alto en matemática. Estos cálculos parecen demasiado elevados para una intervención educativa, lo que apuntan a que probablemente hubo un sesgo de selección, pero resultan muy motivadores para evaluar más rigurosamente este tipo de programas en los países en desarrollo.

Recuadro 8. ¿Qué es la correspondencia del puntaje de propensión?

La correspondencia del puntaje de propensión es una estrategia empírica que se emplea a menudo cuando no es posible usar ninguno de los métodos de evaluación rigurosos mencionados antes en este trabajo. No puede usarse para afirmar que un programa dado provocó determinado impacto. Antes bien, es simplemente un mecanismo para mitigar el problema de evaluación ya mencionado. El método es simple. Los investigadores primero identifican los factores que hacen que los individuos sean más propensos a autoseleccionarse para participar en un grupo de tratamiento y, luego, crean un grupo de control estableciendo correspondencias entre cada individuo “tratado” del grupo de tratamiento con uno o más individuos “no tratados” que tengan características observables comparables. De este modo, los investigadores pueden comparar los resultados de ambos conjuntos de individuos después del tratamiento como si se tratara de un efecto de ese tratamiento. El aspecto negativo de este tipo de estudios es que los investigadores sólo pueden trazar correspondencias entre los individuos sobre aquello que pueden observar. Si hay características no observables que llevaron a que algunos individuos sean más propensos a seleccionarse para participar en un programa, no hay manera de que los investigadores lo sepan, de modo que sus cálculos estarán sesgados.

Estos estudios sugieren que si bien no se sabe qué requisitos de ingreso a la docencia son más eficaces a la hora de predecir la futura efectividad de los docentes, tampoco hay razones convincentes para eliminar por completo estos requisitos. Es necesario saber mejor cuáles requisitos (si acaso alguno) sirven para asegurar un “piso” o nivel mínimo de calidad de enseñanza así como si cabe (y de ser así, cómo) usar más ampliamente los requisitos de ingreso a la docencia sin depender de otros mecanismos quizá más costosos pero también más efectivos para identificar a los docentes eficaces.

Sin embargo, estos estudios se centran en quienes se incorporan a la docencia a través de caminos alternativos que son muy selectivos. Es así que surge la pregunta de si el grueso de los docentes certificados por vías alternativas que ingresan a la profesión a través de mecanismos mucho menos selectivos serán menos eficaces que aquellos autorizados a ejercer la docencia por la vía regular.

Fue precisamente esa pregunta la que llevó a Constantine *et al.* (2009) a comparar el impacto que tienen los docentes certificados por la vía tradicional y por mecanismos alternativos en el rendimiento estudiantil en Estados Unidos en 2003. Para ello, asignaron aleatoriamente a alumnos de la misma escuela y del mismo grado a uno u otro tipo de docentes. El estudio se llevó a cabo con 2.600 estudiantes de 63 escuelas de 20 distritos escolares. Los autores constataron que la diferencia entre los docentes certificados por la vía tradicional y por vías alternativas podía ser mucho menos nítida de lo que se hubiese esperado: 1) en ambos casos variaba sustancialmente la cantidad de horas de formación previas al inicio del ejercicio docente; 2) en promedio, tampoco había diferencias significativas en los puntajes obtenidos por un grupo y otro en los exámenes de ingreso a

la universidad o institución terciaria donde se habían formado, en la selectividad de sus instituciones educativas ni en sus logros académicos; 3) ninguno de los dos grupos resultó, en promedio, más eficaz para mejorar el desempeño de los alumnos, y 4) ni la cantidad ni los contenidos de las materias cursadas por los docentes certificados por alguna vía alternativa marcaban una diferencia en su efectividad. Estos hallazgos sugieren que los docentes certificados a través de programas alternativos pero no muy selectivos no son diferentes de aquellos que ingresaron a la docencia por la vía regular.

Recompensar a los docentes más calificados y con mayor experiencia

Muchos sistemas educativos pagan sueldos más altos a sus docentes por los estudios de posgrado o los años de experiencia acumulados. La fundamentación de estas diferencias salariales es que permiten atraer a los que tienen mejor formación y retener a los docentes con más experiencia. El supuesto implícito en estas políticas es que se trata de docentes más eficaces, de modo que se han realizado investigaciones al respecto para confirmar esta hipótesis.

Hanushek *et al.* (2005) combinaron datos de la base Texas Schools Microdata Panel con datos sobre la asignación de aulas tomados de un distrito escolar urbano grande anónimo del estado de Texas para determinar si ciertas características de los docentes, entre las que se incluyó el tener una maestría o experiencia anterior, predecían un mejor rendimiento estudiantil. El estudio usó como efectos fijos las escuelas y los alumnos. Los autores detectaron que el desempeño de los alumnos de docentes con maestrías no difería, en promedio, del de los alumnos cuyos docentes no habían hecho un posgrado. También constataron que los docentes que se encontraban en el primer año del ejercicio de su profesión tenían, en promedio, un desempeño mucho más bajo que otros docentes (el rendimiento de sus alumnos era 0,12–0,16 desviaciones estándar inferior), aunque los años adicionales de experiencia docente tenían poco impacto sobre el rendimiento estudiantil.

Hasta hace poco, la validez externa de las credenciales tradicionales de los docentes de Estados Unidos también resultaba problemática, ya que la mayoría de los estudios realizados se centraban en un único distrito escolar. En 2012, Kane y Staiger publicaron un informe en el que analizan si los docentes con más experiencia o con maestrías tienen alumnos con mejor desempeño en las pruebas de matemática y comprensión lectora. Al igual que en el estudio de Cantrell y Kane (2013) mencionado anteriormente, el estudio de Kane y Staiger (2012) examinó los distritos escolares de Charlotte-Mecklenburg (Carolina del Norte); Dallas (Texas); Denver (Colorado); condado de Hillsborough (Florida); ciudad de Nueva York y Memphis (Tennessee).

Al no contar con la asignación aleatoria de docentes a alumnos, los autores trataron de explicar los factores que confundían el efecto de las credenciales usando el desempeño promedio de dos secciones diferentes de alumnos a cargo del mismo docente como su variable de resultado. Hallaron que el desempeño de los alumnos de docentes con 12 o más años de experiencia era 0,01 desviaciones estándar más alto en matemática y 0,02 más alto en comprensión lectora que el de los alumnos de docentes con menos de tres años de experiencia. De manera similar, constataron que los alumnos de docentes que tenían una maestría registraron un rendimiento 0,03 desviaciones estándar superior en matemática y 0,02 desviaciones estándar inferior en comprensión lectora.³

Los autores midieron el poder predictivo de estas ventajas tradicionales de los docentes usando un parámetro de calidad que combina el valor agregado del docente, los resultados de la encuesta a alumnos y el puntaje de las observaciones en el aula de otra clase. Detectaron que este parámetro combinado funcionaba mucho mejor para predecir el rendimiento estudiantil: los alumnos de un docente ubicado en el 25% más alto de este parámetro combinado se desempeñaron 0,22 desviaciones estándar mejor en matemática y 0,07 desviaciones estándar mejor en comprensión lectora en las pruebas de cada estado que los alumnos cuyo docente se ubicaba en el 25% más bajo del parámetro combinado.⁴

Juntos, estos estudios ofrecen evidencias convincentes de que las acreditaciones educativas y la experiencia son indicadores muy débiles de la efectividad de los docentes y que los parámetros más minuciosos pueden resultar mejores para predecir la calidad de la enseñanza. De acuerdo con estos hallazgos, sería recomendable que los sistemas educativos no pagaran salarios diferenciales en función de la mayor formación y experiencia para atraer a docentes mejores.

Aumentar el salario docente

Por último, una de las políticas más pregonadas para atraer a los más talentosos a la docencia es el aumento del salario docente. La eficacia de un salario alto para atraer a

³ Debido a la preocupación que existe en Estados Unidos respecto de los exámenes de los distintos estados, los autores también analizaron en qué medida la experiencia docente y los estudios de posgrado hacían predecir un mejor rendimiento estudiantil en pruebas de “auditoría” de matemática y comprensión lectora. Según sus resultados, el potencial predictivo de las maestrías era ligeramente más alto considerando estas pruebas: los alumnos de docentes con maestrías tuvieron un desempeño 0,05 desviaciones estándar superior en matemática y 0,03 desviaciones estándar superior en comprensión lectora al de los alumnos de docentes sin estudios de posgrado. En cambio, el efecto predictivo de la experiencia medido en función de estas pruebas de “auditoría” fue más desigual: el rendimiento de los alumnos de docentes con 12 o más años de experiencia fue 0,05 desviaciones estándar más bajo en matemática y 0,03 desviaciones estándar más alto en comprensión lectora.

⁴ Los resultados fueron similares usando pruebas de “auditoría”: el desempeño del primer grupo de alumnos fue 0,13 desviaciones estándar más alto en matemática y 0,13 desviaciones estándar más alto en comprensión lectora que el del segundo grupo.

docentes mejores es difícil de evaluar porque los docentes no están asignados aleatoriamente a los salarios. Sin embargo, Hoxby y Leigh (2004) observaron que la sindicalización comprimía los salarios. Recurrieron a leyes que favorecían o dificultaban la sindicalización docente para averiguar si la disminución de las aptitudes de los docentes estadounidenses registrada en las últimas décadas puede atribuirse al aumento de las oportunidades que se les ofrece a las mujeres fuera de la profesión docente o bien a los salarios deprimidos que se pagan en la docencia. Así, pudieron valerse de las leyes como “instrumentos” de la compresión salarial en una estimación con variables instrumentales.

La compresión salarial aumentó en 9 puntos porcentuales la proporción de mujeres graduadas de diversos institutos educativos con menos habilidades que se incorporaron a la docencia e hizo que se redujera en 12 puntos porcentuales la proporción de graduadas con mejores aptitudes. Hoxby y Leigh también observaron que la mejora en la paridad salarial redujo la proporción de mujeres dedicadas a la docencia 3,2 puntos porcentuales en el caso de las docentes con mejores aptitudes, mientras que en el caso de las docentes con menos aptitudes la proporción se mantuvo en 0, es decir, no cambió. Estos hallazgos indican que la compresión salarial desempeñó un papel clave en el deterioro de las aptitudes en el ámbito de la profesión docente. La conclusión que se extrae de este estudio es que, en vez de incrementar los salarios universalmente, resulta conveniente pagar mejor a los mejores docentes para atraerlos y retenerlos.

Preparar a los docentes con formación y experiencia útiles

Las investigaciones teóricas y empíricas en el campo de la economía han demostrado que los trabajadores altamente calificados valoran más la capacitación que los trabajadores con menos preparación. Salop y Salop (1976) y, más tarde, Autor (2001) llamaron a este fenómeno “complementariedad” entre competencias y capacitación. Así, ofrecer capacitación puede promover una autoselección positiva respecto de las competencias que, de otro modo, los empleadores no pueden observar. Esto se aplica particularmente al mercado laboral docente, en el que hay muy pocas habilidades observables que permitan determinar quiénes tienen más probabilidades de ser buenos docentes.

Si la teoría del capital humano es correcta y la capacitación permite a las personas adquirir nuevas habilidades, los sistemas educativos deberían subsidiar la formación del docente antes de que comience a ejercer su profesión. Pero, tal como plantea el modelo de Becker (1964), esto es válido sólo si las competencias necesarias para ser un buen docente son específicas y no generales. Según Becker, dado que la capacitación para adquirir habilidades generales hace que los empleados tengan más probabilidades de desempeñarse bien no sólo en la empresa o industria que les proporciona la capacitación sino también en otras, las empresas son renuentes a invertir en capacitación en

habilidades generales. Sin embargo, los gobiernos tienen claros incentivos para subsidiar la formación docente previa al ejercicio de la profesión, dado que tal formación lleva casi invariablemente a un puesto de trabajo, lo que evita muchos de los denominados “problemas contractuales” que tienen las empresas privadas para brindar capacitación en competencias específicas (Prendergast, 1993).

Pero tal como señaló Autor (2001), aun cuando la formación corresponda a un modelo de señalización pura, la capacitación podría ser una inversión valiosa como mecanismo de selección. La capacitación puede generar información sobre las habilidades de los potenciales empleados que no sería posible observar en su ausencia y que las empresas pueden usar para tomar decisiones de contratación. Esto es particularmente pertinente para la política docente, dado que la docencia parece ajustarse muy bien al perfil de las ocupaciones descritas por Terviö (2003), en las que la experiencia inicial adquirida en el trabajo constituye un predictor bastante exacto del desempeño posterior.

Si bien la formación previa al ejercicio de la docencia es una de las vías más comunes a través de las que los gobiernos procuran mejorar las competencias docentes, hay pocas investigaciones rigurosas sobre su impacto. Esto se debe, en gran medida, a la autoselección hacia la capacitación, lo que dificulta la tarea de separar los efectos de selección de los efectos de tratamiento. El pequeñísimo corpus de investigaciones rigurosas sobre esta materia está conformado principalmente por trabajos empíricos que evalúan cambios en las modalidades tradicionales de formación o bien intervenciones implementadas en los comienzos de la carrera docente que podrían complementar la formación previa al ejercicio de la docencia.

En el Cuadro 3 se resumen dos tipos de intervenciones que han sido rigurosamente evaluadas: 1) asignar tutores a los docentes principiantes e 2) incluir un componente de práctica en la formación docente.

Asignar tutores a los docentes principiantes

Numerosos trabajos han notado que los docentes tienen un desempeño más bajo en sus primeros años de ejercicio. Esto ha llevado a algunos sistemas educativos a asignar tutores o entrenadores a los docentes principiantes para acelerar su proceso de aprendizaje en lo que suele llamarse “programas de inducción”. La evidencia sugiere que esta clase de intervenciones puede mejorar el aprendizaje de los alumnos, pero el impacto de la calidad y la dosificación de la tutoría no es fácil de anticipar cuando se diseñan estos programas.

Rockoff (2008) evaluó un programa de tutoría en la ciudad de Nueva York en el 2004. Para medir su impacto, comparó el desempeño de docentes con experiencia previa

(a quienes el programa no incluía) con el de docentes principiantes (destinatarios del programa y, por ende, con más probabilidades de que se les asignara un tutor). Esta estrategia se conoce como análisis de diferencias en diferencias (Recuadro 9). Rockoff constató que los maestros principiantes a quienes se les asignó un tutor tuvieron una probabilidad 4,5% mayor de completar su primer año que los docentes sin tutores, pero mostraron casi la misma probabilidad (o, para decirlo de otro modo, no mostraron una probabilidad mayor) de permanecer en sus escuelas o mejorar el rendimiento estudiantil.

Recuadro 9. ¿Qué es el análisis de diferencias en diferencias?

El análisis de diferencias en diferencias suele usarse cuando no pueden llevarse adelante otras estrategias empíricas rigurosas. Es simple en el sentido de que calcula el impacto de un programa obteniendo la diferencia de los resultados de un grupo de individuos que recibió un tratamiento antes y después del tratamiento (llamada “primera diferencia”), obteniendo la diferencia de los resultados de un grupo similar no sometido al tratamiento (llamado “segunda diferencia”) en el mismo período y luego restando la segunda de la primera. Si bien puede intentarse calcular el impacto de un programa usando sólo la primera diferencia, es probable que otras cosas hayan ocurrido al mismo tiempo que el tratamiento, lo que podría confundir sus efectos. Por lo tanto, la segunda diferencia les permite a los investigadores dar cuenta de cualquier otro evento que puede haber afectado a todos los individuos (denominado “tendencia secular”). Si nada más cambió entre los dos grupos en el momento del tratamiento y la composición de ambos continuó siendo la misma durante el tratamiento, el análisis de diferencias en diferencias ofrece un cálculo no sesgado del impacto de un programa.

Rockoff también usó una estimación con variables instrumentales para analizar si la experiencia, la cantidad de horas dedicadas a cada docente por los tutores y las percepciones de los docentes sobre los tutores marcaban alguna diferencia. Observó que los docentes eran menos proclives a irse de una escuela si en ella tenían tutores con experiencia previa, lo que, según su interpretación, podría señalar que una buena parte de la tutoría podía consistir en impartir conocimientos específicos de esa escuela. También descubrió que la cantidad de horas que el tutor dedicaba al docente tenía un impacto positivo sobre el rendimiento estudiantil: por cada 10 horas de tutoría, los alumnos de los docentes con tutores mejoraron su desempeño 0,05 desviaciones estándar en matemática y 0,04 desviaciones estándar en comprensión lectora.

La importancia de la calidad y dosificación de los programas de tutoría llevaron a Glazerman *et al.* (2010) a evaluar el impacto de dichos programas en 17 distritos urbanos de Estados Unidos. El equipo de investigadores asignó aleatoriamente 418 escuelas primarias a una de las siguientes dos categorías: 1) escuelas con un programa de inducción “integral”, en las que los docentes principiantes recibían capacitación profesional mensual, tenían oportunidades de observar a docentes experimentados y tenían asignado un tutor de tiempo completo con acceso a capacitación y material

didáctico actualizados, o 2) escuelas cuyos maestros principiantes participaban en el programa de inducción habitual y menos intensivo proporcionado por el distrito. En 10 de los 17 distritos, se brindaron estos servicios en las escuelas que participaban en el tratamiento durante un año y en los siete distritos restantes, se brindaron por dos años. Sin embargo, los distritos no fueron asignados aleatoriamente a la duración de los servicios, de modo que los hallazgos de ambos conjuntos de distritos evaluados deben interpretarse por separado.

El 90% de los docentes de los distritos asignados a la inducción integral tuvo un tutor durante su primer año de ejercicio, a diferencia del 72% de los docentes de los distritos que participaron en el programa de inducción habitual. Sin embargo, es interesante el hecho de que los docentes que recibieron la inducción integral no resultaron más efectivos que sus pares de control, tal como se midió a través de observaciones de clases y el rendimiento estudiantil. Más aún, los alumnos cuyos docentes recibieron dos años de inducción integral no mostraron mejor desempeño durante los dos primeros años de trabajo de los docentes, pero hacia el tercer año los alumnos lograron un desempeño 0,11 desviaciones estándar mejor en comprensión lectora y 0,20 en matemática comparados con los alumnos de los docentes del grupo de control. Ninguno de los programas de inducción tuvo un efecto positivo en la retención o satisfacción docente. Este trabajo complementa el estudio de Rockoff porque proporciona una visión más realista de las tutorías, según la cual mejorar su calidad y dosificación no implica solamente aumentar la intensidad del tratamiento.

Incluir un componente de práctica en la formación docente

Las investigaciones señalan que el desempeño docente mejora considerablemente durante los primeros años de ejercicio laboral. Algunos lo han atribuido a la importancia de la experiencia directa de dominar la práctica en el aula, lo que lleva a algunos sistemas educativos a incorporar componentes “clínicos” (basados en la práctica) en sus programas de formación docente.

Una iniciativa en este sentido es el programa Boston Teacher Residency (BTR), que selecciona a graduados talentosos, personas que desean cambiar de profesión y miembros de la comunidad de todas las edades para participar en un programa de capacitación intensiva de verano y una “residencia” de un año de duración en una clase de alguna escuela pública de Boston. Durante ese año, los participantes trabajan con un tutor cuatro días por semana, y una vez por semana asisten a seminarios. Una vez que completan el programa, reciben un título de maestría y una licencia “doble”, que los habilita para desempeñarse en el área de la educación especial. Estos graduados son asignados a escuelas en las que ya hay graduados trabajando y allí reciben un programa de desarrollo

profesional e inducción personalizada. Chicago y Denver adoptaron programas similares en 2002 y 2004, respectivamente.

Papay *et al.* (2011) usaron una combinación de efectos fijos para conducir la primera evaluación independiente de BTR aplicado durante los años lectivos 2001–2010. Los graduados de BTR se caracterizaron por presentar una diversidad racial mayor que otros docentes principiantes de escuelas públicas de Boston, por ser más propensos a enseñar matemática y ciencias y por mostrar más predisposición a permanecer en su profesión después de los primeros cinco años. Los graduados de BTR no resultaron más eficaces en el aula, según las mediciones de su valor agregado en los resultados de las pruebas de rendimiento de sus alumnos, pero su desempeño mejoró rápidamente después de su segundo año de ejercicio y pasado el quinto año, superó incluso al de los docentes experimentados. Estos hallazgos indican que la experiencia clínica puede ofrecer al docente herramientas para mejorar su desempeño más rápidamente mientras ejerce la profesión.

Asignar a los docentes donde más se los necesita

Desde el punto de vista de la economía, la coincidencia con el puesto de trabajo (*job matching*) reduce las ganancias, algo costoso tanto para los empleadores como para los empleados, y mejora lo que se conoce como “eficiencia asignativa” (Hicks, 1939; Kaldor, 1939). La teoría económica ofrece un marco para examinar por qué una escuela que se adecua bien a un docente puede hacer que ese docente sea más productivo. En la “teoría de la búsqueda” o la “teoría del emparejamiento” (Pissarides, 1979), los economistas plantean que es útil pensar en los puestos de trabajo como algo que no difiere de otros bienes que se consumen: al buscar un empleo, los individuos consideran las opciones que tienen y eligen un puesto de trabajo en función de sus preferencias y sus limitaciones.

Sin embargo, hay una diferencia importante entre el consumo de, por ejemplo, los típicos artículos para el hogar y la elección de un empleo (Nelson, 1970). Los artículos para el hogar típicos son “bienes de búsqueda”, que son fáciles de evaluar antes de adquirirlos (Lucas y Prescott, 1974). En el caso de bienes de búsqueda, la falta de correspondencia entre las preferencias y limitaciones de las personas y el bien que más las satisface se debe a problemas de información imperfecta que pueden resolverse mejorando la información. En cambio, los puestos de trabajo son “bienes de experiencia”: no es fácil para las personas evaluar si un empleo determinado es el que mejor se ajusta a ellas en términos productivos a menos que lo acepten y permanezcan en ese empleo durante un tiempo (Johnson, 1978). Esta es la razón por la cual los economistas observan que las personas ganan más cuando cambian de empleo (porque normalmente cambian por uno que las hace más productivas) y cambian menos de empleo cuando se las efectiviza o

ganan experiencia (porque con el tiempo las personas encuentran el empleo “apropiado”, el que mejor se adecua a ellas) (Altonji y Shakotko, 1987; Bartel y Borjas, 1981; Jovanovic, 1979a y 1979b; Topel y Ward, 1992).

Los economistas también justifican la coincidencia con el puesto de trabajo con la “segmentación del mercado de trabajo”. Tal como sostuvieron Reich, Gordon y Edwards (1973), hay numerosos estudios empíricos que indican que las competencias de los trabajadores no son el único factor determinante del salario y que los trabajadores con niveles similares de calificación reciben salarios diferentes en diferentes mercados laborales. Los partidarios de los mercados de trabajo segmentados sostienen que “la ley de un solo precio”—que establece que un mismo producto no puede venderse en dos mercados a precios diferentes porque crearía oportunidades de arbitraje que no tardarían en nivelar los precios de ambos mercados— no se aplica al mercados laboral. Si están en lo cierto, entonces la coincidencia con el puesto de trabajo podría ser funcional para corregir esas fallas de mercado. No obstante, cuáles son los mecanismos para saber cómo debería darse esa coincidencia han sido objeto de mucho debate (Lang y Dickens, 1987).

En el Cuadro 4 se resumen tres tipos de intervenciones que han sido evaluadas rigurosamente: 1) ofrecer una retribución extra a los docentes para atraerlos a escuelas con necesidades; 2) ofrecer una retribución extra a los docentes para que enseñen en áreas temáticas en las que hay un déficit crítico de docentes y 3) mejorar las condiciones laborales.

Ofrecer una retribución extra para enseñar en escuelas con déficits docentes

Steele, Murnane y Willett (2010) evaluaron una experiencia realizada en 2000–2002 en California en la que se ofrecía una retribución extra o “bono” de US\$ 20.000 al momento de la contratación para atraer a docentes principiantes y talentosos a escuelas de bajo desempeño y retenerlos durante al menos cuatro años. Los investigadores aprovecharon el hecho de que esta medida fue sancionada sólo por dos años para llevar adelante una estimación con variables instrumentales. Los beneficiarios de este monto extra habrían estado menos dispuestos a enseñar en escuelas de bajo desempeño que sus pares comparables observables si no hubiera sido por dicha retribución; sin embargo, debido a la retribución extra, la probabilidad de que los beneficiarios enseñaran en escuelas de bajo desempeño aumentó en 28 puntos porcentuales, y el 75% de los docentes que comenzaron a trabajar en escuelas de bajo desempeño, recibieran o no la retribución extra, permanecieron en tales escuelas durante al menos cuatro años.

Ofrecer una retribución extra para enseñar materias con déficits docentes

Clotfelter *et al.* (2008) evaluaron un programa en Carolina del Norte en el 2001 que ofreció una retribución extra anual de US\$ 1.800 para retener a docentes certificados en matemática, ciencias y educación especial que trabajaban en escuelas muy pobres o de muy bajo desempeño. Compararon la rotación de docentes que cumplían y docentes que no cumplían con esas condiciones antes y después del programa para realizar un análisis de diferencias en diferencias. La retribución adicional aumentó la retención interanual de docentes entre el 10% y el 13%. Sin embargo, este efecto se produjo principalmente por los docentes de matemática: aquellos que recibieron la retribución extra se mostraron 18% menos proclives a irse de sus escuelas que quienes no la recibieron. Los docentes de ciencias y de educación especial que recibieron la retribución extra no se mostraron menos proclives a irse de sus escuelas que quienes no la recibieron. Los efectos registrados se concentraron en las escuelas de nivel medio: los docentes de este nivel resultaron ser 27% menos proclives a dejar la escuela si habían recibido la suma extra, pero no se observaron diferencias en términos de la distribución de los beneficiarios en las escuelas secundarias. Considerando que en el primer año de la implementación de esta estrategia la retribución extra representaba sólo entre el 4% y el 5% del salario promedio de un docente, cabe suponer que los incentivos financieros, aunque sean modestos, parecen influir en la decisión de los docentes de permanecer en escuelas difíciles en términos de dotarlas de personal.

Mejorar las condiciones laborales

Una serie de trabajos de investigación recientes arrojaron luz sobre el efecto de las mejoras de las condiciones laborales en la productividad de los docentes.

Jackson (2013) utilizó datos de los alumnos de 3° a 5° grado de Carolina del Norte en 1995–2006 e incluyó efectos fijos de años, escuelas y docentes para calcular los cambios en la productividad de los docentes (medida por su puntaje de valor agregado) como resultado de transferirse de una escuela a otra. Observó que los docentes que cambiaron de escuela se volvieron más eficaces —0,09 desviaciones estándar en matemática y 0,07 en comprensión lectora— que antes del cambio, lo que interpretó como un indicador de “efectos de compatibilidad” (*matching*) entre docente y escuela. Estos efectos dieron cuenta del 10% al 40% de lo que normalmente se considera calidad docente, lo que sugiere que una parte considerable de esa calidad no es trasladable. De hecho, Jackson también constató que ciertos tipos de docentes se desempeñaban mejor en ciertos tipos de escuelas, de manera tal que una coincidencia óptima entre los docentes y las escuelas podría mejorar los resultados de todos los alumnos.

Aunque este estudio apunta a que la compatibilidad entre docentes y escuelas es importante, no indica qué factores elevarían la productividad docente. Sin embargo, Jackson y Bruegman (2009) realizaron un estudio que indica que el aprendizaje inter pares (docentes que aprenden de otros docentes) es un mecanismo importante a través del cual la compatibilidad mejora la productividad docente. Los autores usaron la misma base de datos que Jackson (2013) y emplearon efectos fijos de alumnos, docentes, escuelas y años. Los docentes que experimentaron una mejora de una desviación estándar en características observables de sus pares (por ejemplo, experiencia y autorizaciones para ejercer) fueron asociados a un aumento de 0,10 desviaciones estándar en el puntaje obtenido por sus alumnos en matemática y comprensión lectora. Los docentes que experimentaron un aumento de una desviación estándar en los puntajes de valor agregado de sus pares tuvieron alumnos cuyo desempeño fue 0,05 desviaciones estándar mejor en matemática y comprensión lectora.

Si bien hay muy pocos estudios sobre la compatibilidad entre docentes y escuelas, el mensaje de estos estudios es claro: es un factor importante en la retención y efectividad del docente y los incentivos monetarios han demostrado tener resultados prometedores para asignar a los docentes adonde más se los necesita. Un déficit importante que presenta la literatura es la falta de estudios rigurosos sobre iniciativas de esta naturaleza en países en desarrollo, precisamente donde más se necesitan.

Liderar a los docentes con buenos directores

Si bien los economistas no tienen mucho para decir acerca de la importancia de los administradores, el típico argumento al que suelen recurrir a la hora de justificar la asignación de trabajadores a distintos puestos dentro de una empresa es que cada trabajador posee un “ventaja comparativa”, es decir, tiene mayor aptitud para realizar una determinada tarea y por lo tanto conviene que se especialice en dicha tarea (Sattinger, 1975; Rosen, 1978). Así, los economistas sostienen que los docentes que poseen habilidades excepcionales para la gestión son más aptos para ocupar el puesto de directores, mientras que aquellos que resultan especialmente eficaces en el aula deben ocupar el puesto de maestros de grado o profesores particulares. Algunas teorías basadas en este enfoque también señalan que cuando es necesario asignar un único trabajador a un puesto, se debe seleccionar a trabajadores altamente aptos para ocupar aquellos puestos para los cuales se valora altamente la aptitud (Rosen, 1978).

Las investigaciones acerca de las políticas que repercuten en el trabajo de los directores son muy recientes y sólo se cuenta con unos pocos estudios realizados con métodos rigurosos. No obstante, los resultados de dichos estudios ofrecen indicios del tipo de políticas que podrían contribuir mejor a la eficacia de los directores.

En el Cuadro 5 se resumen tres intervenciones que han sido evaluadas rigurosamente: 1) contratar a directores más eficaces, 2) establecer requisitos para ocupar el puesto de director y 3) darles a los directores mayor autoridad en la contratación de personal.

Contratar a directores más eficaces

Las investigaciones acerca de la eficacia de los directores son muy recientes y aún se encuentran en una etapa de evolución constante. A pesar de ello, hay algunos estudios acerca de si se puede calcular el valor agregado de un director sobre la base del desempeño académico de los alumnos de su escuela (y de ser así, sobre cómo hacerlo).

En la provincia de la Columbia Británica, Canadá, se rota a los directores entre las escuelas de un mismo distrito. Coelli y Green (2012) aprovecharon esto para analizar el desempeño de distintos directores en una misma escuela. Utilizando efectos fijos de escuela, barrio y grupo de pares, detectaron una considerable heterogeneidad en el efecto de la calidad de los directores sobre los puntajes obtenidos en las pruebas de inglés por los alumnos de 12° grado. Según sus estimaciones, contar con un director cuya eficacia es una desviación estándar más alta en la distribución de los efectos de los directores implica una mejora de 2,6 puntos porcentuales en las tasas de graduación (o, aproximadamente, 0,33 desviaciones estándar de la distribución de la tasa de graduación interescolar) y de 2,5 puntos porcentuales en los puntajes de las pruebas de lengua inglesa (lo que equivale aproximadamente a una desviación estándar).

Branch, Hanushek y Rivkin (2012) incorporaron datos longitudinales al cálculo de los efectos generados por los directores. Para determinar su productividad, se basaron en datos recogidos en el estado de Texas de 1995–2001 y utilizaron efectos fijos de director, la escuela y el año. Constataron que, según el método empleado, el impacto anual de un director eficaz era de entre 0,05 y 0,21 desviaciones estándar. Analizaron también un mecanismo potencial a través del cual los directores eficaces pueden mejorar el rendimiento estudiantil: despedir a los docentes ineficaces. En coincidencia con esa hipótesis, detectaron que existía una mayor probabilidad de que los docentes que abandonaban las escuelas a cargo de los directores más eficaces se encontraran entre los menos eficaces de dichas escuelas, a diferencia de lo que ocurría con los maestros que abandonaban las escuelas a cargo de directores menos eficaces. Los autores también analizaron si los directores que cambiaban de escuela eran más eficaces que aquellos que permanecían en una misma escuela, pero no pudieron establecer una relación entre la calidad de los directores y la probabilidad de que se trasladaran. Los directores en el cuartil de desempeño más bajo resultaron tener menos probabilidades de permanecer en su cargo y más probabilidades de abandonar el sistema de educación pública. En cambio,

los directores en el cuartil de desempeño más alto demostraron mayor probabilidad de permanecer en su cargo y menor probabilidad de abandonar el sistema de educación pública que sus colegas ubicados en el cuartil más bajo, aunque tienen mayor probabilidad de abandonar su cargo y menor probabilidad de abandonar el sistema de educación pública que aquellos comprendidos en los dos cuartiles intermedios.

Sin embargo, tal como señalaron Grissom, Kalogrides y Loeb (2011), los resultados dependen del modelo utilizado y pueden variar considerablemente. Los autores usaron datos del condado Miami-Dade, Florida, correspondientes al período 2003–2010 y a los grados 3° a 10° con el objeto de demostrar que, mientras que algunos modelos asignaban a los directores efectos del orden de 0,15 desviaciones estándar en matemática y 0,11 en comprensión lectora, otros identificaban efectos mucho más bajos, de hasta 0,02 desviaciones estándar en ambas áreas temáticas para los mismos directores. Lo interesante es que los modelos menos atractivos desde el punto de vista conceptual, que asignaban a los directores un porcentaje demasiado alto de efectos sobre el desempeño de la escuela, demostraron mayor coincidencia con las mediciones de desempeño no derivadas de los resultados de las pruebas que los criterios que establecían una separación más convincente entre los efectos producidos por los directores y los generados por otros datos escolares. La principal conclusión de este estudio es que, si bien existen indicios de que la calidad de los directores es un factor importante, medir con precisión su eficacia va a requerir más investigación.

Establecer requisitos para ocupar el puesto de director

A pesar de que se ha llegado a un consenso respecto de la importancia de contar con directores eficaces, no queda del todo claro cómo identificarlos. Existen dos estudios sobre la relación entre las características del director y su eficacia, ambos realizados en Estados Unidos, lo que plantea interrogantes acerca de su validez externa.

Clark, Martorell y Rockoff (2009) utilizaron datos de empleo y rendimiento estudiantil de la ciudad de Nueva York del período 1998–2006 y usaron efectos fijos de escuelas. Hallaron escasos indicios de una relación entre el desempeño de una escuela y la selectividad de la institución donde un director cursó sus estudios de grado o posgrado. También fue escasa la evidencia sobre la relación entre el desempeño del director y su experiencia laboral previa. Sin embargo, el desempeño de las escuelas a cargo de directores con muy poca experiencia resultó ser superior cuando éstos habían ocupado el cargo de vicedirectores en la misma escuela. Además, se estableció una relación positiva entre la experiencia del director y el desempeño de la escuela: los alumnos de los directores con tres años de experiencia obtuvieron puntajes 0,04 desviaciones estándar más altos en matemática que los estudiantes a cargo de directores en su primer año en el puesto, mientras que en el caso de los alumnos de los directores con cinco años de

experiencia o más, los puntajes de los alumnos resultaron 0,06 desviaciones estándar más altos. Por último, se obtuvieron resultados dispares respecto de la relación entre los programas de desarrollo profesional y capacitación del director y el desempeño de la escuela.

Grissom y Loeb (2011) utilizaron información del 2008 del condado Miami-Dade, Florida, que incluía los resultados de encuestas a directores y vicedirectores, las calificaciones asignadas por los padres a las escuelas de sus hijos y datos administrativos (por ejemplo, índices de desempeño de las escuelas y pruebas de los estados). Las escuelas cuyos directores poseían mayores habilidades de gestión fueron las que obtuvieron mejor desempeño de sus alumnos: en el caso de todas las escuelas, el aumento de una desviación estándar en la autoevaluación de los directores respecto de su “gestión organizacional” estuvo asociado a un aumento de 0,12 puntos de la calificación obtenida por la escuela en la escala original, o a 0,10 desviaciones estándar. Las autoevaluaciones de los directores respecto de sus habilidades de gestión también estuvieron relacionadas con la calificación asignada por los padres de los alumnos a las escuelas, aunque no coincidieron con el grado de satisfacción de los docentes. Por último, también se constató una relación entre la calificación asignada por los vicedirectores a las capacidades de gestión de los directores y los avances en matemática y comprensión lectora.

Darles a los directores mayor autoridad en la contratación de personal

El estudio más riguroso sobre los efectos de la autoridad de los directores en la contratación de personal se realizó en la ciudad de Nueva York por Rockoff *et al.* (2011). Se evaluó el impacto de un programa piloto durante el año lectivo 2007–2008 en el cual se seleccionaron directores entre un grupo de voluntarios para proporcionarles mediciones del valor agregado de los docentes de sus escuelas y capacitación acerca de los métodos utilizados para elaborar dichas mediciones. El hecho de que la selección de los directores se realizó en forma aleatoria permitió a los autores implementar un estudio de asignación aleatoria. Las opiniones previas de los directores acerca de la efectividad de los docentes coincidían con las mediciones de su valor agregado, y los directores actualizaron su opinión a partir de la recepción de nuevos datos sobre el desempeño docente. De hecho, una vez divulgada esta información, los directores se mostraron más proclives a retener a los docentes con puntajes más altos de valor agregado y menos proclives a retener a aquellos con puntajes bajos, lo cual llevó una leve mejora en la productividad de sus escuelas. Sin embargo, es importante mencionar que la provisión de mediciones del valor agregado no hizo que se desestimara la información sobre la efectividad de los docentes recogida por los directores a través de sus observaciones de clases. Estos resultados sugieren que los directores poseen opiniones acertadas sobre la eficacia de los docentes a su cargo y que pueden utilizar bien la información sobre el desempeño docente para tomar decisiones referidas a la contratación de personal.

Un programa en Chicago en 2004 confirió a los directores la facultad de utilizar un procedimiento acelerado para despedir a docentes que se encontraban en período de prueba (es decir, docentes con menos de cinco años de experiencia) por cualquier motivo. Jacob (2010a, 2010b) recurrió a un análisis de diferencias en diferencias para evaluar esta reforma. Para ello, comparó los cambios en los niveles de ausentismo de los docentes a prueba y los titulares previos y posteriores a la aplicación de esta política. Jacob determinó que la política redujo el ausentismo entre los docentes a prueba en un 10% y entre los maestros con 15 faltas o más en un 20%. También comprobó que la política mejoró el desempeño de los alumnos del nivel primario. Lo interesante es que la reducción del ausentismo de los docentes sólo da cuenta de una parte de esta mejora, lo que implica que la política también repercutió en la conducta de los docentes en clase. Por último, los directores tuvieron en cuenta las ausencias y las mediciones de valor agregado a la hora de decidir a qué docentes despedir. Esto indica que, ante una mayor flexibilidad para despedir a los maestros, los directores pueden tomar decisiones sensatas.

Estos estudios, a pesar de su carácter experimental, generan conclusiones claras. Los directores con experiencia en gestión son los más efectivos. De hecho, los directores (al menos, en los países desarrollados) parecen estar preparados para asumir mayor responsabilidad y autoridad para mejorar el desempeño de la escuela.

Evaluar el aprendizaje y la enseñanza

Los economistas consideran que la evaluación cumple dos objetivos: mantener a los empleados motivados e identificar a empleados con aptitudes para ser ascendidos. En economía, los debates en torno a la evaluación del desempeño han solido centrarse en el problema del agente y el principal. En un principio, la cuestión era inducir al trabajador a hacer su trabajo y el empleador sabía qué era lo que tenía que pedir a los trabajadores que hicieran en un contexto de información asimétrica (Ross, 1973). Más tarde, los economistas reconocieron que los empleadores no siempre saben de qué manera sus empleados pueden alcanzar el máximo nivel de productividad; entonces, se convirtió en una cuestión de, en primer lugar, determinar cuál sería la conducta adecuada del empleado para luego inducirlo (Baker, 1992). Más recientemente, se llegó a la conclusión de que, para algunos puestos de trabajo, es mejor evaluar el desempeño a través de parámetros basados en los resultados antes que en la utilización de insumos (Prendergast, 2002). Esto podría aplicarse a la enseñanza, donde existe una gran controversia acerca de las acciones de los docentes que contribuyen a mejorar el nivel de aprendizaje estudiantil.

Además, los economistas consideran que la evaluación del desempeño resulta útil para la contratación. Cuando las acreditaciones o los mecanismos de selección no alcanzan para predecir la productividad, las empresas pueden minimizar los costos de contratación, al

mismo tiempo que obtienen el máximo nivel de productividad del trabajador, a través de la creación de “mercados laborales internos”: contratar empleados para puestos de menor responsabilidad, evaluar su desempeño y decidir cuáles de ellos deben permanecer en esos puestos y cuáles merecen un ascenso (Williamson, 1975; Osterman, 1984). La manera en la cual muchos sistemas educativos determinan los ascensos de los docentes está arraigada en este enfoque, lo cual sugiere que existe una fuerte convicción de que la experiencia es la clave por la cual los docentes pueden convertirse en directores y los directores pueden obtener conocimientos específicos acerca de su escuela.

En el Cuadro 6 se resumen cuatro tipos de intervenciones que han sido evaluadas rigurosamente: 1) aumentar la participación de los padres y de la comunidad en los asuntos de la escuela, 2) calificar y/o clasificar a las escuelas en base al rendimiento estudiantil, 3) monitorear el esfuerzo de los docentes y 4) monitorear el desempeño de los docentes.

Aumentar la participación de los padres y de la comunidad en los asuntos de la escuela

Las iniciativas orientadas a aumentar la participación de los padres y de la comunidad en la educación han sido objeto de evaluación rigurosa en diversos contextos. En Uttar Pradesh, India, se asignaron aldeas en forma aleatoria a un grupo de control o a una de las siguientes intervenciones: 1) una iniciativa que explicaba el papel de los comités comunales a sus miembros; 2) un programa que capacitaba a voluntarios para tomar evaluaciones de diagnóstico de comprensión lectora, redacción y matemática y comunicar los resultados, o 3) un proyecto que capacitaba a voluntarios locales para tutoría de comprensión lectora fuera del horario escolar. Ninguna de las primeras dos intervenciones repercutió en el nivel de participación de la comunidad ni en el desempeño académico de los alumnos. Sin embargo, la tercera intervención tuvo un efecto positivo en cuanto a la participación de los jóvenes, pues los motivó a postularse como voluntarios para enseñar, lo que a su vez mejoró el puntaje obtenido por los alumnos en las pruebas de comprensión lectora en 0,02 desviaciones estándar aproximadamente. Los autores interpretaron estos resultados como una muestra de que las personas deben enfrentar grandes limitaciones para mejorar la calidad de sus escuelas públicas, incluso cuando les interesa y desean participar.

No obstante, la evidencia surgida de un experimento similar llevado a cabo en la India indica que el formato y el contexto de estas iniciativas también son importantes. Pandey, Goyal y Sundararaman (2009) analizaron una iniciativa implementada en Karnataka, Madhya Pradesh y Uttar Pradesh en 2006, en la que se asignaron aldeas en forma aleatoria a una campaña que difundía información en las comunidades acerca de los roles y responsabilidades en la gestión de las escuelas establecidos por el estado, por un lado, y a un grupo de control, por el otro. La campaña de información aumentó la cantidad de

reuniones de los comités de educación y la participación de los miembros en un 25% en Uttar Pradesh e incrementó la cantidad de padres que hablaban con los docentes acerca de la calidad de la educación en Madhya Pradesh. Además, elevó el índice de presentismo docente en un 11% en Uttar Pradesh y las actividades en las aulas en un 30% en Madhya Pradesh. Sin embargo, la campaña sólo generó un leve aumento del desempeño en comprensión lectora. Estos resultados sugieren que si bien las campañas de información pueden fomentar la participación de los docentes y la comunidad, las perspectivas de que mejoren el nivel de aprendizaje son limitadas.

Estos resultados parecen imitar los de un estudio similar llevado a cabo en un país desarrollado. Avvisati *et al.* (2008) implementaron un estudio de asignación aleatoria con el fin de analizar una iniciativa que convocaba a reuniones con los padres de los alumnos para tratar cuestiones tales como ayudar a sus hijos con la tarea, los boletines de calificaciones y otros temas relacionados con la escuela. El objetivo era aumentar la participación de los padres tanto en la escuela como en el hogar. Los niños de las familias en el grupo de tratamiento desarrollaron conductas y actitudes más positivas en la escuela y presentaron menor cantidad de problemas de lectoescritura. De hecho, se generaron efectos de derrame considerables en los compañeros de clase de los alumnos cuyas familias pertenecían al grupo de tratamiento. Estas conclusiones refuerzan la idea del potencial que encierran las intervenciones destinadas a fomentar la participación de la comunidad con el objeto de influir en la conducta de las familias, aunque plantean el interrogante de hasta qué punto estos cambios mejoran el aprendizaje.

Calificar o clasificar a las escuelas en base al rendimiento estudiantil

Otra iniciativa sometida a evaluación rigurosa es la calificación o clasificación de las escuelas en base al desempeño de los alumnos. El único estudio de asignación aleatoria de esta política, implementado en Punjab, Pakistán, en 2004, evaluó la entrega de informes de desempeño de los alumnos y las escuelas. Andrabi, Das y Khwaja (2009) determinaron que las calificaciones fueron 0,10 desviaciones estándar más altas en las comunidades en las que se entregaron los informes que en las que no, aunque las mejoras en el nivel de aprendizaje se limitaron a las escuelas de bajo rendimiento. De hecho, la entrega de los informes generó un aumento en las calificaciones de 0,34 desviaciones estándar en el puntaje obtenido por las escuelas que inicialmente habían obtenido calificaciones por debajo de la mediana, pero no produjo efectos en aquellas cuyas calificaciones iniciales estaban por encima de la mediana. Por último, la entrega de los informes trajo aparejada una reducción del 18% en los aranceles de las escuelas privadas. Dado que pocos alumnos cambiaban de escuela, la entrega de los informes de desempeño generó competencia entre todas las escuelas para mejorar la calidad de la educación de modo tal que los aranceles resultasen proporcionales.

Un estudio reciente pone de manifiesto otras consecuencias de los intentos de calificar o clasificar a las escuelas que suelen pasar inadvertidas. Feng, Figlio y Sass (2010) aprovecharon la oportunidad que se presentó cuando el estado de Florida modificó repentinamente la forma en la cual calificaba a sus escuelas en 2002 para llevar a cabo un análisis de diferencias en diferencias. Determinaron que los docentes de las escuelas que obtuvieron calificaciones por debajo de lo esperado tenían un 11% más de probabilidades de abandonar su escuela, mientras que aquellos cuyas escuelas obtuvieron calificaciones por encima de lo esperado tenían un 2,3% menos de probabilidades de irse. Estos efectos resultaron aún más evidentes en las escuelas que no sólo obtuvieron calificaciones más bajas que lo esperado sino que además descendieron a nivel de reprobado. Los docentes de estas escuelas presentaron mayores probabilidades de abandonarlas (42%) y de pasar a trabajar en una escuela del mismo distrito (67%). Sin embargo, es importante destacar que los efectos positivos y negativos de esta política sobre la movilidad de los maestros no alteró la composición del plantel docente de las escuelas que obtuvieron una calificación reprobatoria: los docentes que permanecieron en estas escuelas aumentaron su nivel de dedicación, lo cual generó un aumento de 0,05 desviaciones estándar en el desempeño de los alumnos, y aquellos que las abandonaron resultaron ser también, en promedio, más eficaces. Las conclusiones sugieren que el índice de confiabilidad que representan las calificaciones de la escuela puede, en ocasiones, repercutir en la movilidad de los docentes sin por ello mejorar el nivel de aprendizaje de los alumnos.

En 2007, se implementó una política similar en la ciudad de Nueva York con el fin de calificar a las escuelas según una escala que iba de la letra A a la F así como de asociar a la calificación obtenida las consecuencias positivas y negativas, que llegaban hasta el posible cierre de la institución. Una evaluación indicó que muchos de los hallazgos acerca de este tipo de políticas dependen de las particularidades de cada una. Rockoff y Turner (2010) utilizaron un diseño de regresión discontinua para evaluar los efectos de las calificaciones resultantes de esta iniciativa, que fueron divulgadas cuando el año lectivo se encontraba demasiado avanzado para generar cambios significativos en las actividades o el personal de las instituciones. No obstante, los autores verificaron que en las escuelas que obtuvieron una D o una F mejoró el desempeño en matemática y que en aquellas que recibieron una F también mejoraron los puntajes correspondientes a comprensión lectora. Además, los autores comprobaron que las evaluaciones de los padres sobre la calidad de las escuelas mejoraron en el caso de aquellas que recibieron una D o una F. Estos resultados sugieren que las políticas destinadas a evaluar y calificar a las escuelas sin limitarse a hacerlo con la información suministrada por ellas pueden generar mejoras en la calidad de las instituciones educativas.

Por otra parte, como en el caso anterior, existe evidencia de que los detalles específicos de estos tipos de intervenciones son importantes. Mizala y Urquiola (2007) evaluaron una iniciativa vigente en Chile desde 1996 que clasifica a las escuelas según su desempeño,

ajustado según el nivel socioeconómico de los alumnos. El programa ofrece un incentivo económico a las escuelas que alcanzan un nivel de desempeño por encima de un umbral determinado. Los autores aprovecharon este umbral para evaluar los efectos del programa a través de un diseño de regresión discontinua. La política no generó efecto alguno en los niveles de aprendizaje en las escuelas que recibieron el incentivo económico. Esto sugiere que la información pública acerca de la calidad de las escuelas, incluso cuando está asociada a consecuencias positivas y negativas, no siempre genera mejoras en el desempeño estudiantil.

Monitorear el esfuerzo de los docentes

Los datos sobre las iniciativas concebidas específicamente para evaluar el desempeño de los docentes son mucho más claros que la evidencia referida a la clasificación de las escuelas. Duflo, Hanna y Ryan (2010) evaluaron una iniciativa implementada en el 2003 en Rajastán, India, con el fin de monitorear diariamente el ausentismo docente mediante el uso de cámaras y según la cual los salarios de los docentes variaban en función de la asistencia, para lo que asignaron a las escuelas en forma aleatoria al grupo de tratamiento o al grupo de control. La intervención redujo el ausentismo docente del 42% al 23%. Al cabo de un año, los puntajes de las pruebas en las escuelas del grupo de tratamiento resultaron 0,17 desviaciones estándar más altos que en las escuelas del grupo de control y se incrementó el porcentaje de alumnos que pasaron de grado. Estos resultados indican que las iniciativas destinadas a fomentar el nivel de dedicación de los docentes que se centran de manera específica en la conducta que se pretende modificar pueden arrojar resultados sorprendentes.

Monitorear el desempeño de los docentes

Un estudio reciente analizó hasta qué punto la eficacia de un docente en un aula (aula A) puede utilizarse para predecir su valor agregado en otra aula (aula B). Para ello, la eficacia se calculó en función de un conjunto de indicadores, entre los que se encontraban maestrías, experiencia, conocimientos pedagógicos específicos de un área temática, certificación de la NBPTS, calificaciones de los directores, resultados de encuestas realizadas a los alumnos, observaciones de clases y valor agregado. El estudio constató que los docentes identificados como eficaces en el aula A mediante una combinación de tres parámetros (encuestas a los alumnos, observaciones de clases y valor agregado) también tendían a obtener altos puntajes de valor agregado en el aula B (Kane y Staiger, 2010, 2012). De hecho, Kane *et al.* (2013) determinaron —al comparar a docentes de la misma materia, el mismo grado y la misma escuela— que una desviación estándar a nivel de alumno en el parámetro compuesto de eficacia del docente registrada en un año se correspondía aproximadamente con una desviación estándar a nivel de alumno en el valor agregado el año siguiente. Los estudios utilizaron datos correspondientes a los grados 4°

a 8° del período 2009–2011 en Charlotte-Mecklenburg (Carolina del Norte), Dallas (Texas), Denver (Colorado), condado de Hillsborough (Florida), ciudad de Nueva York y Memphis (Tennessee).

Mihaly *et al.* (2013) compararon cuatro maneras de ponderar los tres parámetros que miden la enseñanza eficaz: 1) ponderarlos para obtener la mayor precisión posible en la predicción de las mejoras en las pruebas de los estados del año siguiente (lo que dio como resultado una ponderación del 81% para el valor agregado, una ponderación del 2% para las observaciones de clases y una ponderación del 17% para las encuestas a los alumnos); 2) aplicar una ponderación del 50% al valor agregado y del 25% a las observaciones de clases y a las encuestas a los alumnos; 3) asignar igual ponderación (33%) a cada uno de los tres componentes, y 4) asignar una ponderación del 50% a las observaciones de clases. El primer modelo aprovechó al máximo la capacidad predictiva del índice compuesto, medido en términos de su correlación con el valor agregado del próximo año, mientras que el tercer y el cuarto modelo potenciaron al máximo la confiabilidad (es decir, la proporción de variación de los puntajes que integran el índice compuesto, que reflejan diferencias sistemáticas en la práctica entre cada docente). Los autores no recomendaron un conjunto “ideal” de ponderaciones, pero sostuvieron que debe existir una correspondencia entre las ponderaciones de los distintos parámetros y el objetivo principal del sistema de evaluación docente y el sistema asociado de calificación de escuelas y docentes.

Apoyar a los docentes para que mejoren la enseñanza

Sobre la base del trabajo de Greenwald (1986) y Acemoglu y Pischke (1998), se puede afirmar que cuando las empresas invierten en capacitación en habilidades generales, se genera un “aprendizaje asimétrico”: aprenden más acerca del efecto de la capacitación sobre la productividad de los trabajadores de lo que lo aprenden sus competidores. Así, la decisión de la empresa de invertir en este tipo de capacitación depende por completo del efecto que produce la capacitación en su productividad. Estas consideraciones pueden ofrecer a los sistemas educativos un motivo para invertir en la capacitación de los docentes una vez que ingresan a la profesión. En el caso de la educación, el riesgo de rotación se ve mitigado por el hecho de que el sistema educativo en su conjunto es el que asume el costo de la capacitación y de que gran parte de la competencia para obtener maestros se da entre escuelas, en vez de entre la educación y otras áreas.

En el Cuadro 7 se resumen dos tipos de intervenciones que han sido evaluadas rigurosamente: 1) proporcionar o mejorar el material didáctico y 2) brindar formación continua.

Proporcionar o mejorar el material didáctico

Los esfuerzos realizados para ofrecer material didáctico adicional a los docentes no rindieron los frutos esperados. Muchas de las iniciativas en este sentido intentaron introducir computadoras en las aulas. Barrera-Osorio y Linden (2009) realizaron un estudio de asignación aleatoria con el fin de evaluar un programa implementado en Colombia en 2006 destinado a integrar computadoras donadas por el sector privado a la enseñanza de la lectura en las escuelas públicas. El programa casi no produjo efectos en los puntajes de las pruebas ni generó otros resultados, y no se detectaron diferencias en los resultados por grado, materia o género de los alumnos. Las encuestas realizadas a los docentes sugirieron que el principal motivo del efecto limitado del programa fue que pocos docentes incorporaron efectivamente las computadoras donadas a las actividades en clase. Estas conclusiones indican que es probable que las políticas basadas en los recursos tengan una repercusión limitada en el nivel de aprendizaje de los alumnos mientras no logren modificar los procedimientos pedagógicos habituales.

Sin embargo, no todas las intervenciones fracasaron. Banerjee *et al.* (2007) utilizaron un estudio de asignación aleatoria para evaluar un programa de aprendizaje asistido por computadora implementado en Vadodara, India, en el 2002. En el marco del programa, los alumnos de 4º grado compartían una computadora con un compañero durante dos horas por semana para resolver problemas de matemática mediante juegos con un grado de dificultad adecuado para su edad. El programa mejoró los puntajes de matemática 0,35 desviaciones estándar durante el primer año y 0,47 desviaciones estándar durante el segundo, y demostró el mismo grado de efectividad para todos los alumnos. No obstante, es curioso que el programa no haya generado efectos en los puntajes de comprensión lectora de los estudiantes, lo cual sugiere que los efectos de derrame fueron limitados. Además, los efectos parecieron desaparecer poco después de finalizado el programa. La diferencia entre este programa y el aplicado en Colombia parece residir en que este generó un cambio significativo en la pedagogía en aula.

La evidencia proveniente de países desarrollados parece avalar esta conclusión. Rouse y Krueger (2004) recurrieron en 2001 a un estudio de asignación aleatoria para evaluar el impacto de un programa informático para la enseñanza de la lectura sobre alumnos con dificultades en esta materia en un distrito no identificado de Estados Unidos. Si bien el programa mejoró algunos aspectos de las habilidades de aprendizaje de los alumnos, dichas mejoras no se tradujeron en un mayor índice de adquisición del lenguaje o en una mejora de las habilidades lectoras.

En conjunto, las tres evaluaciones analizadas indican que el impacto del aprendizaje asistido por computadora varía en forma considerable y que depende de los detalles específicos de cada iniciativa.

Otros tipos de políticas basadas en los recursos resultaron incluso menos efectivas para mejorar el nivel de aprendizaje de los alumnos. Borkum, He y Linden (2009) realizaron un estudio de asignación aleatoria en Bangalore, India, en el año 2007 para analizar el impacto de un programa implementado en escuelas primarias a cargo de bibliotecarios calificados. El programa utilizó un esquema de escuelas “proveedoras” y “receptoras”, en el cual las escuelas “proveedoras” eran las que contaban con bibliotecas físicas mientras que las escuelas “receptoras” eran aquellas a las que los bibliotecarios debían trasladarse. El programa no logró aumentar las competencias en lengua. De hecho, los cálculos fueron suficientemente precisos para desestimar efectos superiores a 0,11 y 0,13 desviaciones estándar sobre la base de intervalos de confianza (Recuadro 10). Estos hallazgos indican que es poco probable que mejorar las bibliotecas de las escuelas tenga un impacto positivo en el rendimiento estudiantil.

Recuadro 10. ¿Qué es un intervalo de confianza?

Un intervalo de confianza proporciona el espectro de los valores que podría adoptar una estimación si se realizara la misma evaluación en forma sucesiva. Por ejemplo, supongamos que un estudio determina que una intervención dada tiene un efecto de 0,20 desviaciones estándar. El intervalo de confianza podría indicar que si la evaluación se repitiera muchas veces, el efecto se reduciría a 0,10 desviaciones estándar o aumentaría a 0,30 desviaciones estándar. Si bien es posible que muchos lectores no hayan tenido que interpretar intervalos de confianza en relación con evaluaciones de impacto, probablemente hayan oído hablar de ellos en las encuestas de opinión. Cuando se divulgan los resultados de una encuesta, se suelen mencionar los “márgenes de error”. Estos márgenes no son otra cosa que la amplitud del intervalo de confianza a ambos lados del efecto calculado. Por ejemplo, en el caso mencionado, el margen de error sería de $\pm 0,10$ desviaciones estándar.

En realidad, al menos en los países en desarrollo, incorporar más libros parece producir un efecto leve. Glewwe, Kremer y Moulin (2009) utilizaron un estudio de asignación aleatoria para evaluar un programa en Busia y Teso, Kenia, que suministró libros de texto en forma gratuita a las escuelas. Esta intervención no generó efecto alguno en el promedio de los puntajes de las pruebas ni en el índice de asistencia de los alumnos. Curiosamente, los libros de texto aumentaron 0,06 desviaciones estándar las calificaciones de las evaluaciones de alumnos que ya eran sobresalientes. Pero pocos niños de los grados inferiores (15%–29% de los alumnos de la mediana) eran capaces de leer estos libros en inglés, su tercer idioma. Este estudio sirve para reforzar la idea de que las políticas basadas en los recursos tienden a producir un efecto limitado cuando existen razones para creer que esos recursos no van a utilizarse en forma adecuada.

Un estudio realizado por Glewwe *et al.* (2004) en los distritos de Busia y Teso, Kenia, constató que el impacto limitado de las políticas basadas en los recursos no sólo se refiere

a los alumnos, sino también a los docentes. Los autores utilizaron un estudio de asignación aleatoria para evaluar el impacto de la provisión gratuita de posters en el aprendizaje de los alumnos. Dichos posters incluían gráficos de ciencia, un manual para docentes de ciencias, gráficos sobre salud, una serie de gráficos para matemática y un mapa, en su mayoría adecuados para alumnos de 7° y 8° grado. No hay evidencias de que los posters hayan generado puntajes más altos en las pruebas. Esta conclusión resulta particularmente interesante pues otros estudios menos rigurosos hallaron un efecto en los puntajes de las evaluaciones de 0,20 desviaciones estándar. Esta evaluación ilustra la importancia de las investigaciones rigurosas para tomar decisiones críticas en materia de política.

Ofrecer formación continua

Un tipo de intervención destinada a apoyar la labor de los docentes que se sometió a evaluaciones rigurosas es la formación continua. Jacob y Lefgren (2004b) evaluaron una reforma implementada en Chicago en 1996 que puso en período de prueba académica a 71 de los 489 docentes de escuelas primarias del distrito y les proveyó financiamiento para capacitación. Los autores realizaron un diseño de regresión discontinua aprovechando el hecho de que la participación en el período de prueba estaba determinada por un punto de corte en los puntajes estandarizados de comprensión lectora de los alumnos. Los autores determinaron que los aumentos marginales de la formación continua no generaron efecto alguno en el rendimiento en comprensión lectora ni en matemática. Aunque es necesario llevar a cabo más investigaciones rigurosas sobre el impacto de la formación docente continua, estas conclusiones son un recordatorio de que, si bien los distritos escolares invierten grandes sumas en este tipo de capacitación, existe escasa evidencia rigurosa que avale su efectividad.

De manera similar, en el año 1995, Angrist y Lavy (2001) llevaron adelante un análisis de diferencias en diferencias en Jerusalén, Israel, de un puñado de escuelas públicas que recibieron una inyección de fondos destinados principalmente a formación continua. En el caso de las escuelas seculares (es decir, no religiosas), la formación continua mejoró los puntajes obtenidos por los alumnos en las pruebas 0,20–0,40 desviaciones estándar, mientras que los efectos en escuelas religiosas resultaron menos robustos, lo que significa que dependieron de la forma en la cual los autores modelaron la relación entre el tratamiento y las mejoras de los alumnos. Angrist y Lavy conjeturaron que el hecho de que la intervención comenzara más tarde y se aplicara en menor escala en las escuelas religiosas podría dar cuenta de esta diferencia. Los autores compararon esta intervención con estrategias alternativas para obtener mejoras en las escuelas y llegaron a la conclusión de que se trata de una forma relativamente económica de mejorar el rendimiento estudiantil.

Resulta difícil sacar conclusiones respecto de los estudios rigurosos sobre las distintas formas de brindar apoyo a los docentes porque la evidencia es dispar. Sin embargo, existen razones para creer que las políticas basadas en los recursos (sobre todo las de los países en desarrollo) tienen un impacto limitado sobre el desempeño de los alumnos, especialmente cuando no afectan la enseñanza en el aula, y que el impacto de la formación continua de los docentes parece depender al menos del contexto en el cual se la implementa y probablemente también del formato.

Motivar a los docentes en su desempeño

Normalmente, los economistas piensan en los incentivos en el contexto del problema del agente y el principal. Tal como señalan Mirrlees (1976), Holmstrom (1979) y Shavell (1979), la consideración clave en este marco es que a la hora de elegir la manera en la cual van a establecer la remuneración de su personal, los empleadores deben elegir entre ofrecerles un salario base que asegure un nivel mínimo de remuneración y supeditar el salario a su desempeño para inducirlos a trabajar mucho. Un contrato “eficiente” es aquel que logra un equilibrio entre estas dos metas de “seguro contra todo riesgo” e “incentivos de primera opción”.

Los economistas han señalado que los contratos de incentivos pueden tener varias consecuencias involuntarias. Un concepto clave propuesto por Holmstrom y Milgrom (1991) y Baker (1992) es que el desempeño ponderado de un empleado puede resultar muy distinto de la contribución total del empleado al valor de la empresa y que los esquemas de incentivos que recompensan al primero pueden desalentar su contribución al trabajo de sus pares o los efectos a largo plazo de sus acciones (Gibbons y Waldman, 1999). El otro inconveniente que presentan los contratos de incentivos es el equilibrio entre la motivación intrínseca y la extrínseca: los mecanismos que ofrecen recompensas extrínsecas a los empleados por acciones de las cuales extraen motivación intrínseca pueden llegar a atenuar a esta última (Bénabou y Tirole, 2003).

En el Cuadro 8 se resumen dos tipos de intervenciones que han sido evaluadas rigurosamente: 1) contratar docentes temporarios y 2) remunerar a los docentes por mejorar el desempeño de los alumnos.

Contratar docentes temporarios

En muchos sistemas educativos, los docentes son empleados públicos y, una vez que pasan a formar parte de la planta permanente, sólo se los puede despedir en casos extremos mediante un proceso que suele extenderse en el tiempo. Por lo tanto, en vez de tomar nuevos docentes con categoría de empleados públicos, algunos países

implementaron reformas que permiten incorporar docentes a través de contratos anuales renovables.

Muralidharan y Sundararaman (2010) son los responsables de la evaluación más reciente acerca de este tipo de iniciativa, que se realizó a partir del 2005 en Andhra Pradesh, India. El programa proveyó un docente temporario adicional a cien escuelas primarias estatales seleccionadas de manera aleatoria. Al cabo de dos años, el desempeño de los alumnos de las escuelas que habían incorporado al docente temporario adicional resultó 0,15 desviaciones estándar superior en matemática y 0,13 desviaciones estándar superior en comprensión lectora al de los alumnos de las escuelas utilizadas como grupo de control. Si bien todos los alumnos se vieron beneficiados con el programa, aquellos que cursaban su primer año en la escuela o asistían a escuelas alejadas fueron los que sacaron mayor provecho. Los docentes bajo contrato tuvieron menos inasistencias que los maestros de planta permanente (16% versus 27% del período) y resultó más probable encontrarlos enseñando (49% versus 43%). Por último, aunque los alumnos de los docentes por contrato y los de los docentes de planta permanente demostraron mejoras similares en cuanto a sus logros académicos, los docentes por contrato insumieron sólo una quinta parte del costo de los docentes de planta permanente.

En algunos contextos, los docentes por contrato demostraron ejercer una influencia positiva en el desempeño, incluso en aquellos casos en los que no contaban con capacitación formal. En 2001, Banerjee *et al.* (2007) realizaron un estudio de asignación aleatoria para evaluar una iniciativa en Mumbai y Vadodara, India, que dotaba a escuelas públicas de un docente adicional para ocuparse de los alumnos de 3° y 4° grado con problemas de desempeño académico. Estos docentes, por lo general mujeres jóvenes, provenían de las comunidades de la zona y apenas habían finalizado la escuela secundaria. No obstante, los puntajes de las pruebas de esos alumnos mejoraron 0,14 desviaciones estándar al cabo del primer año y 0,28 al cabo del segundo. Es más, las escuelas del grupo de tratamiento obtuvieron puntajes 0,10 desviaciones estándar superiores a los de las escuelas del grupo de control un año después del programa. Quizás el dato más significativo sea que los niños con las calificaciones más bajas en la distribución inicial de los puntajes y los que recibieron clases de apoyo fueron los que obtuvieron mayores beneficios.

Remunerar a los docentes por mejorar el desempeño de los alumnos

Una de las estrategias más comunes y evaluadas con mayor rigor que vienen utilizando los sistemas educativos para tratar de motivar a los docentes es supeditar parte de su salario al nivel de desempeño de sus alumnos.

En 2007, la ciudad de Nueva York implementó un programa de incentivos docentes en más de 200 escuelas de bajos recursos. El programa asignaba a cada escuela hasta US\$ 3.000 por cada docente de tiempo completo afiliado a un sindicato si cumplía con el objetivo de desempeño anual establecido por el Departamento de Educación de Estados Unidos en los informes de los puntajes de desempeño de la escuela. Dicho objetivo estaba basado en: el desempeño de los alumnos y sus avances en las pruebas del estado para las escuelas primarias y de nivel medio, un examen final para la escuela secundaria, el nivel de presentismo de los alumnos y una encuesta sobre el entorno de aprendizaje realizada a docentes, padres y alumnos. El programa también otorgaba a las escuelas US\$ 1.500 por cada docente de tiempo completo afiliado a un sindicato si alcanzaban al menos el 75% del objetivo estipulado. Las escuelas que recibían la bonificación podían distribuir el dinero a su criterio. Dado que la asignación de las escuelas al programa se realizó en forma aleatoria, Fryer (2011) aprovechó esta situación para evaluar el impacto de tal programa. Constató que no había producido efecto alguno en el desempeño estudiantil y no halló evidencia de que hubiese repercutido en los niveles de retención de los docentes, el presentismo o el entorno de aprendizaje. El autor analizó varias razones por las cuales el programa de incentivos podía haber resultado ineficaz, entre las que cabe mencionar la posibilidad de que el monto de los incentivos no haya sido suficientemente alto, la complejidad del esquema de incentivos, el hecho de que los incentivos grupales no sean eficaces y la probabilidad de que los docentes no sepan cómo mejorar el desempeño de los estudiantes.

Los resultados de un experimento realizado en Nashville, Tennessee, resultaron igualmente desalentadores. Springer *et al.* (2010) evaluaron el impacto del proyecto Project on Incentives in Teaching (POINT), una iniciativa que ofrecía a los docentes de matemática de las escuelas de nivel medio bonificaciones de hasta US\$ 15.000 al año si lograban que sus alumnos obtuviesen mejoras extraordinariamente altas en las calificaciones de la evaluación del estado —el Programa de Evaluación Integral del Estado de Tennessee (TCAP, por su sigla en inglés)— durante los años lectivos 2006–2008. Los autores realizaron un estudio de asignación aleatoria por el cual se asignó a la mitad de los docentes que se ofrecieron voluntariamente a participar en el estudio a un grupo con derecho a recibir la bonificación y la otra mitad a un grupo sin derecho a la bonificación. Springer y sus colaboradores determinaron que, en promedio, no hubo diferencias entre el desempeño de los alumnos de los docentes con derecho a bonificación y los de aquellos sin derecho a recibirla. Mientras que los alumnos de 5° grado de los docentes con derecho a bonificación obtuvieron un desempeño superior al de sus pares en el segundo y tercer año del estudio, este efecto no se mantuvo una vez que los alumnos pasaron al siguiente grado.

La mayor parte de las demás evaluaciones rigurosas de planes de retribución basados en el desempeño docente se realizó en países en desarrollo. Muralidharan y Sundararaman

(2009) llevaron a cabo un estudio de asignación aleatoria con el fin de evaluar un programa de incentivos a docentes implementado en el estado de Andhra Pradesh, India. El programa ofrecía bonificaciones de alrededor del 3% del salario anual de los docentes en base al promedio de mejora del puntaje obtenido por sus alumnos en evaluaciones realizadas en forma independiente. Al cabo de dos años, el programa mejoró el desempeño de los alumnos 0,28 desviaciones estándar en matemática y 0,16 en comprensión lectora. Un dato interesante es que las escuelas que recibieron bonificaciones tuvieron un mejor desempeño en los componentes mecánico y conceptual de las evaluaciones, lo cual sugiere que las mejoras no se debieron a que se preparó a los alumnos “para el examen”. De hecho, los alumnos de las escuelas del grupo de tratamiento también lograron un mejor desempeño en las materias que no estaban alcanzadas por los incentivos, como ciencia y estudios sociales. Por último, los resultados de los incentivos individuales y grupales no registraron diferencias durante los dos primeros años del programa. Este estudio indica que los planes de retribución en base al mérito pueden generar efectos significativos en cuanto a mejorar el nivel de desempeño de los alumnos en los países en desarrollo.

Un estudio en Kenia centró la atención en la importancia de lo que este tipo de programas evalúan y recompensan. Glewwe, Ilias y Kremer (2010) aplicaron un estudio de asignación aleatoria a un programa en Busia y Teso, Kenia, que otorgaba a los docentes de escuelas primarias premios en especie según las calificaciones de sus alumnos en los exámenes distritales. El programa asignaba puntajes bajos a los alumnos que no se presentaban a rendir examen para evitar que las escuelas deliberadamente alentaran a los alumnos de menor rendimiento a no presentarse. Los alumnos de las escuelas del grupo de tratamiento obtuvieron puntajes más altos que los de las de control el primer año. No obstante, la mayor parte de las mejoras se concentraron en los aspectos evaluados por el modelo de recompensa, y el programa no generó efecto alguno en la conducta de los docentes (presentismo, cantidad de tarea para el hogar asignada, métodos pedagógicos). La bonificación hizo que aumentara la preparación para los exámenes en un 4,2% el primer año y un 7,4% el segundo año. Estos hallazgos son ilustrativos de que los programas de incentivos, a menos que estén diseñados en forma adecuada, pueden llevar a los docentes a concentrarse en la preparación específica para un examen pero no a generar cambios que favorezcan un mejor aprendizaje.

En Chile, Rau y Contreras (2009) se valieron del análisis de diferencias en diferencias, el diseño de regresión continua y la correspondencia del puntaje de propensión para evaluar un programa que ofrecía bonificaciones a todos los docentes de las escuelas primarias y de nivel medio públicas o subsidiadas por el Estado sobre la base del rendimiento los alumnos y de los avances en los puntajes de la prueba nacional de desempeño, más otros insumos escolares. La bonificación era baja pero no demasiado (entre 10% y 40% del salario mensual del docente). El programa elevó entre 0,7 y 0,12 desviaciones estándar el

promedio de los resultados de aprendizaje. No obstante, no se halló evidencia de que el hecho de obtener la bonificación generara algún tipo de avance adicional. Estas conclusiones parecen indicar que si bien la implementación del programa puede haber llevado a las escuelas a dedicar mayores esfuerzos a elevar el nivel de aprendizaje de los alumnos mediante simples mejoras de la eficiencia, la complejidad (y, por ende, la previsibilidad) de la bonificación no debe de haberles dado muchos indicios acerca de cómo lograr avances más allá de estas mejoras iniciales.

Algunas de las evaluaciones sobre los incentivos docentes también revelan aspectos que pueden resultar útiles para el diseño de este tipo de iniciativas. Por ejemplo, Lavy (2008, 2009) utilizó un diseño de regresión continua para analizar un programa en Israel que otorgaba bonificaciones a los docentes en base a los promedios de los alumnos en los exámenes de matriculación para el secundario y en los índices de aprobación en función de puntajes predeterminados (ajustados de acuerdo con el nivel socioeconómico de los alumnos). El monto de la bonificación era alto (entre 70% y 300% del salario mensual de un docente) y los docentes tenían probabilidades de recibirlo (alrededor de 48% de ellos recibió algún tipo de premio). Como resultado del programa, los alumnos de las escuelas incluidas en el grupo de tratamiento tuvieron un índice de aprobación 14% superior en matemática y 5% superior en inglés. Estos alumnos también obtuvieron puntajes 10% más altos en matemática y 4% más altos en inglés. Sin embargo, Lavy estableció que si bien los efectos de las bonificaciones no presentaron diferencias en términos de género, las mujeres se mostraron más pesimistas respecto de la probabilidad de recibir las. Este estudio sugiere que incluso cuando las bonificaciones son previsibles, distintos grupos de docentes pueden reaccionar ante dicha previsibilidad de distintas maneras.

La evaluación de un programa de incentivos docentes implementado en México arroja luz sobre la importancia relativa de los distintos aspectos que se premian y la persistencia de los premios en un programa de retribución por mérito. McEwan y Santibáñez (2005) utilizaron un diseño de regresión discontinua para evaluar el programa, que otorgaba una gratificación a los docentes y directores que obtenían un puntaje por encima del punto de corte según un índice que tomaba en cuenta, entre otros factores, su formación, su experiencia y las calificaciones de sus alumnos. La retribución no sólo era considerable, sino que además se mantenía a lo largo de toda la carrera del docente. Sin embargo, no se halló evidencia de que las calificaciones de los alumnos hayan mejorado como resultado de esta reforma. Los autores conjeturaron que quizá se debió a que el grueso de la retribución (alrededor del 80%) estaba determinado por los antecedentes de los docentes y directores antes que por su desempeño laboral. Esto indica que, a la hora de diseñar programas de incentivos, los sistemas educativos deberían garantizar que las bonificaciones estén asociadas a la clase de cambios que desean promover en la conducta y la productividad de sus empleados.

Por último, un estudio realizado en Kenia brinda importantes consideraciones acerca del potencial de los incentivos docentes basados en los recursos antes que en los resultados. Kremer *et al.* (2001) se valieron de un estudio de asignación aleatoria para evaluar un programa de bonificaciones individuales en el nivel preescolar que otorgaba recursos a los directores para que recompensaran a los docentes con buen nivel de presentismo. El monto de la bonificación era alto, pues podía representar hasta un 300% del salario mensual de un docente. Pero los directores distribuyeron la bonificación entre todos los maestros de sus escuelas sin importar el índice de presentismo. Así, la bonificación no tuvo impacto alguno en los niveles de ausentismo de los docentes (en promedio, alrededor de 29%), el nivel de presentismo de los alumnos ni el desempeño. Además de confirmar el riesgo del usufructo del beneficio por parte de quienes no pagan por él (*free riding*) planteado por el estudio sobre las bonificaciones implementado en Nueva York, este análisis parece sugerir que existe poca evidencia de que las bonificaciones basadas en los recursos ofrezcan probabilidades ciertas de incrementar los niveles de aprendizaje de los alumnos.

En suma, los estudios de intervenciones que intentan elevar la productividad y el nivel de dedicación de los docentes aportan al menos dos lecciones claras. Primero, la incorporación de docentes por contrato parece ser una manera promisorio de aumentar la masa de docentes en los países en desarrollo. Segundo, aunque los efectos de los programas de retribución basados en el mérito en los países desarrollados parecen ser dispares, la evidencia proveniente de los países en desarrollo resulta más alentadora. No obstante, las características fundamentales del diseño (por ejemplo, bonificaciones grupales versus individuales, alcance, parámetros de desempeño, procesos de asignación, previsibilidad, monto de las bonificaciones, etc.) son vitales a la hora de arbitrar el impacto de estos planes (Bruns y Santibáñez, 2011).

Interpretación de la teoría y la evidencia: conclusiones y salvedades

En el presente trabajo se ha intentado ofrecer un marco analítico que permita interpretar la evidencia disponible acerca de las políticas docentes en los países desarrollados y en desarrollo y entablar un diálogo con los lectores respecto de qué es lo que se sabe y con qué grado de certeza. A diferencia de muchas revisiones anteriores, en este trabajo no se ha procurado “condensar” las principales lecciones de la evidencia disponible para un público sin conocimientos técnicos. El objetivo ha sido explicar los métodos utilizados en los estudios de las políticas docentes con el fin de fomentar el diálogo sobre la base del conocimiento con todas las partes interesadas: ciudadanos, docentes, funcionarios de gobierno y expertos, entre otros.

En esta sección se ofrecen algunas consideraciones finales acerca de la situación de la evidencia sobre las políticas docentes y se proponen directrices para la investigación futura.

En primer lugar, los datos sobre las políticas docentes son dispares y deben interpretarse con cautela. Se sabe mucho acerca de algunas políticas y poco o nada sobre otras. Esto obedece a diversos factores. Uno de ellos es que existen ciertas áreas en las que no se han presentado oportunidades de realizar investigaciones rigurosas, principalmente debido a que las normas y reglamentaciones dificultan a los investigadores la implementación de algunas estrategias metodológicas analizadas en este estudio. Otra de las razones es que algunas de las intervenciones son relativamente recientes y se necesita tiempo para que los investigadores puedan evaluar los resultados.

A falta de estudios rigurosos acerca de una intervención en particular, puede resultar tentador recurrir al mejor estudio disponible, aun cuando no sea capaz de distinguir el impacto de una intervención independientemente de los demás factores que puedan afectar los resultados de interés. Esta postura es comprensible: tal como se mencionó al principio de este trabajo, todos los días hay que tomar decisiones en materia de políticas y los gobiernos no pueden esperar a que se publiquen los estudios más rigurosos para hacerlo. No obstante, también queda claro que la estrategia de utilizar las investigaciones que funcionan como “segunda alternativa” no siempre resultó de utilidad para la comunidad educativa. Ha habido varios casos de políticas docentes importantes que parecían promisorias en un comienzo hasta que estudios más rigurosos demostraron lo contrario. Antes que basarse en estudios poco rigurosos, los gobiernos deberían concebir nuevas intervenciones susceptibles de ser sometidas a evaluaciones rigurosas. De este modo, una política puede repercutir por sí misma en decisiones futuras.

En segundo lugar, si bien este trabajo se centró en el impacto de las políticas docentes en el desempeño estudiantil, no se tuvieron en cuenta resultados importantes que los gobiernos deben tomar en consideración. Tal como lo demuestran varios estudios sobre los aspectos económicos de la educación, a menudo puede haber efectos “latentes” significativos de las políticas que no pueden identificarse hasta que los niños alcancen la adolescencia o incluso la adultez (relacionados con, por ejemplo, la salud, la criminalidad, la planificación familiar, los ingresos y el espíritu cívico). Los gobiernos necesitan contar con sistemas informáticos sólidos que permitan a los investigadores evaluar estos resultados clave a largo plazo.

En tercer lugar, en el presente trabajo se estableció la necesidad de desarrollar una comprensión más acabada no sólo de los efectos de cada una de las políticas docentes por separado, sino también de sus interacciones. Varios estudios han comenzado a hacerlo mediante el diseño de evaluaciones destinadas a analizar el impacto de distintas

combinaciones de intervenciones (por ejemplo, la reducción de la cantidad de alumnos por clase por sí sola, la reducción de la cantidad de alumnos por clase combinada con la evaluación del docente y la reducción de la cantidad de alumnos por clase combinada con la evaluación del docente y el seguimiento por habilidades). Celebramos estas iniciativas y alentamos a otros a seguir por el mismo camino.

Por último, creemos que existe una necesidad de incluir más consideraciones de la relación costo-beneficio en las evaluaciones de impacto. Suele ocurrir que se dispone de varias intervenciones que pueden resolver un mismo problema en forma eficaz pero, al carecer de información sobre la relación costo-beneficio, resulta difícil determinar cuál de ellas promete un mayor rendimiento. Algunos autores han comenzado a prestar atención a estos asuntos en otras áreas de la educación (Dhaliwal *et al.*, 2011), pero deben redoblar los esfuerzos para evaluar las cuestiones de costos en los análisis de las distintas políticas docentes.

Cuadro 1. Datos sobre las intervenciones destinadas a fijar expectativas claras para los docentes

Brindar andamiaje al trabajo docente				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Clases con guión de contenidos (preescolar, 2° grado), con un currículo estructurado, actividades de lectura, seguimiento por habilidades y lectura obligatoria en el hogar.	Estados Unidos	Borman <i>et al.</i> (2007)	Muy riguroso (EAC)	<ul style="list-style-type: none"> Los efectos abarcaron de 0,21 desviaciones estándar en comprensión de fragmentos a 0,33 desviaciones estándar en reconocer palabras.
Currículo y actividades planificados previamente (preescolar y 1° grado) para comprensión lectora.	Mumbai, India	He, Linden, y MacLeod (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> Se constataron mejoras en el puntaje de los alumnos en comprensión lectora de 0,26–0,70 desviaciones estándar, las mayores en el caso de alumnos de nivel preescolar y de bajo rendimiento. La versión del programa implementada fuera de la jornada escolar fue más efectiva que la que se instrumentó durante el horario de clase, lo que arrojó un efecto adicional de 0,24 desviaciones estándar más que el efecto de 0,26 desviaciones estándar.
Actividades de clase planificadas previamente (grados 1°–5°) a través de una máquina o de tarjetas didácticas (<i>flashcards</i>).	Maharashtra, India	He, Linden, y MacLeod (2007)	Muy riguroso (EAC)	<ul style="list-style-type: none"> Se constataron mejoras en el aprendizaje del idioma inglés de alrededor de 0,30 desviaciones estándar. Se constató que fue más efectivo en el caso los alumnos de mayor edad y menor rendimiento. La versión del programa enseñada por los docentes no sólo mejoró los puntajes de los alumnos en inglés, sino también en matemática, lo que demuestra efectos de derrame positivos.
Maratón de lectura (4° grado) acompañada de capacitación para los docentes y suministro de material de lectura.	Tarlac, Filipinas	Abeberese, Kumler, y Linden (2011)	Muy riguroso (EAC)	<ul style="list-style-type: none"> Aumentó cantidad de libros leídos durante el mes posterior a la implementación del programa (de 2,3 a 9,5). Los puntajes de comprensión lectora aumentaron 0,13 desviaciones estándar.

				<ul style="list-style-type: none"> ▪ Tres meses después de la maratón de lectura, los alumnos incluidos en el grupo de tratamiento seguían leyendo 3,1 libros más que los del grupo de control y sus puntajes en esta materia eran 0,06 desviaciones estándar más altos.
Manual instructivo e informes de desempeño de la escuela (grados 3° y 4°)	Madagascar	Lassibille <i>et al.</i> (2010); Glewwe y Maïga (2011)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Los efectos sobre los puntajes de las pruebas no fueron estadísticamente significativos. ▪ El impacto no varió según el tipo de docentes (de planta permanente, contratados o practicantes). ▪ El programa elevó el índice de asistencia y disminuyó el de repetición sólo cuando se lo combinó con intervenciones a nivel distrital y subdistrital.
Aumentar las horas de clase				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Tutorías después del horario de clase y durante el verano (grados 5°–6°), que ofrecían actividades académicas, de enriquecimiento y de apoyo después de la escuela y durante el verano.	Washington DC	Linden, Herrera, y Grossman (2011)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Puntajes 0,09 desviaciones estándar más altos (comprensión lectora) y 0,12 desviaciones estándar más altos (solución de problemas) al cabo de dos años. ▪ Los alumnos que participaron en el programa eran más propensos a visitar escuelas secundarias, a conseguir información sobre ellas y a conversar con pares y adultos sobre aquella a la cual postularse.
Escuela de verano (3° y 6° grado) obligatoria para todos los alumnos que no aprobaban los exámenes de fin de año para pasar de grado.	Chicago	Jacob y Lefgren (2002)	Riguroso (DRD + EVI)	<ul style="list-style-type: none"> ▪ Se registraron puntajes 0,11–0,13 desviaciones estándar más altos en comprensión lectora y matemática en 3° grado, pero no hubo diferencias en los puntajes de 6° grado.

Reducir la cantidad de alumnos por clase				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Reducción de la cantidad de alumnos por clase (preescolar a 3° grado) mediante clases reducidas o clases regulares con auxiliar docente.	Tennessee	Krueger (1999)	Muy riguroso (EAC)	<ul style="list-style-type: none"> El efecto promedio por alumno fue de alrededor de 0,048 desviaciones estándar en comprensión lectora y matemática. El efecto fue más marcado en el caso de los alumnos pobres y los de bajo nivel socioeconómico.
Reducción de la cantidad de alumnos por clase (1° grado) mediante la incorporación de docentes por contrato, el seguimiento en función del desempeño inicial de los alumnos y/o facultar a los padres para evaluar el desempeño de los docentes.	Busia y Teso, Kenia	Duflo, Dupas, y Kremer (2007)	Muy riguroso (EAC)	<ul style="list-style-type: none"> La reducción de la cantidad de alumnos por clase (en promedio, de 80 a 46), en ausencia de otras reformas, llevó a una disminución en el esfuerzo del docente y a un aumento insignificante de los puntajes de los alumnos en las pruebas. Combinar la reducción de la cantidad de alumnos por clase con mayores incentivos para los docentes llevó a mejoras significativamente más altas en los puntajes, ya sea mediante la incorporación de docentes por contrato o el aumento de la supervisión por parte de los padres (0,19 desviaciones estándar en lectoescritura y aritmética) o a través de la incorporación de docentes por contrato y el seguimiento de los alumnos (0,25–0,31 desviaciones estándar).
Reducción de la cantidad de alumnos por clase (1° a 6° grado) surgida de variaciones idiosincráticas en la población o normas vigentes sobre tope máximo/mínimo.	Connecticut	Hoxby (2000)	Riguroso (EVI + DRD)	<ul style="list-style-type: none"> No generó efecto alguno en el rendimiento estudiantil. Incluso se pudo descartar efectos modestos (0,02–0,04 desviaciones estándar en el caso de una reducción del 10% de la cantidad de alumnos por clase).
Reducción de la cantidad de alumnos por clase (3° a 5° grado) a través de una norma que establece un máximo de 40 alumnos por aula.	Israel	Angrist y Lavy (1999)	Riguroso (DRD)	<ul style="list-style-type: none"> El efecto por alumno fue de 0,017–0,019 desviaciones estándar para 4° grado y de 0,036–0,071 desviaciones estándar para 5° grado, si se combina el rendimiento estudiantil en ambas materias.

Fuente: Elaborado por los autores.

Nota: EVI = estimación con variables instrumentales; EAC = estudio de asignación aleatoria; DRD = diseño de regresión discontinua.

Cuadro 2. Datos sobre las intervenciones destinadas a atraer a los mejores a la docencia

Fijar requisitos de ingreso a la docencia				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Certificación profesional (2° a 5° grado) sobre la base de filmaciones en el aula y trabajos escritos.	Los Ángeles	Cantrell <i>et al.</i> (2008)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Los alumnos de docentes con certificación profesional obtuvieron un puntaje 0,22 desviaciones estándar por encima de los alumnos cuyos docentes se habían postulado para la certificación pero no habían aprobado. Sin embargo, los alumnos de los docentes certificados no resultaron más eficaces que los alumnos de los docentes que no se postularon a la certificación. ▪ Los alumnos de los docentes que fueron certificados también tuvieron mejor desempeño que los alumnos de los docentes que se postularon y no fueron aprobados, pero no tuvieron mejor desempeño que los alumnos de los docentes que no se postularon a la certificación. ▪ Los alumnos de los docentes que obtuvieron calificaciones altas en su certificación tuvieron un mejor desempeño que los alumnos de los docentes que lograron la certificación con calificaciones bajas.
Exámenes sobre conocimiento de contenidos (6° grado) para docentes en ejercicio.	Perú	Meltzer y Woessmann (2010)	Experimental (efectos fijos: alumno, docente y materia).	<ul style="list-style-type: none"> ▪ Un aumento de una desviación estándar en el puntaje del docente en las evaluaciones de matemática y comprensión lectora eleva los puntajes de los alumnos en las pruebas en 0,10 desviaciones estándar.

<p>Certificación docente tradicional (3° a 5° grado), incluidos experiencia docente, puntajes de evaluaciones y autorizaciones para ejercer.</p>	<p>Carolina del Norte</p>	<p>Clotfelter, Ladd y Vigdor (2007a)</p>	<p>Experimental (efectos fijos: alumno, escuela, materia y año)</p>	<ul style="list-style-type: none"> ▪ La experiencia del docente, los puntajes de las evaluaciones y la certificación por vía regular tienen efectos positivos en el aprendizaje de los alumnos, más altos en matemática que en comprensión lectora. ▪ La magnitud de los efectos resulta considerable si se la compara con los efectos generados por la reducción de la cantidad de alumnos por clase o las características socioeconómicas de los alumnos.
<p>Certificación docente (4° a 8° grado).</p>	<p>Ciudad de Nueva York</p>	<p>Kane, Rockoff y Staiger (2006)</p>	<p>Experimental (efectos fijos: escuela, grado y año)</p>	<ul style="list-style-type: none"> ▪ La condición del docente respecto de la certificación tiene escasos efectos sobre el desempeño de los alumnos. ▪ Se detectaron diferencias considerables y persistentes respecto de la efectividad entre docentes con el mismo tipo de certificación. ▪ Incluso los grupos con alta rotación certificados por vías alternativas tendrían que ser sólo ligeramente más eficaces durante su primer año de ejercicio para compensar los efectos de sus altos índices de deserción.
<p>Exámenes sobre conocimiento de contenidos (4° a 8° grado) para docentes en ejercicio.</p>	<p>Charlotte-Mecklenburg (Carolina del Norte), Dallas (Texas), Denver (Colorado), condado de Hillsborough (Florida) ciudad de Nueva York y Memphis (Tennessee)</p>	<p>Cantrell y Kane (2013)</p>	<p>Experimental (correlaciones entre dos secciones de alumnos)</p>	<ul style="list-style-type: none"> ▪ Se correlacionaron los exámenes sobre conocimiento de contenidos para la enseñanza de matemática y comprensión lectora de los docentes con su desempeño en las observaciones de las clases de las materias específicas, pero no con su valor agregado. ▪ No se correlacionaron las evaluaciones con el valor agregado de los docentes en ningún punto de la distribución de valor agregado (es decir, no resultó de utilidad para identificar a los docentes con

Flexibilizar los requisitos de ingreso a la docencia para aspirantes sobresalientes				desempeño bajo, promedio o alto).
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Vía alternativa para incorporarse a la docencia (1° a 5° grado) que atrae a graduados terciarios o universitarios sobresalientes para enseñar en escuelas con muchas necesidades durante dos años.	Estados Unidos	Decker, Mayer y Glazerman (2004)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ El programa atrajo a docentes con mayores probabilidades de asistir a una universidad prestigiosa y obtener un título mientras se dedicaban a la docencia y menores probabilidades de contar con formación específica en educación y experiencia docente que otros en sus escuelas. ▪ Los alumnos de los docentes del programa tuvieron un desempeño 0,15 desviaciones estándar más alto en matemática que los alumnos de los docentes del grupo de control (o 10% de equivalentes de grado escolar), pero no se registraron diferencias en comprensión lectora. ▪ Comparados con los alumnos de otros docentes principiantes, los alumnos de los docentes del programa tuvieron un desempeño 0,26 desviaciones estándar más alto en matemática, pero no hubo diferencias en comprensión lectora.
Vía alternativa para incorporarse a la docencia (preescolar a 5° grado) que reúne a todos los docentes capacitados por vías alternativas.	Estados Unidos	Constantine <i>et al.</i> (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Los docentes certificados por la vía tradicional y los certificados por la vía alternativa mostraron variaciones sustanciales en la cantidad de horas de formación previas al inicio del ejercicio docente. ▪ No se detectaron diferencias significativas en los puntajes obtenidos en los exámenes de ingreso a la institución terciaria o universidad, en la selectividad de estas instituciones educativas ni en los logros académicos entre los docentes certificados

				<p>por la vía tradicional y los certificados por la vía alternativa.</p> <ul style="list-style-type: none"> Los docentes certificados por la vía tradicional y los certificados por la vía alternativa resultaron igualmente eficaces para mejorar el desempeño de los alumnos. Ni la cantidad ni los contenidos de las prácticas de los docentes certificados por la vía alternativa marcaron una diferencia en su eficacia.
<p>Vía alternativa para incorporarse a la docencia (7° a 9° grado) que atrae a graduados terciarios o universitarios sobresalientes para enseñar en escuelas con muchas necesidades durante dos años.</p>	Chile	Alfonso, Santiago y Bassi (2011)	Experimental (CPP)	<ul style="list-style-type: none"> Las escuelas que participaron en el programa obtuvieron un puntaje 0,22–0,51 desviaciones estándar más alto en comprensión lectora y 0,17–0,43 desviaciones estándar más alto en matemática al cabo de un año y 0,75 desviaciones estándar más alto en español y 0,33 desviaciones estándar más alto en matemática transcurridos dos años. Los alumnos de las escuelas que participaron en el programa también registraron mediciones más altas de autoestima, autoeficacia y habilidades intelectuales y meta cognitivas.
Recensar a los docentes más calificados y con mayor experiencia				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
<p>Experiencia docente (4° a 8° grado) de tres a 12 años</p>	Charlotte-Mecklenburg (Carolina del Norte), Dallas (Texas), Denver (Colorado), condado de Hillsborough (Florida), ciudad de	Kane y Staiger (2012)	Experimental (efectos fijos: escuela, grado y materia; dos secciones de alumnos)	<ul style="list-style-type: none"> Los alumnos de docentes con 12 años de experiencia o más alcanzaron logros que fueron 0,01 desviaciones estándar más altos en matemática y 0,02 desviaciones estándar más altos en comprensión lectora en las pruebas de los estados que los alumnos de docentes con menos de tres años de experiencia. Los resultados fueron más bajos en el caso de pruebas

	Nueva York y Memphis (Tennessee).			<p>complementarias.</p> <ul style="list-style-type: none"> Los alumnos de los docentes con una maestría registraron mejoras que resultaron 0,03 desviaciones estándar más altas en matemática y 0,02 desviaciones estándar más bajas en comprensión lectora en las pruebas de los estados comparados con los alumnos de docentes sin maestrías. Los resultados fueron más altos en el caso de pruebas complementarias. Los logros derivados de la experiencia resultaron ser mucho más bajos que los derivados de un parámetro combinado de calidad docente integrado por el valor agregado del docente, los resultados de encuestas a alumnos y el puntaje de las observaciones en el aula.
Posgrados y certificación (3° a 8° grado)	Texas	Hanushek <i>et al.</i> (2005)	Experimental (efectos fijos: escuela y alumno)	<ul style="list-style-type: none"> La calidad de los docentes parece no estar relacionada con el hecho de poseer posgrados o certificación; en cambio, la experiencia en el primer año de ejercicio parece ser relevante. Los buenos maestros tienden a ser eficaces con alumnos de todos los niveles de habilidad, pero agrupar a docentes y alumnos de acuerdo con la raza arroja un valor positivo.
Aumentar el salario docente				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Mejoras en el salario docente (escuela primaria y secundaria) en términos de niveles y distribución.	Estados Unidos	Hoxby y Leigh (2004)	Riguroso (EVI)	<ul style="list-style-type: none"> La compresión salarial aumentó en 9 puntos porcentuales la proporción de mujeres graduadas de diversos institutos educativos con menos habilidades que se incorporaron a la docencia e hizo que se redujera en 12 puntos porcentuales la proporción de

				<p>graduadas con mejores aptitudes incorporadas a la docencia.</p> <ul style="list-style-type: none"> Las mejoras en la paridad salarial redujeron la proporción de mujeres dedicadas a la docencia en 3,2 puntos porcentuales en el caso de las docentes con mejores aptitudes, mientras que en el caso de las docentes de los tres grupos de aptitud más bajos se mantuvo en 0.
--	--	--	--	--

Fuente: Elaborado por los autores.

Nota: EVI = estimación con variables instrumentales; CPP = correspondencia del puntaje de propensión; EAC = estudio de asignación aleatoria.

Cuadro 3. Datos sobre las intervenciones destinadas a preparar a los docentes con formación y experiencia útiles

Fijar requisitos de ingreso a la docencia				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Tutoría obligatoria (4° a 8° grado), según la cual los maestros con menos de un año de experiencia se reúnen en forma semanal con un tutor.	Ciudad de Nueva York	Rockoff (2008)	Riguroso (ADD + EVI)	<ul style="list-style-type: none"> ▪ Los docentes sin experiencia previa a quienes se les asignó un tutor tuvieron una probabilidad 4,5% mayor de completar su primer año que los docentes a los que no se le asignó un tutor. ▪ La retención de los docentes era mayor si en la escuela tenían tutores con experiencia previa en esa escuela. ▪ Por cada 10 horas de tutoría, los alumnos de los docentes con tutores mejoraron su desempeño 0,05 desviaciones estándar en matemática y 0,04 desviaciones estándar en comprensión lectora.
Inducción integral (2° a 5° grado) con apoyo intensivo de tutoría, orientación, desarrollo profesional, observaciones en clase y devoluciones para docentes principiantes.	Estados Unidos	Glazerman <i>et al.</i> (2010)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ El 90% de los docentes principiantes de los distritos asignados a la inducción integral recibió un tutor, comparado con un 72% de los docentes de los distritos que participaron del programa de inducción habitual. ▪ Los docentes que recibieron la inducción integral durante un año no resultaron más eficaces que sus pares de control, tal como se midió en las observaciones de clases y en el desempeño de los alumnos. ▪ Los docentes con dos años de inducción integral elevaron el desempeño de los alumnos 0,11 desviaciones estándar en comprensión lectora y 0,20 desviaciones estándar en matemática durante el tercer año del estudio. ▪ La inducción integral no tuvo efecto alguno en la retención o satisfacción docente.

Incluir un componente de práctica en la formación docente				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Residencia docente (4° a 8° grado) con postulantes locales, duración de un año con cuatro reuniones semanales con el tutor, aumento gradual de responsabilidades en el aula y compromiso por tres años.	Boston	Papay <i>et al.</i> (2011)	Experimental (efectos fijos: escuela, alumno, grado y año)	<ul style="list-style-type: none"> ▪ Los docentes participantes se caracterizan por presentar una diversidad racial mayor que otros docentes principiantes de su distrito, por ser más propensos a enseñar matemática y ciencias y por una tendencia a permanecer en su distrito durante los primeros cinco años de ejercicio. ▪ Inicialmente, estos docentes no fueron más eficaces para mejorar los resultados de las pruebas de lengua de sus alumnos que otros docentes principiantes y fueron menos eficaces en matemática en relación con el mismo grupo, pero mejoraron su desempeño rápidamente de modo tal que durante el cuarto y quinto año de ejercicio superaron incluso a los docentes experimentados.

Fuente: Elaborado por los autores.

Nota: ADD = análisis de diferencias en diferencias; EVI = estimación con variables instrumentales; EAC = estudio de asignación aleatoria.

Cuadro 4. Datos sobre las intervenciones destinadas a asignar a los docentes donde más se los necesita

Ofrecer una retribución extra para enseñar en escuelas con déficits docentes				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
<p>Retribución extra para atraer a docentes principiantes y talentosos a escuelas de bajo rendimiento y retenerlos (todos los grados) durante al menos cuatro años, seleccionados en base a promedios de calificaciones, cartas de recomendación, CVs, trabajos escritos y entrevistas.</p>	California	Steele, Murnane, y Willett (2010)	Riguroso (EVI)	<ul style="list-style-type: none"> ▪ La retribución extra de US\$ 20.000 a la firma del contrato aumentó la probabilidad de que los beneficiarios enseñaran en escuelas de bajo rendimiento en 28 puntos porcentuales. ▪ El 75% de los docentes con o sin retribución extra que comenzaron a trabajar en escuelas de bajo rendimiento permanecieron en esas escuelas durante al menos cuatro años.
Ofrecer una retribución extra para enseñar materias con déficits docentes				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
<p>Retribución extra para retener docentes certificados en matemática, ciencia y educación especial (10° a 12° grado) en escuelas de zonas muy pobres o de muy bajo rendimiento.</p>	Carolina del Norte	Clotfelter <i>et al.</i> (2008)	Riguroso (ADD)	<ul style="list-style-type: none"> ▪ La retención interanual aumentó entre el 10% y el 13%. ▪ Los docentes de matemática que recibieron la retribución extra se mostraron 18% menos proclives a irse de sus escuelas al año siguiente que quienes no la recibieron, pero los docentes de ciencia y educación especial no se mostraron menos proclives a irse. ▪ Los docentes de las escuelas de nivel medio fueron 27% menos proclives a dejar la escuela si habían recibido la bonificación, mientras que los docentes de escuelas secundarias que habían recibido la suma extra no mostraron diferencias estadísticamente significativas en los patrones de deserción.

Mejorar las condiciones laborales				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Coincidencia con el puesto de trabajo (<i>job matching</i>) (3° a 5° grado), medida en función del cambio de los docentes de una escuela a otra.	Carolina del Norte	Jackson (2013)	Experimental (efectos fijos: docente, escuela y año)	<ul style="list-style-type: none"> ▪ Los puntajes de valor agregado de los docentes aumentaron 0,09 desviaciones estándar en matemática y 0,07 desviaciones estándar en comprensión lectora cuando pasaron a desempeñarse en una escuela más afín a ellos. ▪ Se determinó una correlación negativa entre el efecto de coincidencia o afinidad y el cambio de escuelas, efecto no relacionado con el abandono de la profesión, y se determinó que aumenta con la experiencia.
Pares efectivos (3° a 5° grado), medidos según el desempeño previo de los alumnos de los pares de un docente que cambia de escuela.	Carolina del Norte	Jackson y Bruegman (2009)	Experimental (efectos fijos: alumno, docente, escuela y año)	<ul style="list-style-type: none"> ▪ Los docentes que experimentaron una mejora de una desviación estándar respecto de las características observables de sus pares fueron asociados a un aumento de 0,10 desviaciones estándar en el puntaje obtenido por sus alumnos en matemática y comprensión lectora. ▪ Los alumnos de los docentes que experimentaron un aumento de una desviación estándar en los puntajes de valor agregado de sus pares obtuvieron un desempeño 0,05 desviaciones estándar más alto en matemática y comprensión lectora.

Fuente: Elaborado por los autores.

Nota: ADD = análisis de diferencias en diferencias; EVI = estimación con variables instrumentales.

Cuadro 5. Datos sobre las intervenciones destinadas a liderar a los docentes con buenos directores

Contratar directores más efectivos				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Contratar directores más eficaces (3° a 8° grado), medidos según el desempeño promedio de los alumnos de sus escuelas.	Texas	Branch, Hanushek, y Rivkin (2012)	Experimental (efectos fijos: director y escuela)	<ul style="list-style-type: none"> ▪ El impacto anual de contar con un director eficaz se estableció en 0,05–0,21 desviaciones estándar, dependiendo del método utilizado para su estimación. ▪ La mayor probabilidad de que los maestros que abandonaban las escuelas a cargo de los directores más eficaces se encontró entre los maestros menos eficaces de dichas escuelas, a diferencia de lo que ocurría con los maestros que abandonaban las escuelas a cargo de directores menos eficaces. ▪ Los directores más eficaces no siempre son más proclives a permanecer en su puesto y menos propensos a abandonar la educación pública.
Contratar directores más eficaces (3° a 10° grado), medidos según el desempeño promedio de los alumnos de sus escuelas.	Miami	Grissom, Kalogrides, y Loeb (2012)	Experimental (efectos fijos: escuela, barrio y pares)	<ul style="list-style-type: none"> ▪ La elección del modelo utilizado para calcular los efectos de los directores es primordial para la evaluación. ▪ Mientras que algunos modelos adjudicaron a los directores efectos del orden de 0,15 desviaciones estándar en matemática y 0,11 en comprensión lectora, otros identificaron efectos mucho más bajos, de hasta 0,02 desviaciones estándar en ambas materias para los mismos directores.

Contratar directores más eficaces (3° a 10° grado), medidos según el desempeño promedio de los alumnos de sus escuelas.	Columbia Británica, Canadá	Coelli y Green (2012)	Experimental (efectos fijos: escuela, barrio y pares)	<ul style="list-style-type: none"> ▪ Contar con un director con un nivel de desviación estándar más alto en la distribución de efectos de los directores indicó que los índices de graduación se incrementan 0,33 desviaciones estándar y los de lengua inglesa resultan una desviación estándar más altos en 12° grado.
Establecer requisitos para ocupar el puesto de director				
Intervención	Ubicación geográfica	Estudio	Método	Resultado
Encuestas a padres, docentes y vicedirectores sobre la eficacia de los directores	Miami	Grissom y Loeb (2011)	Experimental (efectos fijos: docente y alumno)	<ul style="list-style-type: none"> ▪ El aumento de una desviación estándar en la autoevaluación de los directores respecto de su “gestión organizacional” estuvo asociado a un aumento de 0,10 desviaciones estándar en la calificación de desempeño de la escuela. ▪ También se vincularon las autoevaluaciones de los directores respecto de su “gestión organizacional” con la calificación asignada por los padres de los alumnos a las escuelas, aunque no en forma sistemática en el caso del mejor desempeño en comprensión lectora y matemática. ▪ Las calificaciones de los vicedirectores de las capacidades organizacionales de los directores estuvieron relacionadas con los avances en matemática y comprensión lectora.
Experiencia y formación del director, medidas según la selectividad de la institución donde cursó sus estudios de grado y de posgrado, experiencia laboral previa y	Ciudad de Nueva York	Clark, Martorell, y Rockoff (2009)	Experimental (efecto fijo: escuela)	<ul style="list-style-type: none"> ▪ Hubo escasos indicios de existencia de una correlación entre el desempeño de una escuela y la selectividad de la institución donde el director cursó sus estudios de grado o posgrado. ▪ Se halló escasa relación entre el desempeño y la

desarrollo profesional.				<p>experiencia laboral previa del director.</p> <ul style="list-style-type: none"> ▪ Los alumnos de los directores con tres años de experiencia obtuvieron puntajes 0,04 desviaciones estándar más altos en matemática que los alumnos a cargo de directores en su primer año en el puesto, mientras que en el caso de los alumnos de los directores con cinco años de experiencia o más, los puntajes de matemática resultaron 0,06 desviaciones estándar más altos. ▪ Hubo resultados dispares respecto de la relación entre los programas de desarrollo profesional y capacitación del director y el desempeño de la escuela.
-------------------------	--	--	--	--

Darles a los directores mayor autoridad en la contratación de personal				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Suministrar puntajes de valor agregado (4° a 8° grado), que permiten a los directores recibir medidas de desempeño de los docentes a su cargo basadas en los resultados de las pruebas de los alumnos.	Ciudad de Nueva York	Rockoff <i>et al.</i> (2011)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Se constataron mayores probabilidades de que los docentes con puntajes de desempeño bajos abandonaran la escuela al divulgarse la información, lo que generó una leve mejora en la productividad los docentes de la escuela. ▪ La recepción de los resultados no hizo que se desestimara la información reunida por los directores mediante la observación de clases.
Autoridad discrecional conferida a los directores para despedir a docentes a prueba (3°, 5° y 8° grado) por cualquier motivo en forma expeditiva	Chicago	Jacob (2010a y 2010b)	Riguroso (ADD + emparejamiento)	<ul style="list-style-type: none"> ▪ Reducción del ausentismo entre los docentes a prueba en aproximadamente un 10% y entre los maestros con 15 faltas o más en un 20%. ▪ La política mejoró el desempeño de los alumnos del nivel primario. La reducción del ausentismo de los docentes sólo da cuenta de una parte de los efectos, lo que implica que la política también repercutió en la conducta de los docentes. ▪ Existe evidencia experimental de que los directores tienen en cuenta el nivel de ausentismo de los docentes y su valor agregado, además de varias características demográficas, al decidir a quién despedir.

Fuente: Elaborado por los autores.

Nota: ADD =análisis de diferencias en diferencias; EVI = estimación con variables instrumentales; EAC = estudio de asignación aleatoria.

Cuadro 6. Datos sobre las intervenciones destinadas a evaluar el aprendizaje y la enseñanza

Aumentar la participación de los padres y de la comunidad en los asuntos de la escuela				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Aumentar la participación de la comunidad en la gestión o el apoyo de la escuela (1° a 6° grado) al explicar el papel de los comités educativos comunales a los miembros de la comunidad y capacitar a voluntarios para tomar evaluaciones de diagnóstico o dar clases de apoyo.	Uttar Pradesh, India	Banerjee <i>et al.</i> (2010)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Un programa que explicó el papel de los comités comunales de educación a sus integrantes y a sus comunidades aumentó el nivel conocimiento de los integrantes del comité respecto de su papel en 0,39 desviaciones estándar. ▪ Un programa que capacitó a voluntarios de la comunidad para tomar evaluaciones de diagnóstico de comprensión lectora, escritura y matemática y comunicar los resultados aumentó el nivel de conocimiento del comité respecto de su función en 0,35 desviaciones estándar. ▪ Un programa que capacitó a voluntarios locales para dar clases de apoyo en comprensión lectora fuera del horario escolar aumentó el nivel de conocimiento de los miembros del comité respecto de su función en 0,32 desviaciones estándar y mejoró las calificaciones de comprensión lectora de los alumnos en 0,02 desviaciones estándar aproximadamente.
Brindar información a las comunidades respecto de su papel en la gestión de la escuela (1° a 5° grado) a través de roles y responsabilidades en la gestión de la escuela establecidos a nivel del Estado.	Karnataka, Madhya Pradesh y Uttar Pradesh, India	Pandey, Goyal, y Sundararaman (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Aumentó la cantidad de reuniones de los comités de educación comunales y la participación de los miembros en las inspecciones escolares en un 25% en Uttar Pradesh y se incrementó la cantidad de padres que hablaron con los docentes acerca de la calidad de la educación en Madhya Pradesh. ▪ El programa también elevó el índice de presentismo docente en un 11% en Uttar Pradesh y la actividad en el aula en un 30% en Madhya Pradesh. ▪ El programa mejoró el desempeño en comprensión

				lectora en 3° grado tanto en Uttar Pradesh como en Madhya Pradesh.
Organizar reuniones con los padres (6° grado) para tratar cuestiones tales como ayudar con la tarea, boletines de calificaciones y otros asuntos relacionados con la escuela.	Francia	Avvisati <i>et al.</i> (2008)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Aumentó la participación de los padres tanto en la escuela como en el hogar. ▪ Los niños de las familias incluidas en el grupo de tratamiento desarrollaron conductas y actitudes más positivas en la escuela y presentaron menor cantidad de problemas de lectoescritura. ▪ Se generaron efectos de derrame considerables hacia los compañeros de los alumnos cuyas familias pertenecían al grupo de tratamiento.
Calificar o clasificar a las escuelas en base al rendimiento estudiantil				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Informes de desempeño de alumnos y escuelas (3° grado) con calificaciones y “rankings” en lengua inglesa, matemática y urdú, usando otras escuelas y aldeas con fines comparativos.	Punjab, Pakistán	Andrabi, Das, y Khwaja (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Mejoraron los índices de aprendizaje en 0,10 desviaciones estándar. ▪ Se constataron logros heterogéneos: los informes generaron una mejora de 0,34 desviaciones estándar en escuelas que inicialmente habían obtenido calificaciones por debajo de la mediana, no se generaron efectos en las escuelas con calificaciones iniciales por encima de la mediana y generó un aumento de 0,10 desviaciones estándar en los puntajes de las escuelas estatales. ▪ Se redujeron 18% los aranceles de las escuelas privadas.
Calificar a las escuelas (3° a 10° grado) mediante una escala que va de la A a la F sobre la base de los índices de competencia en comprensión lectora, escritura y matemática.	Florida	Feng, Figlio, y Sass (2010)	Riguroso (ADD)	<ul style="list-style-type: none"> ▪ Los docentes de las escuelas que obtuvieron calificaciones por debajo de lo esperado tenían un 11% más de probabilidades de abandonar la escuela, mientras que aquellos cuyas escuelas obtuvieron calificaciones por encima de lo esperado tenían un 2,3% menos de probabilidades de irse.

				<ul style="list-style-type: none"> ▪ Los docentes de las escuelas reprobadas presentaron mayores probabilidades de irse de la escuela (42%) y de pasar a trabajar en una escuela del mismo distrito (67%). ▪ La eficacia de los docentes que permanecieron en las escuelas reprobadas aumentó el desempeño de los alumnos en 0,05 desviaciones estándar, pero dado que los docentes que abandonaron estas escuelas presentaban mejores niveles de calidad, el efecto neto sobre la composición del plantel docente fue nulo.
Calificar a las escuelas y asociar los resultados a consecuencias positivas y negativas (4° a 8° grado), incluida la posibilidad del cierre de las escuelas.	Ciudad de Nueva York	Rockoff y Turner (2010)	Riguroso (DRD)	<ul style="list-style-type: none"> ▪ Se incrementaron los logros de los alumnos en matemática y lengua inglesa y mejoraron las evaluaciones de los padres respecto de la calidad de la escuela.
Clasificar a las escuelas en base al desempeño de los alumnos (4° 8° y 10° grado), dando cuenta de su nivel socioeconómico y a través del ofrecimiento de incentivos económicos por buen desempeño.	Chile	Mizala y Urquiola (2007)	Riguroso (DRD)	<ul style="list-style-type: none"> ▪ No se detectó un efecto uniforme en los resultados asociados a los niveles de aprendizaje en las escuelas que obtuvieron la mejor clasificación y que recibieron el incentivo.

Monitorear el esfuerzo de los docentes				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Seguimiento del presentismo docente y remuneración acorde (grados 1–6) mediante el uso de cámaras inviolables y variación de los salarios en función de la asistencia..	Rajastán, India	Duflo, Hanna, y Ryan (2010)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Las escuelas rurales que realizaron un seguimiento de la asistencia de los docentes con cámaras y que implementaron un sistema de remuneración según el nivel de asistencia redujeron el ausentismo entre los docentes del 42% al 23%. ▪ Al cabo de un año, los puntajes de las pruebas en las escuelas de tratamiento resultaron 0,17 desviaciones estándar más altos que en las del grupo de control y aumentó el porcentaje de alumnos que pasaron de grado.
Monitorear el desempeño de los docentes				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Combinación de varios parámetros de eficacia del docente (4° a 8° grado), incluidas las observaciones de clases, las encuestas a alumnos y las encuestas a directores.	Charlotte-Mecklenburg (Carolina del Norte), Dallas (Texas), Denver (Colorado), condado de Hillsborough (Florida), ciudad de Nueva York y Memphis (Tennessee)	Kane <i>et al.</i> (2013)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Mediante un índice compuesto de efectividad docente que incorporaba los resultados de las encuestas a alumnos, las observaciones de clases y el historial de las mejoras de desempeño de los alumnos del docente en las pruebas de los estados, se identificó la magnitud de la mejora del desempeño de los alumnos resultante de la eficacia docentes luego de la asignación aleatoria de docentes a los cursos. ▪ El grado de mejora de los logros obtenidos por los docentes identificados como eficaces antes de la asignación aleatoria de los cursos coincidió con su desempeño según el índice compuesto. ▪ Los alumnos de un subconjunto de docentes identificados como eficaces antes de la asignación aleatoria de los cursos también lograron un mejor desempeño en las pruebas “de auditoría” de matemática y comprensión lectora.

Fuente: Elaborado por los autores.

Nota: ADD = análisis de diferencias en diferencias; EAC = estudio de asignación aleatoria; DRD = diseño de regresión discontinua.

Cuadro 7. Datos sobre las intervenciones destinadas a apoyar a los docentes para que mejoren la enseñanza

Proporcionar o mejorar el material didáctico				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Suministrar computadoras a escuelas públicas (3° a 9° grado) para integrarlas a la enseñanza de la lectura	Colombia	Barrera-Osorio y Linden (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Se constató escasa repercusión en los puntajes obtenido por los alumnos en las pruebas y otros efectos. No se detectaron diferencias en los resultados en términos de grado, materia o género de los alumnos. ▪ Según encuestas, pocos docentes realmente incorporaron las computadoras a la planificación de sus clases.
Suministrar computadoras y software de aprendizaje adaptativo (4° grado) para matemática en y después de la escuela (2 horas por semana) con instructores locales.	Vadodara, India	Banerjee <i>et al.</i> (2007)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Se constató una mejora de 0,36 desviaciones estándar en las calificaciones de matemática, pero ningún efecto en el desempeño en comprensión lectora, lo que sugiere efectos limitados en términos de derrame.
Suministrar software de lengua/lectura (3° a 6°).	Estados Unidos	Rouse y Krueger (2004)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Mejoraron algunos aspectos de las habilidades de aprendizaje de los alumnos, pero estas mejoras no se tradujeron en mayores logros en cuanto a la adquisición del lenguaje o la comprensión lectora.
Incorporar bibliotecas (1° a 6° grado) a través de un sistema de escuelas “proveedoras y receptoras” y bibliotecarios capacitados que ayudan a los alumnos.	Bangalore, India	Borkum, He, y Linden (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ El programa no logró aumentar la competencia en términos de lengua.
Suministrar libros de texto (3° a 8° grado) de lengua inglesa, matemática y ciencia.	Busia y Teso, Kenia	Glewwe, Kremer y Moulin (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ No se hallaron efectos en los puntajes de las pruebas de los alumnos ni en el nivel de presentismo. ▪ Los puntajes aumentaron 0,06 desviaciones estándar en el caso de alumnos de alto rendimiento, pero pocos alumnos de los grados inferiores podían leer los textos (en inglés) (15–29% de los estudiantes de la mediana).

Suministrar rotafolios (3° a 8° grado) con gráficos de ciencia, manuales para docentes, gráficos sobre salud, gráficos para matemática y mapas.	Busia y Teso, Kenia	Glewwe <i>et al.</i> (2004)	Muy riguroso (EAC)	<ul style="list-style-type: none"> No se comprobaron efectos en los puntajes de los alumnos.
Ofrecer formación continua				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Proporcionar formación continua (2° a 8° grado) a través de financiamiento para docentes de escuelas de bajo rendimiento.	Chicago	Jacob y Lefgren (2004b)	Riguroso (DRD)	<ul style="list-style-type: none"> No se detectó efecto alguno en los logros de matemática o comprensión lectora.
Proporcionar formación continua (4° grado) en forma semanal, en la escuela y a cargo de instructores externos; focalizada en la enseñanza de hebreo, matemática e inglés.	Jerusalén, Israel	Angrist y Lavy (2001)	Riguroso (ADD + correspondencia)	<ul style="list-style-type: none"> La capacitación para los docentes en ejercicio en escuela seculares genera una mejora del desempeño de los alumnos de 0,20–0,40 desviaciones estándar. No se pudieron definir claramente los resultados en el caso de las escuelas religiosas, quizá porque la capacitación en estas instituciones comenzó más tarde y se implementó a menor escala.

Fuente: Elaborado por los autores.

Nota: ADD= análisis de diferencias en diferencias; EAC = estudio de asignación aleatoria.

Cuadro 8. Datos sobre las intervenciones destinadas a motivar a los docentes en su desempeño

Contratar docentes temporarios				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Contratar docentes temporarios (1° a 5° grado) seleccionados por comités escolares, con contratos anuales, sin beneficios laborales y con un salario más bajo.	Andhra Pradesh, India	Muralidharan y Sundararaman (2010)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Los puntajes de las pruebas fueron 0,15 desviaciones estándar más altos en matemática y 0,13 desviaciones estándar más altos en comprensión lectora. ▪ Los docentes por contrato tuvieron menos inasistencias que los de planta permanente (16% versus 27%). ▪ Los docentes por contrato dedicaron 49% del tiempo a enseñar, mientras que los docentes de planta permanente, 43%. ▪ Si bien los alumnos de los docentes por contrato y los de los docentes de planta permanente demostraron mejoras similares en cuanto a logros académicos, los docentes por contrato representaron una quinta parte del costo de los docentes de planta permanente.
Contratar docentes temporarios (3° a 4° grado) que sacan a los alumnos del aula para darles clases de apoyo.	Mumbai y Vadodara, India	Banerjee <i>et al.</i> (2007)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Los puntajes de las evaluaciones de los alumnos mejoraron 0,14 desviaciones estándar al cabo del primer año y 0,28 al cabo del segundo en relación con los puntajes obtenidos por los alumnos del grupo de control. ▪ Las escuelas incluidas en el grupo de tratamiento obtuvieron puntajes 0,10 desviaciones estándar más altos incluso un año después de finalizado el programa. ▪ Los niños con las calificaciones más bajas en los exámenes iniciales y los que recibían clases de apoyo fueron los que obtuvieron mayores beneficios.

Remunerar a los docentes por mejorar el desempeño de los alumnos				
Intervención	Ubicación geográfica	Estudio	Método	Resultados
Bonificaciones por desempeño y recursos (3° a 12° grado) por asistencia, ambiente y progreso del desempeño estudiantil; las escuelas deciden cómo distribuir las bonificaciones.	Ciudad de Nueva York	Fryer (2011)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ No mejoró el desempeño de los alumnos. ▪ No se encontró evidencia de que los incentivos a los docentes hayan influido en los índices de retención o ausentismo de los docentes ni en el entorno de aprendizaje.
Bonificaciones por desempeño (5° a 8° grado) por contribuir a alcanzar un nivel mínimo de desempeño en matemática.	Nashville, Tennessee	Springer <i>et al.</i> (2010)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Los alumnos de escuelas medias asignados a docentes con derecho a recibir el incentivo por mejoras en el desempeño de los alumnos mostraron un desempeño en promedio similar al de los alumnos de docentes sin derecho a percibir el incentivo. ▪ Los alumnos de 5° grado de los docentes con derecho a bonificación obtuvieron un desempeño superior al de sus pares en el segundo y el tercer año del estudio; este efecto no se mantuvo una vez que los alumnos pasaron al siguiente grado.
Bonificaciones por desempeño individuales y grupales (2° a 5° grado) por contribuir a alcanzar un umbral mínimo y por aumentos adicionales.	Andhra Pradesh, India	Muralidharan y Sundararaman (2009)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ El programa de incentivos mejoró el desempeño de los alumnos 0,28 desviaciones estándar en matemática y 0,16 en comprensión lectora al cabo de dos años. ▪ Las escuelas del grupo de tratamiento obtuvieron un mejor desempeño en los componentes mecánico y conceptual de las evaluaciones, lo que sugiere que no se verificaron consecuencias adversas. ▪ Los alumnos de las escuelas del grupo de tratamiento también alcanzaron un mejor desempeño en las materias que no estaban alcanzadas por los incentivos, tales como ciencia y estudios sociales, lo que indica que se produjeron efectos de derrame positivos. ▪ Los resultados de los incentivos individuales y

				grupales no registraron diferencias estadísticamente significativas durante los dos primeros años del programa.
Bonificaciones grupales por desempeño (4° a 8° grado) según “ranking” de desempeño y mejora.	Busia y Teso, Kenia	Glewwe, Ilias, y Kremer (2010)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ Las escuelas cuyos comités escolares otorgaban bonificaciones a los docentes cuyos alumnos obtenían buenos resultados en las pruebas obtuvieron puntajes 0,14 desviaciones estándar más altos que las escuelas de control al cabo del primer año. ▪ El programa no repercutió en el nivel de presentismo de los docentes, en la cantidad de tarea para el hogar asignada ni en los métodos pedagógicos utilizados. ▪ El programa aumentó el nivel de preparación para las evaluaciones en 4,2% el primer año y en 7,4% el segundo.
Bonificaciones grupales por desempeño e insumos (4°, 8° y 10° grado) basadas en los niveles de aprendizaje de los alumnos y en los insumos docentes o escolares.	Chile	Rau y Contreras (2009)	Riguroso (ADD + CPP + DRD)	<ul style="list-style-type: none"> ▪ Se elevó entre 0,7 y 0,12 desviaciones estándar el promedio de los resultados de aprendizaje. ▪ No se halló evidencia de que el hecho de obtener la bonificación generara algún tipo de mejora adicional.
Bonificaciones individuales por desempeño (10° a 12° grado) basadas en las calificaciones con las que los alumnos aprobaron las pruebas y los promedios de los exámenes de ingreso a la secundaria.	Israel	Lavy (2008, 2009)	Riguroso (DRD + correspondencia)	<ul style="list-style-type: none"> ▪ Los alumnos de las escuelas incluidas en el grupo de tratamiento aprobaron con calificaciones superiores en un 14% y obtuvieron puntajes 10% más altos en matemática, mientras que en lengua inglesa aprobaron con calificaciones superiores en un 5% y obtuvieron puntajes 4% más altos. ▪ Las bonificaciones no generaron diferencias en términos de género, aunque las mujeres se mostraron más pesimistas respecto de la probabilidad de obtenerlas.
Bonificaciones individuales por desempeño e insumos (3° a 6° grado), teniendo en cuenta título, experiencia, capacitación, evaluación de pares, conocimientos y desempeño estudiantil.	México	McEwan y Santibáñez (2005)	Riguroso (DRD)	<ul style="list-style-type: none"> ▪ No se comprobó que la reforma de remuneración basada en el desempeño haya mejorado los puntajes de los alumnos en las pruebas.

Bonificaciones grupales por desempeño y otros insumos (9° a 11° grado), según “ranking” de escuelas basado en créditos, índices de graduación y deserción y logros de los alumnos.	Israel	Lavy (2002)	Riguroso (DRD)	<ul style="list-style-type: none"> ▪ Mejoró 0,13 desviaciones estándar el desempeño de los alumnos y se incrementaron los créditos obtenidos y los porcentajes de alumnos que rinden exámenes de ingreso.
Recompensa individual con recursos (preescolar), incluye ausentismo; los directores deciden cómo asignar las bonificaciones con los recursos que recibieron.	Busia y Teso, Kenia	Kremer <i>et al.</i> (2001)	Muy riguroso (EAC)	<ul style="list-style-type: none"> ▪ No se constató impacto alguno en el nivel de ausentismo de los docentes, el presentismo de los alumnos ni el desempeño.

Fuente: Elaborado por los autores.

Nota: ADD = análisis de diferencias en diferencias; CPP = correspondencia de puntaje de propensión; EAC = estudio de asignación aleatoria; DRD = diseño de regresión discontinua.

Bibliografía

- Abdulkadiroglu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane y P. A. Pathak. 2011. “Accountability and Flexibility in Public Schools: Evidence from Boston’s Charters and Pilots”, en: *Quarterly Journal of Economics* 126: 699–748.
- Abeberese, A. B., T. J. Kumler y L. L. Linden. 2011. *Improving Reading Skills by Encouraging Children to Read: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines*. Cambridge, MA: Innovations for Poverty Action (IPA).
- Acemoglu, D. y J.-S. Pischke. 1998. “Why Do Firms Train? Theory and Evidence” en: *Quarterly Journal of Economics* 113(1): 79–119.
- Akerlof, G. A. 1970. “The Market for ‘Lemons.’ Quality Uncertainty and the Market Mechanism”, en: *Quarterly Journal of Economics* 84(3): 488–500.
- 1982. “Labor Contracts as Partial Gift Exchange”, en: *Quarterly Journal of Economics* 97(4): 543–569.
- Alcázar, L., F. H. Rogers, N. Chaudhury, J. Hammer, M. Kremer y K. Muralidharan. 2006. “Why Are Teachers Absent? Probing Service Delivery in Peruvian Primary Schools”, en: *International Journal of Educational Research* 45:117–136.
- Alfonso, M., A. Santiago y M. Bassi. 2011. *Estimating the Impact of Placing Top University Graduates in Vulnerable Schools in Chile*. Washington DC: Banco Interamericano de Desarrollo.
- Andrabi, T., J. Das y A. Khwaja. 2009. *Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets*. Washington DC: Banco Mundial.
- Angrist, J. D. y J. Guryan. 2008. “Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements”, en: *Economics of Education Review* 27(5): 483–503.
- Angrist, J. D. y V. Lavy. 1999. “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement”, en: *Quarterly Journal of Economics* 114(2): 533–575.
- 2001. “Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools”, en: *Journal of Labor Economics* 19(2): 343–369.
- Arrow, K. J. 1963. “Uncertainty and the Welfare Economics of Medical Care”, en: *American Economic Review* 53(5): 941–973

- Autor, D. H. 2003a. “Lecture Note: Self-Selection – The Roy Model”. MIT 14.661 (14 de noviembre). Massachusetts Institute of Technology.
- , 2003b. “Lecture Note: Monitoring, Measurement and Risk”. MIT 14.661. (13 de noviembre). Massachusetts Institute of Technology.
- Avvisati, F., M. Gurgand, N. Guyon y E. Maurin. 2008. *Quels Effets Attendre d'une Politique d'Implication des Parents d'Élèves dans les Collèges? Les Enseignements d'une Expérimentation Contrôlée*. París: Ecole d'Economie de Paris.
- Baker, G. 1992. “Incentive Contracts and Performance Measurement”, en: *Journal of Political Economy* 100: 598–614.
- Banerjee, A. V., R. Banerji, E. Duflo y M. Walton. 2011. *What Helps Children to Learn? Evaluation of Pratham's Read India Program in Bihar and Uttarakhand*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Banerjee, A. V., S. Cole, E. Duflo y L. Linden. 2007. “Remedying Education: Evidence from Two Randomized Experiments in India”, en: *Quarterly Journal of Economics* 122(3): 1235–1264.
- Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster y S. Khemani. 2010. “Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India”, en: *American Economic Journal* 2(1): 1–30.
- Barlevy, Gadi y Derek Neal. 2011. “Pay for Percentile”. Documento de trabajo del NBER N° 17.194. Cambridge, MA: National Bureau of Economic Research.
- Barrera-Osorio, F. y L. L. Linden. 2009. *The Use and Misuse of Computers in Education: Evidence from a Randomized Controlled Trial of a Language Arts Program*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Becker, G. S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.
- Becker, G. S. y K. Murphy. 2000. *Social Economics*. Cambridge, MA: Harvard University Press.
- Bénabou, R. y J. Tirole. 2003. “Intrinsic and Extrinsic Motivation”, en: *Review of Economic Studies* 70(3): 489–520.
- Borjas, G. 1987. “Self-Selection and the Earnings of Immigrants”, en: *American Economic Review* 77: 531–553.

- Borkum, E., F. He y L. Linden. 2009. *School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Borman, G., R. E. Slavin, A. Cheung, A. Chamberlain, N. A. Madden y B. Chambers. 2007. “Final Reading Outcomes of the National Randomized Field Trial of Success for All”, en: *American Educational Research Journal* 44(3): 701–731.
- Boyd, D., H. Lankford y S. Loeb. 2005. “The Draw of Home: How Teachers’ Preferences for Proximity Disadvantage Urban Schools”, en: *Journal of Policy Analysis and Management* 24(1): 113–132.
- Boyd, D., H. Lankford, S. Loeb y J. Wyckoff. 2011. “Teacher Layoffs: An Empirical Illustration of Seniority versus Measures of Effectiveness”, en: *Education Finance and Policy* 6(3): 439–454.
- Boyd, D., P. L. Grossman, H. Lankford, S. Loeb y J. Wyckoff. 2009. “Teacher Preparation and Student Achievement”, en: *Educational Evaluation and Policy Analysis* 31(4): 416–440.
- Boyd, D., P. Grossman, K. Hammerness, H. Lankford, S. Loeb, M. Ronfeldt y J. Wyckoff. 2010a. “Recruiting Effective Math Teachers: How Do Math Immersion Teachers Compare? Evidence from New York City”. Documento de trabajo del NBER N° 16.017. Cambridge, MA: National Bureau of Economic Research.
- Boyd, D., H. Lankford, S. Loeb, M. Ronfeldt y J. Wyckoff. 2010b. “The Role of Teacher Quality in Retention and Hiring: Using Applications-to-Transfer to Uncover Preferences of Teachers and Schools”. Documento de trabajo del NBER N° 15.966. Cambridge, MA: National Bureau of Economic Research.
- Boyd, D., P. Grossman, M. Ing, H. Lankford, S. Loeb, R. O’Brien y J. Wyckoff. 2011. “The Effectiveness and Retention of Teachers with Prior Career Experience”, en: *Economics of Education Review* 30: 1229–1241.
- Bruns, B., D. Filmer y H. A. Patrinos. 2011. *Making Schools Work: New Evidence on Accountability Reforms*. Washington DC: Banco Mundial.
- Bruns, B. y L. Santibáñez. 2011. “Making Teachers Accountable”, en: B. Bruns, D. Filmer y H. A. Patrinos (Eds.), *Making Schools Work: New Evidence on Accountability Reforms*. Washington DC: Banco Mundial.
- Cantrell, S. y T. J. Kane. 2013. *Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project’s Three-Year Study*. Seattle: Bill and Melinda Gates Foundation.

- Cantrell, S., J. Fullerton, T. J. Kane y D. O. Staiger. 2008. “National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment”. Documento de trabajo del NBER N° 14.608. Cambridge, MA: National Bureau of Economic Research.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach y D. Yagan. 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star”, en: *Quarterly Journal of Economics* 126(4): 1593–1660.
- Clark, D., P. Martorell y J. Rockoff. 2009. “School Principal and School Performance”. Documento de trabajo de CALDER N° 38. Washington DC: National Center for Analysis of Longitudinal Data in Educational Research (CALDER), the Urban Institute.
- Clotfelter, C. T., H. F. Ladd y J. L. Vigdor. 2007a. “Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects”, en: *Economics of Education* 26(6): 673–682.
- 2007b. “Are Teacher Absences Worth Worrying About in the U.S.?”. Documento de trabajo del NBER N° 13,648. Cambridge, MA: National Bureau of Economic Research.
- Clotfelter, C., E. Glennie, H. Ladd y J. Vigdor. 2008. “Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina”, en: *Journal of Public Economics* 92(5-6): 1352–1370.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence.
- Constantine, J., D. Player, T. Silva, K. Hallgreen, M. Grider y J. Deke. 2009. *An Evaluation of Teachers through Different Routes to Certification: Final Report*. Washington DC: Mathematica Policy Research/Institute of Education Sciences.
- Das, J., S. Dercon, J. Habyarimana y P. Krishnan. 2007. “Teacher Shocks and Student Learning: Evidence from Zambia”, en: *Journal of Human Resources* 42(4): 820–862.
- Decker, P. T., D. P. Mayer y S. Glazerman. 2004. *The Effects of Teach for America on Students: Findings from a National Evaluation*. Princeton, NJ: Mathematica Policy Research, Inc.
- Dee, T. S. 2004. “Teachers, Race and Student Achievement in a Randomized Experiment”, en: *The Review of Economics and Statistics* 86(1): 195–210.
- 2007. “Teachers and the Gender Gaps in Student Achievement”, en: *Journal of Human Resources* 42(3): 528–554.

- Deming, D. J. 2011. “Better Schools, Less Crime?”, en: *Quarterly Journal of Economics* 126(4): 2063–2115.
- Dhaliwal, I., E. Duflo, R. Glennerster y C. Tulloch. 2011. *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Dobbie, W. y R. G. Fryer. 2009. “Are High Quality Schools Enough to Close the Achievement Gap? Evidence from a Social Experiment in Harlem”. Documento de trabajo del NBER N° 15.473. Cambridge, MA: National Bureau of Economic Research.
- , 2011. “The Impact of Youth Service on Future Outcomes: Evidence from Teach for America”. Documento de trabajo del NBER N° 17.402. Cambridge, MA: National Bureau of Economic Research.
- Duflo, E., P. Dupas y M. Kremer. 2007. *Peer Effects, Pupil-Teacher Ratios, and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- , 2009. *Additional Resources versus Organizational Changes in Education: Experimental Evidence from Kenya*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Duflo, E., R. Hanna y S. Ryan. 2010. *Incentives Work: Getting Teachers to Come to School*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Feng, L., D. N. Figlio y T. Sass. 2010. “School Accountability and Teacher Mobility”. Documento de trabajo del NBER N° 16.070. Cambridge, MA: National Bureau of Economic Research.
- Ferraz, C. y B. Bruns. De próxima publicación. “Incentives to Teach: The Effects of Performance Pay in Brazilian Schools”. Washington DC: Banco Mundial.
- Fryer, R. G. 2011. “Teacher Incentives and Student Achievement: Evidence from New York City Public Schools”. Documento de trabajo del NBER N° 16.850. Cambridge, MA: National Bureau of Economic Research.
- Ganimian, A. J. y E. Vegas. 2011. “What Are the Different Profiles of Successful Teacher Policy Systems?”. Documento de información del proyecto SABER-Docentes N° 5. Washington DC: Banco Mundial.

- Gibbons, R. y M. Waldman. 1999. "Careers in Organizations: Theory and Evidence", en: O. Ashenfelter y D. Card (Eds.), *Handbook of Labor Economics*, Vol. 3. Amsterdam: Elsevier.
- Glazerman, S. y A. Seifullah. 2012. *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) After Four Years: Final Report*. Washington DC: Mathematica Policy Research.
- Glazerman, S., E. Isenberg, S. Dolfin, M. Bleeker, A. Johnson, M. Grider y M. Jacobus. 2010. *Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study*. Washington DC: Mathematica Policy Research/Institute of Education Sciences.
- Glewwe, P. y E. Maïga. 2011. *The Impacts of School Management Reforms in Madagascar: Do the Impacts Vary by Teacher Type?* Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Glewwe, P., N. Ilias y M. Kremer. 2010. "Teacher Incentives", en: *American Economic Journal: Applied Economics* 2(3): 205–227.
- Glewwe, P., M. Kremer y S. Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya", en: *American Economic Journal: Applied Economics* 1(1): 112–135.
- Glewwe, P., M. Kremer, S. Moulin y E. Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya", en: *Journal of Development Economics* 74: 251–268.
- Goodman, S. y L. Turner. 2010. "Teacher Incentive Pay and Educational Outcomes: Evidence from the NYC Bonus Program". Documento preparado para la conferencia "Merit Pay: Will It Work? Is It Politically Viable?", Programa de Política Educativa y Gobierno, Escuela de Gobierno John F. Kennedy de la Universidad de Harvard, 3 y 4 de junio.
- Greenwald, B. C. 1986. "Adverse Selection in the Labour Market", en: *Review of Economic Studies* 53: 325–347.
- Grissom, J. A. y S. Loeb. 2011. "Triangulating Principal Effectiveness: How Perspectives of Parents, Teachers and Assistant Principals Identify the Central Importance of Managerial Skills", en: *American Educational Research Journal* 48(5): 1–33.
- Hanushek, E. A. 2003. "The Failure of Input-Based Schooling Policies", en: *The Economic Journal* 113: F64–F98.
- Hanushek, E. A. y L. Woessmann. 2007. *Education Quality and Economic Growth*. Washington DC: Banco Mundial.

- Hanushek, E. A., J. F. Kain, D. M. O'Brien y S. G. Rivkin. 2005. "The Market for Teacher Quality". Documento de trabajo del NBER N° 11.154. Cambridge, MA: National Bureau of Economic Research.
- He, F., L. L. Linden y M. MacLeod. 2007. *Helping Teach What Teachers Don't Know: An Assessment of the Pratham English Language Program*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- , 2009. *A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).
- Hicks, J. 1939. "The Foundations of Welfare Economics", en: *The Economic Journal* 49(196): 696–712.
- Holmstrom, B. 1979. "Moral Hazard and Observability", en: *Bell Journal of Economics* 9: 74-91.
- Holmstrom, B. y P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design", en: *Journal of Law, Economics and Organization* 7: 24–52.
- Horng, E. L., D. Klasik y S. Loeb. 2010. "Principal's Time Use and School Effectiveness", en: *American Journal of Education* 116: 491–523.
- Hoxby, C. M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation", en: *Quarterly Journal of Economics* 115(4): 1239–1285.
- Hoxby, C. M. y A. Leigh. 2004. "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States", en: *American Economic Journal* 93(2): 236–240.
- Jackson, C. K. 2013. "Match Quality, Worker Productivity and Worker Mobility: Direct Evidence from Teachers", en: *Review of Economics and Statistics*.
- Jackson, C. K. y E. Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers", en: *American Economic Journal: Applied Economics* 1(4): 85–108.
- Jacob, B. A. 2010a. "The Effect of Employment Protection on Worker Effort: Evidence from Public Schooling". Documento de trabajo del NBER N° 15.655. Cambridge, MA: National Bureau of Economic Research.

- , 2010b. “Do Principals Fire the Worst Teachers?”. Documento de trabajo del NBER N° 15.715. Cambridge, MA: National Bureau of Economic Research.
- Jacob, B. A. y L. Lefgren. 2004a. “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis”, en: *Review of Economics and Statistics* 86(1): 226–244.
- , 2004b. “The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago”, en: *Journal of Human Resources* 39(1): 50–79.
- , 2005. “Principals as Agents: Subjective Performance Measurement in Education”. Documento de trabajo del NBER N° 11.463. Cambridge, MA: National Bureau of Economic Research.
- , 2007. “What Do Parents Value in Education? An Empirical Investigation of Parents’ Revealed Preferences for Teachers”, en: *Quarterly Journal of Economics* 122(4): 1603–1637.
- Jacob, B. A. y S. D. Levitt. 2003. “Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating”, en: *Quarterly Journal of Economics* 118(3): 843–878.
- Johnson, W. 1978. “A Theory of Job Shopping”, en: *Quarterly Journal of Economics* 92: 261–277.
- Kaldor, N. 1939. “Welfare Propositions in Economics and Interpersonal Comparisons of Utility”, en: *The Economic Journal* 49(195): 549–552.
- Kane, T. J., D. F. McCaffrey, T. Millery D. O. Staiger. 2013. *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Seattle: Bill and Melinda Gates Foundation.
- Kane, T. J. y D. O. Staiger. 2010. *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*. Seattle: Bill and Melinda Gates Foundation.
- , 2012. *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*. Seattle: Bill and Melinda Gates Foundation.
- Kane, T. J., J. E. Rockoff y D. O. Staiger. 2006. “What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City”. Documento de trabajo del NBER N° 12.155. Cambridge, MA: National Bureau of Economic Research.

- Kane, T. J., E. S. Taylor, J. H. Tyler y A. L. Wooten. 2010. “Identifying Effective Classroom Practices Using Student Achievement Data”. Documento de trabajo del NBER N° 15.803. Cambridge, MA: National Bureau of Economic Research.
- Kemple, J. J. y C. J. Willner. 2008. *Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment and Transitions to Adulthood*. Nueva York: MDRC.
- Kremer, M. E., D. Chen, P. Glewwe y S. Moulin. 2001. *Interim Report on a Teacher Incentive Program in Kenya*. Cambridge, MA: Harvard University.
- Krueger, A. 1999. “Experimental Estimates of Education Production Functions”, en: *Quarterly Journal of Economics* 114: 497–532.
- Lang, K. y W. T. Dickens. 1987. “Neoclassical and Sociological Perspectives on Segmented Labor Markets”. Documento de trabajo del NBER N° 2.127. Cambridge, MA: National Bureau of Economic Research.
- Lassibille, G., J.-P. Tan, C. Jesse y T. V. Nguyen. 2010. “Managing for Results in Primary Education in Madagascar: Evaluating the Impact of Selected Workflow Interventions”, en: *The World Bank Economic Review* 24(2): 303–329.
- Lavy, V. 2002. “Evaluating the Effect of Teachers’ Group Performance Incentives on Pupil Achievement”, en: *The Journal of Political Economy* 110(6): 1286–1317.
- 2008. “Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Pay Tournaments among Teachers”. Documento preparado para la conferencia “Economic Incentives: Do They Work in Education?”, Programa de Política Educativa y Gobierno, Escuela de Gobierno John F. Kennedy de la Universidad de Harvard, 16y 17 de mayo.
- 2009. “Performance Pay and Teachers’ Effort, Productivity, and Grading Ethics”, en: *The American Economic Review* 99(5): 1979—2011.
- 2010. “Do Differences in School’s Instruction Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries”. Documento de trabajo del NBER N° 16.227. Cambridge, MA: National Bureau of Economic Research.
- Lazear, E. P. y S. Rosen. 1981. “Rank-Order Tournaments as Optimum Labor Contracts”, en: *Journal of Political Economy* 89(5): 841–864.
- Linden, L. L., C. Herrera y J. Grossman. 2011. *Achieving Academic Success After School: A Randomized Evaluation of the Higher Achievement Program*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL).

- Lucas, R. E. y E. C. Prescott. 1974. “Equilibrium Search and Unemployment”, en: *Journal of Economic Theory* 7(2): 188–209.
- McEwan, P. y L. Santibáñez. 2005. “Teacher and Principal Incentives in Mexico”, en: E. Vegas (Ed.), *Incentives to Improve Teaching: Lessons from Latin America*. Washington DC: Banco Mundial.
- Metzler, J. y L. Woessmann. 2010. “The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation”. IZA DP N° 4.999. Bonn, Alemania: Institute for the Study of Labor (IZA).
- Mihaly, K., D. F. McCaffrey, D. O. Staiger y J. R. Lockwood. 2013. *A Composite Estimator of Effective Teaching*. Seattle: Bill and Melinda Gates Foundation.
- Miller, R. T., R. J. Murnane y J. B. Willett. 2007. “Do Teacher Absences Impact Student Achievement? Longitudinal Evidence from One Urban School District”. Documento de trabajo del NBER N° 13.356. Cambridge, MA: National Bureau of Economic Research.
- Mincer, J. 1962. “On-the-Job Training: Costs, Returns and Some Implications”, en: *Journal of Political Economy* 70: 50–79.
- Mirrlees, J. 1976. “The Optimal Structure of Incentives and Authority within an Organization”, en: *Bell Journal of Economics* 7: 105–131.
- Mizala, A. y M. Urquiola. 2007. “School Markets: The Impact of Information Approximating Schools’ Effectiveness”. Documento de trabajo del NBER N° 13.676. Cambridge, MA: National Bureau of Economic Research.
- Montgomery, J. 1991. “Equilibrium Wage Dispersion and Interindustry Wage Differentials”, en: *Quarterly Journal of Economics* 106: 163–179.
- Muralidharan, K. y V. Sundararaman. 2009. “Teacher Performance Pay: Experimental Evidence from India”. Documento de trabajo del NBER N° 15.323. Cambridge, MA: National Bureau of Economic Research.
- , 2010. *Contract Teachers: Experimental Evidence from India*. Washington DC: Banco Mundial.
- Murnane, R. J. 2010. “Progress and Puzzles in Educational Policy Research”, en: *Harvard Education Letter* 26(2).
- Murnane, R. J. y J. B. Willett. 2011. *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Nueva York: Oxford University Press.

- Myung, J., S. Loeb y E. L. Horng. 2011. “Tapping the Principal Pipeline: Identifying Talent for Future School Leadership in the Absence of Formal Succession Management Programs”, en: *Educational Administration Quarterly* 47(5): 695–727.
- Nelson, R. R. y E. S. Phelps. 1966. “Investment in Humans, Technological Diffusion and Economic Growth” en: *The American Economic Review* 56(1/2): 69–75.
- Osterman, P. 1984. *Internal Labor Markets*. Cambridge, MA: MIT Press.
- Pandey, P., S. Goyal y V. Sundararaman. 2009. “Community Participation in Public Schools: Impact of Information Campaigns in Three Indian States”, en: *Education Economics* 17(3): 355-375.
- Papay, J. P., M. R. West, J. B. Fullerton y T. J. Kane. 2011. “Does Practice-Based Teacher Preparation Increase Student Achievement? Early Evidence from the Boston Teacher Residency”. Documento de trabajo del NBER N° 17.646. Cambridge, MA: National Bureau of Economic Research.
- Pissarides, C. A. 1979. “Job Matching with State Employment Agencies and Random Search”, en: *The Economic Journal* 89(356): 818–833.
- Prendergast, C. 1993. The Role of Promotion in Inducing Specific Human Capital Acquisition. *Quarterly Journal of Economics* 108(2): 523–34
- Prendergast, C. 2002. The Tenuous Tradeoff between Risk and Incentives. *Journal of Political Economy*, 110(5): 1071–102.
- Rau, T. y D. Contreras. 2009. “Tournaments, Gift Exchanges, and the Effect of Monetary Incentives for Teachers: The Case of Chile”. Documento de trabajo del Ministerio de Economía N° 305. Santiago: Universidad de Chile.
- Reardon, S. F. 2011. “The Widening Academic Gap between the Rich and the Poor: New Evidence and Possible Explanations”, en G. Duncan y R. J. Murnane (Eds.): *Whither Opportunity? Rising Inequality, Schools and Children’s Life Chance*. Nueva York: Russell Sage Foundation and Spencer Foundation.
- Reich, M., D. M. Gordon y R. C. Edwards. 1973. “A Theory of Labor Market Segmentation”, en: *American Economic Review* 63(2): 359–365.
- Rockoff, J. E. 2008. “Does Mentoring Reduce Turnover and Improve Skills of New Employees? Evidence from Teachers in New York City”. G. Duncan, G. and R.J. Murnane Documento de trabajo del NBER N° 13.868. Cambridge, MA: National Bureau of Economic Research.

- Rockoff, J. E. y L. J. Turner. 2010. “Short Run Impacts of Accountability on School Quality”, en: *American Economic Journal: Economic Policy* 2(4): 119–147.
- Rockoff, J. E., B. A. Jacob, T. J. Kane y D. O. Staiger. 2008. “Can You Recognize an Effective Teacher When You Recruit One?”. Documento de trabajo del NBER N° 14.485. Cambridge, MA: National Bureau of Economic Research.
- Rockoff, J. E., D. O. Staiger, T. J. Kane y E. S. Taylor. 2011. *Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools*. Nueva York: Columbia Business School.
- Rosen, S. 1978. “Substitution and Division of Labor”, en: *Económica* 45: 235–250.
- Ross, S. 1973. “The Economic Theory of Agency: The Principal’s Problem”, en: *American Economic Review* 63(2): 134–139.
- Rouse, C. E. y A. B. Krueger. 2004. “Putting Computerized Instruction to the Test: A Randomized Evaluation of a ‘Scientifically Based’ Reading Program”, en: *Economics of Education Review* 23(4): 323–338.
- Roy, A. 1951. “Some Thoughts on the Distribution of Earnings”, en: *Oxford Economic Papers* 235–246.
- Sattinger, M. 1975. “Comparative Advantage and the Distributions of Earnings and Abilities”, en: *Econometrica* 43: 455–468.
- Schultz, T. W. 1963. *The Economic Value of Education*. Nueva York, NY: Columbia University Press.
- Shapiro, C. y J. Stiglitz. 1984. “Equilibrium Unemployment as a Worker Discipline Device”, en: *American Economic Review* 74: 433–444.
- Shavell, S. 1979. “Risk Sharing and Incentives in the Principal and Agent Relationship”, en: *Bell Journal of Economics* 10: 55–73.
- Spence, A. M. 1974. *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge: Harvard University Press.
- Springer, M. G., D. Ballou, L. Hamilton, V.-N. Le, J. R. Lockwood, D. McCaffrey, M. Pepper y B. M. Stecher. 2010. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*. Pittsburgh: The RAND Corporation.
- Steele, J. L., R. J. Murnane y J. B. Willett. 2010. “Do Financial Incentives Help Low-Performing Schools Attract and Keep Academically Talented Teachers? Evidence

- from California”, en: *Journal of Policy Analysis and Management* 29(3): 451–458.
- Stiglitz, J. 1975. “The Theory of Screening, Education and the Distribution of Income”, en: *American Economic Review* 65: 283–300.
- Suryadarma, D., A. Suryahadi, S. Sumarto y F. H. Rogers. 2006. “Improving Student Performance in Public Primary Schools in Developing Countries: Evidence from Indonesia”, en: *Education Economics* 14(4): 401-429.
- Taylor, E. S. y J. H. Tyler. 2011. “The Effect of Evaluation on Performance: Evidence from Longitudinal Student Achievement Data of Mid-Career Teachers”. Documento de trabajo del NBER N° 16.877. Cambridge, MA: National Bureau of Economic Research.
- Terviö, M. 2003. *Mediocrity in Talent Markets*. Berkeley, CA: Haas School of Business, University of California, Berkeley.
- Tyler, J. H. 2011. “If You Build It, Will They Come? Teacher Use of Student Performance Data on a Web-Based Tool”. Documento de trabajo del NBER N° 17.486. Cambridge, MA: National Bureau of Economic Research.
- Urquiola, M. y E. Verhoogen. 2009. “Class-Size Caps, Sorting and the Regression Discontinuity Design”, en: *American Economic Review* 99(1): 179–215.
- Vegas, E., A. J. Ganimian, N. Goldstein, A. Paglayan, S. Loeb y P. Romaguera. 2011. “What Are Teacher Policy Goals, How Can Education Systems Reach Them and How Will We Know When They Do?”. Documento de información del proyecto SABER-Doctentes N° 2. Washington DC: Banco Mundial.
- Williamson, O. 1975. “Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange”, en: *Bell Journal of Economics* 16: 250–280.