

## CAPÍTULO 5

# Incentivos para influir en la calidad y distribución de los maestros

---

¿Por qué centrarse en los maestros?.....	136
Tabuladores de sueldo e incentivos actuales.....	138
Usar incentivos: Factores técnicos y prácticos.....	141
Evidencia acerca de sistemas de incentivos para maestros para fines determinados.....	153
Conclusión: Infraestructura para apoyar los sistemas de incentivos.....	156
Notas.....	158
Referencias bibliográficas.....	161

### ¿POR QUÉ CENTRARSE EN LOS MAESTROS?

Un extenso y cada vez mayor acervo de investigaciones sobre las escuelas demuestra de manera convincente que la calidad docente es el *factor de la educación* más importante que influye en el logro de los estudiantes.<sup>1</sup> Se calcula que la diferencia entre tener un maestro muy eficiente contra uno muy ineficiente equivale a no menos de un año completo de diferencia en la evolución del aprendizaje de los alumnos. Además, la forma en que influyen las diferencias en la calidad docente agobia el impacto de otras inversiones en la educación, como las reducciones en el tamaño de la clase (Goldhaber *et al.*, 1999; Rivkin *et al.*, 2005).<sup>2</sup> Para ponerlo en perspectiva, Rivkin *et al.* (2005) y Rockoff (2004) calculan que un aumento de una desviación estándar en la calidad docente eleva el logro de los alumnos en lectura y matemáticas en aproximadamente 10% de una desviación estándar; un efecto en el logro que está en el mismo orden de magnitud como reducir el tamaño del grupo de entre 10 a 13 alumnos (Rivkin *et al.*, 2005). Teniendo en cuenta hallazgos como éste, no sorprende que los responsables de la política educativa estén investigando distintas formas de aumentar la calidad de la fuerza laboral docente.

Desafortunadamente, aumentar la calidad de la fuerza laboral docente no es una tarea sencilla; más que nada porque la conexión entre ésta y los atributos docentes fácilmente cuantificables como categoría de licenciatura, rango académico y niveles de experiencia, es débil (Goldhaber, 2002).<sup>3</sup> Aunque gran parte del trabajo sobre la calidad docente se basa en investigación realizada en **Estados Unidos de América (EUA)**, la débil relación entre la calidad y las credenciales de los maestros se repite en la investigación que relaciona los títulos del docente con el logro estudiantil en **México**. Específicamente, Santibáñez *et al.* (2007) descubren que la experiencia docente y el nivel de los estudios universitarios no son factores predictivos estadísticamente significativos del logro estudiantil y, en el caso del desempeño de los maestros en las pruebas de licenciatura, sus hallazgos son contraintuitivos en el sentido de que el mayor resultado en las pruebas de los maestros se relaciona de manera negativa con el logro estudiantil en evaluaciones nacionales.<sup>4</sup>

Pero aunque los hallazgos anteriores constituyen un desafío para formular políticas que aumenten la calidad docente; sabemos que ésta es un bien muy variable: algunos maestros simplemente son mucho mejores que otros. Éste es uno de los hallazgos más sorprendentes de la reciente investigación centrada en los efectos de los maestros. Kane *et al.* (2007), por ejemplo, encuentran mucha más variación en el desempeño docente observado dentro de una determinada ruta hacia la certificación en el aula que entre rutas. Calculan que la diferencia en el valor agregado promedio entre maestros en los cuartiles de desempeño inferior y superior en cada ruta hacia la certificación sea de aproximadamente 30% de una desviación estándar; o más o menos diez veces la diferencia promedio en valor agregado para maestros que ingresan a la docencia mediante distintas rutas de enseñanza.

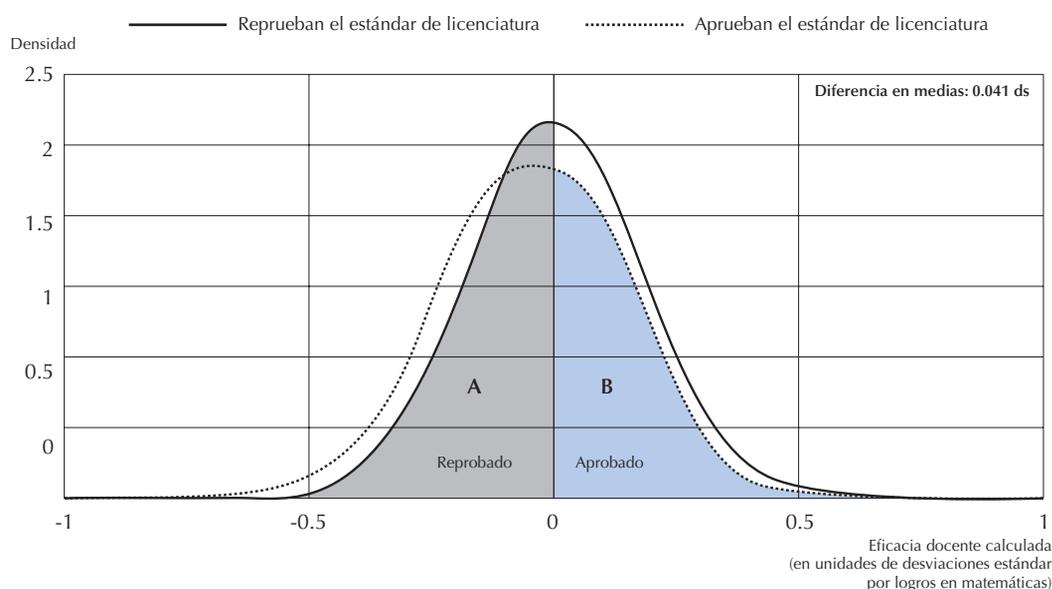
Goldhaber (2007) encuentra resultados similares para el desempeño docente en los exámenes de licenciatura. La Gráfica 5.1, abajo, se deriva de su análisis. Esta gráfica muestra la distribución de eficacia calculada para maestros que aprueban los exámenes de licenciatura exigidos para certificarse totalmente (línea recta) y los que no logran alcanzar ese estándar (línea punteada). Como podría esperarse, en promedio, se encuentra que los maestros que alcanzan el estándar son más eficaces que quienes no lo logran; en aproximadamente 4% de una desviación estándar (la distancia horizontal entre F y P). Sin embargo, también hay una superposición importante en las distribuciones de eficacia superior a 90%. Por ejemplo, todos los maestros mostrados en el área A aprobaron los exámenes de licenciatura pero eran menos eficientes que el maestro típico que no logró alcanzar el estándar; los maestros mostrados en el área B reprobaron, pero eran más eficientes que el maestro típico que había aprobado.

Si los títulos requeridos, como licenciatura, sólo se relacionan débilmente con el logro estudiantil, centrarse entonces en políticas docentes “de entrada” – como exigir que los maestros egresen de un programa aprobado o que ostenten grados educativos específicos – será una forma ineficiente de influir en la calidad de la fuerza de trabajo docente. La razón es que estas políticas de entrada derivarán en cantidades importantes de “positivos

falsos” y “negativos falsos”; es decir, muchos solicitantes que cumplan con los criterios de elegibilidad para el empleo resultan ser maestros ineficientes (positivos falsos), y muchos que no los cumplan habrían sido eficientes en el aula de haberseles permitido ingresar en la fuerza laboral docente (negativos falsos). Los negativos falsos quizá nunca perseveren para convertirse en maestros – una pérdida para la profesión – y quizá sea difícil deshacerse de los positivos falsos una vez que conquisten la seguridad laboral que generalmente brindan las escuelas públicas en la mayoría de los países, México incluido.

Gráfica 5.1

### Distribución de la eficacia docente calculada por el examen oficial de licenciatura



Nota: Las cifras estimadas de eficacia docente son de maestros de primaria con calificaciones válidas en Praxis II en escuelas públicas de Carolina del Norte de los años escolares del periodo 1997-2005, dada la categoría de aprobación que usa el estándar estatal para los Exámenes Praxis II. El error de medición se ha controlado usando un ajuste bayesiano empírico, que atenúa las cifras estimadas hacia la media global en cada año escolar. Véase Goldhaber (2007), Goldhaber y Hasen (2008) para más información.

Fuente: Goldhaber, D. (2007), “Everyone’s Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?”, en *Journal of Human Resources*, Vol. 42, núm. 4, pp. 765-794.

Un argumento similar es válido en el caso de la compensación de los maestros. Las políticas que vinculan el sueldo de los maestros con sus títulos serán una forma burda de recompensar el desempeño si éste sólo se relaciona debilmente con los títulos premiados, como indica la evidencia empírica.

Todo esto da crédito al argumento de que la política debe apartarse del enfoque basado en títulos y dirigirse a otro que dependa de tomar decisiones sobre los maestros una vez que estén en el aula y los responsables de la política educativa puedan opinar sobre su rendimiento. Muchos sostienen que un componente fundamental de ese enfoque es agregar incentivos explícitos destinados a maestros individuales, grupos de docentes (como una escuela) o ciertos tipos de maestros (por ejemplo, de matemáticas).

La compensación, desde luego, es una importante herramienta administrativa para cambiar los incentivos. Como es lógico, ha habido bastante nuevo interés (y se ha renovado) en usar la compensación para premiar la excelencia de los maestros. Por ejemplo, abundan los lugares en todo el mundo donde los responsables de la política educativa han aprobado programas de compensación de los maestros de “pago por desempeño”, aunque éstas tienden a ser la excepción para como suele recompensarse a los maestros de escuelas públicas. Esos programas en muchos casos son polémicos; pero, con todo, la evidencia empírica indica que sus efectos son prometedores.

Este capítulo examina la posibilidad de usar incentivos y la reforma de compensaciones en particular, como una política de la fuerza laboral docente. Empieza con la descripción del sistema de pago que se usa generalmente para retribuir a los maestros en casi todos los países, cómo difiere la estructura de incentivos de los mercados laborales del sector privado y qué incentivos crea la dicotomía entre la forma en que se paga a los maestros, y a otros trabajadores, para los maestros actuales y futuros.

Luego centra su atención en algunos de los temas técnicos y prácticos que deben enfrentar los responsables de la política educativa al diseñar sistemas de incentivos (el capítulo 3 ofrece un tratamiento minucioso de los argumentos teóricos relacionados con los sistemas de incentivos). Estos incluyen, pero no se limitan a, preguntas sobre si las pruebas de logro estudiantil pueden usarse para evaluar el desempeño docente, el nivel al cual deben dirigirse los incentivos (por ejemplo, individual o colectivo); y si los responsables de la política educativa podrían desear usar incentivos, aparte de sólo el sueldo para recompensar a los maestros. Los responsables de la política educativa se enfocan en muchos de estos temas al pensar en sistemas de pago por desempeño (o en cualquier otro sistema basado en incentivos diseñado para premiar el desempeño docente). Sin embargo, como se sostiene más adelante, muchos de los temas abordados son pertinentes para cualquier sistema que se base en incentivos, ya sea que el incentivo se dirija al desempeño o se diseñe para alcanzar alguna otra meta (por ejemplo, una determinada distribución de maestros de todas las áreas geográficas).

A continuación, se describen las evidencias de los sistemas de incentivos para maestros, centrándose en especial en varios sistemas basados en pago por desempeño vigentes en todo el mundo y en qué investigación disponible se ha abordado lo tocante a sus efectos. Esta sección argumenta que uno no puede saber *a priori* si un determinado sistema de incentivos será exitoso, ya que eso dependerá en buena parte de la seriedad de la implementación del programa. Los últimos párrafos ofrecen algunas conclusiones sobre lo que podría necesitarse para apoyar los sistemas de incentivos.

## TABULADORES DE SUELDO E INCENTIVOS ACTUALES

Los responsables de la política educativa a menudo hablan sobre crear incentivos para maestros sin reconocer que ya existen incentivos del mercado laboral; en muchos casos no sólo incentivos que se hayan diseñado con un fin determinado.<sup>5</sup> Y hay excelente evidencia de que los maestros, al igual que los empleados en general (Prendergast, 1999), responden a la variedad de incentivos que se les presenten. Esos incentivos influyen en quien ingresa y permanece en la docencia; e impactan, por ejemplo, la elección inicial de ocupación (Hanushek y Pace, 1995; Stinebrickner, 2001a, 2001b), en la que acabarán los maestros de escuelas (Lankford *et al.*, 2002), y en decisiones secundarias cuando decidan jubilarse (Costrell and Podgursky, 2008).<sup>6</sup>

Una causa determinante clave de si los individuos optarán por la enseñanza es la compensación que puedan tener como maestros contra lo que podrían devengar en una ocupación distinta. Comparar la compensación es más difícil de lo que podría creerse a un primer vistazo como comparaciones simples en, digamos, los sueldos promedio en todos los sectores de una economía; pueden ocultar diferencias importantes en las características y habilidades de quienes trabajan en esos sectores, o importantes factores laborales no pecuniarios; el nivel de

seguridad laboral o seguridad en el lugar de trabajo, por ejemplo. Aun así, los niveles salariales son un lugar importante para empezar, cuando se evalúa la conveniencia de una ocupación.

Al juzgar por el sueldo, los maestros de hecho están bastante bien pagados en muchos países de América Latina y del Caribe. Claro que hay una variación importante en los sueldos en todos los países; y, en cierto grado, las conclusiones dependen de las comparaciones que se hagan. Sin embargo, un análisis de los sueldos de 2005 en 17 países de América Latina y del Caribe, de Hernani-Limarino (2005), revela que los maestros de hecho están bien pagados (en un sentido relativo) en varios de esos países. Por ejemplo, en **Argentina, Chile, Colombia, El Salvador, Honduras, Panamá, Paraguay y Perú**, los maestros están relativamente bien pagados comparados con trabajadores de otros sectores; y los maestros de primaria en **México** reciben salarios más altos que 89% de otros fuera de la educación que tienen de 13 a 15 años de estudios.<sup>7</sup>

Sin embargo, tener sueldos altos no garantiza la calidad docente; al menos, no la calidad que se transforma directamente en éxito escolar. Usar la proporción salarios docentes-producto interno bruto per cápita como una norma de medición de lo bien que se paga a los maestros muestra poca correlación obvia entre los sueldos y el logro estudiantil (Goldhaber, de próxima aparición; Ladd, 2007). Por ejemplo, la proporción para **México** es considerablemente más alta que para el promedio de los países miembros de la OCDE; pero los niveles de logro para los estudiantes mexicanos van a la zaga de la mayoría de los demás países (OCDE, 2005).

Esto no es sorprendente; si bien los sueldos docentes son una importante causa que determina los incentivos relacionados con la elección de buscar empleo en las escuelas, difícilmente son el único factor que influye en la elección.<sup>8</sup> Por ejemplo, Goldhaber (de próxima aparición) demuestra que los sistemas de desarrollo docente (es decir, los sistemas de formación y selección que se requieren para convertirse en maestro) varían mucho de un país a otro y estos influyen en la probabilidad de que quien aspire a un trabajo de maestro pueda conseguir uno. Y otros factores, como el volumen de trabajo, las condiciones laborales, la seguridad laboral y el prestigio de la profesión pueden variar considerablemente de un país a otro; además, todos influirán en los estímulos para convertirse en maestro.<sup>9</sup>

Al mirar más allá de los promedios, es probable que la estructura (lo que recompensa al maestro y los atributos relacionados con el trabajo) de compensación de los maestros (y los incentivos en general) desempeñe una función importante para determinar la calidad y distribución de los maestros. La mayoría de los países latinoamericanos y de la OCDE paga a los maestros conforme a un tabulador de sueldos que establece la compensación, basada casi totalmente en la experiencia y en el nivel educativo del maestro (Hernani-Limarino, 2005; Ladd, 2007; McEwan y Santibáñez, 2005; Murnane y Steele, 2007: OCDE, 2004).<sup>10</sup>

Aunque en muchos casos el sector privado también recompensa esos atributos de los individuos, es mucho más probable que premie otros atributos como el rendimiento en pruebas estandarizadas (Murnane *et al.*, 1995), el criterio de selección de la institución donde se tituló la persona (Brewer *et al.*, 1999) y las medidas de productividad (Bretz y Milkovich, 1989; Hein, 1996; Milkovich y Wigdor, 1991; Podgursky y Springer, 2007).<sup>11</sup> La compensación del sector privado también suele reflejar las condiciones de la oferta y la demanda que influyen en la compensación que pueden tener los individuos con determinadas habilidades (Grogger y Eide, 1995), o el hecho de que las condiciones laborales relacionadas con cierto trabajo puedan exigir el pago de un “ajuste por compensación” (otro incremento al ingreso o al salario) para atraer a los futuros empleados para que acepten un trabajo en un puesto relativamente menos deseable sobre otro que ofrezca condiciones laborales no pecuniarias más favorables (Chambers, 1995; Rosen, 1986).

Esta dicotomía en las estructuras de compensación entre el sector privado y (normalmente) y los sistemas escolares significa que la distribución de los ingresos en la enseñanza en general es más limitada que en el sector privado; un hallazgo reflejado en la investigación sobre la estructura salarial docente en **Chile** (Mizala y Romaguera, 2005).

A su vez, esto crea incentivos en el campo laboral docente; aunque no sean incentivos diseñados especialmente para fomentar la calidad docente. En particular, los individuos con atributos o habilidades que reconoce mejor el sector privado tenderán a hacer mayores sacrificios financieros que otros para enseñar; enfrentan “costos de oportunidad” distintos. La investigación sobre la dicotomía en las estructuras de incentivos entre los mercados laborales docentes del sector público y el mercado laboral del sector privado muestra que la importancia de los diferenciales en los costos de oportunidad entre los individuos con atributos variables puede ser impresionante: cifras estimadas de Goldhaber *et al.* (2008) prueban que los egresados universitarios empleados en **EUA** con estudios técnicos como especialidades en ciencias o matemáticas, y que elijan la enseñanza, parecen sacrificar mucho más financieramente al hacerlo que los empleados sin estudios en esas especialidades.<sup>12</sup> Esta investigación muestra hallazgos similares cuando se trata de medir la competencia académica de los docentes (por ejemplo, sus resultados en los exámenes de ingreso a la universidad).

Teniendo en cuenta estos incentivos, no sorprende que la investigación muestre que los sistemas escolares a menudo tienen mucha dificultad para contratar maestros de matemáticas y ciencias (Hare *et al.* 2000; Harrington 2001; Shields *et al.* (2001), y que hay un desgaste más rápido fuera de la enseñanza en esas áreas de especialidad que en las de otras asignaturas. De modo similar, es probable que las escuelas con condiciones laborales menos deseables enfrenten dificultades en la contratación. En **EUA**, la investigación prueba con claridad que en todas las escuelas se clasifica a los maestros a partir de los atributos de los alumnos; con los maestros que poseen más títulos y experiencia dotando de personal a las aulas de los alumnos más aventajados. Algo de esto se basa en el emparejamiento inicial entre maestros y escuelas (Lankford *et al.*, 2002) y porque es más probable que los maestros dejen las escuelas que atienden a poblaciones escolares más pobres y de minorías más altas. Una explicación para estos hallazgos es que el sistema de compensación de los maestros generalmente no logra reflejar la dificultad de un trabajo.<sup>13</sup>

No es algo sin precedentes para un paquete de compensación de los maestros (es decir, la compensación directa así como otros estímulos financieros) dar razón de las dificultades laborales o de los servicios complementarios percibidos de la ubicación de un empleo. Por ejemplo, a los países latinoamericanos a menudo se les dificulta contratar maestros en las áreas rurales (Urquiola y Vegas, 2005). Algunos de esos países han tratado de usar los incentivos para subsanar esa situación. **Costa Rica**, por ejemplo, les da a los maestros rurales una compensación adicional y créditos extras hacia la jubilación; **Honduras** les otorga un complemento de sueldo y ascenso más rápido en la banda salarial; y **México** les proporciona diversas formas de apoyo para la vivienda (McEwan, 1999). Sin embargo, estos tipos de incentivos no son la norma. Además, vale la pena hacer notar que los tabuladores de sueldos, casi por definición, no logran reflejar todos los factores que influyen en la conveniencia de un determinado trabajo de enseñanza (Goldhaber *et al.*, 2007). En cambio, se cree que los salarios definidos de manera competitiva en el sector privado reflejan mejor todos los factores que influyen en la conveniencia de un trabajo, desde el riesgo y la dificultad hasta los servicios complementarios laborales no financieros (Boyd *et al.*, 2003). Ésta es otra razón por la cual se esperaría una dicotomía entre las estructuras de compensación de los maestros y de quienes no lo son.

Una preocupación importante que surge con respecto a la dicotomía entre los incentivos relacionados con los tabuladores de sueldos docentes y la estructura de incentivos en el mercado laboral en conjunto es la posibilidad de que la estructura de compensación disuada a los individuos más talentosos de que no busquen una carrera docente. Investigación reciente en **EUA** muestra que es menos probable que los egresados universitarios con calificaciones más altas en los exámenes ingresen a la docencia que a otras ocupaciones (Corcoran *et al.*, 2004; Goldhaber y Liu, 2003; Hanushek y Pace, 1995; Henke *et al.*, 1996; Hoxby y Leigh, 2004; LakDawalla, 2001). Eso establece un fuerte contraste con la situación de hace varios decenios cuando muchas de las egresadas académicamente más capaces se convirtieron en maestras. Específicamente, Corcoran *et al.* (2004) descubren la probabilidad de que una maestra que haya estado en el 10% superior (en un examen

de logro estandarizado) entre los estudiantes de secundaria bajó en más de la mitad (aproximadamente a una probabilidad de 20%) en 1964, a más o menos una probabilidad de 10% en el 2000. Los autores sostienen que la principal explicación para esto es que las oportunidades para las mujeres fuera de la enseñanza se han ampliado de manera asombrosa durante las últimas cuatro décadas.

Hoxby y Leigh (2004) ofrecen otra explicación para la disminución en la cantidad de mujeres académicamente talentosas que ingresan en la docencia. Sostienen que el descenso se debe a que a lo largo del tiempo ha habido cambios en la compresión salarial, entre la enseñanza y otras ocupaciones. Atribuyen la mayor compresión en la enseñanza a la creciente influencia de los sindicatos de maestros, y hacen notar que la mayor compresión salarial en la enseñanza significa que un individuo que opte por la docencia tiene mucho menos probabilidades de recibir un sueldo cercano al promedio que si hubiera escogido, en cambio, una profesión ajena a ésta. Por consiguiente, la enseñanza se ha convertido en una ocupación financieramente cada vez más atractiva para los individuos con aptitudes menores cuya compensación (en la enseñanza) tiende a ser arrastrada ascendiendo en la media; y una ocupación menos financieramente deseable para los que tengan aptitudes superiores cuya compensación tiende a ser arrastrada descendiendo en la media. Sus análisis de encuestas sobre egresados universitarios recientes de varios años entre 1961 y 1967 indican que aproximadamente 80% de la disminución que ven en maestros procedentes de universidades muy selectivas puede rastrearse hasta la compresión salarial en la docencia.

Si bien el hallazgo aquí de un descenso en el calibre académico de la fuerza laboral docente se limita a la fuerza laboral de EUA, hay motivos de preocupación de que los incentivos que crearon la situación en EUA también pudieran intervenir en los mercados laborales docentes de América Latina. En particular, la investigación realizada por Hernani-Limarino (2005) sobre sueldo docente en América Latina y el Caribe coincide con algunos de los hallazgos de Hoxby y Leigh. Específicamente, Hernani-Limarino pronostica la compensación que los individuos podrían tener dentro y fuera de la docencia usando una metodología estadística diseñada para dar razón sólo de factores individuales observables, como experiencia en el mercado laboral y niveles educativos que influyen en la compensación; pero también de factores que pasan inadvertidos e influyen en la productividad del individuo y, por ende, en su compensación.<sup>14</sup> Este análisis indica que, en casi todos los países examinados (**Nicaragua** es una excepción notable), los individuos que tienden a ser menos productivos devengarían relativamente más como maestros y quienes posean atributos que los hagan muy productivos tenderían a ganar relativamente menos.<sup>15</sup> Esto concuerda perfectamente con la afirmación de Hoxby y Leigh (2004) de que la compresión salarial en la enseñanza tiende a convertirla en una ocupación relativamente poco deseable para individuos talentosos porque ellos pueden tener una mayor compensación en otra ocupación.

En general, la investigación descrita en esta sección señala de manera convincente que los sistemas de pago en la educación pública que dependen estrictamente de tabuladores de sueldos no reconocen la realidad del mercado laboral de que algunos maestros tendrán más oportunidades competitivas y financieramente atractivas que otros. Tampoco reflejan la realidad de que algunas escuelas y aulas serán más exigentes para el personal por la dificultad del trabajo o de los servicios complementarios propios de éste.

Uno puede ignorar esos incentivos del mercado laboral o intentar diseñar sistemas de incentivos que expliquen de manera explícita las realidades del mercado laboral. Por ejemplo, esos sistemas podrían centrarse en el desempeño de los maestros y diseñarse para redistribuir a los docentes de una escuela o localidad a otra, o en ambas. Los problemas técnicos y prácticos que surgen al diseñar sistemas de incentivos se examinan más adelante.

## USAR INCENTIVOS: FACTORES TÉCNICOS Y PRÁCTICOS

Uno de los problemas que se presentan al considerar los incentivos para maestros es qué tipo de incentivos son los más eficaces para cambiar los comportamientos del mercado laboral; por ejemplo, aumentar la labor de los maestros, persuadirlos para que se cambien de una escuela a otra o mantenerlos en determinadas escuelas. Este

capítulo no ahondará mucho en los argumentos teóricos sobre los distintos tipos de incentivos, ya que éstos se cubren de manera más exhaustiva en los capítulos 2 y 3. Sin embargo, vale la pena hacer notar que existe una amplia variedad de posibles incentivos: incentivos financieros como subsidio de vivienda, subvenciones para colegiatura; bonos únicos relacionados con los resultados de los alumnos; incentivos basados en títulos para el escalafón; y una completa variedad de incentivos no monetarios, desde el reconocimiento especial por desempeño a opciones para dar clases particulares, hasta más tiempo para preparación y oportunidades especiales de desarrollo profesional. Desafortunadamente, pocos de esos incontables incentivos se han estudiado de manera rigurosa para probar su eficacia.

Aun así, el capítulo ofrece dos ideas sobre el tipo de incentivos. La primera es que uno debe considerar la estructura de los sistemas de desarrollo docente (por ejemplo, instituciones de formación docente y prácticas de licenciatura) y las prácticas de la fuerza laboral al considerar el tipo de incentivos que se ofrecen. Específicamente, los incentivos que representan inversiones iniciales cuantiosas pueden ser bastante costosos porque subsidian a individuos que no terminan enseñando (o que los responsables de la política educativa podrían creer que no enseñan). Por ejemplo, algunos países como **Singapur** usan las subvenciones para colegiatura como un incentivo para que los individuos talentosos ingresen a los programas de formación docente (Barber y Mourshed, 2007). Pero uno de los rasgos distintivos del sistema de Singapur es que el conducto para el desarrollo docente se estrecha de manera importante en el punto en que los individuos ingresan a la formación docente, y la mayoría de esos egresados terminan con una larga carrera en la docencia. Si fuera el caso de que el país produjera muchos más individuos egresados de la formación docente que puestos docentes, necesariamente se desembolsaría un subsidio general de costos de formación a muchos individuos que no enseñen (o que quizá no lo hagan durante mucho tiempo).

La segunda idea es que los incentivos estructurados para dar reconocimiento a los títulos docentes más que a los maestros individuales, basados en evaluaciones de su desempeño en el aula, necesariamente se identificarán de manera relativamente deficiente. Como se describe antes, existen abundantes evidencias que demuestran que los maestros que al parecer son similares a partir de títulos en papel (como experiencia o nivel educativo) varían considerablemente en cuanto a eficacia. Por consiguiente, incluso si se muestra el título que se usa como una medición del éxito, en promedio, para pronosticar la eficacia docente, es muy probable que reconocer a los maestros a partir de esto se traduzca en inversiones en maestros que no sean eficaces en demasía debido a los positivos falsos.

Como un ejemplo de esto, en EUA muchos estados y localidades premian a los maestros que consigan un diploma de enseñanza avanzada, certificación del NBPTS. Algunas evidencias encuentran efectos estadísticamente significativos en el logro estudiantil relacionados con tener un maestro certificado por el NBPTS (Cavalluzzo, 2004; Goldhaber y Anthony, 2007). Goldhaber (2006a) descubre que hay una cantidad considerable de duplicación parcial en las distribuciones calculadas de eficacia (similar a la Gráfica 5.1 en este capítulo). De manera específica, aproximadamente 60% de los maestros certificados son más eficientes que sus compañeros no certificados al evaluarlos por logros estudiantiles en pruebas de matemáticas; y más o menos 55% al evaluarlos a partir de logros estudiantiles en pruebas de lectura. Esto sin duda es mejor de lo que se esperaría de echarlo a la suerte para decidir; pero difícilmente es una garantía de que un incentivo basado en ese diploma habitualmente premie a los maestros más eficientes.

Varias preguntas surgen en el contexto de los incentivos destinados al maestro individual. Si bien esas preguntas son especialmente pertinentes para los incentivos de pago por desempeño, de hecho aplican a cualquier caso en que el sistema educativo desee destinar incentivos a maestros individuales. Por ejemplo, los responsables de la política educativa quizá deseen ofrecer una mayor compensación a los maestros con conjuntos de habilidades que tengan demanda en el sector privado; pero podrían no desear recompensar a todos los que tengan esas

habilidades dado que no todos serán maestros eficaces. Del mismo modo, los responsables de la política educativa podrían desear ofrecer incentivos por retención a los maestros que estén en escuelas a las que sea difícil dotar de personal; pero también desear que se les paguen sólo a los que demuestren su eficacia.

**Incentivos y disfunción docente** La individualización de incentivos para maestros es polémica por varias razones. Por ejemplo, algunos indican que la docencia no es muy apropiada para recompensas individualizadas porque eso crea incentivos poco adecuados. El trabajo que realizan los maestros es múltiple y diverso y, quizá, no del todo fácil de evaluar. Como Dixit (2002) hace notar en una perspectiva general de las publicaciones teóricas y empíricas sobre el uso de incentivos al sueldo, la individualización de incentivos (como el pago por mérito) es mucho menos probable en empleos complejos, cuyos objetivos abarquen múltiples facetas y en los que es difícil cuantificar con precisión la productividad de los empleados.

El problema potencial con los incentivos en el caso de escuelas es que los maestros quizá muestren conductas disfuncionales en forma individual, a causa del incentivo. Por ejemplo, podrían centrarse de manera muy limitada en tareas, asignaturas o contenido del programa que sean premiados (lo que se denomina “reducción del currículo” o “enseñar para el examen”) a costa de otras tareas o metas apreciadas por el público (como promover la ciudadanía o reducir el consumo de drogas o la violencia).<sup>16</sup> También preocupa que la individualización pudiera tener un efecto contraproducente en el lugar de trabajo, debilitar la colaboración y la colegialidad docente al introducir la competencia entre los maestros (Murnane y Cohen, 1986).

Cierto trabajo empírico sí apoya la noción de que las recompensas individualizadas en la enseñanza llevan a conductas docentes disfuncionales y contraproducentes. Por ejemplo, Jacob y Levitt (2003) encuentran pruebas de que los incentivos relacionados con el logro estudiantil en pruebas estandarizadas lleva a que se haga trampa (los maestros les dan las respuestas de los exámenes a los alumnos), y tanto Cullen y Reback (2006) como Figlio y Getzler (2002) encuentran pruebas de que los docentes emplean varios mecanismos (como clasificar a estudiantes en educación especial) para excluir las calificaciones de las pruebas de los alumnos de bajo rendimiento usados para fines de rendición de cuentas.

Claro que relacionar los incentivos con el aprendizaje de los alumnos según lo medido por pruebas estandarizadas no es la única manera de hacer evaluaciones sobre los maestros. Por ejemplo, uno podría premiar a los maestros a partir de evaluaciones de sus colegas o supervisores; y hay ejemplos de sistemas que dependen de múltiples indicadores de desempeño distintos a las calificaciones de las pruebas de los alumnos (tema que se analiza de manera más detallada en el capítulo 3).

Es natural preguntarse si los instrumentos de evaluación con exámenes sin estudiantes sí distinguen a los maestros eficaces de quienes no los son. El conjunto relativamente escaso de evidencia empírica que se centra en la correlación entre diversos instrumentos para evaluar a los maestros y medidas de “valor agregado” más directas de la eficacia docente (el valor agregado se describe de manera más detallada adelante) indica que hay formas de evaluar a los maestros que no se alínean razonablemente bien con su contribución al aprendizaje de los alumnos, conforme a lo medido por el rendimiento de los estudiantes en pruebas estandarizadas.

**Colaboración y producción docente conjunta** Quizá incluso más importante que la pregunta de si los incentivos causan disfunción a nivel docente individual es el tema más importante de si la individualización de los incentivos produce disfunción en la escuela, debido a los efectos sobre la colaboración. Éste es un tema especialmente destacado en la secundaria, donde los estudiantes a menudo cambian entre clases y maestros; y es probable que haya aspectos complementarios en asignaturas como matemáticas y ciencias. Koedel (de próxima aparición) aborda directamente el tema de los complementos entre asignaturas y su relación con los incentivos para maestros o ambos. Calcula exámenes y modelos de valor agregado en logro en lectura si en el logro en lectura del estudiante influye no sólo la calidad de sus maestros de lengua sino también la de sus

maestros de matemáticas, estudios sociales y ciencias. Koedel encuentra muy pocos datos que indiquen la producción conjunta (en todas las asignaturas) de logro en lectura de los estudiantes.

Los hallazgos de Koedel no demuestran de manera definitiva que no debemos preocuparnos porque un sistema basado en incentivos lleve a la disfunción; pero sí disminuyen la inquietud ya que es probable que los problemas de colaboración tengan mayores consecuencias en el logro estudiantil, si el aprovechamiento en una asignatura se relaciona fuertemente con la calidad de los maestros en otra.<sup>17</sup> También son alentadoras las pruebas de que los incentivos individualizados (pago por mérito en particular) existen en una variedad de distintos ambientes de trabajo que exigen colaboración, incluido, por ejemplo, el uso mucho más frecuente de pago por desempeño en las escuelas particulares, al menos en **EUA** (Ballou y Podgursky, 1997). Por último, evidencia directa de un sistema de incentivos indica que las preocupaciones sobre la disfunción asociada de las escuelas puede ser exagerada: Lavy (2004) estudia un incentivo docente a nivel individual en Israel y descubre que no hay efectos negativos en el logro estudiantil en asignaturas que no son parte del sistema de incentivos.

Claro que es posible atacar la inquietud de que los incentivos para maestros individualizados podrían crear problemas de colaboración dentro de la escuela al equilibrarlos con incentivos para todo el grupo en una sección o en toda la escuela. Una ventaja de los incentivos basados en la escuela es que ayudan a encarar el problema de cómo evaluar a los maestros en asignaturas no sometidas a prueba, al menos en el caso especial de usar las pruebas de los estudiantes para juzgar el desempeño (un tema enfocado en detalle más adelante). Las pruebas de logro estudiantil generalmente sólo abarcan matemáticas y lectura (y a veces ciencias) y se aplican en grados seleccionados: en **USA**, por ejemplo, es característico que menos de un cuarto de los maestros esté cubierto por grados y asignaturas en los cuales se aplican pruebas estandarizadas a los alumnos. Eso significa que no podría evaluarse directamente a muchos maestros, si la evaluación se basara en el logro de pruebas estandarizadas. Por consiguiente el objetivo principal de la escuela tiene la virtud de ser aplicable a todos los maestros en una escuela, incluso a los de áreas en las que generalmente no se evalúa a los estudiantes a partir de pruebas (música, por ejemplo).

Varios países han experimentado con los incentivos para maestros basados en las escuelas y proporcionan algunos ejemplos de distintos enfoques para comparar escuelas. En **Chile**, por ejemplo, el Sistema Nacional de Evaluación de Desempeño en los Establecimientos Educacionales (SNED) compara escuelas en distintas regiones geográficas del país y todos los maestros en la escuela de máximo desempeño de cada región reciben bonos financieros (Mizala y Romaguera, 2005). Sin embargo, es poco claro si este sistema es justo con todas las escuelas ya que puede haber una variación importante en el tipo de estudiantes (y en las ventajas que tienen) inscritos en las escuelas de cada región del país.

En **EUA**, varios estados y distritos escolares también han usado compensaciones basadas en las escuelas (Clotfelter y Ladd 1996; Meyer 1996; Odden y Kelley 1997). Aunque algunos sistemas dependen de las comparaciones entre escuelas demográficamente similares, otros, a diferencia del SNED de **Chile**, usan cierto tipo de ajustes estadísticos para comparar a las escuelas. Por ejemplo, en **Carolina del Sur** el plan de pago por desempeño basado en el grupo otorga bonos a las escuelas por mejorar en las calificaciones de las pruebas, después de controlar las diferencias en categoría socioeconómica y nivel promedio de educación del maestro. El distrito escolar independiente de **Dallas** evaluó a las escuelas a partir del desempeño promedio en la evaluación estatal, controlando el nivel socioeconómico y el dominio limitado del inglés, así como raza/origen étnico en un proceso estadístico más complejo de dos etapas.

Independientemente del método, la clave al hacer comparaciones escuela-escuela, maestro-maestro es que se hagan de manera que las contribuciones de las escuelas o de los maestros se aislen de otros factores que influyen en el logro estudiantil, como el origen de los estudiantes. Eso es fundamental si uno quiere que las

gratificaciones relacionadas con un sistema de incentivos sean justas en el sentido de que se basen en lo que de hecho influyen las escuelas o los maestros. Bueno, por ejemplo, uno no quisiera comparar directamente el desempeño en una escuela de estudiantes económicamente favorecidos con una escuela de estudiantes desproporcionadamente desfavorecidos desde el punto de vista económico.

Aunque hay aspectos favorables para utilizar los incentivos para maestros basados en la escuela – es decir, atacar el problema de colaboración y abarcar a los maestros en asignaturas o grados no sometidos a examen – también hay desventajas importantes para un sistema de incentivos que se base de manera estricta en la escuela. Una de éstas es que los incentivos son poco precisos, lo que puede dar lugar a una tendencia para que los maestros traten de aprovecharse de los esfuerzos de sus colegas (Hansen, 1997; Prendergast, 1999). Quizá lo más importante, los sistemas basados estrictamente en el grupo no lograrán premiar de manera adecuada a los maestros destacados dentro del grupo seleccionado. La investigación sobre la variación en las calificaciones de las pruebas de los alumnos indica que éste es un problema importante cuando se trata de los maestros; ya que más de la variación en la eficacia docente está dentro de las escuelas que entre escuelas (Rivkin *et al.*, 2005).

Sin embargo, no hay motivo para que uno tenga que elegir entre incentivos individuales y colectivos. El análisis antes citado indica que un sistema híbrido, con una parte del incentivo de un maestro vinculada con el desempeño colectivo y otra vinculada con el desempeño individual podría ser más eficaz. Teniendo esto en cuenta, vale la pena cerrar esta subsección con una consideración práctica sobre la individualización de los incentivos para maestros. Para que los incentivos tengan un efecto, debe ser percibidos como alcanzables y lo suficientemente importantes a la vez para que induzcan cambios de conducta, por ejemplo, gratificaciones monetarias suficientemente grandes (Lavy, 2007). Según se describe más adelante, hay una variación considerable tanto en la proporción de maestros que reciben pagos por incentivo como en la magnitud absoluta y relativa (para los ingresos) de alicientes financieros en todos los esquemas de incentivos. Sin embargo, en las publicaciones empíricas no hay consenso sobre qué constituye la forma correcta para estructurar incentivos, en cuanto a la proporción de maestros que los reciben o su tamaño en relación con la compensación básica de los docentes.

**Métodos de valor agregado para usar las calificaciones de las pruebas de los alumnos como una norma de medición para evaluar a los maestros.** A menudo hay limitaciones institucionales cuando se trata de las evaluaciones subjetivas de los maestros; y éste es uno de los argumentos clave para usar metodologías estadísticas (construcción de modelos de valor agregado) para intentar aislar las contribuciones que los maestros individuales hacen al logro estudiantil.

Pero varios problemas técnicos potencialmente importantes se relacionan con usar las calificaciones de las pruebas de los alumnos como una norma de medición para evaluar la eficacia docente. Esos problemas son complejos, así que sólo puede proporcionarse un panorama somero a lo más. Sin embargo, antes de ahondar en ellos, vale la pena hacer notar que – para muchos – la idea de usar el logro estudiantil en pruebas para hacer cualquier tipo de evaluación sobre las escuelas en sí misma es polémica. Las pruebas estandarizadas están limitadas en cuanto a qué miden. En el mejor de los casos, las pruebas sólo pueden usarse para evaluar un subconjunto de los innumerables objetivos de la educación (por ejemplo, es poco probable que sean pruebas que abarquen la compleja conducta de socialización que se enseña en las escuelas) ya que están limitadas en cuanto a su capacidad para identificar el grado en que los estudiantes han aprendido los temas importantes cubiertos en el aula (Koretz, 2008). Además, incluso los objetivos que cubren las pruebas se medirán de manera imperfecta (por ejemplo, una prueba de matemáticas será una muestra de lo que se enseña en una clase de matemáticas).

En el nivel más profundo, los problemas sobre el uso de pruebas para evaluar a los maestros están estrechamente relacionados con las opiniones propias sobre si son útiles para evaluar lo que los alumnos han aprendido. Si bien es importante tener en cuenta los límites de las pruebas al formular una política, deben sopesarse esos

límites contra otros dos factores. Primero, en muchos países las pruebas de los alumnos habitualmente se elaboran con información de numerosos grupos de interesados que ayudan a impulsar políticas educativas (como la rendición de cuentas escolar) y, por tanto, supuestamente reflejan las metas educativas de un país. Segundo, se ha demostrado que el logro estudiantil en pruebas estandarizadas es un importante factor predictivo de resultados algunos años después en la vida (Grogger y Eide, 1995; Murnane *et al.*, 1995) y se relaciona con las tasas de crecimiento totales del país (Hanushek *et al.*, 2008).

Dejando atrás la pregunta de si las pruebas estandarizadas son una buena medida del logro estudiantil, es necesario confrontar si pueden ser usadas para evaluar el desempeño docente. Los modelos estadísticos, llamados “modelos de valor agregado” (VAM), que se usan para pronosticar los factores que influyen en el aprendizaje de los estudiantes se basan en varias suposiciones sólidas. Además, esos modelos dependen de las pruebas de los estudiantes que, en sí mismas, poseen diversas propiedades psicométricas. Por ejemplo, las pruebas quizá oculten incrementos en aprendizaje para estudiantes de alto o bajo rendimiento, a partir de reactivos específicos de las pruebas incluidos en ellas. Por consiguiente, si se usan para evaluar el desempeño docente posiblemente subestimen o exageren la eficacia docente en diferentes entornos.

También hay problemas relacionados con la cantidad de información verdadera sobre la eficacia del maestro que son expresados hasta por la mejor de las pruebas de los estudiantes. Por ejemplo, uno no quisiera atribuir demasiado significado a los cambios en el resultado de los exámenes si un maestro sólo tiene una pequeña cantidad de estudiantes. La razón es que las pruebas siempre son medidas imprecisas de lo que saben los estudiantes, un alumno que hace la misma prueba dos veces en el mismo día probablemente no obtenga la misma calificación simplemente debido a factores aleatorios. Por tanto, los maestros cuya medida de eficacia se basa sólo en algunas calificaciones de las pruebas de los alumnos es mucho más probable que sea influenciada por esos factores aleatorios. En jerga estadística, nos preocupamos por la cantidad de verdadera “señal” sobre la eficacia docente que se transmite a diferencia del “ruido» estadístico.

Por último, hay preguntas importantes sobre cuáles son las comparaciones correctas entre los maestros. La esperanza es que los modelos estadísticos ayuden a ajustar las diferencias en el tipo de estudiantes que tiene un maestro en el aula; pero uno debe decidir si se compara a los maestros dentro de una escuela, en todas las escuelas de un distrito o en algún nivel más alto, una región o todo el país, por ejemplo.

El interés de la política pública de usar los VAM ha dado por resultado un creciente acervo de publicaciones que examinan todo los problemas antes citados (por ejemplo, OCDE, 2008). De hecho, hay una larga historia de la construcción de modelos de valor agregado en la economía (por ejemplo, Hanushek, 1972) y el término en sí mismo abarca una amplia clase de diferentes especificaciones estadísticas. Es sólo recientemente que los VAM se han usado de manera exhaustiva para tratar de identificar las contribuciones que los maestros individuales hacen a los incrementos del aprendizaje estudiantil.<sup>18</sup>

Los esfuerzos para identificar los efectos causales de los maestros (o de cualquier insumo de la enseñanza) se complican por la complejidad del proceso de aprendizaje. Como Todd y Wolpin (2003) señalan, los VAM dependen de suposiciones sólidas sobre la naturaleza del aprendizaje de los estudiantes a lo largo del tiempo y las violaciones de esas suposiciones del modelo pueden llevar a conclusiones erróneas sobre los efectos de los recursos educativos, como los maestros.

Solamente unos ejemplos: los VAM normalmente suponen el deterioro constante del aprendizaje (la cantidad de aprendizaje que los niños olvidan entre cada prueba aplicada), pero la cantidad de deterioro del aprendizaje de los estudiantes de un grado escolar a otro quizá no sea uniforme en todos los grados (un “grado” en el sistema de **EUA** equivale al nivel de un año en otros sistemas de la OCDE). Las cantidades de deterioro mayores o menores en un grado o en el siguiente llevarían a los estadígrafos a subestimar o a exagerar las contribuciones de los

docentes en esos grados. Los maestros también podrían contribuir al aprendizaje de los estudiantes más allá del grado que imparten. Los VAM normalmente suponen que la totalidad de las contribuciones docentes de un grado al siguiente se refleja en la evolución de las calificaciones de las pruebas entre esos grados (y luego eso declina a un ritmo constante a lo largo del tiempo). Por tanto, en casos en que los maestros enseñan en el grado  $g$  influye en la evolución del aprendizaje de los estudiantes en el grado  $g+1$  o en un grado más alto; el efecto de ese maestro se subestimaría y más bien se atribuiría mal a los maestros en esos grados superiores. Eso podría suceder si, por ejemplo, los maestros del grado  $g$  enseñan habilidades que no refleja la prueba estandarizada del grado  $g$ .

Al calcular los efectos docentes, los estadígrafos también deben tomar decisiones clave sobre qué estudiantes (y los resultados de sus exámenes) se usan para calcular las cifras estimadas de valor agregado del maestro. Los VAM generalmente dependen de los incrementos en medición en el logro estudiantil, también limitan la muestra de estudiantes a los que son examinados al menos dos veces. Por consiguiente, los estudiantes con mucha movilidad – para los cuales no están disponibles los incrementos – no se integrarán en las cifras estimadas de desempeño docente. Incluso si los estudiantes tienen pruebas múltiples, lo que permite cifras estimadas en incrementos, uno debe decidir qué hacer con los que no toman clases un año completo (por ejemplo, pueden haber entrado a un grupo en octubre en vez de en septiembre, pero aun así tomaron una prueba estandarizada aplicada en noviembre).

No queda claro si se han examinado empíricamente las implicaciones de la movilidad de los estudiantes para evaluar los efectos de un maestro individual. Una forma no probada pero intuitiva de atacar este problema es atribuir los incrementos en logro estudiantil a los maestros en proporción con el tiempo que ellos pasan en la clase de un maestro. Eso desde luego es más fácil si el sistema de datos de un estado/región o país permite rastrear a los estudiantes de una escuela a otra.

Otra razón por la que los estudiantes pueden ser excluidos de la factorización en las cifras calculadas de desempeño docente es que la información sobre sus antecedentes se pasa por alto o es limitada. Si la información de los antecedentes de los estudiantes es necesaria para evaluaciones exactas de los maestros, a uno no le gustaría depender de cifras estimadas que no utilizan esa información. Sin embargo, si los antecedentes de los estudiantes no son importantes, uno quisiera incluir el complemento total de estudiantes a los que enseña cada maestro ya que eso constituiría la información más completa sobre cómo influyen los maestros en sus alumnos. McCaffrey *et al.* (2003) exploran este problema usando una serie de simulaciones. Demuestran que los modelos que dependen de información longitudinal sobre el logro estudiantil producen cifras estimadas similares de desempeño docente, incluso a falta de antecedentes detallados de los estudiantes; pero sólo mientras exista una cantidad importante de mezcla de estudiantes de todos los maestros.<sup>19</sup> Si hay poca superposición de estudiantes, entonces es necesario tener información sobre los antecedentes de los estudiantes para hacer valoraciones válidas sobre el desempeño docente.

Ballou (2005a) y Ballou *et al.* (2004) se centran en la importancia de los controles de antecedentes estudiantiles usando modelos basados empíricamente en los cuales calculan los efectos docentes. Descubren que la omisión de controles de antecedentes estudiantiles, como raza y pobreza, en los modelos de efectos docentes de valor agregado rara vez produce errores importantes en las cifras estimadas de desempeño docente cuando las comparaciones entre maestros son dentro del mismo distrito escolar. Sin embargo, es importante hacer notar que esta investigación se basa en un determinado distrito con importante integración racial y económica (Ballou, 2005b).

Otro factor de complicación cuando se trata de usar los métodos del VAM para calcular la eficacia docente es la asignación de estudiantes a los maestros. La asignación de maestros a escuelas y clases, así como las preferencias de los padres por determinados maestros hace muy improbable que los maestros y estudiantes se emparejen aleatoriamente entre sí en EUA.<sup>20</sup> De hecho, si ese emparejamiento fuera al azar, no tendríamos que preocuparnos

mucho sobre incluir controles de antecedentes estudiantiles en los modelos estadísticos porque los maestros no serían asignados de manera sistemática a distintos tipos de estudiantes. Tampoco tendríamos que preocuparnos si tuviéramos información completa sobre los antecedentes de todos los estudiantes y si ésta pronosticara perfectamente la contribución del medio ambiente y del hogar del estudiante (compañeros, etc.) al aprendizaje de los alumnos. Si éste fuera el caso, simplemente podríamos usar la información en modelos estadísticos para ajustar las expectativas de cómo la clase de un maestro debe funcionar para el tipo de estudiantes que tiene. Pero como los estadígrafos no tienen antecedentes completos (por ejemplo, no saben la cantidad exacta de tiempo que los padres de un alumno pasan estimulándolo y ayudándolo con el trabajo escolar) y las calificaciones de las pruebas son medidas imperfectas de la capacidad escolar, un componente de cualquier término de error del VAM será la capacidad escolar no medida. Y, es probable que este error se correlacione con el maestro al cual está asignado un estudiante dado un proceso de emparejamiento no aleatorio entre maestros y estudiantes. Esta correlación viola una suposición fundamental de los modelos estadísticos que es necesaria para asegurar que los hallazgos del modelo, como las cifras estimadas de la eficacia docente, no sean erróneos.

Existe un buen motivo para creer que el proceso de emparejar a estudiantes y maestros es más que sólo una preocupación teórica. Por ejemplo, como se menciona antes en este capítulo, una cantidad importante de investigación en **EUA** muestra que los estudiantes más privilegiados, en cuanto a ingreso familiar y educación de sus padres, tienden a ser asignados a maestros de calidad superior, según lo medido por características como experiencia, nivel universitario y resultados de los exámenes (Lankford *et al.*, 2002; Loeb, 2001). Además, investigación realizada por Clotfelter *et al.* (2006) demuestra que un emparejamiento no aleatorio es probable que produzca cifras estimadas del coeficiente docente con sesgo ascendente en los modelos estadísticos que no logran explicarlo.

Quizá sea posible explicar el proceso de emparejamiento al calcular modelos estadísticos más complejos (a los que se hace referencia como especificaciones del modelo “efectos fijos del estudiante” o de la “escuela”), pero en muchos casos esto es difícil en vista de la información disponible sobre escuelas, maestros y estudiantes.<sup>21</sup> Es por esta razón que algunos investigadores (por ejemplo, Rothstein, 2008a) advierten contra el uso de cifras estimadas del VAM relativas al efecto del maestro para fines de rendición de cuentas en los que hay mucho en juego, como el pago por desempeño. Aunque ésta es una preocupación, en equilibrio, al parecer vale la pena usar el VAM ya que el sistema que normalmente se usa en la actualidad para distribuir los sueldos de los maestros parece que no está bien alineado con la productividad docente. Además, hay formas estadísticas de limitar el probable error relacionado con las medidas del VAM de la eficacia docente. Por ejemplo, Rothstein (2008b) hace notar que los modelos del VAM que incluyen una serie de logro estudiantil anterior tienen menos probabilidades de producir cifras estimadas sesgadas de los efectos docentes. Además, Kane y Staiger (2008) ponen a prueba directamente la posibilidad de sesgo en las cifras estimadas de efectos docentes aprovechando un experimento en que los maestros se emparejan aleatoriamente con las clases. Evalúan diversas especificaciones del VAM y encuentran que la especificación de un modelo que incluya las calificaciones de las pruebas de los alumnos de varios años anteriores (a diferencia de un año anterior único, como es común) produce efectos docentes muy similares a los producidos en condiciones en que los maestros son emparejados en forma aleatoria con sus clases. Por último, según muestran Koedel y Betts (2009), evaluar a los maestros usando información de múltiples años sobre la clase maestro-alumno también es probable que modere el sesgo potencial que surge del proceso de emparejamiento.

Estrechamente relacionado con los problemas generales de la especificación del VAM está el hecho de que el VAM depende de las propiedades psicométricas de los exámenes que quizá existan o no. Lo más importante entre éstos es el requisito de que los incrementos en logro de distintos alumnos – quienes pueden tener distintos niveles de logro o hasta de nivel escolar – puedan compararse entre sí. Eso es posible si la escala con la cual se mide el logro tiene unidades iguales para todos los estudiantes; de modo que el incremento en logro de un

estudiante pueda compararse de manera significativa con otro, incluso si ambos no empiezan en el mismo punto en la distribución del logro. Algunas pruebas se diseñan explícitamente para medir el crecimiento, por lo general emplean la “teoría de respuesta al ítem” para crear calificaciones escalonadas “alineadas verticalmente”.<sup>22</sup> Sin embargo, muchas pruebas no se diseñan de manera que tengan escalas de intervalos iguales dentro y entre todos los niveles escolares sometidos a examen para facilitar las comparaciones de los estudiantes.

Entre las posibles implicaciones de esto es que quizá no se compare con facilidad a los maestros. Por ejemplo, una prueba quizá tenga pocos reactivos de nivel alto, dificultando que los estudiantes de alto rendimiento aumenten sus calificaciones por mucho; por consiguiente, los métodos de valor agregado medirían incorrectamente (al subestimar) las contribuciones de maestros con relativamente más de esos estudiantes en sus grupos.<sup>23</sup> Por otra parte, una prueba que se concentre demasiado en reactivos de alto nivel puede ser inadecuada para identificar las necesidades de estudiantes de bajo rendimiento y pasar por alto los incrementos en aprendizaje de un año al siguiente logrados por esos estudiantes, o ambas cosas.<sup>24</sup> En trabajo reciente (Koedel y Betts, de próxima aparición) se ha abordado si las cifras estimadas de eficacia docente de valor agregado pueden ser sensibles a los topes (como estudiantes que sacan puntajes cerca del máximo mismo de una escala de pruebas) en el instrumento de pruebas; pero otro trabajo empírico breve aborda la gradación de pruebas según una escala y el valor agregado directamente. Un método usado para compensar la falta de alineación vertical en un instrumento de pruebas es colocar los resultados de las pruebas de los alumnos en una norma de medición separada; por ejemplo, normalización estándar de las calificaciones de logro estudiantil y comparar cuando los alumnos con determinados maestros queden dentro de la distribución de un año al siguiente. Es poco claro si este tipo de transformación explica satisfactoriamente la alineación no vertical de las pruebas.

Quienes deseen hacer juicios sobre los maestros individuales a partir de las pruebas de los alumnos sin duda les irá mejor si lo hacen aplicando pruebas graduadas adecuadamente. Sin embargo, la gradación según una escala no es el único asunto psicométrico del que uno debe ocuparse cuando se trata de pruebas usadas para fines de rendición de cuentas. Como se describe más adelante, el error de medición es un problema importante al evaluar el desempeño docente; el error de medición dependerá de la calidad y fiabilidad de los reactivos en la prueba o de pruebas que se usen para ese fin. En este contexto vale la pena hacer notar que un informe reciente de RAND suscita inquietud sobre las propiedades psicométricas de las pruebas de estudiantes que forman parte del programa de **México**, Carrera Ministerial. Observa, por ejemplo, que no hay documentación que describa datos concretos sobre la calidad de los reactivos de las pruebas, y que los reactivos de las pruebas a menudo son modificados de manera informal sin pensar en las consecuencias psicométricas de cambiarlos, por ejemplo, la validez de la prueba (Santibáñez *et al.*, 2007).

Relacionado con el tema del error de medición está el asunto de cómo decidir cuanta información sobre estudiantes-maestros emparejados es necesaria antes de que sea apropiado hacer juicios sobre los maestros. El rendimiento de un estudiante en una prueba reflejará tanto la verdadera capacidad o el conocimiento así como influencias aleatorias (por ejemplo, enfermedad en los días en que se aplican las pruebas) al que se alude como error de medición o “ruido” estadístico. Cuando se trata de evaluar el desempeño laboral, este origen del error de medición de las pruebas significa que un componente de cualquier efecto docente calculado es el origen del error de medición de las pruebas. A esto se agrega, en el caso de los maestros, la suerte del sorteo en los estudiantes que se asignan a los maestros.

Los investigadores han examinado este tema exhaustivamente en el caso de las cifras estimadas de desempeño a nivel de escuela. Hallazgos de **EUA** (Kane y Staiger, 2001) y **Chile** (Chay *et al.*, 2005) demuestran que una cantidad considerable de “ruido” se debe al error de medición de las pruebas y además eso implica que las escuelas pequeñas tienen muchas más probabilidades que las grandes de ser clasificadas como muy eficaces o ineficaces; pero es menos probable que esa clasificación identifique con precisión la eficacia de esas escuelas.<sup>25</sup>

En el caso de maestros individuales, el error de medición suscita mayor inquietud ya que las cifras estimadas del VAM de desempeño laboral docente generalmente se basarán en muestras de estudiantes más pequeñas que son cifras estimadas del desempeño a nivel de la escuela. Eso significa que muchos maestros no se distinguirán uno de otro desde un punto de vista estadístico. Por ejemplo, en un análisis de cifras estimadas de desempeño individual por año y maestro, Ballou (2005b) descubre en la abrumadora mayoría de los casos que los efectos docentes individuales no pueden distinguirse estadísticamente del maestro típico (al nivel de confianza de 90%): en lectura sólo 2.5% de los efectos docentes de primaria y 7.6% de los de secundaria fueron diferentes al promedio; en las cifras correspondientes los efectos fueron ligeramente superiores en 17% y 30% respectivamente para maestros de primaria y de secundaria.

El error de medición necesariamente también significa que los maestros evaluados a partir de cifras estimadas del VAM que estén a cualquier lado de un valor de corte no se distinguirán estadísticamente. Por ejemplo, si un país adoptara una política de incentivos para premiar a los maestros que quedan comprendidos en el quintil superior de desempeño, los maestros del percentil 79 de desempeño calculado podrían presentar fuertes argumentos estadísticos de que ellos son tan eficaces como los maestros que están en el percentil 81 de desempeño calculado.

El tema del error de medición a últimas fechas ha recibido mucha atención en el contexto de los maestros individuales. Por ejemplo, varios investigadores han usado diversos métodos afines para descomponer la varianza total anual de las cifras estimadas de eficacia docente en cifras estimadas de efectos docentes “verdaderos”, “señal” y error de muestreo o “ruido” (véase, por ejemplo, Aaronson *et al.*, 2007; Goldhaber y Hansen, 2008a; Koedel y Betts, 2009a; McCaffrey *et al.*, 2008; Rothstein, 2008a, 2008b); pero encuentran las correlaciones de las cifras estimadas de desempeño laboral del valor agregado del maestro de un año al siguiente que se cree sean “moderadas” (véase Aaronson *et al.*, 2007; Ballou, 2005; Koedel y Betts, 2007; Goldhaber y Hansen, 2008a; McCaffrey *et al.*, 2008); y una proporción no insignificante de variación intertemporal en esos efectos parece deberse a ruido estadístico. Por ejemplo, la correlación intertemporal de las cifras estimadas del efecto docente del VAM está entre los límites de 0.3 a 0.5, y generalmente es más alto si se usa el logro en matemáticas del estudiante como una norma de medición que cuando se usa el logro en lectura (Goldhaber y Hansen, 2008a). Además, clasificar a los maestros a partir de esas cifras estimadas demuestra, a nivel de primaria Koedel y Betts, 2007) como de secundaria a la vez (McCaffrey *et al.*, 2008), que aproximadamente de un cuarto a un tercio de maestros clasificados en el quintil superior de desempeño (basado en los VAM) permanecen en el mismo quintil de un año al siguiente; 10%-15% de los maestros recorren todo el camino desde el quintil inferior al superior; y 10%-15% descienden del quintil superior al inferior.

Definitivamente uno no puede distinguir cuánto de la variabilidad a lo largo del tiempo en los efectos docentes se debe a ruido comparado con cambios verdaderos en el desempeño docente; pero una forma de mejorar la estabilidad de esas cifras estimadas es usar múltiples años para calcular los efectos docentes; y esas cifras estimadas de múltiples años serán menos propensas a ser ruidosas porque un sorteo afortunado en un año es contrarrestado por otro desafortunado en el siguiente (Goldhaber y Hansen, 2008a; Koedel y Betts, 2009). También bien vale la pena poner estos temas en contexto. No hay una norma universal respecto a lo que constituye un nivel de estabilidad aceptable de desempeño laboral, pero las correlaciones intertemporales informadas líneas atrás no son muy distintas de las cifras estimadas de desempeño laboral en sectores de la economía que las consideran para fines en los que hay mucho en juego (como retención laboral y determinación del sueldo) donde las correlaciones de desempeño de un año a otro fluctuaron de 0.26 a 0.55 (véase Hoffman *et al.*, 1992, 1993).

Dos problemas técnicos que se presentan al usar las calificaciones de pruebas para evaluar la eficacia docente se relacionan con preguntas sobre el grupo de comparación contra el cuál han de ser valorados los maestros.

Los maestros pueden ser comparados con otros colegas dentro de varios límites geográficos o institucionales. Por ejemplo, los maestros pueden ser comparados con otros maestros dentro de la misma escuela, distrito o estado.<sup>26</sup> La desventaja de modelos menos localizados que crean comparaciones de maestros entre las escuelas es que combinan efectos docentes con efectos que pueden ser importantes en otros niveles del sistema educativo, por ejemplo escuelas o distritos (Goldhaber y Hansen, 2008a). Todos los maestros de una escuela con un director muy eficiente se beneficiarán de la influencia del director en el logro estudiantil en modelos que permitan las comparaciones entre escuelas.

La desventaja de comparaciones más localizadas es que puede ser bastante difícil crear grupos de comparación contrafactual. Es probable que esto sea especialmente problemático, por ejemplo, en el caso de estructuras de gobernabilidad de escuelas más pequeñas (como los distritos escolares pequeños), ya que éstas necesariamente tendrán menos maestros en quien basar sus evaluaciones de eficacia relativa. Y en casos extremos quizá no haya grupo de comparación aplicable; por ejemplo, si sólo hay un maestro de matemáticas en una determinada región de un país. Además, los maestros que son evaluados por ser muy eficaces en una escuela, como en el caso de cifras estimadas del VAM dentro de la escuela, de hecho pueden ser relativamente ineficaces al ser comparados con un conjunto más amplio de maestros, digamos en un estado en conjunto.

No hay una respuesta exacta definitiva para el nivel correcto de comparaciones de maestros. En un marco de valor agregado, los efectos docentes se identifican en relación con el efecto de otros maestros, no de algún estándar objetivo. Por tanto, una clave para obtener buenas cifras estimadas de la contribución relativa de un maestro a la evolución del logro estudiantil es que un maestro dado se compare contra otros que enseñan a estudiantes comparables en un entorno escolar comparable. En función de las influencias relativas de los antecedentes de los estudiantes contra otros factores escolares, estatales o regionales, uno podría presentar información para cualquier tipo de comparación; pero teniendo en cuenta la abrumadora investigación que muestra la influencia de los antecedentes de un estudiante en el logro, quizá hay más de un argumento para el conjunto de comparaciones más amplio sobre esos que están más localizados.

**Alternativas para el valor agregado usado para evaluar el desempeño docente.** La subsección anterior aborda la idea de usar métodos estadísticos para fraccionar el valor agregado de maestros en el rendimiento de las pruebas de estudiantes, como una norma de medición para el desempeño docente. Aunque intuitivamente atractivo, éste no es el único medio de evaluar a los maestros. Como señalan Toch y Rothman (2008), pocos sistemas de evaluación docente parecen ser muy rigurosos o han sido probados estadísticamente para ver si corresponden a las cifras estimadas del VAM de la eficacia docente.

Hay pocas excepciones. Por ejemplo, el Consejo Nacional de Estándares Profesionales de Enseñanza (NBPTS), con sede en EUA, es una institución específicamente diseñada para evaluar el conjunto de habilidades de los maestros con experiencia. Y como se menciona antes, alguna investigación muestra que el proceso de evaluación del Consejo Nacional sí distingue a los maestros eficaces de los ineficaces, si bien con cierto error.<sup>27</sup> Y una investigación de un determinado sistema de desempeño por escalafón, donde el ascenso de un maestro (que lleva a mayor compensación) se basa en buena parte en evaluaciones de supervisores de la asignatura, muestra que los maestros que avanzan a través del sistema son más eficaces en términos de valor agregado que quienes no ascienden.<sup>28</sup> Este estudio es muy convincente porque se basa en un experimento en que los alumnos se asignaron aleatoriamente a los maestros, eliminando la posibilidad de que los hallazgos sean un subproducto del emparejamiento entre alumnos y maestros.<sup>29</sup>

Por último, Jacob y Lefgren (2008) correlacionan las evaluaciones que los directores hacen de los maestros con las cifras estimadas del efecto del valor agregado y encuentran que los directores generalmente hacen un buen trabajo al identificar a los maestros muy eficaces y a los muy ineficaces; pero son menos capaces de discriminar entre maestros que están en medio de la distribución de eficacia. De manera importante, los autores también

encuentran evidencia de que las evaluaciones que los directores hacen de los maestros pronostican el logro estudiantil e identifican los atributos de los docentes correlacionados con su éxito, que trasciende lo que uno puede detectar simplemente a partir de medidas estadísticas de la eficacia docente.<sup>30</sup>

En resumen, los estudios anteriores indican que uno puede utilizar medios distintos a las calificaciones de las pruebas de los alumnos para evaluar a los maestros. Sin embargo, es una pregunta abierta en cuanto a si los sistemas escolares que dependen de otros medios de hecho distinguen entre los maestros. El problema es que, en muchos lugares, las limitaciones políticas e institucionales (como las relaciones escuela/sindicato de maestros) dificultan hacer evaluaciones rigurosas de los maestros (Toch y Rothman, 2008), lo cual hace que muy pocos maestros reciban una calificación alta (New Teacher Project, 2007). Podría argumentarse que una de las principales razones de que a los responsables de la política educativa les interese la idea de usar (métodos de valor agregado) estadísticos para identificar la eficacia de maestros individuales es sortear la dificultad de conseguir que los directores y otros evaluadores hagan evaluaciones veraces del desempeño docente.

**Resumen sobre cómo medir la eficacia docente.** Indudablemente la investigación antes descrita ofrece lecciones valiosas sobre las formas adecuadas de medir la eficacia docente. Sin embargo, es una exageración señalar que las discusiones sobre cómo medir con precisión las contribuciones docentes al aprendizaje de los estudiantes se hayan calmado. Pero tras haber descrito algunos de los problemas que se presentan en el contexto del valor agregado, parece adecuado señalar que esta metodología no debe evaluarse contra cierto nirvana imaginado sino, más bien, contra el sistema que existe. Y aquí la alternativa aplicable en la mayoría de los países basa la compensación y el empleo docente en títulos que en sí mismos se relacionan débilmente, en el mejor de los casos, con el aprendizaje de los estudiantes.

En algunas formas es adecuado permitir que la teoría de la acción personal para la manera en que los incentivos para maestros individuales podrían influir en la calidad de la fuerza laboral sea una guía para las formas en que se miden los efectos docentes individuales. Probablemente hay ventajas y desventajas entre la exactitud de las cifras estimadas del VAM de los efectos docentes y la transparencia para los maestros acerca de cómo se calculan sus efectos y su opinión sobre el incentivo. Así que, por ejemplo, si la idea es identificar a los maestros más eficaces y proponerlos como modelos de conducta para que los demás maestros aprendan de ellos; entonces es muy importante que se identifique a los maestros correctos. Esto indica un modelo complejo de valor agregado. Sin embargo, si la teoría es que los maestros respondan a los incentivos para el desempeño; quizá sea más importante que entiendan cómo serán evaluados que las cifras estimadas del VAM sean tan exactas como sea posible. El desempeño docente varía de un año a otro, entonces aunque es probable que las cifras estimadas de múltiples años de los efectos docentes sean más exactas son un incentivo más débil porque los años de alto valor agregado contrarrestan los años de bajo valor agregado y viceversa (Rivkin, de próxima aparición). Por consiguiente, cabe la posibilidad de que las políticas de incentivos más eficaces podrían sacrificar alguna exactitud por incentivos más marcados.

**El tamaño y el costo de los incentivos para maestros individuales** Dos problemas importantes que surgen cuando se trata de vincular las medidas de eficacia docente con los incentivos para el desempeño docente son el tamaño de los incentivos (por ejemplo, qué porcentaje del paquete de compensación total del maestro está ligado al incentivo) y el costo de un sistema de incentivos.

Claro que una crítica de muchos sistemas de incentivos para maestros es que el tamaño del incentivo es muy pequeño para motivar cambios en la conducta de los maestros. Por ejemplo, investigación sobre el SNED en Chile, muestra que sólo un porcentaje muy pequeño de maestros en el país son elegibles para un bono de desempeño; y que el componente del desempeño en los sueldos es pequeño en relación con la compensación total y otros factores (como experiencia) que influyen en ellos; por ejemplo, en el 2003, menos de un punto porcentual del sueldo de un maestro se basaba en el desempeño (Mizala y Romaguera, 2005).

La teoría económica indicaría que incluso un incremento de sueldo muy pequeño estructurado como incentivo debe tener efectos marginales (Prendergast, 1999); pero, en la práctica, los programas de incentivos por desempeño que se han puesto a prueba en escuelas públicas suelen tener incentivos que son muy pequeños en comparación con los que hay afuera de las escuelas públicas. Ballou (2001), que investiga los incentivos de pago por desempeño en escuelas de **EUA**, encuentra que el incentivo docente promedio en las escuelas particulares es de aproximadamente 10% del sueldo total; mientras que sólo es de más o menos 2% en las escuelas públicas con incentivos. Estudios del sector privado indican que la parte del sueldo que se basa en el incentivo varía considerablemente entre empresas e industrias diferentes (Murphy, 1999; Prendergast, 1999); pero, en general, la parte del sueldo ligada al incentivo es mucho más alta que incluso el 10% observado en las escuelas particulares.

La siguiente sección describe la evidencia empírica de los efectos de distintos sistemas de incentivos. En general, los datos presentados no parecen mostrar una clara correspondencia entre la distribución o el tamaño de los sistemas de incentivos para maestros y los efectos de esos sistemas. Esto no es una tremenda sorpresa ya que relativamente hay pocos estudios que evalúen los efectos de los incentivos de desempeño docente.

Otro factor importante que los responsables de la política educativa deben considerar al diseñar sistemas de incentivos para maestros es su costo. Hay investigación que ahonda en el costo de distintos diseños, pero por lo menos hay dos buenas razones para creer que esos sistemas deben aumentar los costos educativos (aunque es de esperar que con un incremento más que compensatorio en el logro estudiantil) al menos en el mediano plazo. La primera es obvia: serán más costosos de administrar ya que exigirán evaluaciones de los maestros individuales y capacidad administrativa para rastrear los pagos que se les hagan. La segunda es menos obvia pero posiblemente más importante. En el grado en que los incentivos representen una parte importante de la compensación total de los maestros, los maestros tendrían que recibir una prima de riesgo ya que su incertidumbre sobre la compensación total sería mayor de la que tendrían en un sistema basado en un tabulador de sueldos.

## EVIDENCIA ACERCA DE SISTEMAS DE INCENTIVOS PARA MAESTROS PARA FINES DETERMINADOS

Al parecer hay un creciente interés en diseñar incentivos para maestros para atraer talento a la enseñanza, premiar a maestros de alto desempeño y redistribuir maestros en diversas escuelas y áreas geográficas. Esto aplica en **EUA** donde el gobierno federal hace poco instituyó el Fondo de Incentivos para Maestros (TIF), que da subvenciones a los estados o localidades para desarrollar alternativas de sueldo al tabulador tradicional (Goldhaber, 2006b), así como en varios países europeos (Sclafani and Tucker, 2006). **México**, como parte del programa Carrera Magisterial, fue uno de los primeros países en vincular de manera explícita la compensación de los maestros con el desempeño docente en todo el territorio, aunque los incentivos para maestros basados en el desempeño también se usan en otros países como Singapur (Sclafani, 2008).

A pesar del interés en usar incentivos para influir en los maestros; en definitiva, se sabe relativamente poco sobre el tipo correcto de incentivos para influir en la distribución o desempeño docentes. Esto, a la vez, es porque pocos países usan en forma generalizada los incentivos para maestros y porque estudiar los efectos de los incentivos en forma rigurosa requiere datos que a menudo no están disponibles. La inquietud al evaluar los programas es que a menudo no hay grupos adecuados de comparación contrafactual para maestros que estén supeditados a un programa de incentivos y, en esas circunstancias, la investigación puede combinar el efecto de un incentivo con el efecto de quienes reciben un incentivo. Por ejemplo, algunos programas de desempeño, como el de **México**, Carrera Magisterial, están diseñados de tal manera que la participación en el sistema de incentivos es voluntaria (McEwan y Santibáñez, 2005). En este caso, los mismos maestros eligen ingresar en la categoría del programa y uno podría preocuparse, por ejemplo, de que maestros más motivados que creen que

tendrían un buen desempeño decidan ingresar de manera desproporcionada al programa. Si fuera este el caso, una comparación sencilla entre los participantes del programa y quienes no participan puede imputar mal la repercusión del tipo de maestros que decidan participar en el efecto del mismo programa de incentivos. Aquí se describen algunos de los estudios sobre incentivos que se han diseñado de manera explícita para subsanar el problema de quienes se eligen ellos mismos para algo; para un análisis más extenso sobre los distintos sistemas de incentivos para maestros y sus efectos, véase el capítulo 3.

La mayor parte de los estudios sobre incentivos se centran en los incentivos financieros previstos para el desempeño docente. Lavy (2002) and Glewwe *et al.* (2003) se centran en los incentivos por desempeño a nivel de la escuela. La investigación de Lavy utiliza un enfoque de “discontinuidad en la regresión» para comparar los resultados de estudiantes en 62 escuelas en **Israel** que están sujetas a incentivos por desempeño con los de escuelas que acababan de perder la oportunidad de ser parte del sistema de incentivos (el tratamiento); es probable que estos dos grupos de escuelas sean muy similares en cuanto a factores docentes inadvertidos (y otros) en virtud del hecho de que la diferencia entre ser parte del grupo de tratamiento o del grupo de control, en términos aleatorios, probablemente es de alrededor del valor de corte que determina la elegibilidad para participar en el programa de incentivos.<sup>32</sup>

Los hallazgos de Lavy demuestran que participar en el programa de incentivos tiene impactos positivos estadísticamente significativos en varios resultados medidos: calificaciones de las pruebas de los alumnos, número de créditos tomados y estudios terminados para el examen de ingreso a la enseñanza superior. Además, los efectos fueron más grandes en los estudiantes de origen más desfavorecido. Este hallazgo es coherente con la investigación sobre incentivos por desempeño docente en EUA (Figlio y Kenny, 2007) que encuentra efectos más importantes de los incentivos en escuelas que atienden a estudiantes desfavorecidos. Señala que los responsables de la política educativa que están inseguras sobre la eficacia de esos incentivos en general podrían desear implementarlos primero en las escuelas más necesitadas ya que se cree que éstas son las que se benefician más de ellos.

Glewwe *et al.* (2003) se centran en un experimento real en **Kenya** donde maestros de 50 escuelas, escogidas al azar entre 100, se asignaron de manera aleatoria a un grupo experimental, sujeto a incentivos por desempeño, y se compararon con maestros en un grupo no experimental que no formaban parte del sistema de desempeño. Descubren que los maestros incentivados aumentaron la cantidad de sesiones para preparación de exámenes que ofrecían después de la escuela; pero sólo hubo un aumento de un año en las tasas de aprobación de los estudiantes en el examen estatal. Esos hallazgos pueden interpretarse en el sentido de que los maestros aumentaron sus niveles de esfuerzo, pero los autores también hacen notar que uno podría considerarlos como maestros que tratan de manipular al sistema al centrarse de manera limitada en la preparación de exámenes. Esta preocupación es puesta de relieve por el hecho de que no hallaron un efecto en las ausencias de los maestros ni en las tácticas pedagógicas; y que los incrementos en logro estudiantil no persistieron después del programa.<sup>34</sup>

Varios estudios rigurosos también se centran en los efectos del pago por desempeño a nivel de maestro individual. Lavy (2004) usa un enfoque de discontinuidad en la regresión para evaluar un programa de incentivos por desempeño docente en preparatoria en **Israel**.<sup>35</sup> Descubre efectos positivos del tratamiento estadísticamente importantes (en las tasas de inscripción en preparatoria); y, además, la importancia de los efectos indica que el incentivo por desempeño a nivel docente individual es más rentable que los bonos basados en las escuelas o que el tiempo de instrucción extra (y más o menos comparable con los bonos por incentivos pagados directamente a los estudiantes por logro).

Santibáñez *et al.* (2007) investigan los efectos de la Carrera Magisterial de **México**, un programa de incentivos para maestros de largo alcance introducido en 1992, con la intención de mejorar el desempeño de los maestros

mediante el desarrollo de habilidades así como de incentivos financieros explícitos por resultados en logro estudiantil.<sup>36</sup> Al igual que Lavy (2002, 2004), utilizan un enfoque de discontinuidad en la regresión para tratar de aislar las repercusiones del efecto del programa. Sin embargo, a diferencia de Lavy, esta investigación informa hallazgos mixtos en las repercusiones: se descubrió que el programa tiene efectos negativos (aunque no estadísticamente importantes) en los estudiantes de primaria y algunos efectos positivos en estudiantes de secundaria (en el orden de magnitud de 3-15% de una desviación estándar); pero esos resultados positivos sólo se hallaron para los maestros – alrededor del 4% de su muestra – que enfrentaban incentivos muy fuertes. Además, hallaron disminuciones en el rendimiento de los estudiantes posteriores a que los maestros recibieran los bonos o fueran promovidos como parte del programa; quizá debido a la reversión en la media en las calificaciones de pruebas o debido a que los incentivos para que los maestros mejoren las calificaciones de las pruebas de los alumnos sean mucho más débiles una vez que los maestros se integran en sistema Carrera Magisterial y son promovidos, o ambas cosas (Santibáñez *et al.*, 2007).

Aunque no definitivos, podría interpretarse que los hallazgos de Lavy y Santibáñez *et al.* muestran la importancia de la percepción que tienen los maestros sobre el sistema de incentivos. Los incentivos débiles, por ejemplo los que son pequeños en magnitud o poco precisos porque aplican a grupos, quizá no tengan el efecto de los incentivos que se consideran son más importantes y se basan en el desempeño individual. Es indudable que esta conclusión es coherente con la investigación sobre incentivos por desempeño docente en **EUA** realizada por Figlio y Kenny (2007), quienes descubren que los incentivos por desempeño no influyen en el logro estudiantil cuando son recibidos por un porcentaje grande de maestros en una escuela; pero si se otorgan en forma selectiva se correlacionan de manera positiva con el logro estudiantil. Por tanto, como es el caso con muchas intervenciones educativas, los detalles de la implementación quizá sean tan importantes para el éxito de la intervención como la naturaleza de la intervención misma.

Por último, Muralidharan y Sundararamin (2008) informan sobre los hallazgos de un experimento muy importante sobre intervenciones para incentivos para maestros y recursos escolares en una muestra representativa de 500 escuelas primarias rurales dirigidas por el gobierno, en **India**. Encuentran efectos muy importantes (0.19 en pruebas de matemáticas y 0.12 en pruebas de lengua; efectos más grandes en los hallados en los estudios antes descritos) relacionados con estar en una escuela con tratamiento de incentivos.<sup>37</sup> Además, los beneficios del tratamiento se hallaron tanto en los reactivos de pruebas de bajo nivel como en más conceptuales, y los alumnos de escuelas con tratamiento superaron a los de escuelas sin tratamiento en asignaturas que no eran el objetivo del incentivo (ciencias y estudios sociales). Estos hallazgos indican que: 1) los maestros no se centran en estrategias de pruebas limitadas que se hayan diseñado para incrementar el logro estudiantil; y, 2) los incentivos, en lugar de hacer que los maestros sustituyan asignaturas que no abarca el incentivo, produjeron efectos indirectos positivos en esas asignaturas.

A diferencia de Lavy, Muralidharan y Sundararamin no encontraron ninguna diferencia estadísticamente importante entre los diseños de incentivos individuales y colectivos, pero sus hallazgos son coherentes con Lavy, quien descubrió que el incentivo docente es más rentable que otras intervenciones de recursos escolares.<sup>38</sup> Un aspecto especialmente interesante del estudio es que analizan encuestas que se aplicaron a los maestros, en las que se les preguntó como asignarían un aumento en el presupuesto de 15% entre pago por desempeño y un aumento general de sueldos. Al mantener constante el sueldo esperado, los investigadores encuentran una relación positiva estadísticamente significativa entre las preferencias de los maestros por el pago por desempeño y los incrementos promedio de sus alumnos (la encuesta se aplicó antes de que los maestros conocieran esos incrementos). Eso indica que los maestros eficaces saben quiénes son y que desean ser recompensados por su desempeño; y, además, que los sistemas de pago basado en el desempeño pueden producir beneficios en la productividad de la fuerza laboral relacionados con más clasificación, es decir, que personas con talento para enseñar decidan ingresar a la docencia y que los maestros más eficaces decidan quedarse en la profesión. Esta idea es congruente

con investigación más general sobre el mercado laboral; por ejemplo, según Lazear (2000), aproximadamente la mitad de los incrementos por el pago por desempeño se deben a mayores esfuerzos entre los empleados actuales y la mitad a trabajadores productivos que están clasificados en un sistema basado en el desempeño.

Hay estudios mucho menos rigurosos sobre el efecto de otro tipo de incentivos financieros (como apoyo para la vivienda o exoneraciones de colegiatura) o no financieros; o estudios de incentivos diseñados para atraer maestros con determinados conjuntos de habilidades (como matemáticas o ciencias) o para distribuirlos en ciertas escuelas. Gran parte de las publicaciones empíricas en esta área se centran en las inferencias que se extraen al observar el comportamiento del mercado laboral a falta de programas de incentivos: qué tipo de personas deciden ingresar y quedarse en el mercado laboral docente y si optan por enseñar.<sup>39</sup> Aunque sin duda esta investigación es valiosa, uno debe ser cauteloso respecto a sacar conclusiones sólidas de ésta ya que las cifras estimadas de diferenciales compensatorios necesarios, por ejemplo, pueden confundirse por la heterogeneidad inadvertida entre quienes deciden proseguir una determinada profesión o cambiarse de una escuela a otra.<sup>40</sup>

Los estudios que se centran en programas verdaderos basados en incentivos son más creíbles. Uno de ellos (Clotfelter *et al.*, 2008) se centra en el efecto de un aliciente financiero para enseñar en escuelas para las que es difícil encontrar personal. Este estudio analiza un programa en **Carolina del Norte** diseñado para retener a maestros de ciencias y matemáticas en escuelas públicas económicamente desfavorecidas (en Carolina del Norte). Esta investigación encuentra que un bono relativamente pequeño (USD 1 800, que valía aproximadamente 5% del sueldo de los maestros típicos) reducía las tasas de rotación apenas en más de 15%. En fechas más recientes, Steele *et al.* (2009) se centran en un programa de incentivos (en **California**) diseñado para persuadir a maestros principiantes talentosos académicamente para que acepten puestos y permanezcan (mínimo cuatro años) en escuelas de bajo desempeño.<sup>41</sup> Este programa proporcionó incentivos mucho más grandes que los de Carolina del Norte, USD 5 000 por cada año de servicio (hasta cuatro años), lo que equivalía a más o menos 15% de los sueldos iniciales promedio. Quizá no sea inesperado, en vista de la magnitud de los incentivos, los autores encuentran que el incentivo aumentó en aproximadamente 30 puntos porcentuales la probabilidad de que sus beneficiarios decidieran enseñar en escuelas de bajo desempeño.<sup>42</sup>

Si bien hay poco trabajo empírico para orientar a los responsables de la política educativa sobre la eficacia de los incentivos para maestros no financieros, en conjunto, la investigación descrita en esta sección señala con claridad que los incentivos pueden tener efectos importantes en la conducta de los maestros. Pero, como describo con brevedad en la siguiente sección, el éxito de los sistemas de incentivos es probable que dependa más del mismo sistema.

## CONCLUSIÓN: INFRAESTRUCTURA PARA APOYAR LOS SISTEMAS DE INCENTIVOS

Este capítulo ha descrito algunas de las deficiencias de la adhesión estricta al tabulador único para sueldos como la principal estructura para la compensación de los maestros. Una de las reformas que ha captado una atención generalizada es ligar los incentivos a las cifras estimadas de valor agregado del desempeño docente (o escolar). La investigación sobre las pruebas y requisitos de información necesarios para hacerlo, como que las cifras estimadas de desempeño sean descripciones exactas del verdadero desempeño docente o escolar, es relativamente nueva, así que no hay directrices claras para cómo deben medirse precisamente esos efectos. En el mejor de los casos, sabemos que las medidas de desempeño, especialmente las que se centran en los maestros individuales, es probable que sean ruidosas y que necesariamente sólo reflejen el material que cubren esas evaluaciones de las pruebas. Esto no necesariamente significa que el valor agregado no deba ser utilizado, ya que otras formas de evaluar a los maestros tampoco han sido probadas. Más bien, probablemente es una mala idea que los responsables de la política educativa basen los incentivos exclusivamente en medidas de valor agregado cuando también podrían hacerlo en indicadores como las evaluaciones de supervisores o colegas.

Tener un programa de incentivos no basta para garantizar el éxito. Una de las notables excepciones a los hallazgos generalmente positivos sobre los incentivos por desempeño docente, analizado líneas atrás, es el programa de incentivos de **México**. La investigación de Santibáñez *et al.* (2007) y Hopkins *et al.*, (2007) encuentra que, a pesar del programa Carrera Magisterial, los incentivos por desempeño para muchos maestros mexicanos son bastante débiles. Por último, quizá más importante que el diseño de un incentivo docente específico es la implementación de programas de incentivos y de apoyo para los mismos. Por ejemplo, uno de los hallazgos importantes cuando se trata de reformas de incentivos es que los maestros en muchas ocasiones no entienden cómo funcionan los programas o incluso tampoco saben qué incentivos existen (Clotfelter *et al.*, 2008; Goldhaber, 2009; 2006b; Jacob y Springer, 2007); esto sin duda no es favorable para que los programas tengan los efectos deseados. Dicho eso, aunque la investigación sobre la reforma a la compensación de los maestros apenas es lo suficientemente definitiva para recomendar el uso de reformas específicas a los sueldos para alcanzar metas específicas, los pocos estudios cuantitativos que sí existen indican que un uso más estratégico de la compensación de los maestros podría llevar a una asignación más equitativa de los maestros entre los estudiantes y a un mayor logro estudiantil.

## NOTAS

1. En este capítulo el término “calidad docente” se refiere a la capacidad de los maestros para contribuir en formas medibles a los incrementos de los estudiantes en las pruebas estandarizadas (los términos “calidad docente”, calidad del maestro” y “eficacia del maestro” se usan en forma intercambiable). Aunque ayudar a los alumnos a aprender para que rindan mejor en las pruebas en cierto modo es una noción limitada de lo que hacen los maestros, es útil para definir la calidad docente en esta forma porque ofrece un medio para medir a los maestros a partir de los resultados medibles de los estudiantes. Como paso a describir más adelante, no se ha encontrado que las definiciones de calidad basadas en insumos (como los títulos de un maestro) se relacionen fuertemente con el logro estudiantil. Además, se ha demostrado que el rendimiento de los estudiantes en las pruebas es un importante factor predictivo de consecuencias más adelante en la vida (Grogger y Eide, 1995; Murnane *et al.*, 1995) y se relaciona con las tasas de crecimiento totales del país (Hanushek *et al.*, 2008); y, en muchos países, ayuda a impulsar políticas educativas (por ejemplo, la rendición de cuentas escolar).
2. Como es lógico, con el paso del tiempo la influencia de los maestros puede ser sustancial. Por ejemplo, según trabajo de Sanders y Rivers (1996), los estudiantes con una secuencia de tres maestros consecutivos que estén en el grupo del quintil más alto sacan entre 52-54 puntos porcentuales más alto en la distribución del logro que los estudiantes que tengan tres maestros consecutivos en el quintil más bajo.
3. Hanushek (1986, 1997) sostiene que los títulos observables de los maestros se correlacionan sólo debilmente con la calidad docente. En fechas más recientes, algunos estudios se valen de modelos de efectos fijos docentes para calcular la contribución de características docentes observables en relación con factores docentes no observables (Aaronson *et al.*, 2007; Goldhaber *et al.*, 1999; Koedel y Betts, 2007). Generalmente encuentran, basados en la mejora progresiva en R2, que la extensa mayoría de la variación total en la calidad docente (más de 95%) se explica mediante las características docentes no observables más que por las observables como niveles universitarios y de experiencia.
4. Encuentran relaciones en cierto modo más positivas entre las evaluaciones docentes por expertos, participación en el desarrollo profesional y el logro estudiantil.
5. Además, hay muchos tipos de incentivos, desde la motivación intrínseca hasta la seguridad laboral, que probablemente influyan en las preferencias del mercado laboral.
6. Para un análisis minucioso, véase Vegasy Umansky (2005).
7. En cambio, en función del grupo de comparación, se encontró que los maestros de Bolivia, Brasil, la República Dominicana, Ecuador y Nicaragua están relativamente mal pagados.
8. Además, en algunos países las prestaciones por salud y jubilación constituyen un porcentaje relativamente importante de la compensación total; y, en general, esas prestaciones no se consideran al comparar la compensación entre la enseñanza y otras ocupaciones.
9. Hay hallazgos heterogéneos sobre la relación entre los sueldos promedio y la calidad docente y el logro estudiantil. Por ejemplo, véase Ballou y Podgursky (1997), Goldhaber (de próxima aparición) y Hanushek (1986), quienes señalan que hay poca conexión entre el sueldo y la calidad docente o los resultados de los estudiantes; y Figlio (1997) y Loeb y Page (2000), quienes encuentran conexiones más fuerte entre los dos.
10. Algunos países, como Corea (Oh, 2008), informan tener sistemas de incentivos, pero casi todos los maestros reciben – por ejemplo – una gratificación por desempeño; así que, de hecho, no hay un verdadero incentivo. Una excepción notable a esto es Singapur, que ofrece a los maestros bonos cuantiosos (equivalente a uno a tres meses de sueldo) a partir de evaluaciones subjetivas tanto del verdadero desempeño docente como del potencial futuro (Sclafani, 2008).
11. Bretz and Milkovich (1989) calculan que 93-99% de las empresass del sector privado en EUA premian a las personas asalariadas a partir de cierta medida de su desempeño individual.
12. Específicamente, la diferencia calculada en el sueldo por cuatro años fuera de la universidad entre maestros y no maestros que tienen formación en matemáticas y ciencias es de USD 13 469, sólo de USD 6 811 para quienes no tienen ese tipo de preparación.
13. De hecho, pruebas recientes (Player, de próxima aparición) indican que los maestros en demanda son recompensados en el mercado laboral, no mediante una compensación sino mediante asignación a estudiantes más favorecidos.
14. Para comparar individuos con niveles similares de heterogeneidad inadvertida, Hernani-Limarino calcula, a partir de cifras estimadas de la Desviación Absoluta Mínima, diferenciales de sueldo condicionales de diversos cuantiles a lo largo de la distribución salarial condicional.

15. Véase la Gráfica 3.12 en la página 95 y el Cuadro 3.9 en la página 96 en Hernani-Limarino (2005).
16. Los economistas se refieren a esto como el «problema de las multitareas», por medio del cual al incentivar actividades que son más fáciles de supervisar hace que los empleados se centren en éstas en vez de hacerlo en otras actividades valiosas que son menos fácilmente supervisadas.
17. Los hallazgos de Koedel también indican que los métodos de valor agregado podrían aplicarse en el nivel de secundaria para aislar las contribuciones de los maestros en determinadas asignaturas. El aislamiento de contribuciones docentes sería mucho más difícil en el caso de aspectos complementarios fuertes entre asignaturas.
18. Para antecedentes sobre este tema, véase – por ejemplo – a Ballou (2005); Ballou *et al.*; McCaffrey *et al.* (2004); Rothstein (2008a, 2008b); y Todd y Wolpin (2003).
19. Por ejemplo, para obtener cifras estimadas exactas de la eficacia de los maestros A, B y C, al menos algunos estudiantes tienen que ser comunes a los maestros A y B y B y C; o a los maestros A y C y B y C. Al excluir covariables estudiantiles, la identificación de diferencias entre la población estudiantil depende de la relación cruzada entre los estudiantes. Es esta mezcla lo que garantiza que las cifras estimadas de correlación cruzada son válidas.
20. Emparejar a maestros y estudiantes puede darse oficialmente mediante la capacidad de rastrear a los estudiantes o, de manera más informal, por ejemplo si los “buenos” maestros son premiados asignándoles grupos escogidos o si los padres influyen para conseguir que sus hijos estén en clases escogidas (Player, de próxima aparición).
21. Si las especificaciones de efectos fijos controlan plenamente el proceso de emparejamiento entre alumnos y maestros, entonces los modelos estadísticos deben producir cifras estimadas sin sesgo de los efectos docentes individuales; sin embargo, si el proceso de emparejamiento se basa en los atributos inadvertidos de alumnos o maestros, entonces es probable que las cifras estimadas del efecto docente tengan sesgo. Rothstein (2008a), al advertir los problemas de emparejamiento no aleatorio entre alumnos y maestros, desarrolla pruebas de falsificación de las suposiciones que apoyan métodos usualmente utilizados para identificar el efecto de los maestros individuales usando información no experimental. Específicamente, al usar información longitudinal del grupo de expertos de Carolina del Norte, él demuestra que los maestros de quinto grado tienen efectos estadísticamente importantes en el logro estudiantil del cuarto grado en las especificaciones de los VAM sobre efectos fijos del estudiante. Claramente el efecto del maestro de quinto grado no puede ser causal (ya que los alumnos aún no han alcanzado ese grado), así que sus hallazgos implican que tanto los alumnos no son asignados aleatoriamente a los maestros como la inclusión de efectos fijos de los estudiantes no explican satisfactoriamente el proceso de emparejamiento entre maestros y alumnos (nótese que esta prueba de falsificación es una prueba para saber si existe sesgo, no del alcance del sesgo). Esto puede ocurrir si, por ejemplo, el proceso de emparejamiento entre maestros y estudiantes es “dinámico” e inadvertido por los estadígrafos.
22. Véase Ballou (2002) para más información sobre esta metodología.
23. Las pruebas con muchos reactivos de nivel más bajo detectarán con más facilidad los incrementos en logro para estudiantes de bajo rendimiento; incluso si esos estudiantes no están aumentando considerablemente sus niveles de aprendizaje.
24. Para un análisis más minucioso de la relación entre la gradación de pruebas y la construcción de modelos de valor agregado, véase Ballou (de próxima aparición).
25. La regresión a la media es un problema relacionado con el error de medición. Específicamente, un componente del logro estudiantil es la suerte, pero como es aleatoria los estudiantes que inusualmente salen bien o mal en las pruebas en un año (o aplicación de pruebas) es probable que lo hagan menos bien al siguiente simplemente porque un componente de su rendimiento, sea alto o bajo, se relacionó con la suerte y es improbable que ésta persista en pruebas repetidas. Por tanto, los maestros con alumnos de logro muy alto en un año es probable que parezcan menos eficaces en términos de valor agregado porque sus alumnos no tienen probabilidades de tener tanta suerte cuando vuelvan a hacer las pruebas.
26. Por ejemplo, los modelos que incluyen efectos fijos a nivel de la escuela generan comparaciones docentes dentro de la misma.
27. Véase a Goldhaber y Anthony (2007) para una descripción más detallada de las formas en que el Consejo Nacional evalúa a los maestros y de los hallazgos sobre maestros que obtienen ese diploma.
28. Más adelante se incluye un análisis sobre los efectos docentes de valor agregado; pero la idea básica es usar modelos estadísticos para aislar la contribución que hacen los maestros a los incrementos en logro estudiantil en las pruebas estandarizadas.
29. Varios estudios muestran que este proceso de emparejamiento puede producir cifras estimadas del coeficiente con sesgo; véase, por ejemplo, Clotfelter *et al.* (2006), Goldhaber (de próxima aparición) y Cantrell *et al.* (2008).
30. Específicamente, descubren que las evaluaciones del director son factores predictivos estadísticamente importantes del logro estudiantil incluso en modelos que incluyen cifras estimadas del valor agregado anterior del maestro de un estudiante.

31. Además, la investigación apenas está empezando a evaluar cómo perciben los maestros los incentivos financieros y cómo reaccionan a éstos frente a otros tipos de inversiones educativas sin compensación (por ejemplo, ofrecer a los maestros grupos más pequeños y oportunidades especiales de desarrollo profesional) que podrían usarse como alicientes Goldhaber *et al.*, 2008a). Un estudio (Figlio y Kenny, 2007) encuentra que los incentivos financieros docentes influyen en el rendimiento de los estudiantes pero los incentivos para maestros no financieros no lo hacen.
32. La idea fundamental de la discontinuidad en la regresión es que la clasificación en las categorías de tratamiento o de control equivale a estar en términos aleatorios alrededor del valor de corte, arriba del cual está el tratamiento y debajo del cual está el control (o viceversa).
33. Además, teniendo en cuenta lo que se dijo sobre los recursos necesarios para implementar exitosamente los incentivos para maestros, en la sección anterior del capítulo, eso significaría que las escuelas más necesitadas estarían identificadas con fondos extras para la educación.
34. Para otros ejemplos de estudios sobre incentivos en toda la escuela, véase Clotfelter y Ladd (1996) y Ladd (1999).
35. Él también confirma sus hallazgos basado en una metodología de emparejamiento de la puntuación de propensión.
36. El programa también generó diferenciales salariales para trabajo en áreas rurales de México, pero este aspecto del programa no ha sido evaluado rigurosamente, hasta donde estoy enterado.
37. No deja de ser interesante, los resultados no fueron una consecuencia de las diferencias en el ausentismo docente (un problema importante en India) o en conductas fácilmente observables; sin embargo, los maestros de escuelas en tratamiento informan que asignan más tarea a los alumnos y que ofrecen más clases complementarias después de clase que los maestros de escuelas sin tratamiento.
38. Para otros ejemplos de estudios en sistemas de incentivos para maestros en los que hay mucho en juego, véase, por ejemplo, Brooke (2008) sobre Brasil; Mizala y Romaguera (2005) sobre Chile; y Figlio y Kenny (2007) sobre EUA.
39. Véase, por ejemplo, Boyd *et al.* (2003), Goldhaber *et al.* (2007), Hanushek *et al.* (2004) y Scafidi *et al.* (2007).
40. Como ejemplo, las cifras estimadas de la investigación sobre el costo de inducir a los maestros a la enseñanza o a permanecer en escuelas desfavorecidas, a partir de cambios observados entre escuelas, quizá sean exageradas teniendo en cuenta que la transición que hace un maestro de una escuela a otra se hará sólo si los beneficios de cambiarse a la nueva son superiores a los beneficios de permanecer en la antigua y a los costos de transacción del cambio.
41. Una cantidad limitada de incentivos estuvo disponible y los beneficiarios de éstos se determinaron a partir de un proceso competitivo que evaluaba sus trayectorias académicas.
42. Los autores encuentran poca diferencia en la retención entre beneficiarios de los incentivos y no beneficiarios comparables de manera observable, aunque son cautelosos para hacer notar que estos hallazgos son meramente descriptivos (si bien pudieron usar técnicas econométricas para tratar de aislar el efecto del incentivo sobre la probabilidad de que los individuos optaran por aceptar puestos en escuelas de bajo desempeño, no pudieron utilizar esas mismas técnicas para establecer si el incentivo tenía un efecto causal en la retención).

# Referencias bibliográficas

- Aaronson, D., L. Barrow y W. Sanders** (2007), "Teachers and Student Achievement in the Chicago Public High Schools", en *Journal of Labor Economics*, Vol. 25, No. 1, pp. 95-135.
- Ballou, D.** (2001), "Pay for Performance in Public and Private Schools", en *Economics of Education Review*, Vol. 20, Issue 1, february, pp. 51-61.
- Ballou, D.** (2002), "Sizing Up Test Scores", en *Education Next*, Vol. 2, No. 2., pp. 10-15.
- Ballou, D.** (2005a), "Value-Added Assessment: Controlling for Context with Misspecified Models", documento presentado en la Urban Institute Longitudinal Data Conference, marzo.
- Ballou, D.** (2005b), "Value-Added Assessments: Lessons from Tennessee", en R. Lissitz (ed.), *Value-Added Models in Education: Theory and Applications*, JAM Press, Maple Grove, MN, pp. 272-297.
- Ballou, D.** (próxima publicación), "Test Scaling and Value-Added Measurement", en *Education Finance and Policy*.
- Ballou, D. y M. Podgursky** (1997), *Teacher Pay and Teacher Quality*, W. E. Upjohn Institute for Employment Research, Kalamazoo, MI.
- Ballou, D., W. Sanders y P. Wright** (2004), "Controlling for Student Background in Value-Added Assessment of Teachers", en *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 1, pp. 37-66.
- Barber, M. y M. Mourshed** (2007), *How the World's Best-Performing School Systems Come Out on Top*, McKinsey&Company, Nueva York, <http://www.closingtheachievementgap.org/cs/ctag/view/resources/111>, consultado en enero de 2009.
- Boyd, D., H. Lankford, S. Loeb y J. Wyckoff** (2003), "Analyzing the Determinants of the Matching Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets", en *NBER Working Paper No. 9878*, National Bureau of Economic Research, Cambridge, MA.
- Bretz, R. y G. Milkovich** (1989), "Performance Appraisal in Large Organisations: Practice and Research Implications", en *Working Paper No. 89-17*, Centre for Advanced Human Resource Studies, Cornell University, Ithaca, NY.
- Brewer, D., E. Eide y R. Ehrenberg** (1999), "Does it Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type On Earnings", en *Journal of Human Resources*, Vol. 34, No. 1, pp. 104-123.
- Brooke, N.** (2008), "Educational Accountability in Brazil", en *Revista Iberoamericana de Evaluación Educativa*, Vol. 1, No. 1.
- Cantrell, S., J. Fullerton, T. Kane y D. Staiger** (2008), "National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment", en *NBER Working Paper No. 14608*, National Bureau of Economic Research, Cambridge, MA.
- Cavalluzzo, L.** (2004), "Is National Board Certification An Effective Signal Of Teacher Quality?", The CNA Corporation, Washington, D.C.
- Chambers, J. y W. Fowler** (1995), *Public School Teacher Cost Differences Across the United States*, National Center for Education Statistics, NES 95-758, U.S. Department of Education, Washington, D.C.
- Chay, K., P. McEwan y M. Urquiola** (2005), "The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools", en *American Economic Review*, Vol. 95, No. 4, pp. 1237-1258.
- 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*, Vol. 92, No.5-6, pp. 1352-1370.
- Clotfelter, C. y H. Ladd** (1996), "Recognizing and Rewarding Success in Public Schools", en H. Ladd (ed.), *Holding Schools Accountable: Performance-based Reform in Education*, The Brookings Institution, Washington, D.C.

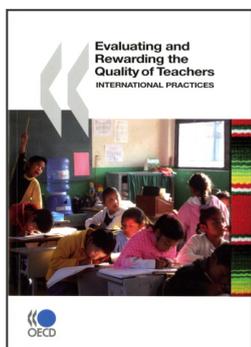
- Clotfelter, C., H. Ladd y J. Vigdor** (2006), "Teacher-Student Matching and the Assessment of Teacher Effectiveness", en *Journal of Human Resources*, Vol. 41, No. 4, pp. 778-820.
- Corcoran, S., W. Evans y R. Schwab** (2004), "Changing Labor-Market Opportunities for Women and the Quality of Teachers, 1957-2000", en *American Economic Review*, Vol. 94, No. 2, pp. 230-235.
- Costrell, R. y M. Podgursky** (2008), "Peaks, Cliffs, and Valleys: The Peculiar Incentives of Teacher Pensions", en *Education Next*, Vol. 8, pp. 22-28.
- Cullen, J. y R. Reback** (2006), "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System", en *NBER Working Paper No. 12286*, National Bureau of Economic Research, Cambridge, MA.
- Dixit, A.** (2002), "Incentives and Organisations in the Public Sector: An Interpretative Review", en *Journal of Human Resources*, Vol. 37, No. 4, pp. 696-727.
- Figlio, D.** (1997), "Teacher Salaries and Teacher Quality", en *Economics Letters*, Vol. 55, Issue 2, pp. 267-71.
- Figlio, D. y L. Getzler** (2002), "Accountability, Ability, and Disability: Gaming the System", en *NBER Working Paper 9307*, National Bureau of Economic Research, Cambridge, MA.
- Figlio, D. y L. Kenny** (2007), "Individual Teacher Incentives and Student Performance", en *Journal of Public Economics*, Vol. 91, Issues 5- 6, pp. 901-914.
- Glewwe, P., N. Ilias y M. Kremer** (2003), "Teacher Incentives", en *NBER Working Paper 9671*, National Bureau of Economic Research, Cambridge, MA.
- Goldhaber, D.** (2002), "The Mystery of Good Teaching", en *Education Next*, Vol. 2, No. 1, pp. 50-55.
- Goldhaber, D.** (2006a), "National Board Teachers Are More Effective, But Are They in the Classrooms Where They're Needed the Most?", en *Education Finance and Policy*, Vol. 1, No. 3, pp. 372-382.
- Goldhaber, D.** (2006b), "Teacher Pay Reforms: The Political Implications of Recent Research", Center for American Progress, Washington, D.C.
- Goldhaber, D.** (2007), "Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?", en *Journal of Human Resources*, Vol. 42, No. 4, pp. 765-794.
- Goldhaber, D.** (2009), "Politics of Teacher Pay Reform", en M. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Goldhaber, D.** (próxima publicación a), "Lessons from Abroad: Exploring Cross-Country Differences in Teacher Development Systems and What They Mean for U.S. Policy", en D. Goldhaber y J. Hannaway (eds.), *Creating a New Teaching Profession*, Urban Institute Press, Washington, D.C.
- Goldhaber, D.** (próxima publicación b), "Licensure Tests: Their Use and Value for Increasing the Quality of the Teacher Workforce", en M. Kennedy (ed.), *A Handbook on Teacher Assessment and Teacher Quality*, Jossey-Bass, Nueva York.
- Goldhaber, D. y E. Anthony** (2007), "Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching", en *Review of Economics and Statistics*, Vol. 89, No. 1, pp. 134-150.
- Goldhaber, D., D. Brewer y D. Anderson** (1999), "A Three-Way Error Components Analysis of Educational Productivity", en *Education Economics*, Vol. 7, No. 3, pp. 199-208.
- Goldhaber, D., M. DeArmond, A. Liu y D. Player** (2008), "Returns to Skill and Teacher Wage Premiums: What Can We Learn By Comparing the Teacher and Private Sector Labor Markets?", en *SFRP Working Paper*, No. 8, Center on Reinventing Public Education, Seattle, WA.
- Goldhaber, D., K. Destler y D. Player** (2007), "Teacher Labor Markets and the Perils of Using Hedonics to Estimate Compensating Differentials in the Public Sector", en *SFRP Working Paper*, No. 17, Center on Reinventing Public Education, Seattle, WA.
- Goldhaber D. y M. Hansen** (2008a), "Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance", en *CRPE Working Paper No. 2008-5*, Center on Reinventing Public Education, Seattle, WA.

- Goldhaber D.** y **M. Hansen** (2008b), "Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions", en *CRPE Research Brief*. Center on Reinventing Public Education, Seattle, WA.
- Goldhaber, D.** y **A. Liu** (2003), "Occupational Choices and the Academic Proficiency of the Teacher Workforce", en W. Fowler (ed.), *Developments in School Finance 2001-02*, National Center for Education Statistics, Washington, D.C., pp. 53-75.
- Grogger, J.** y **E. Eide** (1995), "Changes in College Skills and the Rise in the College Wage Premium", en *Journal of Human Resources*, Vol. 30, No. 2, pp. 280-310.
- Hansen, D.** (1997), "Worker Performance and Group Incentives: A Case Study", en *Industrial and Labor Relations Review*, Vol. 51, No. 1, pp. 37-49.
- Hanushek, E.** (1972), *Education and Race: An Analysis of the Educational Production Process*, D.C. Heath and Company, Lexington, MA.
- Hanushek, E.** (1986), "The Economics of Schooling: Production and Efficiency in Public Schools", en *Journal of Economic Literature*, Vol. 24, No. 3, pp. 1141-77.
- Hanushek, E.** (1992), "The Trade-off Between Child Quantity and Quality", en *Journal of Political Economy*, Vol. 100, No. 1, pp. 84-117.
- Hanushek, E.** (1997), "Assessing the Effects of School Resources on Student Performance: An Update", en *Educational Evaluation and Policy Analysis*, Vol. 19, No. 2, pp. 141-164.
- Hanushek, E., D. Jamison, E. Jamison** y **L. Wößmann** (2008), "Education and Economic Growth: It's Not Just Going to School But Learning That Matters", en *Education Next*, Vol. 8, No. 2, pp. 62-70.
- Hanushek, E., J. Kain** y **S. Rivkin** (2004), "Why Public Schools Lose Teachers", en *Journal of Human Resources*, Vol. 39, No. 2, pp. 326-354.
- Hanushek, E.** y **R. Pace** (1995), "Who Chooses to Teach (and Why?)", en *Economics of Education Review*, Vol. 14, No. 2, pp. 101-17.
- Harrington, P.** (2001), "Attracting New Teachers Requires Changing Old Rules", en *College Board Review*, Vol. 192, No. 6, pp. 6-11.
- Hare, D., J. Nathan, J. Darland** y **S. Laine** (2000), "Teacher Shortages in the Midwest: Current Trends and Future Issues", North Central Regional Educational Laboratory, Oak Brook, IL.
- Hein, K.** (1996), "Raises Fail, But Incentives Save the Day", en *Incentive*, Vol. 170, p. 11.
- Henke, R., S. Geis, J. Giambattista** y **P. Knepper** (1996), "Out of the Lecture Hall and into the Classroom: 1992-1993 College Graduates and Elementary/Secondary School Teaching: With an Essay on Undergraduate Academic Experiences", U.S. Department of Education, Office of Educational Research and Improvement, Washington, D.C.
- Hernani-Limarino, W.** (2005), "Are Teachers Well Paid in Latin America and the Caribbean? Relative Wage and Structure of Returns of Teachers", in E. Vegas (ed.), en *Incentives to Improve Teaching: Lessons from Latin America*, The World Bank, Washington, D.C., pp. 63-96.
- Hoffman, D., R. Jacobs** y **S. Gerras** (1992), "Mapping Individual Performance Over Time", en *American Psychological Association*, Vol. 77, No. 2, pp. 185-195.
- Hoffman, D., R. Jacobs** y **J. Baratta** (1993), "Dynamic Criteria and the Measurement of Change", en *Journal of Applied Psychology*, Vol. 78, No. 2, pp. 194-204.
- Hopkins, D., E. Ahtaridou, P. Matthews** y **C. Posner** (2007), "An Analysis of the Mexican School System in Light of PISA 2006", London Centre for Leadership in Learning, Institute of Education, University of London.
- Hoxby, C.** y **A. Leigh** (2004), "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States", en *American Economic Review*, Vol. 94, Issue 2, pp. 236-240.
- Jacob, B.** y **L. Lefgren** (2008), "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education", en *Journal of Labor Economics*, Vol. 26, No. 1, pp. 101- 136.

- Jacob, B.** y **S. Levitt** (2003), "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating", en *Quarterly Journal of Economics* Vol. 118, No. 3, pp. 843-78.
- Jacob, B.** y **M. Springer** (2007), "Teacher Attitudes on Pay For Performance: A Pilot Study", en *Working Paper 2007-06*, National Center on Performance Incentives, Vanderbilt University, Nashville, TN.
- Kane, T., J. Rockoff** y **D. Staiger** (2007), "Photo Finish: Certification Doesn't Guarantee a Winner", en *Education Next*, Vol. 7, No. 1, pp. 60-67.
- Kane, T.** y **D. Staiger** (2001), *Improving School Accountability Measures*, National Bureau of Economic Research, Cambridge, MA.
- Kane, T.** y **D. Staiger** (2008), "Are Teacher-Level Value-Added Estimates Biased? An Experimental Validation of Non-Experimental Estimates", documento presentado en la National Conference on Value-Added Modeling, University of Wisconsin, Madison, WI.
- Koedel, C.** (próxima publicación), "An Empirical Analysis of Teacher Spillover Effects in Secondary School", en *Economics of Education Review*.
- Koedel, C.** y **J. Betts** (2007), "Re-Examining the Role of Teacher Quality in the Educational Production Function", en *Working Paper WP 07-08*, University of Missouri-Columbia, Columbia, MO.
- Koedel, C.** y **J. Betts** (2009), "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique", en *Working Paper WP 09-02*, University of Missouri-Columbia, Columbia, MO.
- Koedel, C.** y **J. Betts** (próxima publicación), "Value-added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation", *Education Finance and Policy*.
- Koretz, D.** (2008), *Measuring Up: What Educational Testing Really Tells Us*, Harvard University Press, Cambridge, MA.
- Ladd, H.** (1999), "The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes", en *Economics of Education Review*, Vol. 18, No. 1, pp. 1-16.
- Ladd, H.** (2007), "Teacher Labor Markets in Developed Countries", en *The Future of Children*, Vol. 17, No. 1, pp. 201-217.
- Lakdawalla, D.** (2001), "The Declining Quality of Teachers", en *NBER Working Paper No. 8263*, National Bureau of Economic Research, Cambridge, MA.
- Lankford, H., S. Loeb** y **J. Wyckoff** (2002), "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis", en *Educational Evaluation and Policy Analysis*, Vol. 24, No. 1, pp. 37-62.
- Lavy, V.** (2002), "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement", en *Journal of Political Economy*, Vol. 110, No. 6, pp. 1286-317.
- Lavy, V.** (2004), "Performance Pay and Teachers' Effort, Productivity and Grading Ethics", en *NBER Working Paper No. 10622*, National Bureau of Economic Research, Cambridge, MA.
- Lavy, V.** (2007), "Using Performance-Based Pay to Improve the Quality of Teachers", en *The Future of Children*, Vol. 17, No. 1, pp. 87-110.
- Lazear, E.** (2000), "Performance Pay and Productivity", en *American Economic Review*, Vol. 90, No. 5, pp. 1346-1361.
- Loeb, S.** (2001), "Teacher Quality: Its Enhancement and Potential for Improving Pupil Achievement", en D. Monk, H. Walberg y M. Wang (eds.), *Improving Educational Productivity*, Information Age Publishing, Greenwich, CT.
- Loeb, S.** y **M. Page** (2000), "Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-pecuniary Variation", en *The Review of Economics and Statistics*, Vol. 82, Issue 3, pp. 393-408.
- McCaffrey, D., D. Koretz, J. Lockwood** y **L. Hamilton** (2003), *Evaluating Value-Added Models for Teacher Accountability*, RAND Corporation, Santa Monica, CA.
- McCaffrey, D., T. Sass** y **J. Lockwood** (2008), "The Intertemporal Stability of Teacher Effect Estimates", en *NCPI Working Paper No. 2008-22*, National Center on Performance Incentives, Nashville, TN.

- McEwan, P.J.** (1999), "Recruitment of Rural Teachers in Developing Countries: An Economic Analysis", en *Teaching and Teacher Education*, Vol. 15, No. 8, pp. 849-859.
- McEwan, P.** y **L. Santibáñez** (2005), "Teacher and Principal Incentives in Mexico", en E. Vegas (ed.), *Incentives to Improve Teaching: Lessons from Latin America*, The World Bank, Washington, D.C., pp. 213-253.
- Meyer, R.H.** (1996), "Comments on Chapters Two, Three, and Four", en H. Ladd (ed.), *Holding Schools Accountable: Performance-Based Reform in Education*, The Brookings Institution, Washington, D.C.
- Milkovich, G.** y **A. Wigdor** (eds.) (1991), *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*, National Academies Press, Washington, D.C.
- Mizala, A.** y **P. Romaguera** (2005), "Teachers' Salary Structure and Incentives in Chile", en E. Vegas (ed.), *Incentives to Improve Teaching: Lessons from Latin America*, The World Bank, Washington, D.C., pp. 103-150.
- Muralidharan, K.** y **V. Sundararamin** (2008), "Teacher Incentives in Developing Countries: Experimental Evidence from India", en *NCPI Working Paper Number 2008-13*, National Center on Performance Incentives, Nashville, TN.
- Murnane, R.J.** y **D. Cohen** (1986), "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive", en *Harvard Educational Review*, Vol. 56, pp. 1-17.
- Murnane, R., J. Singer, J. Willett, J. Kemple y R. Olsen** (1991), *Who Will Teach? Policies That Matter*, Harvard University Press, Cambridge, MA.
- Murnane, R.** y **J. Steele** (2007), "What is the Problem? The Challenge of Providing Effective Teachers for All Children", en *The Future of Children*, Vol. 17, No. 1, pp. 15-44.
- Murnane, R., J. Willett y F. Levy** (1995), "The Growing Importance Of Cognitive Skills In Wage Determination", en *The Review of Economics and Statistics*, Vol. 77, No. 2, pp.251-66.
- Murphy, K.** (1999), "Executive Compensation", en O. Ashenfelter y D. Card (eds.), *Handbook of Labor Economics*, Vol. 3, North Holland, Amsterdam, pp. 2485-2563.
- New Teacher Project** (2007), "Hiring, Assignment, and Transfers in Chicago Public Schools," <http://www.tntp.org/files/TNTPAnalysis-Chicago.pdf>, consultado en abril de 2009.
- Odden, A.** y **C. Kelley** (1997), *Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools*, Corwin Press, Thousand Oaks, CA.
- OCDE** (2004), "The Quality of the Teaching Workforce", *Policy Brief* (February), OCDE, París.
- OCDE** (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, (Spanish version forthcoming) OCDE, París.
- OCDE** (2008), *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added in Schools*, OCDE, París.
- Oh, H.** (2008), "The Teacher Training System and Teaching Profession in Korea", documento preparado para el Urban Institute's Gates Human Capital Reform Project, Washington, D.C.
- Player, D.** (próxima publicación), "Nonmonetary Returns to Quality in the Public Teacher Labor Market", en *Education Finance and Policy*.
- Podgursky, M.** y **M. Springer** (2007), "Teacher Performance Pay: A Review", en *Journal of Policy Analysis and Management*, Vol. 26, No. 4, pp. 909-950.
- Prendergast, C.** (1999), "The Provision of Incentives in Firms", en *Journal of Economic Literature*, Vol. 37, No. 1, pp. 7-63.
- Rivkin, S.** (próxima publicación), "The Estimation of Teacher Value-added as a Determinant of Performance Pay", en D. Goldhaber y J. Hannaway (eds.), *Creating a New Teaching Profession*, Urban Institute Press, Washington, D.C.
- Rivkin, S., E. Hanushek y J. Kain** (2005), "Teachers, Schools, and Academic Achievement", en *Econometrica*, Vol. 73, No. 2, pp. 417-458.

- Rockoff, J.** (2004), "The Impact of Individual Teachers on Students' Achievement: Evidence from Panel Data", en *American Economic Review*, Vol. 94, Issue 2, pp. 247-252.
- Rosen, S.** (1986), "The Theory of Equalizing Differences", en O. Ashenfelter y R. Layard (eds.), *Handbook of Labor Economics*, Vol. 1, North Holland, Amsterdam.
- Rothstein, J.** (2008a), "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement", en *NBER Working Paper No. 14442*, National Bureau for Economic Research, Cambridge, MA.
- Rothstein, J.** (2008b), "Student Sorting and Bias in Value Added Estimation: Selection on Observables and Unobservables", en *NBER Working Paper No. 14666*, National Bureau for Economic Research, Cambridge, MA.
- Sanders, W. y J. Rivers** (1996), *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*, University of Tennessee Value-Added Research and Assessment Center, Knoxville, TN.
- Santibáñez, L., J. Martínez, A. Datar, P. McEwan, C. Setodji y R. Basurto-Dávila** (2007), "Breaking Ground Analysis of the Assessment System and Impact of Mexico's Teacher Incentive Program 'Carrera Magisterial'", en *RAND Technical Report*, RAND Corporation, Santa Monica, CA.
- Scafidi, B., D. Sjoquist y T. Stinebrickner** (2007), "Race, Poverty, and Teacher Mobility", en *Economics of Education Review*, Vol. 26, No. 2, pp. 145-159.
- Sclafani, S.** (2008), *Rethinking Human Capital in Education: Singapore As A Model for Teacher Development*, The Aspen Institute Education and Society Program, Washington, D.C.
- Sclafani, S. y M. Tucker** (2006), *Teacher and Principal Compensation: An International Review*, Center for American Progress, Washington, D.C.
- Shields, P., D. Humphrey, M. Wechsler, L. Riehl, J. Tiffany-Morales, K. Woodworth, V. Young y T. Price** (2001), *The Status of the Teaching Profession 2001*, The Center for the Future of Teaching and Learning, Santa Cruz, CA.
- Steele, J., R. Murnane y J. Willet** (2009), "Do Financial Incentives Help Low-Performing Schools Attract and Keep Academically Talented Teachers? Evidence from California", en *NBER Working Paper No. 14780*, National Bureau for Economic Research, Cambridge, MA.
- Stinebrickner, T.** (2001a), "Compensation Policies and Teacher Decisions", en *International Economic Review*, Vol. 42, Issue 3, pp. 751-779.
- Stinebrickner, T.** (2001b), "A Dynamic Model of Teacher Labor Supply", en *Journal of Labor Economics*, Vol. 19, Issue 1, pp. 196-230.
- Toch, T. y R. Rothman** (2008), "Rush to Judgment: Teacher Evaluation in Public Education", en *Education Sector Report*, Washington, D.C.
- Todd, P. y K. Wolpin** (2003), "On the Specification and Estimation of the Production Function for Cognitive Achievement", en *Economic Journal*, Vol. 113, pp. F3-F33.
- Urquiola, M. y E. Vegas** (2005), "Arbitrary Variation in Teacher Salaries: An Analysis of Teacher Pay in Bolivia", en E. Vegas (ed.), *Incentives to Improve Teaching: Lessons from Latin America*, The World Bank, Washington, D.C., pp. 187-212.
- Vegas, E. y I. Umansky** (2005), "Improving Teaching and Learning through Effective Incentives", en E. Vegas (ed.), *Incentives to Improve Teaching: Lessons from Latin America*, The World Bank, Washington, D.C., pp. 1-20.



From:  
**Evaluating and Rewarding the Quality of Teachers:  
International Practices**

**Access the complete publication at:**

<https://doi.org/10.1787/9789264034358-en>

**Please cite this chapter as:**

OECD (2010), “Incentivos para influir en la calidad y distribución de los maestros”, in *Evaluating and Rewarding the Quality of Teachers: International Practices*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264079717-7-es>

El presente trabajo se publica bajo la responsabilidad del Secretario General de la OCDE. Las opiniones expresadas y los argumentos utilizados en el mismo no reflejan necesariamente el punto de vista oficial de los países miembros de la OCDE.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to [rights@oecd.org](mailto:rights@oecd.org). Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at [info@copyright.com](mailto:info@copyright.com) or the Centre français d'exploitation du droit de copie (CFC) at [contact@cfcopies.com](mailto:contact@cfcopies.com).