

¿Mejoran las evaluaciones la eficacia docente?

El sistema IMPACT en Washington, DC

Por Tamara Ortega Goodspeed and T’Nia Crutchfield



Los estudios demuestran que los profesores eficaces pueden tener un impacto profundo sobre el éxito de sus estudiantes dentro y fuera de la sala de clases. Al mismo tiempo, una docencia inadecuada limita a los estudiantes y genera carencias que se agudizan a través del tiempo. Por ello, es fundamental que las escuelas y sistemas educativos puedan distinguir entre profesores eficaces e ineficaces y que ayuden a todos los docentes a mejorar sus prácticas.

Sin embargo, los intentos de evaluar el desempeño docente sobre la base del aprendizaje estudiantil usualmente generan gran resistencia, especialmente cuando estas evaluaciones están ligadas a decisiones sobre salarios o seguridad laboral. Sus detractores argumentan que existen factores fuera del control de los docentes que inciden en el aprendizaje de los estudiantes, que las evaluaciones con fuertes consecuencias para los docentes no miden los aspectos más importantes y que están

sujetas a errores (o a la manipulación de sus resultados) y que al final no ayudan a los profesores a mejorar. Sus partidarios responden que sin evaluaciones docentes asociadas al aprendizaje, es imposible detectar problemas, ayudar a los maestros a mejorar o responsabilizarlos por la calidad de la instrucción.

¿Puede la evaluación docente asociada al desempeño de los estudiantes mejorar la eficacia de los profesores? ¿Cuáles son los riesgos y los beneficios de tal sistema?

Desde 2009, los docentes de escuelas públicas en Washington, DC, han sido evaluados utilizando un sistema de evaluación con fuertes consecuencias y en base a incentivos, llamado IMPACT. Este sistema califica a los profesores en base a múltiples observaciones en la sala de clases, de su aporte al aprendizaje estudiantil medido a través de pruebas de rendimiento, de su contribución a la comunidad educativa y de su nivel de profesionalismo. IMPACT establece expectativas claras,

Las Notas de Política PREAL proporcionan resúmenes no técnicos de temas fundamentales en el campo de la política educativa. En esta edición Tamara Ortega Goodspeed, Senior Education Associate del Diálogo Interamericano y la pasante T’Nia Crutchfield describen el sistema de evaluación docente IMPACT de Washington, DC, sus principales características, éxitos y desafíos, y sus posibles implicancias para las políticas educativas docentes en América Latina.

Se puede encontrar más información sobre IMPACT en el [sitio web de DCPS](#) y en los vínculos incluidos en el documento.

La preparación de la presente nota fue posible gracias al apoyo del Fondo de Fortalecimiento de la Capacidad Institucional (ICSF) del Banco Interamericano de Desarrollo, establecido con el aporte del Gobierno de la República Popular China. Sin embargo, los contenidos son responsabilidad absoluta de las autoras y del Diálogo Interamericano.

mide el grado de cumplimiento de esas expectativas y entrega retroalimentación junto con ideas para el desarrollo profesional a los docentes evaluados. Los profesores más eficaces reciben mejores sueldos y reconocimiento y los docentes ineficaces son retirados del aula. Resultados de investigaciones preliminares sugieren que el sistema ha logrado incentivar a los menos calificados a mejorar o renunciar sus puestos, y ha ayudado a retener a los docentes mejor calificados y a mejorar su desempeño. Las evaluaciones docentes forman parte de un conjunto más amplio de reformas educativas en Washington, DC que han contribuido a mejorar los logros educativos de los estudiantes.

dificultades del sistema de evaluación docente IMPACT y extrae posibles implicancias para las políticas educativas docentes en América Latina.

Un sistema escolar que confronta bajos niveles de aprendizaje estudiantil

El distrito escolar de Washington, DC (DCPS, por sus siglas en inglés) ha estado por años entre los distritos de Estados Unidos con más bajo desempeño en pruebas nacionales de rendimiento escolar. DCPS atiende a más de

45.000 estudiantes en más de 100 escuelas. Sólo el 12% de los estudiantes es de raza blanca y más de tres cuartos provienen de familias cuyo bajo nivel de ingresos, les da derecho a alimentación escolar gratis o a bajo costo.¹ A pesar de que han

los puntajes mejorado en los últimos años, menos del 30% de los estudiantes de cuarto grado obtuvo una

calificación de competente o superior en matemáticas y menos de un cuarto obtuvo esa calificación en lectura en la Prueba Nacional de Progreso Educativo (National Assessment of Educational Progress—NAEP) realizada el 2013.² En contraste, más del 40% de los estudiantes de cuarto grado de las escuelas públicas a nivel nacional, alcanzó una evaluación de competente o superior en matemáticas, mientras que un 34% alcanzó dicho nivel en lectura. El porcentaje de estudiantes de octavo grado en DC que obtuvo una calificación de competente o superior fue incluso menor. Aún siendo alarmantes, estas cifras representan una mejoría respecto de 2007, cuando apenas el 14% de los estudiantes de cuarto grado alcanzaron la calificación de competente en lectura y matemáticas.

En 2007, la entonces administradora del sistema de educación pública de DC, Michelle Rhee, comenzó una amplia reestructuración del sistema educativo diseñado para abordar el bajo desempeño crónico, cerrando escuelas de bajo rendimiento y contratando nuevos directores. Ella también desarrolló un plan para mejorar la calidad de la instrucción basado en un nuevo Marco de Enseñanza y Aprendizaje (TLF, por sus siglas en inglés), el cuál define los comportamientos esperados para todos los docentes en tres áreas – planificación, enseñanza y aumento de la eficacia.

Actualmente los docentes son únicamente evaluados de acuerdo a los criterios de “Enseñanza”,

“Resultados de investigaciones preliminares sugieren que el sistema ha logrado incentivar a los menos calificados a mejorar o renunciar sus puestos, y ha ayudado a retener a los docentes mejor calificados y a mejorar su desempeño.”

La presente nota de políticas resume las principales características, logros y

que aborda nueve prácticas pedagógicas clave:³

1. Dirigir clases bien organizadas y con objetivos específicos.
2. Explicar los contenidos de forma clara.
3. Involucrar a los estudiantes de distintos niveles de conocimiento en tareas alcanzables y desafiantes.
4. Entregar a los estudiantes múltiples formas de involucrarse con los contenidos y avanzar hacia el dominio de la materia.
5. Verificar que los alumnos estén comprendiendo.
6. Responder ante el nivel de comprensión de los alumnos.
7. Desarrollar niveles altos de comprensión a través de cuestionamientos efectivos.
8. Maximizar las horas de instrucción.
9. Crear un ambiente de apoyo en el aula, orientado a aprender.

En consonancia con el Marco de Enseñanza y Aprendizaje, Rhee instauró evaluaciones docentes que ligaban el desempeño (medido principalmente a través de observaciones de clases y el rendimiento escolar) con los salarios y la seguridad laboral. Iniciadas en 2009, las evaluaciones anuales llamadas IMPACT continúan hasta hoy con algunas modificaciones. Siendo uno de los primeros sistemas de evaluación del país en utilizar esta estrategia, sus resultados

son seguidos de cerca. ¿Cómo funciona IMPACT?

Los elementos básicos de IMPACT

IMPACT es un sistema de evaluación en base a incentivos, con fuertes consecuencias para los docentes, que califica su desempeño en áreas clave, incluyendo prácticas en el aula y aprendizaje de los estudiantes. Los docentes son categorizados en siete grupos, siendo los Grupos 1 y 2 los que abarcan a la mayoría de los docentes de educación general en primaria y secundaria.⁴ El Grupo 1 comprende profesores de lengua (cuarto a décimo año) y matemáticas (cuarto a octavo año) cuyos estudiantes rinden cada año la prueba Sistema de Evaluación Comprensivo de DC (DC CAS, por sus siglas en inglés) —evaluación anual de conocimientos de los estudiantes basada en los Estándares Comunes de Aprendizaje (Common Core Standards)⁵—y que por ende generan información sobre el valor agregado al desempeño de los estudiantes, que puede asociarse a un profesor específico. Actualmente cerca del 15% de los docentes del sistema público de Washington están en esta categoría. El Grupo 2, por su parte, comprende a profesores de educación general primaria y secundaria cuyos alumnos no rinden la prueba DC CAS en sus respectivas asignaturas y por ende no generan información

sobre el valor agregado al rendimiento de los estudiantes.

Los docentes de los Grupos 1 y 2 son evaluados en cuatro áreas:

- Observación en aula de las prácticas docentes definidas en el ámbito de la “Enseñanza” del TLF;
- Valor agregado al rendimiento de los estudiantes basado en la prueba DC CAS (Grupo 1) y/o mejoras en el rendimiento medidas según otras pruebas rigurosas (Grupos 1 y 2);
- Compromiso con la comunidad educativa (CSC, por sus siglas en inglés); y
- Nivel de profesionalismo (CP, por sus siglas en inglés).

Cada componente tiene una ponderación distinta, las prácticas de aula tienen el peso máximo mientras que el compromiso con la comunidad educativa tiene la menor ponderación. El nivel de profesionalismo no tiene una ponderación específica y se mide en una escala de 10 a 20 puntos que se pueden rebajar del puntaje total de IMPACT. A continuación se describe brevemente cada componente:

PRÁCTICAS DE AULA

Los criterios de “Enseñanza” del Marco de Enseñanza y Aprendizaje (TLF) proveen una guía de cómo medir las habilidades de instrucción docente a través de la observación de clases. El TLF define las prácticas de un profesor efectivo según DCPS, y los puntajes basados en los

criterios de “Enseñanza” del TFL representan el 40% de la calificación total de IMPACT para los profesores del Grupo 1 y el 75% para el Grupo 2 (cuyo puntaje IMPACT no incluye datos individuales de valor agregado). El puntaje del TFL se obtiene sobre la base de cuatro observaciones formales al año: dos a cargo de autoridades administrativas del establecimiento (director o subdirector) y dos a cargo de un educador experto independiente que provee comentarios y orientación para el desarrollo profesional del docente a través de reuniones y reportes escritos después de las observaciones.⁶

“DCPS es uno de los pocos distritos escolares de Estados Unidos que retira prontamente de sus aulas a los docentes evaluados como ineficaces.”

Los profesores nuevos y los que se encuentran en niveles básicos de su desarrollo profesional también reciben apoyo adicional a través de observaciones informales hechas por las autoridades administrativas de la escuela, las que no se contabilizan para el puntaje IMPACT. Los profesores que se encuentran en los niveles avanzados de desarrollo profesional y que obtienen

consistentemente altos puntajes IMPACT, pueden ser observados menos de cuatro veces.⁷

RENDIMIENTO ESCOLAR

Durante el año escolar los profesores de los Grupos 1 y 2 son evaluados de acuerdo a los avances en el aprendizaje de sus estudiantes. Para docentes del Grupo 1, este componente representa el 50% del puntaje IMPACT e incluye el puntaje individual por concepto de valor agregado (IVA, por sus siglas en inglés) que se deriva del desempeño de sus alumnos en las pruebas DC-CAS de matemáticas y lectura (35%) y contra metas de aprendizaje en evaluaciones rigurosas distintas a DC-CAS (15%) establecidas por los docentes y sus administradores (TAS, por sus siglas en inglés). A los profesores del Grupo 2 se les evalúa únicamente usando TAS (15%) para este componente.

Para calcular el puntaje individual por concepto de valor agregado (IVA), el distrito compara el aprendizaje real de los estudiantes con una estimación del efecto promedio del profesor, medida según el puntaje que obtengan sus estudiantes de un año al otro. El cálculo lo realiza una compañía nacional independiente especializada en la investigación, aislando el aporte del profesor de otros factores que puedan incidir sobre el

aprendizaje; por ejemplo, el puntaje del alumno en la prueba DC CAS del año anterior; si califican para recibir comida gratuito o a bajo costo en la escuela, si reciben educación diferencial; si tienen un dominio limitado del inglés; su asistencia durante el año anterior; si el estudiante fue transferido de otra escuela a mitad de año; el puntaje promedio en las pruebas de la clase del estudiante el año anterior; y el rango de variación de los puntajes en la clase del estudiante durante el año anterior. La compañía combina los puntajes individuales de cada estudiante, por cada profesor según la asignatura que imparten, a qué estudiantes les dieron clases y el tiempo dedicado a cada uno, con lo cual se llega a un “puntaje promedio probable” para cada clase enseñada por el profesor.⁸

Al finalizar el año este puntaje se compara con el promedio real de cada curso, lo que permite calcular el valor agregado en base a la siguiente ecuación:

$$\text{Puntaje real prueba DC CAS} - \text{Puntaje probable prueba DC CAS} = \text{Valor agregado}$$

Para recibir un puntaje de valor agregado (IVA), el profesor debe tener un mínimo de 15 estudiantes con puntajes en las pruebas DC-CAS, tanto en el año actual como en el anterior. Los puntajes IVA son calculados para lectura y matemáticas por separado, y los profesores que imparten ambas materias reciben un promedio de los dos puntajes.

TAS mide el aprendizaje de los estudiantes sobre la base de otros instrumentos igualmente rigurosos que DC CAS y que están alineados con los estándares de contenidos que define el sistema de educación pública de DC. Al inicio del año los profesores y los administrativos acuerdan el o los instrumentos a utilizar, su ponderación y metas de rendimiento, y a final de año las comparan con los resultados reales de los estudiantes. A partir de esto, se le asigna un puntaje IMPACT a cada profesor, en una escala de 1 a 4, sobre la base de las metas de rendimiento definidas. Aunque la escala varía según el instrumento y las circunstancias de cada establecimiento, los manuales de DCPS para los profesores de los Grupos 1 y 2 recomiendan que, en general, la nota 1 represente una mejoría promedio en el aprendizaje menor a un año o un dominio contenidos menor al 70%, y que la nota 4 represente un aprendizaje excepcional (1,5 años o más de mejoría o 90% de dominio de contenidos).⁹ Los puntajes TAS deben ser aprobados y validados por la dirección y durante el año se monitorea el estado de avance hacia las metas, para permitir hacer ajustes basados en datos reales sobre la instrucción.

COMPROMISO CON LA COMUNIDAD EDUCATIVA (CSC) Y NIVEL DE PROFESIONALISMO (CP)

CSC representa el 10% del puntaje IMPACT para todos los grupos de profesores y lo determina la dirección del establecimiento a través de evaluación del aporte del docente a las iniciativas del

colegio, la colaboración con los estudiantes, las familias y los colegas, y los esfuerzos de promover expectativas altas. El profesionalismo lo evalúan los directores y mide la observancia de los siguientes principios: no tener ausencias injustificadas; no llegar tarde sin justificar; respetar las normas y procedimientos del establecimiento y del sistema de educación pública; y tratar de manera respetuosa a colegas, estudiantes, familias e integrantes de la comunidad. Existen solamente tres calificaciones: cumple con lo requerido, levemente por debajo de lo requerido o significativamente por debajo de lo requerido. El profesionalismo no tiene ponderación propia; sólo se utiliza para restar 10 puntos del puntaje IMPACT a los calificados como “levemente por debajo de lo requerido” en alguno de los cuatro principios y 20 puntos a los calificados como “significativamente por debajo de lo requerido” en alguno de los criterios.

Calificando la eficacia docente y atribuyendo consecuencias

El puntaje IMPACT total de un profesor, entonces, se compone del promedio ponderado de todos los componentes, menos descuentos por falta de profesionalismo. Los puntajes van en una escala de 100 a 400, donde el profesor queda en una de cinco posibles categorías:

ineficaz (menos de 200 puntos); mínimamente eficaz (200-249); en desarrollo (250-299); eficaz (300-349) y altamente eficaz (350 o más).

Los profesores calificados como altamente eficaces reciben un reconocimiento especial y acceden a oportunidades de ascenso, oportunidades de liderazgo y aumentos de salario. Estos aumentos puede llegar hasta US \$25,000 para profesores que están en el Grupo 1 y se desempeñan en escuelas de sectores pobres con un historial de bajo rendimiento.¹⁰ Los docentes del Grupo 1 reciben bonos mayores que los profesores del Grupo 2, para reconocer el desafío y la rigurosidad que implica tener el 50% de su puntaje IMPACT basado en información sobre el desempeño de sus estudiantes.

Los docentes ineficaces quedan sujetos a despidos inmediatos, al igual que los evaluados como mínimamente eficaces por dos años seguidos y los calificados como en desarrollo durante tres años consecutivos. Las dos últimas categorías no califican para ascensos ni a aumentos salariales hasta no ser evaluados como eficaces o altamente eficaces.

El distrito escolar ofrece diversas oportunidades y recursos para que profesores de todos los niveles mejoren sus prácticas, entre ellas orientación y capacitación para profesores nuevos, instructores educativos no evaluativos, maestros consultores (profesores de excelencia que permiten a otros

colegas observar sus clases y responden sus preguntas) y una biblioteca de materiales digitales que permiten observar a destacados maestros en acción.¹¹

anual de unos US \$6-8 millones.¹³ Financiados en sus inicios por donantes externos, los costos de los bonos fueron asumidos por el sistema a partir del año escolar 2012-2013.¹⁴

por sí solas.¹⁵ La observación de clases realizada por educadores expertos y directivos del establecimiento provee comentarios objetivos y constructivos que muchos maestros de DCPS reportaron valorar y agradecer.¹⁶ Al mismo tiempo, el sistema IMPACT también usa más de una medida de desempeño escolar, y considera aspectos no académicos de la función docente que son igualmente importantes para construir un buen ambiente escolar.

Por cierto, ningún método de medición es perfecto. Si bien la crítica más fuerte del sistema se centra en el aspecto valor agregado de la evaluación (ver sección siguiente), cabe agregar que algunos profesores han reclamado que inconsistencias en la forma de realizar las observaciones y en los resultados entregados por distintos evaluadores, dificulta entender lo que deben hacer para mejorar. Otros estiman que media hora de observación no basta para incorporar todos los criterios incluidos en el ámbito de la “Enseñanza” de la TLF y que los observadores no siempre conocen ciertas circunstancias especiales de cada aula; por ejemplo, que pueda haber un número importante de niños con necesidades especiales. En respuesta, las autoridades de DCPS señalan que existe un espacio para informar sobre circunstancias especiales y aseguran que observar todas las prácticas en el tiempo asignado es perfectamente posible. El convenio colectivo actual entre la administración y el gremio

“Uno de los problemas de las evaluaciones con fuertes consecuencias es que las personas involucradas tienen incentivos a manipular los resultados para hacerlos ver favorablemente o para evitar consecuencias negativas.”

En el año escolar 2012-13, cerca del 30% de los profesores de las escuelas públicas de Washington, DC fueron calificados como altamente eficaces y el 45% como eficaces. Sólo un 5% fueron evaluados como mínimamente eficaces y un 1% como ineficaces.¹²

DCPS es uno de los pocos distritos escolares de Estados Unidos que retira prontamente de sus aulas a los docentes evaluados como ineficaces. Desde los inicios del programa en 2009, han sido despedidos cerca de 400 maestros (de un total que supera los 4.000) basados en sus resultados en IMPACT. Al mismo tiempo, un número importante de maestros altamente eficaces han recibido bonos en dinero -unos 500 profesores el primer año y unos 1.000 en 2013- a un costo

Prometedor, pero no carente de posibles obstáculos

MÚLTIPLES MEDIDAS DE LA EFICACIA DOCENTE

Incluso los detractores de IMPACT coinciden en que una de las virtudes del programa es medir la eficacia docente de distintas maneras. Esta estrategia concuerda con estudios recientes hechos en Estados Unidos que muestran que el uso de una medición múltiple (con observación de clases y evaluación del rendimiento, entre otros) genera una evaluación más imparcial y confiable sobre eficacia docente que las mediciones de desempeño escolar

docente no permite a los evaluadores discutir sus observaciones con los mentores que colaboran con los profesores en mejorar su práctica (los profesores sí pueden compartir las observaciones si así lo desean), lo que complica el uso de la evaluación como método para mejorar las prácticas docentes.¹⁷

DCPS ha estado abierto a recibir comentarios, y el sistema ha llevado a cabo una serie de ajustes y mejoras para abordar preocupaciones surgidas en el curso de la implementación del programa. Por ejemplo, se redujo la ponderación de las pruebas estandarizadas que medían valor agregado (de 50 a 35% del puntaje total para el Grupo 1), se autorizó a los directores a solicitar una exención de un año para maestros sujetos a despido que a su juicio podían mejorar, y algunas observaciones de aula se hicieron optativas para profesores con un desempeño consistente de excelencia.

DATOS DE VALOR AGREGADO

Una de las principales características de IMPACT es el uso de medidas de valor agregado para evaluar el desempeño docente. Estas medidas se han convertido en una de las áreas más polémicas del debate sobre la evaluación docente en Estados Unidos. Sus partidarios manifiestan que el criterio de valor agregado permite evaluar la eficacia docente de manera más apropiada que usando puntajes promedio en pruebas únicamente, ya que estas últimas no consideran el contexto de los

estudiantes o su desempeño académico previo.

Los detractores del sistema, por su parte, hacen al menos cuatro afirmaciones en contra del uso de criterios de valor agregado para mejorar la calidad docente:

- Primero, que el puntaje individual de valor agregado (IVA) subvalora al profesor porque no mide—ni puede medir—un proceso tan complejo como la eficacia docente. Por ejemplo, un educador creativo que use técnicas de instrucción innovadoras podría recibir un puntaje IVA bajo—sin importar cuán buenas puedan ser sus prácticas docentes—si sus alumnos no obtienen el aumento esperado en los puntajes de la prueba DC CAS.
- Segundo, el cálculo de IVA es complejo y por ende padres, profesores y directivos tienen dificultades para entenderlo o utilizarlo para evaluar o mejorar el desempeño. Estos cálculos complejos también son susceptible a errores, con consecuencias que pueden ser serias.
- Tercero, los puntajes de valor agregado pueden fluctuar ampliamente año a año debido a errores de medición en las pruebas, a estudiantes que ese día rindieron un mal examen, y en algunos casos, a la manipulación de los resultados (ver más abajo).

- Cuarto, sólo se generan datos de valor agregado para los profesores cuyos alumnos rinden una prueba estandarizada en la asignatura que ellos enseñan durante dos años consecutivos. Es decir, la mayoría de los criterios de valor agregado rigen sólo para un número reducido de docentes en las asignaturas y grados escolares evaluados; a los demás se les evalúa bajo otros criterios. Dado que los criterios de valor agregado pueden ser más difíciles de cumplir, esto plantea inquietudes en materia de equidad entre los profesores.

“Una de las virtudes de IMPACT es medir la eficacia docente de distintas maneras.”

El programa IMPACT busca minimizar este tipo de riesgos ajustando la ponderación del IVA en el cálculo de la calificación y balanceándolo con otros factores; entregando información clara y precisa sobre cómo opera el sistema; orientando a los profesores en cuanto a cómo mejorar; haciéndose asesorar por destacados investigadores para el diseño, cálculo y estudio de los puntajes individuales, y entregando bonos en dinero más altos a los profesores altamente eficaces pertenecientes al Grupo 1. Los maestros pueden además revisar la lista de estudiantes

para verificar que se les evalúe únicamente en función de los estudiantes a los que efectivamente enseñaron.¹⁸ Al mismo tiempo, los administradores del sistema están firmemente comprometidos con la idea que la mejor medida de la eficacia docente es saber si los niños están aprendiendo o no. Sin embargo, la controversia sobre el valor agregado sigue generando debate e investigaciones sobre la mejor forma de medir el efecto de los profesores sobre el aprendizaje de sus estudiantes, por lo que sin duda la experiencia de Washington continuará siendo observada con interés.

MANIPULACIÓN DE LOS RESULTADOS

Uno de los problemas de las evaluaciones con fuertes consecuencias es que las personas involucradas tienen incentivos a manipular los resultados para hacerlos ver favorablemente o para evitar consecuencias negativas. De hecho, en 2011 denuncias sobre una manipulación generalizada de los resultados de la prueba DC CAS, llevaron a una investigación a cargo de la Oficina del Inspector General del Departamento de Educación nacional. La raíz de la controversia fue una inusitada cantidad de correcciones detectadas (que cambiaban las respuestas incorrectas por respuestas correctas) en los exámenes entre los años 2008 y 2010, lo que hizo sospechar que profesores y directivos podrían haber ayudado de forma indebida a sus estudiantes.¹⁹

El sistema de educación pública de Washington hizo público una evaluación independiente de la prueba DC CAS del año 2009 que no detectó faltas a la probidad y determinó que había explicaciones plausibles al alto número de correcciones.²⁰ Aún así, los detractores del sistema siguen sosteniendo que no se actuó con la suficiente rapidez y el rigor que las denuncias ameritaban.²¹

Para tomar precauciones ante nuevos incidentes, DCPS ha redoblado las medidas de seguridad en sus exámenes. Una firma independiente verifica anualmente la integridad del examen, y los exámenes en que se identifican problemas de seguridad son invalidados y excluidos del análisis de valor agregado.²² Evaluaciones posteriores detectaron irregularidades por parte de tres profesores en 2011 y en un total de 18 aulas en 2012.²³ En 2013 los evaluadores encontraron evidencia de irregularidades serias en cuatro escuelas e infracciones de seguridad moderadas o menores en otras 13.²⁴ Si bien estas cifras no muestran una cultura sistémica de conducta indebida, sugieren que las medidas de responsabilización docente deben considerar las consecuencias no intencionales de ofrecer incentivos asociados al desempeño, y cómo prevenir la manipulación de resultados cuando estas medidas son implementadas.

RETENCIÓN Y DESEMPEÑO DOCENTE

Un estudio realizado en 2013 sobre IMPACT sugiere que el programa está cumpliendo al menos uno de sus objetivos: retener a los buenos profesores y remover del aula a los docentes ineficaces. El estudio, que abarcó los primeros dos años de implementación de IMPACT, señala que en dicho período se logró retener a cerca del 90% de los maestros calificados como altamente eficaces. Todos los docentes calificados como ineficaces fueron despedidos, y un 30% de los calificados por primera vez como mínimamente eficaces renunciaron voluntariamente. El estudio constató además una mejora en el desempeño de los profesores de buen rendimiento, que estuvieron en el umbral de recibir un bono y de docentes de menor rendimiento que permanecieron en el sistema.²⁵ El resultado neto fue un aumento general de la calidad docente.

Sin embargo, dado que el estudio se centró en los profesores en el umbral de ser despedidos o de hacerse merecedores de un bono, el estudio no provee información sobre los efectos en los profesores “promedio” (los situados en el medio de la distribución). Tampoco deja en claro cómo los cambios recientes en el sistema IMPACT podrían afectar estas conclusiones más adelante. Aún así, los resultados iniciales son promisorios.

En cualquier sistema de evaluación con fuertes consecuencias, existe el peligro de que, por error, se despidan a un

buen profesor o retenga a uno malo. Por ejemplo, en el año escolar 2012-2013 una falla de programación en el cálculo de los puntajes de valor agregado por la firma contratista llevó a dar a 44 profesores—alrededor del 1% del total—calificaciones IMPACT erróneas. La mitad de los docentes recibió puntajes demasiado altos y la mitad demasiado bajos. Un profesor—en un sistema que emplea 4.000—fue despedido como consecuencia de este error. La empresa contratista y el sistema de educación pública de DC reconocieron y enmendaron públicamente el error en cuanto se detectó. DCPS decidió dejar los puntajes que se habían sobre-estimado intactos y aumentar los puntajes que se habían sub-estimado. Al maestro despedido por error se le ofreció el reintegro y una compensación por el error.²⁶

RESULTADOS EN LAS PRUEBAS DE APRENDIZAJE DE LOS ESTUDIANTES

Un fenómeno interesante ha sido el aumento de los puntajes obtenidos por los estudiantes de las escuelas públicas de Washington, DC en las pruebas nacionales rendidas desde 2007. Resultados del NAEP del 2013 muestra que los alumnos de cuarto y octavo grado registran aumentos estadísticamente significativos en los puntajes de matemáticas y lectura.²⁷ En promedio, el distrito registró aumentos de dos dígitos en los

puntajes de matemáticas de los estudiantes de 4to y 8vo grado, y de casi 10 puntos en lectura comparado con 2007, logros que en su mayoría se produjeron desde 2009 en adelante. Los puntajes promedio nacionales subieron en no más de cinco puntos en el mismo período.²⁸ Y a diferencia de la prueba DC CAS, los resultados del NAEP, que no suponen fuertes consecuencias para los docentes, no han sido cuestionados por posibles manipulaciones.

“Los países que consideren implementar reformas similares a IMPACT deberán asegurarse de contar de antemano con los recursos humanos, institucionales y financieros necesarios.”

Es cierto que correlación no implica causalidad, por lo que no se puede decir con certeza si los logros obtenidos se deben a la reforma de evaluación docente, a otros cambios ocurridos en el sistema, o bien a una combinación de factores. Sin embargo, los resultados proveen indicios promisorios de progreso y se requieren de estudios adicionales sobre los vínculos entre reformas específicas y el aprendizaje para beneficiarse al máximo de posibles lecciones.

Implicancias en materia de política educativa

Varios aspectos de esta experiencia son de interés para el contexto latinoamericano:

1. **Privilegiar claramente el aprendizaje y la transparencia:** Los administradores de las escuelas públicas de Washington, DC insistieron desde el inicio que la mejor forma de evaluar la calidad de la enseñanza es en base de cuánto aprenden los estudiantes, postura que mantuvieron pese a los obstáculos y críticas de distintos sectores. Establecieron un marco que define en que consiste una buena enseñanza y han trabajado fuertemente para explicarlo a los profesores, haciendo claro qué se espera que hagan y cómo hacerlo. Manuales sobre cómo

se hacen las evaluaciones, cómo se calculan los puntajes y qué constituye un buen desempeño en distintas categorías están de fácil acceso en el sitio web del Distrito. El distrito ha desarrollado también una videoteca virtual que permite observar a profesores destacados en acción. Pocos países latinoamericanos han definido una buena enseñanza con tal claridad o convertido

el aprendizaje en la principal medida del éxito.

2. ***Un compromiso serio con un sistema evaluativo de calidad estructurado sobre la base de evidencia de estudios rigurosos, personalizado para DC y sostenido en el tiempo.***

DCPS se ha esforzado para que el diseño, evaluación y puesta en práctica del sistema IMPACT sean coherentes con los mejores estudios existentes, y han destinando recursos humanos y económicos considerables para lograr que así sea.²⁹

Además, se ha puesto a cargo a destacados profesionales y se ha logrado que la reforma se mantenga de una administración a la otra. De hecho, cuando el alcalde que designó a Michelle Rhee perdió la reelección -en parte debido los conflictos que generaba el estilo de Rhee- su sucesor puso en su lugar a Kaya Henderson, principal colaboradora de Rhee, quien hizo algunos ajustes a IMPACT pero dejó intacto lo fundamental. En América Latina, la continuidad en materia de políticas públicas no es común, especialmente cuando hay cambios de administraciones. Es más, conseguir los profesionales calificados y los recursos necesarios para gestionar un sistema del tipo IMPACT bien podría ser difícil en países de menores recursos. Los países que consideren implementar reformas similares a IMPACT deberán asegurarse de contar de antemano con los recursos

humanos, institucionales y financieros necesarios.

3. ***Liderazgo fuerte y colaboración.***

El programa IMPACT fue posible en parte porque la administradora de las escuelas públicas de Washington, DC tiene facultades para evaluar a los profesores sin tener que negociar cada detalle con el gremio.³⁰ Sin embargo, el estilo conflictivo de Rhee generó múltiples inconvenientes de implementación y relaciones públicas. El actual alcalde y la nueva administradora han continuado el programa, pero con mayor participación y cooperación de parte de los profesores. El presidente del Sindicato de Profesores de Washington dijo en 2012 que, aunque sigue considerando IMPACT como un sistema imperfecto, las enmiendas realizadas hasta la fecha han logrado mejorarlo y el Sindicato está capacitando a sus miembros para trabajar dentro del sistema.³¹ El régimen de bonos para el buen desempeño fue articulado en forma conjunta entre el Sindicato y los administradores del sistema de educación pública de distrito, y forma parte del convenio colectivo.³² Es decir, las autoridades de DCPS lograron vencer la resistencia inicial y mantenerse firmes en materia de consecuencias laborales y salariales, al tiempo de mantener “la disposición a hacer ajustes estratégicas al sistema” que incluyen la participación de

los profesores³³. Este modelo puede sugerir estrategias para la gestión de cambios en América Latina, donde la resistencia a la evaluación docente será probablemente mayor.

4. ***Respuesta rápida y responsable ante los errores.***

Ningún sistema de evaluación docente es perfecto desde su inicio, especialmente los sistemas que integran complejas fórmulas de valor agregado con evaluaciones de fuertes consecuencias para los docentes. Esto significa que las autoridades deben estar dispuestas y tener la capacidad de hacer ajustes en respuesta a los problemas que surjan durante su implementación. También significa reconocer que algunos errores son inevitables, y contar con planes para resolverlos en cuanto se manifiesten. La estrategia de DCPS de reconocer y resolver rápidamente las erróneas calculaciones de valor-agregado de una forma que trataba de manera justa a los profesores afectados ayudó mantener la confianza en la reforma. En América Latina, donde la desconfianza ante la corrupción es alta y la confianza pública en los gobiernos es baja, construir y mantener una reputación de abordar los errores de manera transparente y justa, será un componente importante para cualquier reforma.

NOTAS

¹ Para mayores antecedentes sobre el sistema de educación pública de Washington, ver District of Columbia Public Schools, "[DCPS at a Glance.](#)"

² NAEP es una prueba nacional estandarizada que mide el nivel de aprendizaje de estudiantes de cuarto, octavo y decimosegundo grado. Aunque mide el desempeño nacional en una variedad de asignaturas, para grandes zonas urbanas y al nivel de estado sólo se entrega datos de lectura y matemáticas de cuarto y octavo grado anualmente. Además, cada cuatro años recoge datos de ciencias y expresión escrita a nivel estatal y de grandes distritos urbanos. Aparte de entregar puntajes promedios, también da cuenta del porcentaje de alumnos que se ubican en los niveles menor que básico, básico, competente y avanzado. Para mayor información, visite el sitio web del [National Center for Education Statistics.](#)

³ Para mayor información sobre el *Marco de Enseñanza y Aprendizaje*, ver [Teaching and Learning Framework](#) 2013-2014 District of Columbia Public Schools.

⁴ Además hay guías adicionales para calcular los puntajes IMPACT de los profesores de educación preescolar, diferencial, inglés como segunda lengua, visitantes y de asignaturas especiales y de profesionales de apoyo al estudiante y especialistas en medios. Ver guías para los Grupos 1, 2 y demás categorías [aquí.](#)

⁵ Los Common Core Standards fueron elaborados por las autoridades de educación y gobernadores de 48 estados del los Estados Unidos y definen objetivos claros para la preparación para la universidad y la carrera, en las asignaturas de inglés y matemáticas de la educación preescolar (kindergarten), primaria y secundaria. Han sido adoptadas voluntariamente en 43 estados, donde están trabajando en su implementación. Para más información, ver el sitio web del [Common Core.](#)

⁶ Para más información, ver [IMPACT – The DCPS Effectiveness Assessment System for School-Based Personnel](#) District of Columbia Public Schools.

⁷ Para más información ver District of Columbia Public Schools, "[Leadership Initiative For Teachers \(LIFT\).](#)"

⁸ Para más información sobre cómo se calcula valor agregado en IMPACT ver District of Columbia Public Schools, "[Individual Value-Added Student Achievement Data;](#)" y Isenberg, E., y Walsh, E., "[Measuring Teacher Value Added in DC, 2012-2013 School Year.](#)" Mathematica Policy Research, September 11, 2013.

⁹ Ver District of Columbia Public Schools, "[General Education Teachers, Group 1.](#)" 2013-2014.

¹⁰ Los profesores calificados como altamente eficaces reciben un bono de US \$10.000 si trabajan en escuelas donde el 60% o más de los estudiantes califican para comidas gratuitas o a bajo costo, US \$10.000 adicionales si la escuela está entre el 40% de menor rendimiento, y otros US \$5.000 si están en el Grupo 1. Los docentes que trabajan en escuelas con una menor proporción de estudiantes que califican para comidas gratuitas o a precios reducidos reciben bonos más bajos (US \$2.000), aunque los profesores pertenecientes al Grupo 1 tienen derecho a US \$1.000 adicionales. Para más información, ver [este resumen por DCPS.](#)

¹¹ Para mayor información sobre oportunidades de apoyo y desarrollo profesional, ver "[Ensuring Teacher Success](#)" District of Columbia Public Schools.

¹² Ver Anderson, N. "[D.C. Schools Gave 44 Teachers Mistaken Job Evaluations.](#)" Washington Post, December 23, 2013.

¹³ Costo que fue ostensiblemente menor el primer año. Según DCPS, en los últimos cuatro años se han entregado unos US \$25 millones en bonos.

¹⁴ Ver Brown, E. "[98 D.C. Teachers Fired for Poor Performance](#)". The Washington Post, August 1, 2012; y Hanushek, E., "[Evidence-based Debates on Teacher Quality.](#)" Education Next, March 24, 2014.

¹⁵ Bill and Melinda Gates Foundation, "[Ensuring Fair and Reliable Measures of Effective Teaching.](#)" 2013.

¹⁶ Ver Headden, S., "[Inside IMPACT: D.C.'s Model Teacher Evaluation System.](#)" Education Sector, 2013.

¹⁷ *Ibíd.*

¹⁸ Esta es una de las sugerencias para prevenir los errores identificados por el estudio Gates sobre la evaluación docente en EEUU. Ver Bill and Melinda Gates Foundation, "[Ensuring Fair and Reliable Measures of Effective Teaching.](#)" 2013, p. 13.

¹⁹ Ver Toppo, G, "[Memo warns of rampant cheating in D.C. public schools.](#)" USA Today, April 11, 2013; and Gillum, J. and M. Bello, "[When standardized test scores soared in D.C., were the gains real?](#)" USA Today, March 30, 2011.

²⁰ District of Columbia Public Schools, "[DC Public Schools Releases All 2009 DC CAS Security Investigation Reports.](#)" March 31, 2011.

²¹ Ver, por ejemplo, Strauss, V., "[Why not subpoena everyone in D.C. cheating scandal – Rhee included? \(update\).](#)" The Washington Post, April 12, 2013; y Sanchez, L., "[Long-Lost Memo Stirs Allegation Of Cheating In D.C. Schools.](#)" NPR, May 6, 2013.

²² EdCORE, "[Evaluation of the DC Public Education Reform Amendment Act \(PERAA\).](#)" July 15, 2013.

²³ Brown, E. "[Teachers in 18 D.C. Classrooms Cheated on Tests Last Year, Probe Finds.](#)" The Washington Post, April 12, 2013.

²⁴ Brown, E. "[Review Finds Serious Test-taking Violations in Four D.C. Schools.](#)" The Washington Post, March 19, 2014.

²⁵ Ver Dee, T. & Wyckoff, J., "[Incentives, Selection, and Teacher Performance: Evidence from IMPACT.](#)" National Bureau of Economic Research, 2013.

²⁶ Ver Anderson, N. "[D.C. Schools Gave 44 Teachers Mistaken Job Evaluations.](#)" Washington Post, December 23, 2013 y National Council on Teacher Quality, "[The value-add of value-add.](#)" PDQ: Pretty Darn Quick Blog, December 24, 2013.

²⁷ Ver National Center for Education Statistics, "[2013 Mathematics Assessment Report Card;](#)" y "[2013 Reading Assessment Report Card.](#)"

²⁸ Ver District of Columbia Public Schools, "[2013 DC NAEP Student Achievement Results.](#)"

²⁹ Para más información sobre la experiencia de Washington y las características que la hace distinta a otros esfuerzos de evaluación docente, ver Hess, R., "[Making Sense of the Dee-Wyckoff IMPACT Study.](#)" Education Week, October 18, 2013.

³⁰ Hess, R., "[Making Sense of the Dee-Wyckoff IMPACT Study.](#)" Education Week, October 18, 2013.

³¹ Ver Brown, E. "[98 D.C. Teachers Fired for Poor Performance](#)". The Washington Post, August 1, 2012.

³² Ver District of Columbia Public Schools, "[IMPACTplus.](#)"

³³ Hess, R., "[Making Sense of the Dee-Wyckoff IMPACT Study.](#)" Education Week, October 18, 2013.



1211 Connecticut Ave., NW, Suite 510

Washington, DC, 20036

Phone: 202-822-9002 • Fax: 202-822-9553

Email: education@thedialogue.org

Web: thedialogue.org & prealblog.org