



¿Qué funciona en educación?

Evidencias para la mejora educativa

1

julio de 2015

Qué funciona en educación: la pregunta necesaria

Miquel Àngel Alegre Canosa

¿Es recomendable implantar incentivos salariales para el profesorado vinculados al rendimiento académico de los estudiantes?

J. Oriol Escardíbul

“La educación se ha basado durante demasiado tiempo en inercias y tradiciones, y los cambios educativos en intuiciones o creencias no fundamentadas. El movimiento «qué funciona» irrumpe en el mundo de la educación con un objetivo claro: promover políticas y prácticas educativas basadas en la evidencia. Ivàlua y la Fundación Jaume Bofill han creado una alianza para impulsar este movimiento en Cataluña.”



¿Qué funciona en educación?

Evidencias para la mejora educativa

Qué funciona en educación: la pregunta necesaria



Miquel Àngel Alegre Canosa

Analista del Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua). Experto en evaluación de políticas educativas.

En casi todas las sociedades, la agenda de la educación está orientada por motivaciones de ideología política, saberes convencionales, folclore e ilusiones (wishful thinking), en tanto que se esfuerza por abordar las necesidades e intereses de la economía, la empresa, los trabajadores, la ley y el orden, la sociedad civil, las elecciones familiares, así como, de forma menos retórica, los niños, los jóvenes y los adultos que componen la comunidad educativa (...). Buena parte de este impulso representa el triunfo del deseo sobre la razón, del sentimiento sobre la demostración de efectividad, de la intuición sobre la evidencia.

Davies, "What is Evidence-Based Education?", junio de 1999 [1].

Corren buenos tiempos para las reformas educativas basadas en la evidencia. Gracias sobre todo a la iniciativa Investing in Innovation y al Institute of Education Sciences, está creciendo el número de programas que se han probado efectivos en todas las materias y cursos. (...) [Esto] abre nuevas oportunidades para la política y la práctica en educación. El programa de ayudas School Improvement Grants ofrece opciones a aquellas escuelas que deciden adoptar modelos contrastados de reforma comprensiva. Otras áreas de la política educativa comenzarán pronto a fomentar o a incentivar el uso de programas basados en evidencia robusta.

Slavin, "Are Proven Educational Innovations Ready for Prime Time?", octubre de 2014 [2].

En la primera cita, Philip Davies se refiere a la situación global de la educación en los países desarrollados (incluidos los Estados Unidos); en el segundo fragmento, Robert E. Slavin pone el foco sobre los Estados Unidos. Entre las palabras de uno y otro distan quince años.

Y, en cambio, nos atreveríamos a decir que el panorama actual de la política educativa en Cataluña queda mejor descrito en las palabras de Davies que en las de Slavin. En efecto, en Cataluña, tanto como en España y en otros países de nuestro entorno, las decisiones sobre las políticas educativas —sobre su lanzamiento, su mantenimiento, su reforma, su supresión— raramente se basan en evidencias empíricas sólidas sobre su efectividad; de hecho, no es infrecuente que detrás de estas decisiones no exista evidencia empírica de ningún tipo.

¿Es positivo reducir los años de comprensividad del sistema educativo? ¿Es positivo ampliar la cobertura pública de la educación de 0 a 3 años? ¿Es efectivo que las familias puedan escoger con total libertad la escuela que quieren para sus hijos? ¿Es negativo ampliar la ratio máxima de alumnos por aula? ¿Es bueno ampliar el número de horas que los alumnos pasan en la escuela a diario? ¿Funciona la jornada continua en la educación secundaria? ¿Y en la primaria? ¿Funciona un programa de becas a la continuidad posobligatoria modulado por renta? ¿Y basado en el mérito? ¿Funcionan los incentivos económicos al profesorado? ¿Los rankings escolares? Y todo ello, ¿funciona para qué? ¿O para quién?

Muchos de estos interrogantes han sido tratados y debatidos en foros muy diversos. Al mismo tiempo, todos ellos entroncan con ámbitos de decisión que la política educativa ha afrontado en los últimos años. Sin embargo, no podemos decir que el planteamiento del debate público (más o menos especializado) o el abordaje de las reformas en cuestión hayan tenido demasiado en cuenta la evidencia empírica acumulada sobre la efectividad de las distintas opciones.

Algo parecido sucede cuando lo que hay en juego son programas, actuaciones o prácticas educativas de carácter más “micro” (decisiones pedagógicas, curriculares, de organización de los centros, etc.).

¿Funciona el uso de las TIC en la enseñanza? ¿Funcionan los grupos de nivel? ¿Las agrupaciones flexibles? ¿Los grupos cooperativos? ¿El trabajo por proyectos? ¿Las tutorías intensivas 1 por 1? ¿Es malo poner deberes? ¿O hacer repetir curso? ¿Son beneficiosos los programas de verano? ¿Y los programas familia-escuela? También aquí sería necesario preguntarnos: todas estas iniciativas, ¿en relación con qué objetivos y colectivos deberíamos valorarlas?

En Cataluña, tanto como en España y en otros países de nuestro entorno, las decisiones sobre las políticas educativas —sobre su lanzamiento, su mantenimiento, su reforma, su supresión— raramente se basan en evidencias empíricas sólidas sobre su efectividad.



De nuevo, los debates y las decisiones prácticas que se han formulado en torno a estas cuestiones, raras veces han sido respaldados por ejercicios serios de recogida y análisis de evidencias, ejercicios susceptibles de informar de la capacidad de impacto de unas y otras actuaciones.

Con todo esto no queremos decir que no exista conocimiento sobre la efectividad que han demostrado en distintos países o contextos escolares tanto las políticas educativas de índole más “estructural” como las intervenciones más “micro”. El problema es más bien otro, y suele estar relacionado con otros tipos de dificultad:

¿cómo identificamos el conocimiento más robusto sobre una determinada materia? ¿De dónde sacamos el tiempo para buscarlo y revisarlo? ¿Cómo lo traducimos y conseguimos aprovecharlo para nuestro contexto más inmediato? Y, sobre todo, ¿cómo defendemos su valor en un campo de juego (el de la formación de programas y políticas) donde otros factores, como los intereses o compromisos políticos, los grupos de interés, las inercias institucionales, determinados valores culturales, presiones mediáticas, etc. suelen tener un papel preponderante?

Con esta nueva publicación, Ivàlua y la Fundación Jaume Bofill quieren aportar un pequeño grano de arena a la confrontación de estos retos. Por tanto, nos proponemos hacer de ***Qué funciona en educación*** una publicación capaz de aportar evidencia científica al debate y a la práctica educativa, y hacerlo en un formato susceptible de combinar rigor, divulgación y conexión con los problemas educativos que se le plantean actualmente a nuestro país.

Un poco de contexto: “qué funciona” en educación

¿Qué funciona en educación? Esta es la pregunta que orienta esta publicación, y que los distintos números que la componen darán respuesta focalizando su atención en ámbitos de intervención educativa de especial relevancia. Enlazamos, por tanto, con un movimiento, denominado genéricamente “qué funciona” (*what works*), que tiene su razón de ser en la apuesta por avanzar hacia decisiones y políticas basadas en la evidencia (*evidence-based policy making*).

Está fuera de las posibilidades de esta introducción profundizar en el origen, sentido y desarrollos del movimiento “qué funciona” en su conjunto, un movimiento que arranca en el ámbito sanitario a principios de la década de los setenta [3]. En lo que respecta al ámbito educativo, podríamos considerar el impulso “qué funciona” como la última evolución de la denominada *school effectiveness research* (SER) o investigación sobre la efectividad escolar. Esta línea de investigación nació a principios de los años ochenta como reacción a una literatura que, amparándose en las conclusiones del conocido informe Coleman (1966) [4] situaba la composición socioeconómica y étnica de las escuelas como el principal condicionante del rendimiento académico de los estudiantes, más allá de los recursos de que disponen las escuelas y lo que hacen con ellos. Los estudios de la SER, por el contrario, tratan de demostrar que los centros educativos, la manera en que trabajan y se organizan, pueden contribuir a mejorar los resultados de sus alumnos, llegando incluso a reducir las desigualdades académicas entre unos y otros [5].

El enfoque “qué funciona” puede leerse como una derivada de esta escuela de investigación, con la que comparte el objetivo de analizar la efectividad de las intervenciones educativas y la finalidad última de aportar evidencia empírica a los procesos de toma de decisiones. Ambas perspectivas, sin embargo, se distinguen por un elemento central: el tipo de evidencia que tienden o aspiran a generar. Los estudios de la SER, basados en datos observacionales y en el uso de técnicas estadísticas multivariantes, proporcionan evidencia sobre el grado de asociación existente entre variables independientes (intervenciones o determinadas características escolares) y variables dependientes (por ejemplo, rendimiento académico de alumnos y escuelas) [6]. La perspectiva “qué funciona”, en cambio, aspira a evidenciar relaciones causales entre la intervención y los *outcomes* considerados. Y difícilmente pueden inferirse causalidades si no es a través de evaluaciones experimentales (con asignación aleatoria entre grupo de tratamiento y grupo de control) y, en ciertas circunstancias, de estudios cuasi experimentales.¹

Hay que decir que el campo de las políticas educativas ha sido tradicionalmente poco sensible a los avances y alertas de la perspectiva “qué funciona”. Ha sido en los últimos tiempos, sobre todo en el mundo anglosajón, cuando ha ido cristalizando en aquello que algunos han denominado una “revolución tranquila” en la relación entre educación y evidencia [7]. Las referencias que se muestran en la [tabla 1](#) dan cuenta de algunos de estos avances.

El movimiento “qué funciona” (*what works*) tiene su razón de ser en la apuesta por avanzar hacia decisiones y políticas basadas en la evidencia.



Nos gustaría resaltar dos iniciativas especialmente paradigmáticas, ejemplos de institucionalización de la perspectiva “qué funciona” en el ámbito de la política educativa. El primero lo situamos en los Estados Unidos, en la iniciativa [What Works Clearinghouse](#) (WWC), impulsada por el Institute of Education Sciences del Gobierno federal de los Estados Unidos. La iniciativa nació en el año 2002, vinculada al programa No Child Left Behind y al énfasis que pone en la necesidad de que las actuaciones educativas sufragadas con financiación pública dispongan de —y también generen— pruebas de efectividad. Así, el WWC revisa las evaluaciones existentes sobre un amplio conjunto de intervenciones educativas, valora la calidad metodológica de esta evidencia de acuerdo con estándares muy exigentes y, sobre esta base, establece el grado de efectividad de los programas en cuestión². Toda esta información se pone al alcance de responsables políticos, investigadores y miembros de la comunidad educativa, con el objetivo de promover decisiones y prácticas basadas en la evidencia.

- 1 No pasamos por alto que este nuevo foco sobre la causalidad ha contado, en buena medida, con el acicate de la entrada de la disciplina económica (de la econometría, en particular) en el análisis y la evaluación de la efectividad de las políticas educativas, espacios que hasta hace pocos años estaban dominados principalmente por investigadores vinculados a las ciencias de la educación.
- 2 Únicamente los estudios de efectividad basados en diseños experimentales tienen opción de recibir la cualificación más alta del WWC (“Meets the WWC Standards Without Reservations”).
- 3 Así, la Education Endowment Foundation forma parte de la denominada What Works Network, iniciativa del Cabinet Office del Gobierno británico dirigida a impulsar el uso de evidencias en la toma de decisiones sobre los servicios públicos. La red *what works* la componen, además de la EEF, cinco centros especializados, respectivamente, en salud y protección social, vejez, desarrollo local, atención a la infancia y criminalidad.

Un segundo ejemplo lo encontramos en el Reino Unido, en la constitución de la Education Endowment Foundation (EEF) en el año 2011. Esta fundación, reconocida por el gobierno británico como el centro *what works* especializado en el ámbito educativo, concentra su misión en la generación de evidencias en torno a lo que funciona (y no funciona) en la lucha contra las desigualdades sociales en educación. Entre las principales herramientas que desarrolla la EEF están los denominados Teaching and Learning Toolkit (referido a la educación primaria y secundaria) y Early Years Toolkit (referido a la educación infantil). En esas plataformas se resume y califica el conocimiento existente sobre la efectividad de distintas líneas de actuación educativa, tomando como base la evidencia aportada por estudios y revisiones sistemáticas de especial relevancia. Todo ello otorga a la EEF, y también al WWC una función de interfaz o mediación entre el mundo académico y el mundo de la decisión política, facilitando la adaptación del tempo y del lenguaje típico del primero a las necesidades de inmediatez y claridad del segundo.

En Cataluña, el campo de la política educativa está todavía lejos de lo que representan estos avances y, como decimos, creemos que vale la pena tratar de invertir esta situación.

Instituciones como el What Works Clearinghouse (Estados Unidos) o la Education Endowment Foundation (Reino Unido) tienen una función de interfaz o mediación entre el mundo académico y el mundo de la política y la práctica educativa.



Tabla 1.
Principales repositorios y publicaciones periódicas de impacto en educación

Repositorios “qué funciona” en educación	Institución	Publicaciones periódicas “qué funciona en educación”
What Works Clearinghouse http://ies.ed.gov/ncee/wwc/	Institute of Education Sciences, U.S. Department of Education, Estats Units	Practice Guides / Intervention Reports / Single Study Reviews / Quick Reviews http://ies.ed.gov/ncee/wwc/publications_reviews.aspx
Teaching and Learning Toolkit http://educationendowmentfoundation.org.uk/toolkit/	Education Endowment Foundation, Regne Unit	(Material <i>online</i> disponible en el sitio web de la institución: varios formatos)
Best Evidence Encyclopedia (US) http://www.bestevidence.org/	Center for Data-Driven Reform in Education, Johns Hopkins University, Estats Units	Better: Evidence-based Education Magazine http://www.betterevidence.org/ Best Evidence in Brief http://education.jhu.edu/research/crre/newsletter.html
Best Evidence Encyclopedia (UK) http://www.bestevidence.org.uk/	Institute for Effective Education, University of York, Regne Unit	
Evidence 4 Impact, http://www.evidence4impact.org.uk/index.php		
Impact Evaluations in Education http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E	The World Bank	(Material <i>online</i> disponible en el sitio web de la institución: varios formatos)

Fuente: elaboración propia

Qué recogemos: evidencias sólidas de impacto

Queda claro que la pregunta “¿qué funciona?” nos sitúa ante la preocupación por la efectividad de las políticas, por su impacto sobre la realidad que aspiran a mejorar. Nos preguntamos, por tanto, qué sabemos sobre la capacidad que tienen las políticas y las intervenciones educativas para producir impactos relevantes en dimensiones clave de la realidad educativa. ¿Y qué entendemos por impacto? Pues justamente eso: aquellos cambios en la realidad intervenida que son exclusivamente atribuibles a la intervención en cuestión.

No se nos escapa que identificar el impacto de un programa (educativo o de otro tipo) no es fácil. Por una parte, la dimensión sobre la que se pretende inducir un cambio (por ejemplo, el nivel de abandono educativo) está sometida a la influencia de muchos otros factores (de contexto político, social, laboral, económico, etcétera) que son contemporáneos al programa. Así, lo que les sucede a los participantes después de recibir el programa no tiene por qué ser necesariamente efecto de la intervención en sí.

Por otra parte, la estimación del impacto tampoco puede basarse en la comparación simple y directa de los resultados de los participantes en el programa con los resultados de quienes no han participado en él. Podría perfectamente darse el caso de que participantes y no participantes (ya se trate de escuelas, alumnos o profesores) fueran distintos en ciertas características que podrían influir sobre los resultados de interés. Estas diferencias son consecuencia de lo que se conoce como el “sesgo de selección” en el acceso al programa: bien los participantes son seleccionados activamente en base a determinados requisitos objetivos o valoraciones subjetivas, o bien son los propios participantes los que se autoseleccionan (podrían ser los que están más motivados o los que tienen mejores disposiciones escolares). Por tanto, una comparación directa entre la situación de participantes y no participantes suele resultar en una estimación sesgada del impacto de la intervención.

Así pues, la pregunta “¿qué funciona?” no admite cualquier respuesta: tan solo las que se fundamentan en metodologías sólidas, es decir, aquellas que establecen de forma robusta las relaciones de causalidad entre la política y los cambios en la problemática abordada, superando así los dos obstáculos antes mencionados (factores contemporáneos y sesgo de selección). Esta premisa restringe la atención a aquella evidencia de impacto que se haya obtenido mediante métodos experimentales o de tipo cuasiexperimental.

La pregunta “¿qué funciona?” nos sitúa ante la preocupación por la efectividad de las políticas, por su impacto sobre la realidad que aspiran a mejorar y se fijan en aquella evidencia de impacto que haya sido obtenida mediante métodos experimentales o de tipo cuasiexperimental.



En pocas palabras: el objetivo de estos métodos es llegar a identificar un grupo de control (no participantes en el programa o no beneficiarios de la intervención educativa) lo más parecido posible al grupo de participantes. Idealmente, participantes y no participantes deberían ser idénticos en todas aquellas características que podrían

tener relación con la probabilidad de éxito del programa. De esta manera, podremos afirmar que las diferencias que llegamos a observar entre un grupo y otro una vez finalizada la intervención representan su impacto.

La estrategia más robusta pasa por aleatorizar (mediante un sorteo) quién, entre el conjunto de candidatos elegibles (escuelas, alumnos, profesores...), recibe el programa y quién queda asignado al grupo de control. Este procedimiento es la piedra angular del diseño experimental en evaluación, diseño que algunos autores han calificado como el *gold standard* a la hora de aportar evidencias sobre qué funciona [8]. Cuando la aplicación de este diseño, por unos u otros motivos, no es viable, pueden entrar en juego determinados diseños de evaluación cuasiexperimental con contrafactual, en particular, aquellos diseños susceptibles de minimizar los sesgos de selección que hemos mencionado antes.

No podemos extendernos aquí explicando las características y aplicaciones de los métodos experimentales y cuasiexperimentales en la evaluación del impacto de las políticas educativas. Esta explicación se aborda con detalle en la guía “Cómo evaluar el impacto de las políticas educativas” (de la colección Ivàlua de guías prácticas de evaluación de políticas públicas) [9].

Cómo recogemos y analizamos las evidencias

En cualquier caso, no es función de los artículos de esta publicación evaluar directamente los impactos de los distintos programas. Su propósito es más bien recoger y analizar lo que otros estudios, empleando métodos rigurosos, han dicho sobre su capacidad de impacto. Son diversos los métodos que pueden utilizarse cuando de lo que se trata es de elaborar un cierto balance del conocimiento acumulado sobre un determinado tema o problemática, y hay que tener en cuenta que su nivel de exhaustividad y sistematicidad puede ser muy dispar.

La tabla 2 sintetiza las características de los principales métodos reconocidos de recolección y revisión de evidencias: del método menos sistemático y menos protocolizado (revisión ad hoc de la literatura) al método más exhaustivo (revisión sistemática múltiple).

El propósito de los artículos de esta publicación es recoger y analizar lo que otros estudios, empleando métodos rigurosos, han dicho sobre el impacto de las políticas y prácticas educativas. Se priorizará el procedimiento de “revisión de revisiones” como estrategia de recogida y revisión de evidencias.



Tabla 2.
Procedimientos habituales de recogida y revisión de evidencias

Procedimiento de recogida/revisión	Naturaleza	Definición	Limitaciones	Duración
Revisión de la literatura ad hoc (<i>literature review</i>)	No sistemática	Recopilación de los estudios más relevantes o ilustrativos No explicitan el proceso de búsqueda bibliográfica	Existe un riesgo alto de sesgo de selección: estudios disponibles o que tienen resultados positivos Adecuado en contextos de recursos limitados	de 1 semana a 2 meses
Revisión de corto alcance (<i>quick scoping review</i>)	No sistemática	Elabora un “mapa” de la literatura disponible sobre un tema	Responde a preguntas muy descriptivas Busca en pocas bases de datos Busca con pocas palabras clave	de 1 semana a 2 meses
Revisión de revisiones (<i>review of reviews</i>)	Sistemática	Elabora síntesis de las revisiones realizadas, más que de materiales primarios	Puede utilizarse solamente cuando hay suficiente evidencia acumulada y revisada La calidad de las revisiones seleccionadas puede ser diversa	de 2 a 4 meses
Revisión rápida de la evidencia (<i>rapid evidence assessment</i>)	Sistemática	Búsqueda restringida sobre la efectividad de una política Cuenta con un protocolo de búsqueda Puede trazarse y replicarse	Hay que delimitar una pregunta sobre efectividad muy concreta La ventana temporal es limitada y reciente	de 2 a 6 meses
Revisión sistemática completa (<i>full systematic review</i>)	Sistemática	Revisión amplia de la literatura sobre la efectividad de una política Cuenta con un protocolo sistemático Busca en bases de datos académicas como literatura gris	El proceso de recopilación documental puede ser muy artesanal Requiere mucho tiempo Puede requerir recursos humanos adicionales, internos o externos	de 8 a 12 meses
Revisión sistemática múltiple (<i>multi-arm systematic review</i>)	Sistemática	Revisión amplia de la literatura atendiendo a distintas preguntas de efectividad de una política Ídem “revisión sistemática completa”	Incrementa las limitaciones de la “revisión sistemática completa”	Más de 12 meses

Fuente: adaptación de <http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is>

De forma general, los artículos que se incluirán en la serie **Qué funciona en educación** priorizarán el procedimiento de “revisión de revisiones” como estrategia de recogida y revisión de evidencias. Esta apuesta, que se justifica atendiendo a los estándares de rigor, cobertura y agilidad a los que aspiramos, no excluirá la posibilidad de incluir referencias a evaluaciones o estudios primarios que puedan tener una relevancia especial (por el foco temático, por la metodología utilizada, por la proximidad al contexto geográfico o político catalán, etc.). El recurso a la revisión de estudios primarios será obligado especialmente en las contribuciones que aborden ámbitos de intervención sobre los que no haya suficiente evidencia sólida acumulada y ya revisada.

En qué resultados (*outcomes*) nos fijamos

Como no podría ser de otra manera, la preocupación por “qué funciona” se ha orientado hacia múltiples ámbitos y resultados (*outcomes*) (véase la [tabla 3](#)): resultados cognitivos (rendimiento competencial) y no cognitivos (habilidades sociales, emocionales, metaregulativas), logros formativos (titulaciones conseguidas), transiciones educativas (por ejemplo, itinerarios más allá de la etapa obligatoria), resultados laborales (inserción y condiciones laborales) o incluso en el ámbito de la salud (física y mental) o de la seguridad (prácticas delictivas). En otras palabras, las evidencias de efectividad de unos programas y otros pueden revisarse atendiendo a múltiples variables de interés.

Tabla 3.
Outcomes habituales en la evaluación de impacto de políticas educativas

Ámbito	Outcomes	Instrumentos de medida
Cognitivo	Rendimiento en competencias académicas clave (lengua, matemáticas, ciencias, etc.)	Evaluaciones escolares (cuantitativas y cualitativas) Pruebas estandarizadas (internas o externas) Pruebas ad hoc de competencias (específicas del programa)
No cognitivo	Desarrollo social y emocional Seguridad y autonomía personal Actitudes, disposiciones y expectativas escolares	Evaluaciones escolares (cuantitativas y cualitativas) Cuestionarios estandarizados (escalas métricas homologadas) Cuestionarios ad hoc (específicos del programa)
Logros	Niveles de estudios conseguidos Niveles de graduación/repetición en etapas clave	Registros administrativos (educación) Paneles de datos longitudinales
Transiciones	Itinerarios seguidos entre etapas educativas	Registros administrativos (educación/laboral) Paneles de datos longitudinales
Laboral	Inserción laboral (velocidad de acceso a la ocupación) Condiciones laborales (estabilidad y salarios) Idoneidad laboral (correspondencia según cualificación y tipo de formación)	Registros administrativos (laboral) Paneles de datos longitudinales Cuestionarios ad hoc (específicos del programa)
Salud y seguridad	Estado de salud, física y mental Relación con prácticas delictivas o de riesgo	Registros administrativos (servicios sociales y salud) Paneles de datos longitudinales Cuestionarios ad hoc (específicos del programa)

Fuente: elaboración propia

De hecho, los distintos repositorios de evidencias que mostrábamos en la [tabla 1](#) ofrecen distintas alternativas de lectura:

1. *Del outcome al programa.*

Se identifican programas efectivos (y no efectivos) según el tipo de *outcome* seleccionado. Este es el esquema propio del espacio What Works Clearinghouse del Institute of Education Sciences (Estados Unidos).

2. *Del tipo de programa al outcome de interés.*

Identificados un conjunto de espacios y tipos de intervención, se sintetiza su efectividad sobre unos *outcomes* u otros. Por ejemplo, los Toolkits de la Education Endowment Foundation (Reino Unido) revisan la capacidad de impacto de distintas líneas de intervención, tanto en la educación infantil como en la primaria y secundaria, sobre el rendimiento de los alumnos social y académicamente más desfavorecidos.

La perspectiva que adoptamos en esta publicación comparte el esquema y el foco de la EEF. Así, los artículos que irán nutriendo la serie partirán de una entrada temática, por tipo de intervención, y se preguntarán entonces sobre su capacidad de repercutir positivamente en el progreso y oportunidades educativas de los alumnos, en particular, de los más vulnerables. Obviamente, el alcance y las características de los ámbitos de intervención que irán tratándose condicionarán el perfil final de este foco; esto es, si la preocupación se fija en un tipo de vulnerabilidad o en otro, en unas etapas educativas o en otras, en unos contextos escolares o en otros.

Los artículos de la serie se preguntarán sobre la capacidad que tienen las intervenciones para repercutir positivamente en el progreso y oportunidades educativas de los alumnos, en particular, de los más vulnerables.



Y comenzamos por los programas de incentivos al profesorado

La definición de las condiciones salariales y de incentivación económica de los docentes es un elemento central de las políticas de profesorado, junto con otros ámbitos de atención, como puedan ser los sistemas de selección y acceso a la función docente, las condiciones de trabajo y carreras profesionales, la formación inicial y continuada, los mecanismos de evaluación docente, el apoyo a la profesionalización y el liderazgo de los centros, etc. En Cataluña, buena parte de estos ámbitos ha sido objeto de intervenciones y reformas a lo largo de los últimos años.

En lo que respecta a los incentivos económicos, en los últimos años la administración catalana ha introducido algunas reformas en la definición de los méritos docentes que, en algún sentido, representan un tímido avance hacia la introducción de una cierta lógica de pago por resultados. Este es, por ejemplo, el caso de las últimas medidas tomadas en relación con la concesión de los estadios de promoción (y de los complementos salariales vinculados a ellos) ([Acuerdo GOV/29/2012](#)); en

particular, la consideración de la implicación del docente en la mejora de los resultados de los centros como criterio básico para meritarse la superación de los distintos estadios.

En todo caso, aún estamos lejos del nivel de desarrollo que los sistemas y mecanismos de incentivación económica han experimentado en otros países occidentales. Un caso paradigmático es el de los Estados Unidos, no solamente por la diversidad de esquemas de incentivación del profesorado que se ha desplegado allí, sino también por la cantidad de evidencia empírica (a menudo evidencia de tipo experimental) que se ha generado en torno a la efectividad de estos esquemas. Enseguida tendremos ocasión de comprobarlo.

En Cataluña, el debate sobre las bondades y maldades de los sistemas de incentivación económica del profesorado ha sido un debate marcadamente ideologizado, cargado de resistencias y apriorismos, un debate que ha dejado poco margen a la generación y aprovechamiento del conocimiento científico acumulado sobre esta cuestión.

Somos conscientes de que nos encontramos ante un tema muy delicado. Y podemos compartir una parte importante de los recelos que afloran fácilmente ante el planteamiento de determinados esquemas de incentivos económicos: infravaloración de la motivación intrínseca del profesorado (vocación docente), introducción de dinámicas competitivas en entornos intencionadamente colaborativos, estímulo de prácticas de *teaching to the test*, etc. Entendemos, sin embargo, que vale la pena conocer si distintos esquemas de incentivos han funcionado o no en otros contextos, y en qué grado lo han hecho, es decir, si han sido capaces o no de lograr impactos significativos en los resultados educativos de los alumnos. Entendemos que vale la pena aprender de este conocimiento acumulado, de tal manera que, si alguna vez nos planteamos la oportunidad de ampliar o de contener los avances hacia unas fórmulas u otras de incentivación económica, lo podamos hacer considerando, entre otras cosas, lo que sabemos con certeza sobre su efectividad (o inefectividad).

Por tanto, ¿cómo son de efectivos los sistemas de incentivación económica del profesorado? ¿Funcionan o no funcionan? Más aún, ¿qué recorrido o expectativa de efectividad podrían tener en un contexto social y educativo como el catalán? Estos son los interrogantes que se plantea Oriol Escardíbul en el siguiente artículo, y a los cuales trata de responder huyendo de apriorismos y prejuicios, confiando básicamente en lo que nos dice la evidencia empírica.

Queremos conocer si distintos esquemas de incentivos han funcionado o no en otros contextos y en qué grado lo han hecho.



Referencias

- [1] Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47(2), 108–121. http://www.jstor.org/stable/3122195?seq=1#page_scan_tab_contents
- [2] Slavin, R. E. (2014). Are Proven Educational Innovations Ready for Prime Time? [The Huffington Post]. http://www.huffingtonpost.com/robert-e-slavin/are-proven-educational-in_b_6074382.html
- [3] Cochrane, A. L. (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. Nuffield Provincial Hospitals Trust. http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/Effectiveness_and_Efficiency.pdf
- [4] Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: US Congressional Printing Office. <http://files.eric.ed.gov/fulltext/ED012275.pdf>
- [5] Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. Londres: Falmer Press.
- [6] Luyten, H., Visscher, A., & Witziers, B. (2005). School Effectiveness Research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, 16(3), 249-279. <http://www.tandfonline.com/doi/abs/10.1080/09243450500114884?journalCode=nnes20>
- [7] Collins, K. (2013). "Show me the evidence" - *A quiet revolution - the activity and impact of the EEF*. Institute of Education, Londres.
- [8] Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. A.T.P. Schultz i J.A. Strauss (Eds.) *Handbook of Development Economics - Vol. 4* (pp. 3895-3952). Amsterdam: North-Holland. <http://economics.mit.edu/files/806>
- [9] Alegre, M.A. (2015). *Com avaluar l'impacte de les polítiques educatives*. Col·lecció Ivàlua de guies pràctiques sobre avaluació de polítiques públiques, n. 11. Barcelona: Institut Català d'Avaluació de Polítiques Públiques. <http://www.ivalua.cat/generic/static.aspx?ID=669&>

¿Es recomendable implantar incentivos salariales para el profesorado vinculados al rendimiento académico de los estudiantes?



J. Oriol Escardíbul

Profesor de Economía aplicada en la Universidad de Barcelona, e investigador del Instituto de Economía de Barcelona (IEB) y del Grupo Interdisciplinar de Políticas Educativas (GIPE).

Motivación

En la agenda política de los gobiernos se encuentra aumentar el rendimiento académico de los estudiantes. La evidencia empírica sugiere que es mejor invertir en la calidad del profesorado que en su cantidad [1] [2]. A este respecto, uno de los ámbitos que ha recibido mucha atención recientemente, en especial en los Estados Unidos, ha sido el establecimiento de programas de incentivos salariales denominados *merit pay* o *pay per performance* (pago por rendimiento, PPR). El PPR consiste en evaluar la actuación de los docentes y ofrecer un incentivo monetario a los que logran determinados niveles. En estas evaluaciones cada vez tiene un mayor peso el análisis de la actuación del profesorado vinculada al rendimiento de los estudiantes en pruebas externas, ya que se quieren premiar las acciones que comportan mejoras en su aprendizaje.

Una veintena de países desarrollados tienen incentivos salariales basados en la evaluación del profesorado, y más de la mitad de los estados de los Estados Unidos, aunque no todos incluyen el rendimiento de los estudiantes con la misma intensidad. En este sentido, Cataluña acaba de incluir el PPR en su legislación, considerando parcialmente el rendimiento de los alumnos como factor de evaluación. El número de estudios es insuficiente, y muy centrado en los Estados Unidos, pero permite concluir que a pesar de que un sistema de PPR puede mejorar el rendimiento de los estudiantes, el éxito no está garantizado.

Preguntas que guían la revisión

En esta revisión se consideran distintos programas existentes de incentivos de pago por rendimiento. Estos pueden vincularse a *inputs* o a *outputs* de la función de producción educativa. Entre los primeros se consideran las tareas que desarrollan los profesores en relación con la práctica docente. Entre los segundos, se valora la aportación del profesorado a partir de la actuación de los estudiantes, medida normalmente en términos de resultados, o de mejoras en los resultados, en pruebas externas. En este análisis se consideran principalmente indicadores basados en *outputs*, ya que han centrado la atención de las administraciones educativas recientemente y resultan más controvertidos. La revisión quiere responder a cuatro interrogantes básicos: a) ¿Qué impacto tienen los métodos de PPR sobre los resultados de los alumnos?; b) ¿Cuál es su incidencia sobre la asignación de profesorado a los centros?; c) ¿Cómo valora el profesorado el uso de estos métodos?; d) ¿Cuál sería la aplicabilidad de un sistema de incentivos en Cataluña?

Revisión de la evidencia

Objeto de la revisión

El objetivo de esta revisión es conocer si la implantación de un sistema de incentivos económicos para el profesorado, basado en la evaluación de su actuación, mejora el rendimiento de los estudiantes. No se desarrolla un metaanálisis sofisticado porque los estudios revisados incorporan sistemas de pago por rendimiento (PPR) muy variados, como también lo son las variables relacionadas con la definición de los resultados; además, la cantidad de estudios es insuficiente. Ahora bien, sí permite poder formular determinadas conclusiones. Se analizan tanto evaluaciones de programas concretos como revisiones de las mismas.¹ Antes de considerar las evaluaciones se exponen los formatos de los incentivos de PPR, que son muy diversos:

El objetivo de esta revisión es conocer si la implantación de un sistema de incentivos económicos para el profesorado, basado en la evaluación de su actuación, mejora el rendimiento de los estudiantes.



- Índice de rendimiento basado en *inputs*, *outputs* o una mezcla de ambos.
- En el caso de los *outputs*, puede tenerse en cuenta el resultado de los alumnos o la mejora en los resultados. En el último caso, normalmente se usan dos tipos

¹ Diversas instituciones hacen revisiones de estudios singulares de PPR, pero casi no hay metaanálisis, sino valoraciones de la calidad metodológica de la investigación examinada. En este documento se cita preferentemente el repositorio on-line donde estas instituciones depositan las evaluaciones mencionadas, aunque también se han revisado los estudios primarios (originales). Los repositorios con documentación respecto al PPR son los siguientes:

- What Works Clearinghouse (WWC), del Institute of Education Sciences, del Departamento de Educación del Gobierno de los Estados Unidos. <http://ies.ed.gov/ncee/wwc/>
- Teaching and Learning Toolkit, de la Education Endowment Foundation en el Reino Unido. <https://educationendowmentfoundation.org.uk/toolkit/>
- Impact Evaluation in Education, del Banco Mundial. <http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E>

de métodos. El primero compara las mejoras de resultados de los estudiantes en pruebas externas con las mejoras esperadas, que se estiman teniendo en cuenta los resultados de los estudiantes en pruebas anteriores y sus características socioeconómicas. Así, se compara la mejora obtenida por un alumno con la mejora que cabría esperar dadas sus circunstancias y trayectoria académica anterior; la diferencia o “valor añadido” se asigna al profesor. El segundo tiene en cuenta la mejora de cada estudiante en los resultados respecto al resto de alumnos (o solamente los más similares), de manera que se analiza la mejora en términos comparados. Se asigna a cada profesor una mejora media derivada de la mejora relativa de los resultados de cada uno de sus estudiantes.

- Individual (el incentivo se otorga a cada profesor a partir del rendimiento de sus estudiantes) o colectivo. Este último otorga el premio a un grupo de profesores de un curso, de una materia o bien a todo el centro escolar si logran los objetivos. Posteriormente, el grupo (o escuela) decide el reparto entre los docentes.
- Premio puntual o que se convierte en una mejora salarial permanente.
- De un importe determinado o que varía en función del resultado obtenido.
- Abierto (cualquier profesor puede ganar si consigue los objetivos) o cerrado (tipo competición o *tournament*) en el que solo pueden ganar unos cuantos.
- Con evaluación externa (hecha por una autoridad educativa), interna (desarrollada por la dirección del centro con posible participación del profesorado) o mixta.

En primer lugar se revisa un grupo de estudios que evalúan programas PPR con metodologías de análisis óptimas (es decir, que permiten inferir causalidad). Se basan en diseños experimentales, donde existe selección aleatoria entre el grupo donde se ha aplicado una política que se evalúa y el grupo de control (o de comparación, similar al anterior, pero al que no se aplica la política). En segundo lugar, se considera un conjunto de estudios con métodos cuasiexperimentales, donde, como en el caso anterior, existe grupo de tratamiento y de control, pero no selección aleatoria, de manera que intentan eliminar el sesgo de selección con determinadas técnicas econométricas [3]. En cada grupo se exponen primero los resultados de las evaluaciones para los Estados Unidos y, posteriormente, para el resto de países (el [cuadro 1](#) recoge las principales características de cada programa revisado).

Tabla 1.
Evaluaciones de programas de pago por rendimiento (PPR) (N=19)

Programa	Nivel del incentivo	Índice simple	Input/Output	Valoración output comparado	Importe (\$)		Resultado	Desviación estándar
					Mín.	Máx.		
Diseños experimentales (N=7)								
Nueva York (SPBP)	Escuela	No	15% <i>input</i> 85% <i>output</i> (60% v. añadido)	Sí	1.400	3.600	No significativo	
Chicago Heights	Individual y grupo docente	Sí	(valor añadido)	Sí	0	8.000	Solamente miedo a perder es positivo	0,2-0,4
Nashville (POINT)	Individual	Sí	<i>Output</i> (valor añadido)	Sí	5.000	15.000	No significativo	
Texas (Round Rock)	Grupo docente	No	<i>Output</i> (valor añadido)	No	5.400	5.900	No significativo	
Kenya	Escuela	Sí	<i>Output</i> (puntos y valor añadido)	Sí	26	51	Positivo dudoso	0,13-0,22
India	Individual y escuela	Sí	<i>Output</i> (valor añadido)	No	2,25	450	Positivo. Mejor individual	0,17-0,27 i 0,35-0,52
México	Individual y grupo	Sí	<i>Output</i> (valor añadido)	No	0	2.000	Positivo	0,2-0,6
Diseños cuasiexperimentales (N=12)								
Carolina del Norte	Escuela	Sí	<i>Output</i> (valor añadido)	No	750	1.500	Positivo	0,09-0,13
Distrito de Columbia (IMPACT)	Individual	No	45%/85% <i>input</i> 55%/15% <i>output</i> (v. añadido)	No	5.000	25.000	Positivo	0,24-0,27
Austin (REACH)	Individual, grupo y escuela	No	<i>Input</i> y <i>output</i> (valor añadido)	No	1.000	3.000	Positivo	0,13-0,18
Chicago (TAP)	Grupo	No	50% <i>input</i> 50% <i>output</i> (valor añadido)	No	1.100	6.400	No significativo	
Denver (ProComp)	Individual y grupo	Sí	<i>Inputs</i> y <i>output</i> (valor añadido)	Sí	400	2.500	Positivo	0,09-0,18
Arkansas, Little Rock	Individual	Sí	<i>Output</i> (valor añadido)	No	350	7.600	Positivo	0,15-0,22
Israel (2002)	Escuela	No	<i>Output</i> (puntos prueba y otros)	No	1.000	2.500	Positivo	0,02-0,04
Israel (2009)	Individual	Sí	<i>Output</i> (v. añadido y otros)	No	1.750	15.000	Positivo	0,04-0,09
Reino Unido	Individual	No	<i>Input</i> y <i>output</i> (valor añadido)	No	2.000	2.000	Positivo	0,40
México (Carrera M)	Individual	No	<i>Input</i> y <i>output</i> (puntos prueba)	No	Positivo parcialmente	0,03-0,15
Portugal	Individual	No	<i>Input</i> y <i>output</i> (puntos prueba)	No	Negativo	-0,04 a -0,54
Chile (SNED)	Escuela	No	<i>Input</i> y <i>output</i> (puntos y valor añadido 65%)	Sí	370	439	Positivo	0,14-0,25

El impacto de los incentivos económicos en los resultados de los alumnos

En el primer grupo, los programas evaluados en los Estados Unidos son muy distintos, aunque todos vinculan, de manera importante o exclusivamente, el PPR a los resultados de los alumnos (normalmente como mejoras en pruebas estandarizadas externas en matemáticas y comprensión lectora). Los tres programas expuestos a continuación muestran una falta de resultados positivos. En primer lugar, el programa New York City's Schoolwide Performance Bonus Program (SPBP) contiene un incentivo para escuelas de bajo rendimiento en la ciudad de Nueva York que considera los resultados de los alumnos en pruebas estatales (25 % del total del valor del indicador) y el progreso en los resultados (60 %). Las valoraciones son de tipo comparado, en parte con escuelas del conjunto de la ciudad, pero especialmente con centros similares.

En segundo lugar, el Project on Incentives in Teaching (POINT), en Nashville (Tennessee), para cursos de primaria y secundaria, establece un premio individual para el profesorado condicionado a que sus alumnos obtengan mejoras en pruebas estatales de matemáticas superiores a las obtenidas por otros estudiantes de igual puntuación en el curso anterior. El premio varía en función de que el profesor se sitúe en el percentil 80, 90 y 95 de la distribución de resultados. En tercer lugar, el distrito de Round Rock (Texas) establece un programa de PPR que, a diferencia de los anteriores, está dirigido a escuelas de buenos resultados.

En este caso, el premio se otorga en función del valor añadido de un equipo (pluridisciplinar) de profesores en diversas asignaturas (el grupo debe situarse en el primer tercio para ser premiado). Se trata, por tanto, de un premio de grupo, aunque un profesor de un equipo ganador podría no recibirlo si obtiene un resultado individual bajo. El valor añadido se calcula como la diferencia entre el rendimiento de los estudiantes del grupo en pruebas externas y el rendimiento esperado (que se obtiene a partir de una estimación que considera tanto los resultados en años anteriores como las circunstancias personales y socioeconómicas).

En cambio, tiene un éxito parcial el programa aplicado en Chicago Heights, una localidad de renta baja de Illinois, que crea un sistema de PPR con cuatro tipos de incentivos para el profesorado (de secundaria): dos de tipo individual y dos de tipo colectivo, condicionados a que los estudiantes obtengan un buen resultado relativo (superior a la media de resultados del resto de centros) en pruebas externas. El programa quiere comparar el comportamiento diferenciado del profesorado respecto a la voluntad de ganar un premio y el miedo a perderlo. Así, para cada tipo de incentivo (individual y de grupo), un sistema permite ganar hasta 8000 \$, mientras que el otro entrega la mitad de la retribución por adelantado, pero puede perderse si los estudiantes no consiguen los objetivos (o incrementarse en caso positivo); es lo que se conoce como un esquema de “miedo a la pérdida”. Los resultados indican que el rendimiento de los estudiantes mejora solamente en el esquema de “miedo a la pérdida” y, además, se obtienen efectos más grandes cuando el método de recompensa es colectivo [4] [5].

Cuadro 1.

La importancia del incentivo (I): miedo a perder frente a ganas de ganar

Roland G. Fryer analizó, durante el curso 2010-2011, un programa de incentivos financieros al profesorado vinculados al rendimiento de los estudiantes en Chicago Heights, un distrito educativo de renta baja del estado de Illinois. Son premios (individuales y de grupo) condicionados a que los estudiantes obtengan un buen resultado relativo (por encima de la media) en pruebas test estandarizadas. La investigación considera casi ciento cincuenta profesores en diversos centros, que participan en cuatro grupos de tratamiento y uno de control. Dos grupos (uno con retribución individual y otro colectiva) deben ganar el premio. Dos grupos distintos (también de retribución individual o grupal) reciben por adelantado la mitad del importe máximo (4000 \$), pero pueden perderlo (totalmente o de manera parcial) si sus estudiantes no consiguen los objetivos establecidos (asimismo, lo pueden incrementar hasta conseguir un máximo de 8000 \$). Así pues, el estudio permite comparar el comportamiento respecto a la voluntad de ganar y el miedo a perder.

Para poder comparar el resultado se genera un grupo de diez alumnos con resultados similares antes de la intervención, donde uno es del profesor evaluado y el resto de otras escuelas. Se calcula la posición de cada alumno en su grupo a partir de la variación anual de la nota en pruebas test del estado. El valor añadido del profesor es el valor medio obtenido a partir de la posición que ocupa cada uno de sus alumnos en sus grupos. En los premios de equipo, el resultado depende del rendimiento de los estudiantes del conjunto de docentes del mismo curso y asignatura de cada centro.

Fryer concluye que solamente en los grupos sometidos al miedo a perder y, en especial, en el que otorga premios colectivos, los estudiantes mejoran los resultados.

Más información en:

Roland G. Fryer, Jr., Steven D. Levitt, John List, i Sally Sadoff (2012). "Enhancing the efficacy of teacher incentives through loss aversion: A field experiment". *NBER Working Papers 18237*. Cambridge, MA: National Bureau of Economic Research.

En conjunto, los programas evaluados presentan resultados bastante decepcionantes. Hay que decir, sin embargo, que algunos elementos pueden explicar estos malos resultados. En Nueva York y Texas inciden negativamente tanto la complejidad del indicador (que hace difícil saber a priori el esfuerzo que debe realizar el profesorado para recibir el incentivo) como que buena parte de los docentes no eran partidarios del PPR como medio de pago. En Nueva York y Nashville, todas las escuelas de bajo rendimiento tenían un incentivo para mejorar por el riesgo de sanciones y cierre que podían sufrir después de la aprobación de la ley denominada No Child Left Behind de 2002 (de manera que todos los profesores tenían un aliciente para mejorar los resultados). Finalmente, los objetivos en Nashville fueron quizá demasiado elevados para servir como factor motivador [6].

En conjunto, los programas evaluados presentan resultados bastante decepcionantes. Hay que decir, sin embargo, que algunos elementos pueden explicar estos malos resultados.



Más allá de los Estados Unidos destacan las evaluaciones experimentales en tres países en desarrollo: Kenia, la India y México. En el primer caso, un programa de incentivos a nivel de escuela, aplicado a 50 centros de primaria y secundaria, donde se premia a un determinado número de establecimientos que han obtenido los mejores resultados, o los mayores progresos en los resultados, muestra una mejora del rendimiento del

alumnado. Ahora bien, esta mejora solamente se produce en las asignaturas cubiertas por el programa de incentivos y mientras este está en marcha; por tanto, los resultados pueden interpretarse de manera que el PPR no mejora el aprendizaje, sino que, simplemente, el profesorado ayudó a los alumnos a preparar las pruebas. La segunda experiencia se desarrolla en escuelas de educación primaria en una región de la India. En este caso, el incentivo es por mejoras superiores al 5 % en el rendimiento de los estudiantes en pruebas de matemáticas y lenguaje. Los resultados muestran el éxito del programa, a corto y largo plazo, y un efecto más grande de los incentivos individuales que de los incentivos a nivel de centro [5]. Finalmente, destaca el Programa de Alineación de Incentivos, en 88 colegios de secundaria superior en México, que ofrece tres tipos de incentivos: solamente a estudiantes (según sus resultados en una prueba de matemáticas realizada a final de curso); solamente a profesores (en función del rendimiento de sus alumnos); y a alumnos y profesores (en el primer caso dependiendo de sus resultados y de la puntuación de sus compañeros de clase, y en el segundo en función del resultado de sus alumnos y del resto de alumnos de matemáticas del colegio). El premio depende de que los estudiantes mejoren de nivel (establecen tres) y existe una penalización si caen al más bajo. Los resultados muestran un efecto positivo del programa en el primer caso y especialmente en el último, cuando el estímulo va dirigido tanto a estudiantes como a profesores y depende de los resultados del conjunto de alumnos del centro [7].

Un segundo grupo de estudios lo constituyen las evaluaciones con métodos cuasiexperimentales. Se recogen investigaciones metodológicamente rigurosas, pero solamente permiten inferir causalidad aceptando los supuestos de su elaboración. Como en el caso de los estudios experimentales, en primer lugar se muestran experiencias de los Estados Unidos y, posteriormente, del resto de países.

Dos programas de larga duración se evalúan con técnicas de regresión discontinua. Por una parte, el programa ABC School-Wide Bonus Program, en Carolina del Norte, que otorga un premio a los profesores (de primaria y secundaria) si el conjunto de estudiantes de la escuela mejora los resultados de matemáticas y lengua en una cantidad esperada, y el doble si se supera esta. El resultado esperado se calcula en función de una predicción que tiene en cuenta el rendimiento anterior de los alumnos y su situación socioeconómica. La evaluación muestra un efecto positivo del programa (especialmente en matemáticas) [8]. Por otra parte, el programa IMPACT (en el distrito de Columbia) también consigue mejorar el resultado de los estudiantes de centros públicos en un

Más allá de los Estados Unidos destacan las evaluaciones experimentales en tres países en desarrollo: Kenia, la India y México. En el primer caso, un programa de incentivos a nivel de escuela, aplicado a 50 centros de primaria y secundaria, donde se premia a un determinado número de establecimientos que han obtenido los mejores resultados, o los mayores progresos en los resultados, muestra una mejora del rendimiento del alumnado.



distrito muy conflictivo. Este programa establece cinco categorías de eficiencia del profesorado, con ciertas peculiaridades: los profesores del nivel superior obtienen un premio significativo, que puede consolidarse en la nómina si lo consiguen dos años consecutivos; los profesores del nivel inferior son despedidos y los que están en el tramo inmediatamente superior al más bajo pueden serlo si no mejoran los resultados en un curso. El programa evalúa a los alumnos de profesores en torno a la categoría superior y penúltima (las más relevantes) y muestra cómo se incrementa su rendimiento en ambos casos. Hay que decir que el incentivo se mide de forma multidimensional, de modo que el peso del *output* es superior en las asignaturas sobre las que existe información de los resultados en pruebas estatales y se calcula de manera similar al programa anterior [9].

Cuadro 2.

La importancia del incentivo (II): premios y castigos

Thomas Dee y James Wyckoff analizan el programa IMPACT, aplicado en el distrito de Columbia, una zona de bajo rendimiento educativo. El programa nació en 2009 y los autores evalúan los resultados después de tres años de aplicación.

IMPACT establece cinco categorías de profesor, denominadas “ineficaz”, “mínimamente eficaz”, “en desarrollo”, “eficaz” y “muy eficaz”. Si se consiguen los dos niveles más altos, pueden obtenerse ganancias retributivas. Si el docente se mantiene dos años consecutivos en el nivel más alto, el complemento se transforma en un aumento de sueldo permanente (que varía en función de la categoría laboral y las características del centro, pero que puede llegar a suponer un incremento del 30 % del salario). El profesorado en la última categoría es despedido, mientras que el del penúltimo grupo puede serlo si no mejora en el plazo de un año. El premio se mide a partir de la puntuación otorgada al profesorado mediante el índice multidimensional que deriva de cuatro componentes. El primero se basa en la observación del profesorado en el aula (por parte de la dirección del centro y un experto externo). El segundo depende de la actuación de los alumnos en un test estatal. Dado que solamente hay datos para algunas asignaturas y cursos, para el resto se genera una valoración en función del grado de cumplimiento (según la valoración del director) de objetivos de aprendizaje fijados entre profesores y directores. El tercer elemento depende de la valoración del director de la participación del profesorado en actividades del centro. El último factor depende de la mejora en las pruebas test estatal del conjunto del centro. Finalmente, una mala valoración respecto a la profesionalidad (medida en función del cumplimiento de las normas del centro) puede restar puntos al índice. Para el profesorado con alumnado que se evalúa en test externos, el primer componente cuenta un 35 %, el segundo un 50 %, el tercero un 10 % y el cuarto un 5 %. Para el resto de profesores, los criterios son, respectivamente, 75 %, 10 %, 10 % y 5 %. La evaluación del programa considera dos áreas en torno a dos puntos relevantes: los alumnos de profesores en el grupo de “mínimamente eficaces” y de “muy eficaces”. Los resultados muestran que en ambos casos mejoran significativamente los resultados de los alumnos. Además, buena parte del profesorado en el nivel mínimamente eficaz abandona la carrera docente.

Más información en:

Dee, T., i Wyckoff, J. (2013). “Incentives, selection, and teacher performance: Evidence from IMPACT”. *NBER Working Papers 19529*. Cambridge, MA: National Bureau of Economic Research.

Tres estudios pretenden superar el sesgo de selección mediante técnicas de emparejamiento (*matching*) de centros similares (en unas características observables). El Programa REACH, en el distrito escolar de Austin, contiene diversos sistemas de incentivos (individuales y colectivos, de *input* y *output*), para escuelas de primaria y secundaria. Una evaluación muestra un incremento del valor añadido del profesor, que se mide comparando el resultado obtenido por sus estudiantes respecto al resultado posible teniendo en cuenta sus características socioeconómicas. La mejora se produce solamente el primer año, aunque el nivel más elevado conseguido el primer curso se mantiene en el segundo [10]. En cambio, la evaluación de la aplicación del programa Teaching Advancement Program (TAP) en centros de primaria de Chicago no muestra mejoras en los resultados. En este caso se remunera a los profesores (de manera colectiva) a partir de un método paritario entre la observación de la actuación del docente en el aula y el crecimiento de los resultados de los estudiantes en una prueba estatal (también se comparan los resultados reales y los esperados) [5]. Una experiencia diferente la constituye el programa ProComp de Denver, iniciado en 2006. Este programa no trata de ofrecer incentivos, sino de cambiar el sistema de retribución, de manera que los incrementos salariales (no el salario base) dependan totalmente de estímulos vinculados con indicadores de *inputs* y *outputs*. En el caso de los segundos, un factor para obtener una mayor retribución se vincula con la mejora de los resultados de los alumnos: se obtiene si como mínimo el 50 % de los estudiantes está en el percentil 55 o superior de crecimiento de los resultados de los estudiantes del estado en las áreas de matemáticas y lengua. La evaluación de este programa muestra una mejora en los resultados de los alumnos vinculados a profesores bajo el método ProComp respecto a los que no quisieron pasar a este sistema (para los contratados antes de 2006, la adscripción a este nuevo sistema era voluntaria) [11].

Finalmente, un análisis de diferencias en diferencias, que analiza un programa de PPR en Little Rock (Arkansas), muestra la mejora de los resultados de los alumnos de centros de primaria participantes (respecto a los no participantes), en especial de aquellos con peores resultados antes de aplicarse el programa. En este caso, los incentivos (individuales) dependen de los resultados de los alumnos en la prueba estatal de matemáticas y comprensión lectora; se considera la mejora de cada alumno en un curso, de manera que el premio aumenta a medida que el incremento en los resultados es más grande, y el profesorado recibe el incentivo en función de la mejora observada en el conjunto de sus alumnos [12].

Fuera de los Estados Unidos, destacan por su interés metodológico dos estudios realizados en Israel. El primero, de 2002, considera el efecto de un programa desarrollado en centros de secundaria superior con alumnado de bajo rendimiento. El premio es para el conjunto de la escuela y lo reciben solamente los profesores de los centros situados en el tercio superior de mejora de diversos indicadores vinculados a pruebas nacionales necesarias para acceder a la universidad. El análisis del programa permite constatar que los alumnos de las escuelas participantes mejoraron en resultados y en reducción del abandono de los estudios. La segunda evaluación, de 2009, considera un programa similar al anterior, pero aquí el incentivo lo reciben los profesores de manera individual. La evaluación del programa es positiva, tanto en lo que respecta a la probabilidad de que los alumnos se presenten al examen nacional al terminar el

bachillerato como en lo que atañe a las notas obtenidas. Hay que decir que en ambos casos se mide el rendimiento teniendo en cuenta la diferencia entre el resultado obtenido y el esperado en función de las características de la escuela en el primer caso y de los alumnos de los profesores en el segundo [5].

Otros estudios metodológicamente menos precisos se encuentran en Inglaterra, México, Portugal y Chile. Los programas son diferentes, pero todos incorporan un sistema de PPR individual donde el progreso de los alumnos (resultados o valor añadido en pruebas externas) es solamente uno de los elementos, a menudo menor, junto con indicadores de impulso del proceso educativo (como la actividad docente o el desarrollo profesional, vinculado normalmente a la formación). Asimismo, buscan establecer mejoras salariales permanentes (no solamente un premio por actuación). En los dos primeros casos se constatan efectos positivos del sistema de PPR sobre la mejora en los resultados de los estudiantes, si bien solo son cuantitativamente importantes en Inglaterra. En cambio, el programa de mejora salarial en Portugal empeora los resultados de los estudiantes [5] [13] [14]. Finalmente, en Chile destaca el programa de incentivos iniciado en 1997 y denominado Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos (SNED), en el que compiten colegios de cada región similares desde el punto de vista social. En la evaluación se tiene en cuenta la media del centro en un índice que incluye medidas educativas de *input* y *output*; estos últimos son tanto el resultado de los estudiantes en matemáticas y lengua como su mejora a lo largo del tiempo (equivalen al 65 % del valor total del índice). El incentivo se da a un colegio que esté en el 25 % superior del valor del indicador. El programa ha mejorado los resultados de los estudiantes de manera significativa [15].

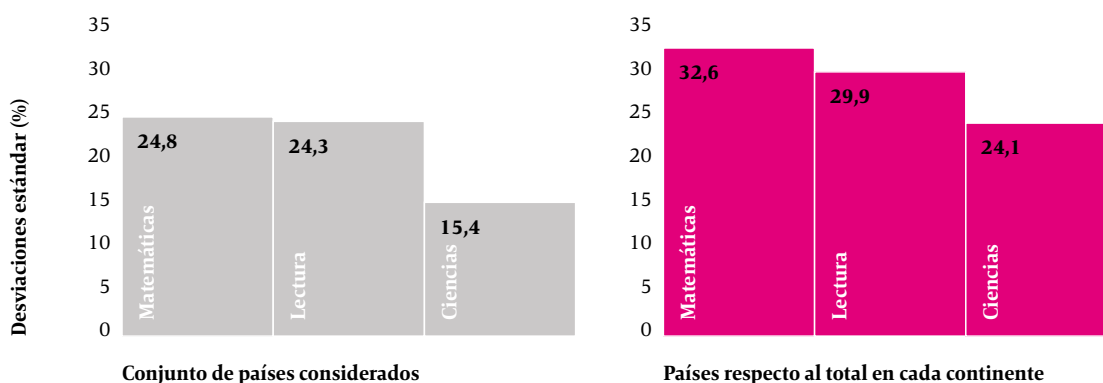
Por último, se incluye en la revisión un análisis de regresión con datos de PISA de 28 países de la OCDE, donde 13 implementan algún tipo de PPR. Aunque la metodología impide inferir causalidad, el análisis es valioso porque considera varios países conjuntamente. Los resultados muestran una relación positiva entre disponer de un sistema de PPR y el resultado de matemáticas y comprensión lectora: los resultados aumentan en torno al 25 % de una desviación estándar de la puntuación en matemáticas y comprensión lectora (15 % en la prueba de ciencias). Los efectos son superiores si se comparan los países solamente con los que están dentro del mismo continente (véase el gráfico 1)² [16].

En conclusión, las evaluaciones experimentales en los Estados Unidos muestran la práctica no efectividad de los programas de PPR. En cambio, la evidencia en países en desarrollo presenta unos mejores resultados. Asimismo, las evaluaciones cuasiexperimentales, tanto en los Estados Unidos como en otros países, tienden a mostrar resultados positivos. Por tanto, el PPR puede tener éxito en la mejora de resultados de los estudiantes, pero no está garantizado a priori.

2 Este efecto del 25 % es relativamente importante: un país en una situación intermedia mejoraría unas 10 posiciones en la clasificación de PISA; asimismo, por ejemplo, el efecto representa el 72 % del impacto de repetir un curso en primaria. Esta incidencia es superior a la detectada en la mayoría de evaluaciones hechas en los Estados Unidos, cuando los resultados son positivos, y similares al efecto medio del resto de países (véase el cuadro 1).

Gráfico 1. Pago por rendimiento y mejoras en los resultados en PISA

Mejora en los resultados de los alumnos de 15 años en la evaluación de PISA 2003 en países con sistemas de pago por rendimiento



Fuente: Elaboración propia a partir de Ludger Woessmann (2011). "Cross-country evidence on teacher performance pay". *Economics of Education Review*, vol. 30, p. 404-418.

Uso de PPR para asignar maestros a centros

La asignación de maestros de más calidad (definida en función de las cualificaciones, conocimientos, experiencia) a centros de bajo rendimiento (a menudo con un alto porcentaje de alumnos de renta baja y minorías étnicas) es un elemento de preocupación en muchos países en vista de la desigual distribución del profesorado por centros [17] [18]. La evidencia muestra que más remuneración no es garantía de atraer (o mantener) a los mejores profesores hacia los centros más necesitados, ya que este elemento no resulta prioritario para muchos docentes, en términos de movilidad y permanencia en los centros. Los docentes declaran estar más preocupados por las condiciones de trabajo, los recursos existentes, la carga docente, el liderazgo de la dirección o el lugar de residencia [19]. En este debate, se ha incorporado el análisis sobre si los sistemas de PPR pueden atraer al profesorado más eficiente (en este caso definido como maestros de alumnos con más mejoras en pruebas externas) hacia los centros que más lo necesitan. No existe prácticamente evidencia al respecto, aunque un reciente experimento aleatorio en los Estados Unidos muestra resultados solo parcialmente positivos y de efecto modesto [20]. Por tanto, no parece que los incentivos monetarios (de PPR u otros) sean un instrumento que ayude demasiado a distribuir los mejores maestros entre los centros de forma equitativa.

Actitud de los propios docentes frente a los PPR

En los Estados Unidos hay estudios sobre el grado de aceptación de los PPR entre los docentes y el conjunto de la sociedad. Los resultados muestran que son poco deseados entre los primeros, pero claramente queridos por el segundo. Entre el profesorado, la mayor oposición la generan los métodos basados en el cálculo de valor añadido de los docentes mediante los resultados de los estudiantes en pruebas estandarizadas [21]. Ahora bien, también influye en la valoración la estructura del PPR: que los docentes perciban justa la evaluación y que los objetivos a conseguir sean realistas favorece su aceptación. En cualquier caso, algunos maestros (y

sindicatos) lo rechazan por la propia filosofía de los PPR, que no la consideran adecuada en el ámbito educativo. Ahora bien, las encuestas realizadas en los Estados Unidos muestran que el porcentaje de profesorado favorable al PPR ha aumentado con el tiempo (podría situarse actualmente en torno al 40 %), siendo mayor la aceptación en ámbitos donde estos ya existen, así como entre los profesores con salarios más bajos o que pertenecen a minorías étnicas [22].

Limitaciones

La literatura existente no es suficiente para llegar a conclusiones definitivas sobre la eficacia de los programas de pago por rendimiento. Son necesarios más estudios (en especial experimentos aleatorios que muestren causalidad) que, posteriormente, permitan hacer metaanálisis. Además, la investigación debe poder constatar si los aumentos de resultados son mejoras reales de aprendizaje y no se deben a que el profesorado favorece la resolución de las pruebas (*teaching to the test*) o al uso de prácticas que alteran el resultado de las pruebas (seleccionar a los alumnos que se presentan, facilitar la copia, etc.), para poder conseguir los incentivos [23]. Asimismo, resulta necesario incluir en la evaluación los impactos a largo plazo, tanto en relación con la continuidad de los alumnos en el sistema educativo como con sus consecuencias en el mercado de trabajo [24] [25]. Finalmente, conviene analizar experiencias en más países, en especial europeos, donde los experimentos de PPR son prácticamente inexistentes, para poder extraer conclusiones de entornos más cercanos al caso de Cataluña.

Resumen

La evidencia empírica presentada no permite llegar a resultados concluyentes respecto a la eficacia de los programas de pago por rendimiento (PPR). Los estudios más recientes, basados en experimentos aleatorios, indican mayoritariamente que los PPR no mejoran los resultados de los estudiantes en los Estados Unidos, pero sí en países en desarrollo. Asimismo, evaluaciones con otras metodologías sí muestran resultados mayoritariamente positivos. Respecto a las características de los PPR, tampoco puede concluirse qué mecanismo es mejor, tanto en lo que respecta al nivel al que se ofrece el incentivo (individualmente a los profesores, a colectivos de docentes o al conjunto del centro educativo) como al importe del premio o si incluye solamente indicadores de *output*. Sí parece más eficaz un sistema de PPR con un indicador simple, y que la evaluación de resultados no sea comparada (teniendo en cuenta a otros alumnos), si bien son necesarios más estudios para establecer conclusiones definitivas al respecto.

Los elementos anteriores, por tanto, solamente permiten afirmar que el éxito de un programa PPR no está garantizado. Esto no implica que no se pueda conseguir, sino que, si se implanta, habrá que hacer un seguimiento muy preciso del programa para ir introduciendo los cambios necesarios para que sea eficaz.

Tabla 2.
Sistemas de pago por rendimiento (PPR) basados en el rendimiento de los estudiantes.

A favor	En contra
Pueden generar mejoras reales en los resultados de los estudiantes.	Los resultados académicos de los estudiantes también dependen de factores ajenos a los docentes. Además, una prueba solamente recoge una parte del aprendizaje.
Es un indicador objetivo. Ayuda a tomar decisiones, ya que los directores detectan bien la actuación de los profesores extremos (mejores y peores), pero les cuesta distinguir entre los de niveles intermedios.	No hay resultados de pruebas externas para todos los alumnos de todos los cursos ni asignaturas para aplicar métodos objetivos. Es difícil incentivar a docentes de asignaturas que no se evalúan externamente.
La docencia es una de las profesiones con menor dispersión salarial. El PPR puede suponer incrementos retributivos de los más eficaces.	Se supone que los resultados de los alumnos son por el trabajo hecho solamente por el profesor del curso, y que un docente no incide en el aprendizaje futuro de los estudiantes.
Puede influir en la selección y retención de docentes, atrayendo a profesores más eficaces y haciendo abandonar a los menos competentes.	Los estudiantes pueden esforzarse de diferente manera en función de la importancia que tiene un determinado curso en su progreso académico.
Puede mejorar la imagen de los centros públicos al dar más importancia a los resultados de los estudiantes.	Pueden generar estrategias en el profesorado para mejorar los resultados de los estudiantes sin mejorar el aprendizaje.
Son más coste-eficientes que medidas vinculadas a dotar al sistema de más profesores por alumno.	Los resultados de los estudiantes en pruebas externas pueden tener variaciones anuales importantes no asignables a los docentes.
Los incentivos de grupo pueden fomentar la colaboración entre profesores. Puede generarse una cultura de mejora continua y de participación de los docentes en la toma de decisiones.	El PPR individual puede reducir la cooperación entre profesores. Los colectivos pueden generar problemas de no implicación (<i>free-rider</i>) de algunos docentes, en especial en centros con una importante cantidad de profesores.
El método de valor añadido considera la diferencia de rendimiento de los alumnos derivada de su situación socioeconómica.	El coste de la gestión del sistema de incentivos puede resultar elevado.
Puede ser más exitoso si se contempla como un plan a largo plazo y no como una prueba piloto.	Posible no predisposición de los docentes al PPR y oposición sindical que conduzca al conflicto.

Implicaciones para la práctica

La revisión de los estudios sobre el pago por rendimiento (PPR) muestra que este tipo de incentivo puede mejorar los resultados de los estudiantes, pero no está garantizado que sea así. Por tanto, si se desea implantar un programa de PPR debería incorporarse la evaluación desde el inicio para ir mejorando el diseño y la aplicación de la política para lograr los resultados deseados.

Asimismo, los estudios revisados tampoco permiten concluir rotundamente sobre qué características pueden facilitar el éxito de la implantación de un sistema de PPR, si bien ciertos elementos sí parecen favorecer el logro de los objetivos de mejora del rendimiento de los alumnos y pueden facilitar su implantación.

Teniendo en cuenta todas estas cautelas, entendemos que la eventual profundización en el planteamiento de un sistema de PPR en nuestro país debería tener en cuenta las siguientes consideraciones.

1. En primer lugar, el indicador de resultados debería ser relativamente sencillo, de manera que el profesorado pudiera valorar el esfuerzo que debe realizar para obtener el incentivo.

2. En segundo lugar, debería plantearse la conveniencia de incorporar elementos de *input* y de *output*. Con el primer tipo de indicador, los docentes pueden constatar que se valoran ciertos elementos que dependen de ellos completamente. Con el segundo se vinculan los incentivos a los resultados de los estudiantes (este podría tener un peso creciente con el tiempo).
3. En tercer lugar, habría que reflexionar sobre el peso que deberían tener los elementos de valoración colectiva (por ejemplo, a nivel de centro), para evitar que se deteriore la necesaria colaboración entre docentes en el proceso educativo, y el que podría asignarse a los elementos individuales.
4. En cuarto lugar, convendría considerar si la concesión del incentivo se establece bien a partir de un determinado umbral de resultados (todo el que lo supere recibe el incentivo) o bien según comparación de resultados (reciben el incentivo los profesores o los centros que mejores resultados obtengan).
5. En quinto lugar habría que valorar el grado de estabilidad que se otorga al sistema (si se plantea como una prueba piloto o como un esquema permanente) y la dotación económica del incentivo en sí. Idealmente, la decisión que se tomara sobre este último punto (el importe del incentivo) debería basarse en un análisis coste-efectividad de la medida, es decir, teniendo en cuenta los impactos que se esperan conseguir y lo que son capaces de lograr otras políticas de profesorado de igual coste o inferior.
6. Finalmente, sería preciso buscar la complicidad del profesorado y de sus representantes, proporcionando información y formación sobre el sistema de PPR a los primeros y abriendo la participación en el diseño del sistema a los segundos.

Hay que señalar que la reciente normativa aprobada en Cataluña respecto a los estadios de promoción docente (ORDEN ENS/330/2014, de 6 de noviembre) prevé que, a partir del curso 2015-2016, los docentes puedan conseguir incrementos salariales en función de su implicación en la mejora de los resultados del centro, así como mediante la evaluación individual del ejercicio docente. En el primer caso, se establece un indicador basado en resultados (*outputs*) mientras que en el segundo predominan los *inputs*. Por tanto, se está abriendo la puerta a la implantación de un sistema de PPR que, además, contiene buena parte de los elementos señalados anteriormente. A pesar de que en el esquema actual los elementos vinculados con la evaluación no son requisitos imprescindibles para conseguir bonificaciones y el sistema es de aplicación únicamente para los centros públicos, no deja de ser significativo que la nueva normativa ya haya recogido un sistema de PPR.

Obviamente, no solamente con incentivos salariales se conseguirá aumentar la motivación de los docentes; hay que mejorar la formación, selección, remuneración y prestigio social de la profesión docente. Al mismo tiempo, hemos insistido en que la evidencia sobre su efectividad es mixta. Por tanto, habrá que observar de cerca cualquier apuesta que pueda hacerse en la línea de avanzar hacia un sistema de PPR en Cataluña, evaluando su diseño, implementación e impacto, reteniendo sus componentes potencialmente efectivos y descartando los más estériles.

Bibliografía

- [1] Hanushek, E. A. (2011). "The economic value of higher teacher quality". *Economics of Education Review*, vol. 30, p. 466-479.
[http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202011%20EER%2030\(3\).pdf](http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202011%20EER%2030(3).pdf)
- [2] Calero, J., Escardíbul, J.O. (2014). *Recursos escolares y resultados de la educación*. Madrid. Fundación Europea Sociedad y Educación.
<http://issuu.com/efse/docs/re-calero-escardibul?e=3027563/9412255>
- [3] Blasco, J. Casado, D. (2009). *Avaluació d'impacte*. Col·lecció Ivàlua de guies pràctiques sobre avaluació de polítiques públiques, n. 5. Barcelona: Institut Català d'Avaluació de Polítiques Públiques.
http://www.ivalua.cat/documents/1/01_03_2010_11_33_12_Guia5_Impacte_Setembre2009_revfeb2010_massavermella.pdf
- [4] Institute of Education Sciences. "What Works Clearinghouse" (WWC).
<http://ies.ed.gov/ncee/wwc/topic.aspx?sid=17>
- [5] The World Bank (Banc Mundial). *Impact Evaluations in Education*.
<http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E>
- [6] Neal, D. (2011). "The design of performance pay in education". *NBER Working Papers 16710*, Cambridge, MA: National Bureau of Economic Research. <http://www.nber.org/papers/w16710.pdf>
- [7] Behrman, J. R., Parker, S. W., Todd, P. E., Wolpin, K. I. (2011). "Aligning learning incentives of students and teachers: Results from a social experiment in Mexican high schools". *Working Paper Penn Institute of Economic Research (PIER)*, n. 13-004. Philadelphia: University of Pennsylvania.
http://economics.wustl.edu/files/economics/imce/todd_paper.pdf
- [8] Vigdor, J. L. (2008). *Teacher salary bonuses in North Carolina*. Nashville, Tn: National Center on Performance Incentives. Vanderbilt University Peabody College.
https://my.vanderbilt.edu/performanceincentives/files/2012/10/200803_Vigdor_TeacherBonusesNC.pdf
- [9] Dee, T., Wyckoff, J. (2013). "Incentives, selection, and teacher performance: Evidence from IMPACT". *NBER Working Paper 19529*. Cambridge, MA: National Bureau of Economic Research.
http://curry.virginia.edu/uploads/resourceLibrary/16_De-IMPACT.pdf
- [10] Balch, R., Springer, M. G. (2015). "Performance pay, test scores, and student learning objectives". *Economics of Education Review*, vol. 44, p. 114-125.
<https://my.vanderbilt.edu/matthewspringer/files/2015/02/Balch-and-Springer-2015.pdf>
- [11] Goldhaber, D., Walch, J. (2012). "Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative". *Economics of Education Review*, vol. 31(6), p. 1067-1083.
[http://www.cedr.us/papers/working/WP%202011-3%20Procomp%20Strategic%20Compensation%20\(9-28\).pdf](http://www.cedr.us/papers/working/WP%202011-3%20Procomp%20Strategic%20Compensation%20(9-28).pdf)
- [12] Winters, M., Greene, J., Ritter, G., Marsh, R. (2008). *The effect of performance-pay in Little Rock, Arkansas on Student achievement*. Nashville, Tn: National Center on Performance Incentives. Vanderbilt University Peabody College.
https://my.vanderbilt.edu/performanceincentives/files/2012/10/200802_WintersEtAl_PerfPayLittleRock.pdf
- [13] Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., Wilson, D. (2009). "Evaluating the impact of performance-related pay for teachers in England". *Labour Economics*, vol. 16, p. 251-261.
<http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp113.pdf>
- [14] Martins, P. S. (2009). "Individual teacher incentives, Student achievement and grade inflation". *IZA Discussion Paper Series No. 4051*. Bonn: The Institute for the Study of Labor (IZA).
<http://ftp.iza.org/dp4051.pdf>
- [15] Contreras, D., Rau, T. (2012). "Tournament incentives for teachers: Evidence from a scaled-up intervention in Chile". *Economic Development and Cultural Change*, vol. 61(1), p. 219-246.
http://eco3147.facea.puc.cl/economia_puc/images/stories/Instituto/Profesores/tournament.pdf
- [16] Woessmann, L. (2011). "Cross-country evidence on teacher performance pay". *Economics of Education Review*, vol. 30, p. 404-418. <http://repec.iza.org/dp5101.pdf>
- [17] Akiba, M., LeTendre, G. K., Scribner, J. P. (2007). "Teacher quality, opportunity gap, and national achievement in 46 countries". *Educational Researcher*, vol. 36(7), p. 369-387.
http://www.researchgate.net/profile/Gerald_Letendre/publication/43901719_Teacher_Quality_Opportunity_Gap_and_National_Achievement_in_46_Countries/links/02e7e52fce8c6aadd7000000.pdf
- [18] Berry, B., Eckert, J. (2012). *Creating teacher incentives for school excellence and equity*. Boulder, CO: National Education Policy Center. <http://files.eric.ed.gov/fulltext/ED528716.pdf>

- [19] Murnane, R. J., Steele, J. L. (2007). "What is the problem? The challenge of providing effective teachers for all children." *The Future of Children*, vol. 17(1), p. 15-43.
http://www.princeton.edu/futureofchildren/publications/docs/17_01_02.pdf
- [20] IES (2014). "Do disadvantaged students get less effective teaching? Key findings from recent Institute of Education Sciences studies". *NCEE Evaluation Brief*, vol. January. Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance.
<http://ies.ed.gov/ncee/pubs/20144010/pdf/20144010.pdf>
- [21] Glazerman, S., A. Protik, B. Teh, J. Bruch, Max, J. (2013). *Transfer incentives for high-performing teachers: Final results from a multisite experiment (NCEE 2014-4003)*. Washington, DC: National Center for Education Evaluation and Regional Assistance. <http://ies.ed.gov/ncee/pubs/20144003/pdf/20144003.pdf>
- [22] Leigh, A. (2013). "The economics and politics of teacher merit pay". *CESifo Economic Studies*, vol. 59 (1), p. 1-33. <http://andrewleigh.org/pdf/PerformancePay.pdf>
- [23] Podgursky, M., Springer, M. G. (2007). "Teacher performance pay: A review". *Journal of Policy Analysis and Management*, vol. 26(4), p. 909-950.
<http://faculty.missouri.edu/~podgurskym/wp-content/uploads/2008/rep10.pdf>
- [24] Chetty, R., Friedman, J. N., Rockoff, J. E. (2014). "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood". *American Economic Review*, vol. 104(9), p. 2633-79.
http://obs.rc.fas.harvard.edu/chetty/value_added.pdf
- [25] Lavy, V. (2015). "Teachers' pay for performance in the long-run: Effects on students' educational and labor market outcomes in adulthood," NBER Working Papers 20983, Cambridge, MA: National Bureau of Economic Research. <http://www2.warwick.ac.uk/fac/soc/economics/staff/vlavy/w20983.pdf>

Otras lecturas

- Alger, V. E. (2014). *Teacher incentive pay that works: A global survey of programs that improve student achievement*. Vancouver, Canadá: The Fraser Institute-Barbara Mitchell Centre for Improvement in Education.
<https://www.fraserinstitute.org/uploadedFiles/fraser-ca/Content/research-news/research/publications/teacher-incentive-pay-that-works.pdf>
- Lavy, V. (2007). "Using performance-based pay to improve the quality of teachers". *The Future of Children*, vol. 17(1), p. 87-109. http://www.princeton.edu/futureofchildren/publications/docs/17_01_05.pdf
- OECD (2009). *Evaluating and rewarding the quality of teachers: international practices*. París: OECD Publishing.
http://www.oxydiane.net/IMG/pdf_8709031E.pdf

Primera edición: septiembre de 2015
© Fundación Jaume Bofill, Ivàlua, 2015
fbofill@fbofill.cat, info@ivalua.cat
<http://www.ivalua.cat>
<http://www.fbofill.cat>

Autores: Miquel Àngel Alegre, J. Oriol Escardíbul
Traductora: Cristina Yugo Cano
Edición: Fundació Jaume Bofill y Ivàlua
Diseño y maquetación: Enric Jardí
ISBN: 978-84-944534-0-3
DL: B 24880-2015

Esta obra está bajo la licencia Creative Commons de Reconocimiento-No comercial-SinObraDerivada (by-nc-nd). Se permite la reproducción, distribución y comunicación pública de la obra siempre que se reconozca la autoría. No se permite el uso comercial de la obra ni la generación de obras derivadas.



Obra Social "la Caixa"