

Estudios sobre la Economía Española - 2015/13

La evaluación docente basada en el resultado como vía de mejora del sistema educativo*

Pau Balart

Universidad Carlos III y UCL Economics

Antonio Cabrales

Universidad Carlos III, UCL Economics y FEDEA

* El presente trabajo ha sido publicado como parte del volumen titulado Reflexiones sobre el sistema educativo español, coordinado por Mercedes Esteban y editado por la Fundación Ramón Areces y la Fundación Europea Sociedad y Educación y se reproduce aquí con el permiso de ambas Fundaciones. El volumen completo se puede descargar aquí: <http://www.sociedadyleducacion.org/site/wp-content/uploads/Reflexiones-sobre-elsistema-educativo-espanol.pdf>

fedea

ÍNDICE

PRESENTACIÓN	7
RESUMEN EJECUTIVO	8
DOCUMENTO DE TRABAJO	13
SUMARIO	14
INTRODUCCIÓN	15
LA IMPORTANCIA DE LA CALIDAD DOCENTE EN EL PROCESO EDUCATIVO	16
Las características observables del profesor y su relación con la calidad docente	16
Los efectos fijos y el valor añadido del profesor: el papel del profesor importa	19
¿CÓMO MEDIR LA CALIDAD DE LA DOCENCIA?	21
UN EJEMPLO: LA EVALUACIÓN DOCENTE BASADA EN RESULTADOS	23
Datos	24
Ejemplo 1: Cálculo del valor añadido de un grupo de profesores	25
Ejemplo 2: Cálculo del valor añadido del profesor	34
CRÍTICAS Y DIFICULTADES DEL MÉTODO DEL VALOR AÑADIDO DEL PROFESOR	42
CONCLUSIONES	43
REFERENCIAS	45

PRESENTACIÓN

Reflexiones sobre el sistema educativo español es el título de un proyecto que ha impulsado la Fundación Europea Sociedad y Educación a lo largo de más de dos años. Su objetivo es profundizar en las relaciones entre rendimiento educativo y crecimiento económico, desde dos perspectivas fundamentales: por una parte, la que vincula el desarrollo y la competitividad de una sociedad con medidas que inciden directamente en la mejora y competencia del capital humano a través de la acción educativa; por otra, la que analiza el diseño de políticas educativas de éxito y sus efectos, a partir de evidencias procedentes de estudios comparados.

En el trabajo “La evaluación docente basada en el resultado como vía de mejora del sistema educativo”, los profesores Pau Balart y Antonio Cabrales se detienen en el papel del profesorado como elemento clave de una correcta gestión educativa y como el mayor componente presupuestario de los centros. Por lo tanto, resulta fundamental asignar el capital humano más adecuado para realizar la labor educativa y llevar a cabo una evaluación del profesorado que permita crear los incentivos necesarios para asegurar la efectividad docente. Este documento forma parte de una obra colectiva publicada en 2015 por las fundaciones Areces y Europea Sociedad y Educación, en la que un grupo de expertos, especializados en economía de la educación, aborda aspectos clave de las políticas educativas y sus efectos en el logro escolar.

Este proyecto, dirigido por la Fundación Europea Sociedad y Educación, y que ha contado con el apoyo de la Fundación Ramón Areces, pretende ofrecer algunas bases para una fundamentada reflexión sobre nuestro sistema educativo, apoyada en los resultados de la investigación y en la evidencia empírica, y dirigida tanto al lector no especialista como a los responsables de la toma de decisiones y a los profesionales de la educación.

Reconociendo y poniendo en valor los logros alcanzados, cada uno de los trabajos refleja un compromiso hacia la mejora de la educación en España y procura, desde su propia perspectiva de análisis, responder a una serie de cuestiones básicas que marquen el rumbo de reformas duraderas y estables en nuestro sistema educativo: ¿qué tipo de educación y qué tipo de competencias predicen un crecimiento sostenible y una más equitativa distribución de oportunidades para el capital humano de nuestro país?, ¿qué tipo de reformas educativas resultan exitosas y responden mejor a los desafíos que plantea hoy la educación?, ¿qué factores de contexto deben ser tenidos en cuenta para implantar políticas educativas que redunden en el desarrollo y competitividad de la sociedad española?

FUNDACIÓN EUROPEA SOCIEDAD Y EDUCACIÓN
EUROPEAN FOUNDATION SOCIETY AND EDUCATION

RESUMEN EJECUTIVO

Este resumen destaca las ideas principales que desarrollan los economistas Pau Barta y Antonio Cabrales en el trabajo *La evaluación docente basada en el resultado como vía de mejora del sistema educativo*, que se integra en el proyecto Reflexiones sobre el sistema educativo español, promovido por la Fundación Europea Sociedad y Educación. Se ha publicado en 2015 como parte de una obra colectiva y cuenta con el patrocinio de la Fundación Ramón Areces.

LA IMPORTANCIA DEL PROFESOR EN EL PROCESO EDUCATIVO

Una correcta gestión educativa requiere entender qué elementos intervienen en el proceso educativo y en qué medida cada uno de ellos contribuye a la eficacia de dicho proceso. El tamaño de las aulas, los recursos materiales o el entorno familiar son algunos de los aspectos que han sido largamente estudiados. Pero de entre todos estos elementos, hay uno que ha sido objeto de especial atención: el papel del profesorado. Dado que la educación es una actividad intensiva en capital humano, el profesorado constituye un elemento clave y el mayor componente presupuestario de los centros educativos. Por lo tanto, dicha atención está más que justificada. Al mismo tiempo, tanto padres y madres de alumnos como los propios estudiantes otorgan una relevancia especial al papel del profesor (Hanushek, 1986). Por lo tanto, desde el punto de vista de la gestión educativa, será clave responder a la siguiente pregunta: ¿qué grado de importancia tiene el papel del profesor dentro del proceso educativo?

La docencia tiene hoy un papel más importante del que se había pensado tradicionalmente. Por lo tanto, será importante asignar el capital humano más adecuado para realizar la labor educativa y llevar a cabo una evaluación del profesorado que permita crear los incentivos necesarios para asegurar la efectividad docente. Por otra parte, los *inputs* y características observables del profesor no son indicativos del grado de eficacia de la labor educativa.

La imposibilidad de vincular la efectividad docente a las características observables del profesor hace que en estos estudios se haya empleado el valor añadido del profesor como medida de la calidad docente. Esta medida consiste en usar los residuos o parte no explicada de los resultados docentes como medida de la efectividad docente. Numerosos estudios han utilizado esta metodología para estudiar la importancia del profesorado, Aaronson, *et al.* (2003), Hanushek (1971, 1992), Hanushek *et al.* (2005), Harbison y Hanushek (1992), Rivkin *et al.* (2005) o Rockoff (2004), entre otros. Estos estudios llegan a las mismas conclusiones que ya había-

mos comentado acerca del escaso efecto de las características observables o *inputs* de los profesores. En cambio, los efectos fijos asociados a los profesores explican gran parte de la variación en el resultado de los alumnos. De esta forma, esta literatura encuentra que las diferencias en el valor añadido individual de los profesores son sustanciales. En uno de sus estudios pioneros, Hanushek (1992) mostraba cómo, en Estados Unidos, la diferencia en el aprendizaje entre una docencia de calidad y una de mala calidad llegaba a ser equivalente al conocimiento adquirido durante todo un curso académico. Hanushek y Rivkin (2010) ofrecen una revisión de esta literatura mostrando resultados más modestos, aunque todavía importantes. La diferencia entre tener un profesor situado en el percentil 25 de la distribución de la efectividad docente y tenerlo en el percentil 75 es de alrededor de 0,2 desviaciones estándar en el resultado en matemáticas de un solo año. Los autores señalan que el tamaño del efecto es comparable al de una reducción de diez alumnos en el tamaño de los grupos.

¿CÓMO EVALUAR AL PROFESORADO?

Dada la relevancia de la labor educativa en los resultados educativos de los estudiantes surge una pregunta esencial para la gestión educativa: ¿cómo detectar un buen profesor o profesora?

Tradicionalmente, se han utilizado medidas basadas en los *inputs* del profesor, como puede ser su nivel de estudios, su participación en cursos y seminarios o su experiencia para valorar la labor educativa.

Una evaluación educativa basada en los *inputs* del profesor corresponde con la que observamos de forma generalizada en el sistema educativo español (por ejemplo, haciendo depender la remuneración de los profesores de su nivel de estudios o de la acumulación de sexenios). Sin embargo, existen razones de peso para pensar que esta no es la mejor manera de evaluar la labor educativa. En primer lugar, las características observables de los profesores en las que tradicionalmente se ha basado tanto la evaluación como la selección del profesorado no están asociadas a unos mejores resultados educativos. En segundo lugar, incluso en el caso que los *inputs* educativos tuvieran poder explicativo sobre la efectividad docente, una evaluación basada en características como el nivel de estudios u otros méritos previos no proporcionará ningún incentivo a la mejora de la efectividad docente. Por lo tanto, este tipo de mecanismos no parecen los ideales a la hora de incentivar la mejora docente.

Otro método común de evaluación del profesorado, sobre todo en las instituciones de Educación Superior, son las encuestas de satisfacción docente. A diferencia del método anterior, este presenta la ventaja de basarse en *outputs* de la actividad

educativa en lugar de en los *inputs*. Sin embargo, no existe evidencia sobre la relación entre las valoraciones que los estudiantes hacen de la labor educativa de un profesor y su efectividad a la hora de transmitir conocimientos. Cabe la posibilidad de que los alumnos no estén basando su evaluación necesariamente en la capacidad educativa del profesor sino en otros aspectos presentes en la interacción profesor-alumno, como podrían ser su simpatía, amabilidad, agrado o dificultad de la asignatura, e incluso laxitud a la hora de calificar a los alumnos. Será necesario profundizar en la relación entre los resultados docentes reales y las encuestas para validar la idoneidad de estas últimas como método de evaluación docente. La literatura en este campo es todavía incipiente, sin embargo, Watchel (1998) muestra la poca confianza de muchos departamentos en este tipo de evaluación.

Este trabajo proporciona un ejemplo práctico a partir de datos de la Universidad Carlos III de Madrid, que ilustra cómo llevar a cabo una evaluación docente basada en el resultado de los alumnos. Se discuten algunas dificultades que pueden aparecer en la evaluación y se relaciona su aplicación con otros métodos, como las encuestas de satisfacción docente. El estudio muestra cómo la disponibilidad de datos facilita la elaboración de un sistema de evaluación que permite identificar a aquellos profesores con mayor efectividad docente. Al mismo tiempo, pone de manifiesto cómo los resultados de las encuestas de evaluación del profesorado no parecen explicar la efectividad docente del profesor.

CONCLUSIONES

La investigación en Economía de la Educación ofrece dos resultados referidos al profesorado, con importantes implicaciones para la gestión educativa. En primer lugar, el papel del profesor resulta clave en el rendimiento de los estudiantes. En segundo lugar, la eficacia docente no se explica por las características observables del profesor, tales como su nivel de estudios o la experiencia más allá del primer año de ejercicio.

1. A la vista de los resultados, una buena gestión educativa requerirá identificar correctamente aquellas personas que puedan llevar a cabo la actividad docente de manera más eficaz. Para evaluar la labor docente, existen métodos más avanzados como el que se presenta en este trabajo, que permiten calcular el valor añadido del profesor basándose en el rendimiento futuro de los estudiantes.

2. La falta de datos educativos impide llevar a cabo este tipo de evaluación en gran parte del sistema educativo español. Centralizar y publicar los datos procedentes de la introducción de pruebas estandarizadas no supondría un coste elevado y permitiría sentar las bases para una mejora del sistema educativo.
3. A pesar de su eficacia en la evaluación docente, el método del valor añadido no es aplicable a la hora de decidir la contratación de profesores noveles. Sin embargo, teniendo en cuenta la importancia de la experiencia durante el primer año en la mejora de la efectividad docente, parece lógico intentar establecer un sistema de entrada de nuevos profesores que facilite, por un lado, la adquisición de experiencia por parte de los nuevos profesores y que, al mismo tiempo, limite la exposición de los estudiantes a los profesores que todavía se encuentran en su primer año de ejercicio.
4. Finalmente, aplicar el método del valor añadido a la evaluación del profesorado presenta ciertas limitaciones como las críticas metodológicas y de aplicación. En este sentido, algunos autores proponen una mayor autonomía de los centros como una alternativa o complemento a la evaluación basada en el valor añadido del profesorado. Esta propuesta se basa en la evidencia documentada sobre la capacidad de los directores de los centros para identificar cuáles son los profesores con mayores y menores niveles de efectividad docente.
5. Los malos resultados educativos de España en pruebas internacionales de conocimiento, como el *International Student Assessment Programme*, advierten sobre la necesidad de mejoras urgentes del sistema educativo. La creación de un sistema de selección, evaluación y provisión de incentivos al profesorado deberá contribuir a la mejora que necesita el sistema educativo español.

DOCUMENTO DE TRABAJO

REFLEXIONES SOBRE EL SISTEMA EDUCATIVO ESPAÑOL

**LA EVALUACIÓN DOCENTE BASADA
EN EL RESULTADO COMO VÍA DE MEJORA
DEL SISTEMA EDUCATIVO**

PAU BALART

ANTONIO CABRALES

SUMARIO

La investigación en Economía de la Educación ofrece dos resultados referentes al profesorado con importantes implicaciones para la gestión educativa. En primer lugar, el papel del profesor resulta esencial para el éxito del proceso educativo. En segundo lugar, la eficacia docente no se explica por las características observables del profesor, como su nivel de estudios o la experiencia más allá del primer año de ejercicio.

A la vista de estos resultados, una buena gestión educativa requerirá identificar correctamente aquellas personas que puedan llevar a cabo la actividad docente de manera más eficaz. La imposibilidad de identificar aquellos profesores con unas mejores aptitudes docentes a partir de sus características observables (*inputs*) recomienda la sustitución de métodos de

evaluación y compensación del profesorado, basados en características como su antigüedad o asistencia a cursos, por métodos más avanzados, basados en los resultados alcanzados por sus alumnos (*output*).

Este capítulo proporciona un ejemplo práctico a partir de datos de la Universidad Carlos III de Madrid que ilustra el modo de llevar a cabo una evaluación docente basada en el resultado de los alumnos. Se discuten algunas dificultades que pueden aparecer en la evaluación y se relaciona su aplicación con otros métodos como las encuestas de satisfacción docente. El estudio permite ver cómo, en caso de disponer de los datos necesarios, se puede llevar a cabo un sistema de evaluación docente más avanzado que permita una mejora del sistema educativo español.

INTRODUCCIÓN

Una correcta gestión educativa requiere entender qué elementos intervienen en el proceso educativo y en qué medida cada uno de ellos contribuye a la eficacia de dicho proceso. El tamaño de las aulas, los recursos materiales o el entorno familiar son algunos de los aspectos que han sido largamente estudiados. Pero, de entre todos estos elementos, hay uno que ha sido objeto de especial atención: el papel del profesorado. Dado que la educación es una actividad intensiva en capital humano, el profesorado constituye su principal factor productivo y el mayor elemento presupuestario de los centros educativos. Por lo tanto, dicha atención está más que justificada. Al mismo tiempo, tanto padres y madres de alumnos como los propios estudiantes otorgan una relevancia especial al papel del profesor (Hanushek, 1986). Por lo tanto, desde un punto de vista de la gestión educativa, será clave responder a la siguiente pregunta: ¿qué grado de importancia tiene el papel del profesor dentro del proceso educativo?

La investigación en Economía de la Educación ha estudiado esta cuestión. Los avances en la disciplina han hecho cambiar notablemente la respuesta que se ha dado a esta pregunta. A pesar de que los primeros resultados en este campo parecían indicar que la actividad docente tiene un papel más bien reducido en la formación del alumno, véase entre otros Coleman *et al.* (1966), otros estudios más recientes, representados principalmente por los trabajos de Erik Hanushek y sus coautores, consideran al profesor como un elemento esencial. La notable diferencia en las conclusiones de estos estudios radica en el enfoque utilizado a la hora de evaluar la importancia del profesorado. Mientras los primeros llevaban a cabo una evaluación basada en los *inputs*, como el nivel de estudios, las acreditaciones u otras características observables del profesor, la literatura más reciente ha utilizado un

enfoque basado en el *output* o resultado del proceso educativo, es decir, en medidas del nivel de conocimientos adquiridos por los estudiantes. Este segundo enfoque entiende como efectividad docente o calidad de la docencia la capacidad de un profesor para conseguir que sus estudiantes aprendan.

Una revisión de los trabajos de Hanushek y sus coautores revela dos conclusiones claras. En primer lugar, la docencia tiene un papel más importante del que se había pensado tradicionalmente. Por lo tanto, será importante asignar el capital humano más adecuado para realizar la labor educativa y llevar a cabo una evaluación del profesorado que permita crear los incentivos necesarios para asegurar la efectividad docente. En segundo lugar, los *inputs* y características observables del profesor no son indicativos del grado de eficacia de la labor educativa. Esta conclusión complica el establecimiento de criterios estandarizados para la contratación de buenos docentes y, al mismo tiempo, desaconseja la utilización de medidas basadas en el *input* como los sexenios, el nivel de estudios o las credenciales a la hora de evaluar y compensar al profesorado.

Este capítulo responde a la pregunta sobre la importancia del profesor a partir de una revisión de la literatura. En segundo lugar, ilustra, con datos obtenidos en la Universidad Carlos III de Madrid, cómo las técnicas usadas por esta literatura permiten llevar a cabo métodos de evaluación del profesorado basados en el resultado.

LA IMPORTANCIA DE LA CALIDAD DOCENTE EN EL PROCESO

EDUCATIVO

LAS CARACTERÍSTICAS OBSERVABLES DEL PROFESOR Y SU RELACIÓN CON LA CALIDAD DOCENTE

Los primeros estudios orientados a evaluar la importancia del profesorado en el proceso educativo partían de la base de que una labor educativa eficaz debía poder explicarse a partir de características observables de los profesores o *inputs*. El principal trabajo dentro de este campo fue el Informe Coleman (Coleman et al., 1966), encargado por el gobierno de Estados Unidos. Este trabajo y los numerosos estudios que lo siguieron, se basaban en lo que se conoce como modelo de la función de producción educativa. Dicho modelo consiste en representar el rendimiento académico de los estudiantes como una función de diversos *inputs* observables. Siguiendo a Hanushek y Rivkin (2006), formalmente se puede describir de la siguiente manera:

$$R_g = f(X_g, G_g, P_g)$$

donde, omitiendo un subíndice para cada estudiante, R_g es el resultado académico de cada estudiante en el curso g (generalmente sus notas en tests estandarizados de conocimiento, aunque también puede ser el nivel de estudios alcanzado o incluso su renta una vez finalizado el proceso educativo), y X_g , G_g y P_g son vectores que contienen valores observables de *inputs* educativos relacionados con la familia y el entorno (por ejemplo, nivel socioeconómico, tipo de unidad familiar o la zona de residencia), el grupo, clase y efectos de pares (por ejemplo, tamaño del grupo o la composición por género) y los profesores (como por ejemplo, su experiencia, titulación o acreditaciones) respectivamente.

Uno de los principales problemas de esta metodología radica en la escasa posibilidad de interpretar como causales la mayoría de las relaciones que puedan existir entre los *inputs* educativos y el rendimiento académico cuando la asignación entre estos no es aleatoria, (Rothstein, 2010). Si, por ejemplo, aquellos profesores con mayor experiencia son capaces de elegir escuelas con mejores estudiantes, esto provocaría un sesgo de simultaneidad a la hora de hacer una interpretación causal de la importancia de la experiencia docente. Por esta razón, en muchos casos, el análisis será llevado a cabo dentro de una misma escuela.

Encontramos otro de los problemas en la posible omisión de variables y su consecuente sesgo en las estimaciones. Entre ellas destaca la habilidad de los estudiantes. Por esta razón, una de las extensiones más utilizadas en esta literatura consiste en utilizar calificaciones anteriores de los estudiantes para controlarlos por su habilidad. De esta forma la función de educación anterior se extendería para incluir el resultado del curso anterior, R_{g-1} .

$$R_g = f(X_g, G_g, P_g, R_{g-1})$$

Generalmente los estudios de este tipo estiman especificaciones lineales de la función anterior:

$$R_g = \mu_0 + \mu_1 R_{g-1} + \mu_2 X_g + \mu_3 G_g + \mu_4 P_g + \varepsilon$$

Donde, μ_0 , μ_1 , μ_2 , μ_3 y μ_4 son vectores de parámetros desconocidos y ε un término de error aleatorio. En la literatura se pueden encontrar distintas especificaciones respecto al parámetro μ_1 . Es habitual encontrar especificaciones donde se supone igual a 1 y se utiliza como variable dependiente la diferencia en el resultado entre un curso y el siguiente¹.

1. Este enfoque es especialmente adecuado en aquellos casos en que se dispone de exámenes anuales que hacen una evaluación acumulativa del conocimiento de los estudiantes. Véase por ejemplo

Una de las principales conclusiones a las que se llega, después de revisar los numerosos estudios basados en la metodología de la función de producción educativa, es que las características del profesorado tienen un papel muy discreto en la efectividad docente. Véase Hanushek (1997 y 2003) para un exhaustivo resumen estadístico de las conclusiones de estos estudios. Algunos de los *inputs* analizados por la literatura han sido el nivel educativo del profesor, la experiencia docente y las acreditaciones.

La investigación en este campo no ha encontrado evidencia robusta acerca de la existencia de efectos del nivel educativo del profesorado sobre la efectividad docente. Hanushek (1997) ofrece un resumen estadístico donde se aprecia, como en la mayoría de estudios, que la formación del profesor aparece como no significativa e incluso como negativamente asociada con la eficacia docente en más de la mitad de los casos donde resulta significativa. Incluso altos niveles educativos, como másteres, no garantizan un mejor resultado de la labor educativa.

Por otro lado, la experiencia adquirida ejerciendo la docencia parece tener un cierto impacto en la eficacia educativa aunque de forma no lineal (Murnane y Philip, 1981). En concreto, estudios más recientes encuentran que la experiencia tan sólo tiene un impacto sobre la calidad docente durante el primer año, véase Chingos y Peterson (2011), Hanushek *et al.* (2005) y Rivkin *et al.* (2005). Es decir, al parecer, la efectividad docente requiere de una serie de habilidades que se adquieren ejerciendo la docencia. Sin embargo su adquisición es rápida y se lleva a cabo, sobre todo, durante el primer año de ejercicio.

Por último, la literatura ofrece resultados diversos a la hora de valorar el papel de las acreditaciones sobre la efectividad docente. Kane *et al.* (2008) encuentran que las certificaciones tienen un papel discreto sobre la efectividad docente, mientras que Clotfelter *et al.* (2007), Goldhaber y Brewer (2000) y Darling-Hammond *et al.* (2001) encuentran que las certificaciones en matemáticas de los profesores mejoran el resultado de los alumnos en pruebas estandarizadas. Sin embargo Jepsen y Rivkin (2002) encuentran que estos efectos no son importantes una vez controlados los efectos no lineales de la experiencia docente comentados en el párrafo anterior.

Una mala interpretación de algunos de estos resultados, sobre todo del Informe Coleman, supuso la generalización de una visión que minimizaba la relevancia del profesor en el éxito educativo. De esta forma se centró la atención en otros aspectos como el estatus socioeconómico familiar o el tamaño del aula, dejando de lado el papel del profesor. Sin embargo, que las características observables del profesor no permitan explicar diferencias en el resultado educativo, en ningún caso

Sass *et al.* (2014).

significa que no haya diferencias notables en su efectividad docente. En primer lugar, porque generalmente las bases de datos acostumbra a ofrecer una cantidad reducida de variables observables del profesor. En segundo lugar, y de forma más importante, el proceso educativo se lleva a cabo en un aula, donde solo estudiantes y profesor son testigos del proceso formativo. Por lo tanto pueden existir variables no observables que expliquen en gran parte las diferencias en la calidad docente.

LOS EFECTOS FIJOS Y EL VALOR AÑADIDO DEL PROFESOR: EL PAPEL DEL PROFESOR IMPORTA

Evitando una mala interpretación de los resultados comentados en el apartado anterior, otro grupo de trabajos muestra cómo, a pesar de que las características observables de los profesores no tienen un impacto relevante sobre el resultado educativo de los estudiantes, la influencia del profesor es esencial.

Estos trabajos se basan en la inclusión de efectos fijos que recogen las diferencias en las calificaciones de los estudiantes según su profesor. Formalmente, este enfoque consiste en añadir a las funciones de producción educativas variables ficticias que recojan el efecto de cada profesor:

$$R_g = \mu_0 + \mu_1 R_{g-1} + \mu_2 X_g + \mu_3 G_g + \mu_4 P_g + \pi_j + \varepsilon$$

donde π_j captura el efecto fijo del profesor, por lo tanto, mide el valor añadido de cada uno de ellos². Un buen profesor será aquel que genere un mayor valor añadido en sus alumnos, es decir, un mayor valor de π_j indicará una mayor calidad docente, en términos de rendimiento o de conocimientos adquiridos.

Como detallan Hanushek y Rivkin (2006), este enfoque ofrece varias ventajas. En primer lugar, no está limitado por la necesidad de restringir el efecto del profesor a unas características concretas, y por lo tanto, queda menos sujeto a la disponibilidad de datos de dichas características. En segundo lugar, no requiere de un conocimiento previo sobre la forma funcional en la que actúan las distintas características. En tercer lugar, ofrece un marco para medir la calidad docente que permitirá la comparación y evaluación de políticas educativas.

Numerosos estudios como los de Aaronson, et al. (2003), Hanushek (1971, 1992), Hanushek et al. (2005), Harbison y Hanushek (1992), Rivkin et al. (2005) o

2. También se pueden evaluar grupos de profesores, como veremos en el apartado “Un ejemplo: la evaluación docente basada en resultados”.

Rockoff (2004), entre otros, han utilizado esta metodología para analizar la importancia del profesorado. Estos estudios llegan a las mismas conclusiones que ya habíamos comentado acerca del escaso efecto de las características observables o *inputs* de los profesores³. En cambio, los efectos fijos asociados a los profesores captan gran parte de la variación en el resultado de los alumnos. De esta forma, esta literatura encuentra que las diferencias en el valor añadido individual de los profesores son sustanciales. En uno de sus estudios pioneros, Hanushek (1992) mostraba cómo, en Estados Unidos, la diferencia en el aprendizaje entre una docencia de calidad y una de mala calidad llegaba a ser equivalente al conocimiento adquirido durante todo un curso académico. Hanushek y Rivkin (2010) ofrecen una revisión de esta literatura mostrando resultados más modestos, aunque todavía importantes. La diferencia entre tener un profesor situado en el percentil 25 de la distribución de la efectividad docente y tenerlo en el percentil 75, es de alrededor de 0,2 desviaciones estándar en el resultado en matemáticas de un solo año.

El enfoque del valor añadido del profesor requiere de la disponibilidad de datos longitudinales tanto a nivel de alumno como de profesor. Por un lado, los datos longitudinales a nivel de alumno permitirán tener un registro de los resultados previos (R_{g-1}) y controlar así la influencia de factores como la habilidad y los conocimientos previos del alumno. Por otro lado, la disponibilidad de datos longitudinales de cada profesor permitirá comprobar la persistencia de los efectos fijos o valor añadido del profesor, evitando el ruido introducido por las características de los grupos impartidos en un curso concreto.

La principal dificultad, a la hora de estimar e interpretar el valor añadido del profesor, será la posible existencia de variables omitidas como podría ser la auto-selección de estudiantes en clases o escuelas. Rothstein (2010) muestra cómo el sesgo introducido por este problema puede producir resultados erróneos. Si tanto los buenos profesores como los buenos estudiantes son atraídos hacia las mismas escuelas o grupos, se sobrestimaré el valor añadido del profesor. Por esta razón, muchos de los estudios se han basado en la estimación del valor añadido del profesor a nivel de escuela, documentando que gran parte de las diferencias se dan dentro de un mismo centro, (véase Aaronson et al. (2007), Jacob y Lefgren (2008) o Nye et al. (2004) entre otros). La posible formación no aleatoria de grupos dentro de la escuela también ha sido estudiada en distintos trabajos. Hanushek et al. (2005) y Rivkin et al. (2005) encuentran diferencias importantes en el valor añadido del profesor una

3. Una excepción la encontramos en el caso de la Educación Superior con respecto al mérito investigador de los profesores. Rodríguez y Rubio (2013) encuentran que el mérito en investigación medido en publicaciones está relacionado con una mayor efectividad docente.

vez considerada la posible formación no aleatoria de grupos dentro de una misma escuela, aunque su valor es sensiblemente inferior al de los estudios que no controlan esta posibilidad. Kane and Staiger (2008) estiman efectos algo mayores y más cercanos a los encontrados en estudios que no controlan por la influencia de los problemas de selección dentro de los centros. Hanushek y Rivkin (2010a) distinguen entre escuelas susceptibles de estar configurando los grupos de manera aleatoria de aquellas que lo hacen de manera no aleatoria, para el distrito de Texas en Estados Unidos, observando que también existen notables diferencias de valor añadido dentro de aquellas escuelas que configuran los grupos de forma aleatoria. Finalmente, Rockoff (2004) trata el problema de selección en escuelas y en formación de grupos a partir de la introducción de efectos fijos tanto a nivel profesor como estudiante, detectando importantes diferencias en la efectividad docente de los profesores.

En resumen, aunque las conclusiones sobre el tamaño del efecto del profesorado varían dependiendo de las especificaciones y del tratamiento que se dé a los distintos problemas de selección, éstos muestran dos conclusiones claras: en primer lugar, la calidad de la docencia explica notables diferencias en el éxito del proceso educativo, y en segundo lugar, tal y como se apreciaba en el enfoque sin efectos fijos, estas diferencias en el resultado no se pueden explicar a partir de las características observables de los profesores.

¿CÓMO MEDIR LA CALIDAD DE LA DOCENCIA?

Dada la relevancia de la labor educativa en los resultados educativos de los estudiantes surge una pregunta esencial para la gestión educativa: ¿cómo detectar un buen profesor o profesora?

Tradicionalmente, se han utilizado medidas basadas en los *inputs* del profesor como puede ser su nivel de estudios, su participación en cursos y seminarios o su experiencia para valorar la labor educativa. Otros métodos habituales, sobre todo en instituciones de Educación Superior, se basan en la realización de encuestas sobre la satisfacción docente. Sin embargo, la imposibilidad de vincular la efectividad docente a las características observables del profesor, explicada en el apartado anterior, aconseja la utilización de medidas de evaluación docente basadas en el valor añadido del profesor, es decir, en el *output* del proceso educativo.

Una evaluación educativa basada en los *inputs* del profesor corresponde con la que observamos de forma generalizada en el sistema educativo español (por ejemplo, haciendo depender la remuneración de los profesores de su nivel de estudios o de la acumulación de sexenios). Sin embargo, existen razones de peso para pensar

que esta no es la mejor manera de evaluar la labor educativa. En primer lugar, como hemos visto en el apartado anterior, las características observables de los profesores en las que tradicionalmente se han basado tanto la evaluación como la selección del profesorado no están asociadas a unos mejores resultados educativos. En segundo lugar, incluso en el caso que los *inputs* educativos tuvieran poder explicativo sobre la efectividad docente, una evaluación basada en características como el nivel de estudios u otros méritos previos no proporcionará ningún incentivo a la mejora de la efectividad docente. Por lo tanto, este tipo de mecanismos no parecen los ideales a la hora de incentivar la mejora docente.

Otro método común de evaluación del profesorado, sobre todo en las instituciones de Educación Superior, son las encuestas de satisfacción docente. A diferencia del método anterior, éste presenta la ventaja de basarse en *outputs* de la actividad educativa en lugar de en los *inputs*. Sin embargo, no existe evidencia sobre la relación entre las valoraciones que los estudiantes hacen de la labor educativa de un profesor y su efectividad a la hora de transmitir conocimientos. Cabe la posibilidad de que los alumnos no estén basando su evaluación necesariamente en la capacidad educativa del profesor, sino en otros aspectos presentes en la interacción profesor-alumno como podrían ser su simpatía, amabilidad, agrado o dificultad de la asignatura e, incluso, la laxitud a la hora de calificar a los alumnos. Será necesario profundizar en la relación entre los resultados docentes reales y las encuestas para validar la idoneidad de este último como método de evaluación docente. La literatura en este campo es todavía incipiente, sin embargo, Watchel (1998) muestra la poca confianza de muchos departamentos en este tipo de evaluación. A partir de datos de la Universidad Carlos III de Madrid, en Rodríguez y Rubio (2013) y en el apartado de este documento “Un ejemplo: la evaluación docente basada en resultados” también se cuestiona la validez de las encuestas de satisfacción.

El enfoque económico del valor añadido del profesor, explicado en el apartado anterior, podrá ser utilizado también para evaluar la actividad docente. Sustituyendo la evaluación educativa tradicional basada en los *inputs* por una evaluación educativa basada en el *output*, se evitaría juzgar la labor educativa mediante características para las cuales no existe ninguna evidencia que explique su intervención directa en la efectividad docente.

Como hemos comentado en el apartado anterior, llevar a cabo una evaluación de la docencia basada en el valor añadido del profesor requerirá de datos longitudinales que permitan hacer un seguimiento del resultado académico de los estudiantes y sus profesores. Desafortunadamente, en la actualidad, este tipo de datos se encuentran rara vez disponibles en el sistema educativo español. Sin embargo, tal y como se ilustra a continuación gracias a datos cedidos por la Universidad Carlos

III de Madrid, la disponibilidad de este tipo de información permitiría llevar a cabo una evaluación docente basada en el valor añadido del profesorado.

UN EJEMPLO: LA EVALUACIÓN DOCENTE BASADA EN RESULTADOS

Tal y como se desprende de los modelos del valor añadido, la labor del profesor resulta clave en el proceso educativo. Por lo tanto, una buena gestión educativa requerirá identificar aquellas personas que puedan llevar a cabo la actividad docente de manera más eficaz. La imposibilidad de identificar aquellos profesores con unas mejores aptitudes docentes a partir de sus características observables o *inputs* recomienda el uso de métodos de evaluación basados en el resultado a partir del valor añadido del profesor. Por esta razón, será interesante contemplar la posibilidad de emplear los modelos del valor añadido como herramienta para la evaluación y posterior compensación o promoción del profesorado.

La falta de datos educativos es el principal obstáculo a la hora de llevar a cabo este tipo de análisis en el sistema educativo español. Sin embargo, encontramos algunas excepciones. En este caso la posibilidad de disponer de datos longitudinales facilitados por la Universidad Carlos III de Madrid permitirá ilustrar el funcionamiento de una evaluación basada en el resultado.

El trabajo reciente de Rodríguez y Rubio (2013) utiliza los mismos datos que el presente estudio. A diferencia de la propuesta aquí presentada, su trabajo se centra en analizar si existe relación entre la calidad docente de los profesores y sus méritos investigadores medidos por la cantidad y calidad de sus investigaciones. Estos autores encuentran que los méritos investigadores están positivamente asociados a una mejor efectividad docente. Aunque para realizar su análisis también llevan a cabo un estudio del valor añadido del profesor, existen notables disparidades con las especificaciones utilizadas aquí⁴. A diferencia del trabajo de Rodríguez y Rubio (2013), el presente estudio se centrará en ilustrar el procedimiento y discutir las posibles dificultades que se desprenden de la aplicación de dichos métodos. Por esta razón, se estudiarán distintas especificaciones para medir el valor añadido y comparar sus resultados.

4. Una primera diferencia es que el presente estudio utiliza tanto un enfoque del valor añadido individual del profesor como del grupo de profesores. En segundo lugar, existe también una importante diferencia técnica en la manera de estimar el valor añadido de los profesores. Mientras que Rodríguez y Rubio (2013) estiman el valor añadido de cada profesor a partir de la constante de una regresión que incluye únicamente los estudiantes del profesor en cuestión, en este trabajo las regresiones incluyen tanto los estudiantes del profesor como de los demás profesores usando la diferencia para el cálculo del valor añadido.

DATOS

La Universidad Carlos III de Madrid se fundó en el año 1989. La comunidad universitaria está formada por más de 18.000 alumnos y dividida en tres campus diferentes situados en Colmenarejo, Getafe y Leganés.

El curso académico consta de dos cuatrimestres lectivos y las asignaturas son de tres horas semanales. La mitad de las horas se dedican a clases de teoría y se imparten en grupos magistrales. La otra hora y media restante se dedica a clases prácticas y de resolución de problemas en grupos reducidos. Para la elaboración de este estudio únicamente se dispone de los datos referentes a los profesores de los grupos reducidos. Los estudiantes de la Universidad Carlos III de Madrid disponen de cuatro convocatorias para aprobar una asignatura. En este trabajo se ha tenido en cuenta únicamente la primera convocatoria de cada uno de los estudiantes.

La organización de los grupos en la Universidad Carlos III de Madrid presenta un elemento que resulta especialmente propicio a la hora de plantear un análisis del valor añadido del profesor. El criterio para asignar los estudiantes en grupos en su primer año de grado es alfabético. Como muestran Bagues, Berniell y Cabrales (2012) con estos mismos datos, no existe correlación entre el orden de los apellidos y las características observables de los estudiantes, por lo tanto habrá una formación cuasi-aleatoria de grupos. Esto elimina uno de los principales problemas a la hora de llevar a cabo el cálculo del valor añadido del profesor que se ha comentado en la anterior revisión de la literatura: la posible selección endógena de los grupos. Al tratarse de una asignación alfabética se elimina la posibilidad de estar clasificando a los estudiantes según sus resultados evitando así posibles sesgos en las estimaciones del valor añadido.

Para la realización del presente estudio se dispone de datos individualizados que hacen un seguimiento de los estudiantes de grado en Administración de Empresas (ADE), Derecho (DER), Economía (ECO) y Finanzas y Contabilidad (FYCO), así como del doble grado de Derecho y Administración de Empresas para los cursos académicos comprendidos entre 2008 y 2012. Estos grados son impartidos en los campus de Getafe y de Colmenarejo (C), excepto en el caso de los grados de Economía y FYCO que solo se imparten en Getafe. Los datos incluyen información detallada sobre características previas del estudiante como la zona de residencia, género, grado que cursa, nota y vía de acceso; información detallada del expediente del estudiante como notas en las asignaturas (medidas en escala de cero a diez), número de convocatorias e información del grupo en que se cursa la asignatura como el profesor, el tamaño y la composición del grupo, y las evaluaciones docentes que recibe el profesor.

Uno de los primeros pasos necesarios para calcular el valor añadido del profesorado será elegir qué variable usar para medir la eficacia educativa de los profesores. Mucha de la literatura anteriormente mencionada dispone de datos sobre exámenes estandarizados que se utilizan a la hora de valorar el conocimiento adquirido. Las diferencias experimentadas por los estudiantes en este tipo de exámenes constituyen una buena medida del valor añadido, ya que tanto su planteamiento como corrección siguen un proceso homogeneizado. En el ejemplo aquí analizado no se dispone de exámenes estandarizados. A menudo pueden existir diferencias tanto en el tipo de exámenes como en los criterios de corrección de los profesores. Este hecho desaconseja el uso de la nota en cada una de las asignaturas como variable dependiente. Además, emplear la nota de la propia asignatura a la hora de instaurar un programa de evaluación docente crearía un obvio problema de incentivos perversos. Los profesores tendrían incentivos para sobrevalorar las calificaciones de sus estudiantes. El *output* sobre el cual se evalúe la labor docente deberá estar fuera del control del profesor. Por esta razón, el presente estudio tendrá en cuenta las notas de los estudiantes en asignaturas posteriores a la de los profesores evaluados a la hora de calcular el valor añadido del profesor.

Una de las principales dificultades que plantean los datos de la Universidad Carlos III de Madrid usados en este estudio es el posible solapamiento del efecto de dos o más profesores. Es decir, dado que los estudiantes se asignan a los distintos profesores por orden alfabético, los grupos compartirán un mismo profesor en varias de las asignaturas, lo cual dificulta distinguir entre el efecto de los distintos profesores. El hecho de contar con datos de varios grupos y que los mismos profesores impartan más de un grupo de cada una de las asignaturas minimiza el impacto de este posible solapamiento. Aun así, para eliminar completamente este problema llevaremos a cabo un primer ejemplo de evaluación docente basado en el grupo de profesores. De este modo, se evita confundir los resultados entre profesores a partir de evaluar al conjunto de profesores, en lugar de a un solo profesor. De manera similar, a veces interesará evaluar la calidad docente de una escuela en lugar de la de un solo profesor.

EJEMPLO 1: CÁLCULO DEL VALOR AÑADIDO DE UN GRUPO DE PROFESORES

En este primer ejemplo, se calculará la efectividad docente de los profesores de primer curso sobre los resultados en las asignaturas de segundo curso. Esto supone un total de 3.971 estudiantes y de 169 grupos reducidos. Una vez desagregado el nú-

mero total de estudiantes para cada una de las asignaturas consideradas se dispone de un total de 30.299 observaciones.

Los datos disponibles ofrecen 30 combinaciones distintas de profesores en el grado de ADE en Getafe, 14 en el grado de ADE en Colmenarejo, 31 en el grado de Derecho en Getafe, 8 de Derecho en Colmenarejo, 16 en el doble grado de ADE y Derecho en Getafe, 14 en el doble grado de ADE y Derecho en Colmenarejo, 36 en el grado de Economía y 20 en el grado de FYCO. Dado que los grupos reducidos se forman alfabéticamente, estos serán fijos. Por lo tanto, se podrá identificar cada uno de los grupos reducidos de primer curso como un grupo de profesores distinto. Dado que las combinaciones de profesores no coinciden entre un año y otro, en este caso no podremos usar los distintos años disponibles para evaluar la estabilidad del valor añadido ni controlar por cohortes.

Como se ha mencionado en la revisión de la literatura anterior, con el fin de calcular el valor añadido de cada profesor deberemos estimar un modelo de efectos fijos. Una manera habitual de estimar los efectos fijos será introducir una variable ficticia para cada uno de los grupos de profesores (excepto uno que será tomado como referencia). De esta forma en el Modelo 1 estimaremos:

$$\text{Modelo 1: } R_{ij} = \mu_0 + \mu_1 N_i + \mu_2 X_i + \mu_3 E_i + \mu_4 M_j + \mu_5 S_{ij} + \pi I_i + \varepsilon_{ij}$$

Donde R_{ij} será la nota del estudiante i en la asignatura j de segundo curso; N_i corresponderá a la nota de acceso del estudiante i ; X_i informará de otras características personales del estudiante (sexo, zona geográfica, vía de entrada, si está matriculado a tiempo completo y si ya está en posesión de un título universitario anterior); E_i será un vector de cinco variables ficticias que capturarán las distintas combinaciones de grado y campus posibles y donde el grupo de referencia será el grado en Economía en Getafe; M_j será un vector de dos variables con las características del grupo al que pertenece el estudiante i : idioma en el que se imparte la asignatura y número de estudiantes⁵; S_{ij} será un vector con una variable ficticia para cada una de las asignaturas de segundo curso y, finalmente, I_i será un vector compuesto por 161 variables ficticias, una para cada grupo de profesores exceptuando un grupo de referencia para cada combinación de grado y campus. Por lo tanto, el vector de coeficientes π capturarán las diferencias en el valor añadido de cada grupo de profesores respecto al grupo de referencia de cada grado.

5. Otra variable que puede afectar significativamente al resultado académico es el horario del curso. Aunque las estimaciones aquí presentadas no incluyen el horario de la asignatura como variable explicativa, los resultados no se ven afectados al incluirlos.

Con el fin de controlar posibles características no observables de los estudiantes se estimará el modelo mediante efectos aleatorios. Esta técnica permitirá controlar por las características no observables, siempre y cuando no estén correlacionadas con el resto de variables explicativas introducidas en el modelo. Rockoff⁶ (2004) combina efectos fijos de estudiante y profesor para controlar cualquier posible fuente de heterogeneidad entre estudiantes. Sin embargo, en nuestro caso, cada estudiante tendrá asociado un solo grupo de profesores lo que impide combinar efectos fijos para estudiante y profesores.

Estimar el valor añadido mediante el Modelo 1 requerirá tomar un grupo de profesores como referencia dentro de cada grado. De este modo, la interpretación de los coeficientes del vector π será respecto al grupo que se tome como referencia. Valores altos de π_g indicarán una mayor efectividad docente del grupo de profesores g con respecto al grupo de referencia dentro de su grado y campus. Se estimará el modelo dos veces, una tomando como referencia al grupo de profesores de mayor valor añadido y otra tomando al de menor valor añadido.

Tal y como vemos en la Tabla 1, podemos encontrar varios grupos de profesores con valores significativamente distintos a los mejores y a los peores grupos de profesores dentro de cada grado. Podemos ver que en los grados de Economía y Derecho existe un mayor solapamiento y aun así se pueden distinguir grupos de profesores de primer curso que tienen efectos significativos tanto en términos de un mayor como de un menor resultado en las asignaturas de segundo curso.

6. Una de las reglas habituales a la hora de elegir entre un modelo convencional de mínimos cuadrados y uno con efectos aleatorios es el test de Breusch and Pagan que, en este caso, recomienda el uso de efectos aleatorios.

TABLA 1. VALOR AÑADIDO DEL GRUPO DE PROFESORES.

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

ADE				ADE(C)				DER				DER(C)			
MENOR VA		MAYOR VA		MENOR VA		MAYOR VA		MENOR VA		MAYOR VA		MENOR VA		MAYOR VA	
17	0,137	1	-0,177	7	0,373	1	-0,238	3	0,441	2	-0,319	5	0,502	1	-0,157
	(0,484)		(0,320)		(0,427)		(0,385)		(0,443)		(0,357)		(0,343)		(0,338)
18	0,498	2	-0,233	8	0,589	2	max	4	0,531	5	-0,254	6	min	2	-0,120
	(0,430)		(0,306)		(0,425)	3	-0,0985		(0,425)		(0,334)				(0,341)
19	0,567	3	-0,438	9	0,645		(0,310)	6	0,701	8	-0,376			3	-0,508
	(0,426)		(0,310)		(0,467)	5	-0,296		(0,428)		(0,334)				(0,354)
20	min	4	-0,431	12	0,675		(0,333)	7	0,548	9	-0,431			4	-0,189
			(0,320)		(0,424)	6	-0,167		(0,425)		(0,351)				(0,358)
21	0,673			13	min		(0,340)	10	0,656	11	-0,570			5	-0,566
	(0,461)	6	-0,325						(0,441)		(0,348)				(0,356)
			(0,319)	14	0,712			12	0,690	12	-0,533			7	-0,0772
		7	max		(0,457)				(0,444)		(0,358)				(0,372)
		8	-0,231					14	0,522	13	-0,291			8	max
			(0,319)						(0,454)		(0,369)				
		12	-0,380					18	0,131	15	-0,243				
			(0,323)						(0,478)		(0,379)				
		14	-0,523					19	0,550	16	-0,227				
			(0,327)						(0,440)		(0,388)				
		15	-0,338					20	0,569	17	-0,0957				
			(0,342)						(0,455)		(0,365)				
		24	-0,528					21	0,274	24	-0,360				
			(0,350)						(0,449)		(0,357)				
		27	-0,00737					22	0,514	25	-0,160				
			(0,383)						(0,419)		(0,374)				
		29	-0,110					23	min	27	-0,230				
			(0,443)								(0,346)				
								24	0,722						
									(0,441)	30	-0,158				
								26	0,472		(0,415)				
									(0,486)	31	max				
								29	0,262						
									(0,434)						

[CONTINÚA EN PÁGINA SIGUIENTE]

TABLA 1. VALOR AÑADIDO DEL GRUPO DE PROFESORES.

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

DER-ADE				DER-ADE(C)				ECO				FYCO			
MENOR VA		MAYOR VA		MENOR VA		MAYOR VA		MENOR VA		MAYOR VA		MENOR VA		MAYOR VA	
5	0,441	1	-0,371	1	0,554	3	-0,381	7	0,693	3	-0,0455	6	0,664	10	-0,0484
	(0,343)		(0,286)		(0,353)		(0,304)		(0,460)		(0,331)		(0,435)		(0,472)
8	min	2	max	2	0,221	4	0,0821	13	0,684	4	-0,306	7	0,584	17	max
9	0,301	3	-0,369		(0,384)		(0,291)		(0,482)		(0,324)		(0,412)	18	-0,436
	(0,326)		(0,322)	12	0,324	6	max	18	0,472	5	-0,243	8	0,523		(0,450)
11	0,527	4	-0,349		(0,366)	8	-0,512		(0,470)		(0,330)		(0,410)	19	-0,457
	(0,366)		(0,290)	13	min		(0,371)	22	0,503	6	-0,525	11	min		(0,415)
12	0,510	6	-0,358	14	0,168	9	-0,287		(0,508)		(0,350)		(0,444)		
	(0,345)		(0,307)		(0,347)		(0,311)	24	min	10	max	13	0,516		
14	0,215					10	-0,292	25	0,156	11	-0,130		(0,428)		
	(0,322)						(0,317)		(0,511)		(0,378)				
15	0,428							26	0,287	12	-0,285				
	(0,343)								(0,468)		(0,357)				
16	0,504							27	0,253	14	-0,169				
	(0,372)								(0,483)		(0,381)				
								28	0,629	15	-0,365				
									(0,463)		(0,355)				
								30	0,481	16	-0,155				
									(0,538)		(0,349)				
								31	0,580	17	-0,193				
									(0,494)		(0,368)				
								35	0,400	19	-0,129				
									(0,500)		(0,382)				
								36	0,185	20	-0,259				
									(0,565)		(0,322)				
										23	-0,292				
											(0,384)				
										34	-0,507				
											(0,376)				

La Tabla 1 divide los grupos de profesores en cada uno de los grados y campus en dos sub-columnas. La primera sub-columna muestra aquellos grupos de profesores con valores añadidos que no son significativamente distintos del grupo de profesores con un menor valor añadido dentro de un mismo grado y campus, mientras que la sub-columna de la derecha muestra aquellos que no son significativamente distintos del grupo de profesores con un mayor valor añadido dentro de un mismo grado y campus. Por ejemplo, al fijarse en el caso del doble grado en ADE y Derecho en Colmenarejo (DER-ADE(C)), se puede ver cómo los profesores del grupo identificado con el número 13 son los que obtienen un menor valor añadido al compararlos con el resto de grupos de profesores del mismo grado y campus, mientras que los grupos de profesores identificados con los números 1, 2, 12 y 14 no obtienen un efecto significativamente distinto al del grupo 13. En cambio, el resto de grupos de profesores del doble grado en Derecho y Administración y Dirección de Empresas cursado en Colmenarejo generan resultados significativamente mayores en sus estudiantes.

Alternativamente se podrá estimar un modelo más sencillo en términos de interpretación. Este consistirá en estimar separadamente una regresión para cada uno de los grupos de profesores considerados. Llamaremos a esta especificación Modelo 2:

$$R_{ij} = \mu_0^r + \mu_1^r N_i + \mu_2^r X_i + \mu_3^r E_i + \mu_4^r M_j + \mu_5^r S_i + \mu_6^r A_{ir} + \pi^r I_i^r + \varepsilon_{ij}^r \text{ con } r=1, \dots, 169$$

Una de las ventajas de este modelo reside en que, al no incluir una variable ficticia para cada uno de los grupos, permitirá controlar por la habilidad media del grupo, concretamente incluye la variable A_{ir}^r . A diferencia del modelo anterior, ahora I_i^r será un escalar que tomará el valor 1 cuando el estudiante pertenezca al grupo r . El resto de variables explicativas serán las mismas que se habían incluido en el Modelo 1, con el único cambio de que ahora el superíndice r nos informará del grupo para el cual se está estimando el efecto fijo. Como antes, $r=1, \dots, 169$ denotará cada uno de los grupos de profesores para los cuales queremos evaluar el valor añadido. Por lo tanto, se realizarán un total de 30 regresiones para los profesores en el grado de ADE en Getafe, 14 en el grado de ADE en Colmenarejo, 31 para el grado de Derecho en Getafe, 8 para Derecho en Colmenarejo, 16 para el doble grado de ADE y Derecho en Getafe, 14 para el doble grado de ADE y Derecho en Colmenarejo, 36 para el grado de Economía y 20 para el grado de FYCO.

Nótese que mediante este procedimiento los coeficientes de las variables explicativas podrán variar en cada una de las regresiones. Idealmente, estos coeficientes deberían ser fijos como en el Modelo 1. Sin embargo, la variación de estos es mínima y, por lo tanto, este hecho no representará un problema para el modelo. En cambio, la estimación del coeficiente del valor añadido, π^r variará notablemente. Este modelo permitirá una interpretación mucho más intuitiva de la evaluación de los profesores. En este caso, valores positivos de π^r estarán asociados con valores añadidos del profesor por encima de la media dentro del mismo grado y campus.

En la Tabla 2 se presentan de nuevo aquellos grupos de profesores que habían sido identificados con unos mejores y peores resultados en el modelo anterior. Sin embargo, ahora solo se incluyen las estimaciones de aquellos coeficientes significativamente distintos de cero. Se puede observar cómo, estableciendo un criterio de identificación basado únicamente en un valor añadido significativamente mayor o menor a la media, se reducen notablemente los grupos de profesores que aparecen como destacados. Esto sucede especialmente en aquellos grados donde había una mayor cantidad de grupos de profesores con resultados que no eran significativamente distintos de los mejores y peores resultados, como ocurría en el grado en Derecho en el campus de Getafe o con el grado en Economía. Al tomar como ejemplo el grado de Derecho en Getafe se puede observar cómo de los dieciséis grupos de profesores que producían peores resultados, tan solo el grupo de profesores identificado con el número 29 está asociado con un resultado en segundo curso significativamente menor a la media⁷. En concreto, se observa cómo los estudiantes que cursaron las asignaturas de primer curso con el grupo de profesores 29, obtuvieron de media 0.726 puntos menos (con puntuaciones medidas en escala del 0 al 10) en las calificaciones de segundo curso que el resto de alumnos que cursaron Derecho en el campus de Getafe. De manera parecida, los estudiantes que cursaron la asignatura de primero de Derecho en Getafe con el grupo de profesores 5 y 24 obtuvieron un resultado en segundo curso más de cuatro décimas por encima de la media de sus compañeros.

7. Como se puede observar, el grupo de profesores con unos resultados significativamente inferiores a la media de los demás grupos no siempre coincide con aquel que tenía un menor coeficiente en la estimación del Modelo 1. Esto se puede deber tanto a diferencias en el error estándar del coeficiente del grupo de profesores en cada una de las distintas regresiones como a los cambios inducidos al introducir una nueva variable explicativa que controle la habilidad media del grupo.

TABLA 2. VALOR AÑADIDO DEL GRUPO DE PROFESORES.

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

ADE				ADE(C)				DER				DER(C)			
MENOR VA		MAYOR VA		MENOR VA		MAYOR VA		MENOR VA		MAYOR VA		MENOR VA		MAYOR VA	
17	-0,984	1	0,434	7	1			3	2			5	1		
	(0,331)		(0,234)	8	2	0,667		4	5	0,430		6	-0,889	2	
18	-0,713	2	0,395	9		(0,233)		6		(0,210)			(0,254)	3	
	(0,248)		(0,214)	12	3	0,590		7	8					4	
19	-0,754	3		13	-0,903	(0,240)		10	9					5	
	(0,240)	4			(0,321)	5	0,452	12	11					7	
20	-1,074	6		14		(0,270)		14	12						
	(0,352)	7	0,626		6	0,523		18	13						
21	-0,504		(0,239)			(0,278)		19	15						
	(0,299)	8	0,584					20	16						
			(0,229)					21	17						
		12						22	24	0,425					
		14						23		(0,245)					
		15						24	25						
		24						26	27						
		27						29	-0,726	30					
		29							(0,225)						
										31					

[CONTINÚA EN PÁGINA SIGUIENTE]

TABLA 2. VALOR AÑADIDO DEL GRUPO DE PROFESORES.

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

DER-ADE			DER-ADE(C)			ECO			FYCO				
MENOR VA	MAYOR VA		MENOR VA	MAYOR VA		MENOR VA	MAYOR VA		MENOR VA	MAYOR VA			
5	1	0,360	1	3		7	3	0,609	6	10	0,947		
8	-0,579	(0,217)	2	-0,463	4	0,633	13	-0,479	(0,201)	7	(0,344)		
9	(0,252)	2	0,722	(0,281)	(0,194)	(0,265)	4		8	17	0,931		
11		(0,217)	12	-0,583	6	0,475	18		5	0,593	11	-0,913	(0,348)
12		3	(0,260)	(0,244)	22	-0,514	(0,209)	(0,290)	18				
14	-0,666	4	-0,933	8	(0,270)	6		13	19				
	(0,209)	6	0,421	(0,266)	9	24	-0,807	10	0,610				
15		(0,238)	14	-0,663	10	(0,398)	(0,240)						
16			(0,235)			25		11					
						26	-0,405	12					
						(0,242)							
						27	-0,453	14					
						(0,270)							
						28		15					
						30		16					
						31	-0,668	17					
						(0,246)							
						35		19					
						36		20	0,411				
						(0,248)							
						23							
						34							

Como se ha mencionado anteriormente, al evaluar el valor añadido del grupo de profesores se evita el problema de tener que distinguir entre el efecto de los distintos profesores. Por ejemplo, la asignatura de Teoría Microeconómica, que se imparte en segundo en todos los grados incluidos en este estudio excepto en el de Derecho, requiere tanto de conocimientos de teoría económica de la asignatura Microeconomía como de conocimientos que se imparten en la asignatura de Matemáticas II. Al disponer tan solo de datos de cuatro cursos distintos podría no haber suficientes combinaciones entre profesores de Matemáticas II y Microeconomía, lo cual podría llevar a confundir el efecto de un profesor con el otro. El enfoque basado en el grupo de profesores reconoce este problema y evalúa directamente los grupos de profesores. Sin embargo, el enfoque basado en el grupo de profesores presenta otro problema. Al calcular el valor añadido del grupo de profesores, podríamos estar capturando diferencias que emanen de los propios estudiantes en lugar de ser atribuibles a la labor del profesor. Aunque la asignación alfabética de los alumnos pueda considerarse como aleatoria, al estar tratando con grupos relativamente pequeños, de un máximo de cuarenta estudiantes, no se puede descartar que algunos de ellos estén formados por grupos de distinta habilidad. Este problema se solventa parcialmente al incluir en el Modelo 2 un control de la habilidad media del grupo basada en las notas de selectividad. No obstante, si existen otros motivos que provoquen diferencias en el rendimiento del grupo de estudiantes como pueda ser el ambiente en clase o la cooperación entre ellos, este enfoque no estará evaluando solamente la labor del profesor sino también el rendimiento del grupo de estudiantes.

Este problema no estará presente en el siguiente ejemplo de cálculo del valor añadido individual del profesor, ya que cada uno de los profesores imparte docencia en varios grupos. Además, en este caso, será posible comprobar la estabilidad de las estimaciones en distintos cursos académicos lo cual dotará de mayor robustez las evaluaciones.

EJEMPLO 2. CÁLCULO DEL VALOR AÑADIDO DEL PROFESOR

Este segundo ejemplo ilustrará cómo se podría llevar a cabo una evaluación individual de los profesores. En este caso se evaluará a los profesores de la asignatura de “Matemáticas para la Economía I” impartida en el primer cuatrimestre en los grados de Economía y Administración de Empresas, así como del doble grado en Derecho y Administración de Empresas (impartidos en Getafe y en Colmenarejo). Esto supone un total de 4.170 estudiantes.

TABLA 3. VALOR AÑADIDO DEL PROFESOR DE MATEMÁTICAS SOBRE LAS NOTAS DE LAS ASIGNATURAS DEL SEGUNDO CUATRIMESTRE (ESTADÍSTICA I, INTRODUCCIÓN A LA CONTABILIDAD, MATEMÁTICAS II Y MICROECONOMÍA)

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

PROFESOR	EFECTOS ALEATORIOS (VA1)		MCO (VA2)		PRINCIPIOS DE ECONOMÍA (VA3)		FUNDAMENTOS ADMON. EMP. (VA4)		DERECHO CIVIL (VA5)	
	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR
1	0,262**	(0,128)	0,264***	(0,082)	0,408***	(0,105)	0,169	(0,126)	0,297**	(0,129)
2	-0,0120	(0,168)	-0,00261	(0,106)	-0,191	(0,139)	-0,0473	(0,125)	-0,0749	(0,122)
3	-0,299*	(0,157)	-0,296***	(0,104)	-0,230*	(0,135)	-0,305**	(0,119)	0,424***	(0,118)
4	-0,444**	(0,213)	-0,363***	(0,140)	-0,487***	(0,149)	-0,531***	(0,153)	0,444***	(0,151)
5	-0,0828	(0,108)	-0,102	(0,070)	-0,0420	(0,085)	0,0789	(0,129)	0,157	(0,115)
6	-0,0938	(0,164)	-0,0746	(0,104)	-0,0225	(0,126)	-0,457***	(0,153)	-0,116	(0,135)
7	0,113	(0,104)	0,105	(0,069)	0,104	(0,084)	0,227**	(0,092)	0,123	(0,077)
8	0,107	(0,190)	0,149	(0,126)	0,156	(0,149)	-0,0886	(0,202)	-0,276	(0,200)
9	0,110	(0,119)	0,0979	(0,082)	0,0526	(0,097)	0,198*	(0,106)	0,206**	(0,105)
10	-0,0331	(0,137)	-0,0175	(0,087)	-0,0526	(0,096)	(.)	(.)	(.)	(.)
11	0,00899	(0,114)	0,00676	(0,077)	-0,0548	(0,126)	0,123	(0,098)	0,0836	(0,099)
12	0,137	(0,133)	0,151*	(0,084)	0,0631	(0,098)	-0,0731	(0,186)	0,151	(0,112)
13	-0,0476	(0,143)	-0,0694	(0,091)	-0,199**	(0,097)	-0,377	(0,252)	-0,527**	(0,238)
14	-0,373***	(0,143)	-0,369***	(0,090)	-0,548***	(0,105)	-0,198	(0,129)	-0,397**	(0,162)
15	0,112	(0,132)	0,125	(0,085)	0,527***	(0,134)	-0,0674	(0,132)	-0,0674	(0,119)
16	0,133	(0,110)	0,158**	(0,073)	0,319***	(0,090)	0,263***	(0,099)	-0,0109	(0,094)
17	-0,405*	(0,237)	-0,415***	(0,149)	-0,0986	(0,172)	(.)	(.)	(.)	(.)
18	0,230*	(0,134)	0,193**	(0,085)	0,158	(0,112)	0,150	(0,130)	0,182	(0,116)
19	-0,489**	(0,191)	-0,504***	(0,123)	0,652***	(0,132)	-0,531***	(0,135)	0,654***	(0,139)
20	-0,117	(0,109)	-0,121*	(0,069)	0,368***	(0,088)	-0,178*	(0,097)	-0,146	(0,107)
21	0,0329	(0,091)	0,0253	(0,058)	0,360***	(0,076)	0,334**	(0,139)	0,305**	(0,121)
22	-0,308***	(0,109)	-0,313***	(0,072)	0,470***	(0,092)	-0,124	(0,100)	0,280***	(0,086)
23	-0,127	(0,118)	-0,152**	(0,077)	-0,168**	(0,082)	-0,0204	(0,143)	0,00941	(0,121)
24	0,351	(0,246)	0,377**	(0,155)	0,391**	(0,166)	0,387	(0,286)	0,286	(0,287)
25	0,128	(0,316)	0,119	(0,223)	-0,120	(0,231)	0,00552	(0,235)	0,00627	(0,249)
26	0,280*	(0,166)	0,328***	(0,111)	0,303**	(0,133)	0,0315	(0,160)	0,00667	(0,159)
27	0,247**	(0,106)	0,246***	(0,070)	0,200*	(0,104)	0,202	(0,127)	0,476***	(0,132)
28	-0,0137	(0,178)	-0,0212	(0,113)	0,808	(0,581)	0,0108	(0,169)	-0,0409	(0,152)
29	0,182	(0,158)	0,207*	(0,106)	0,384***	(0,135)	0,000736	(0,227)	0,592	(0,974)

(CONTINUA EN PAGINA SIGUIENTE)

TABLA 3. VALOR AÑADIDO DEL PROFESOR DE MATEMÁTICAS SOBRE LAS NOTAS DE LAS ASIGNATURAS DEL SEGUNDO CUATRIMESTRE (ESTADÍSTICA I, INTRODUCCIÓN A LA CONTABILIDAD, MATEMÁTICAS II Y MICROECONOMÍA)

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

PROFESOR	EFECTOS ALEATORIOS (VA1)		MCO (VA2)		PRINCIPIOS DE ECONOMÍA (VA3)		FUNDAMENTOS ADMON. EMP. (VA4)		DERECHO CIVIL (VA5)	
	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR
30	0,0314	(0,219)	-0,00778	(0,157)	0,424**	(0,189)	0,0898	(0,195)	0,319	(0,207)
31	0,126	(0,153)	0,136	(0,103)	-0,140	(0,141)	-0,00418	(0,152)	0,0835	(0,141)
32	-0,0656	(0,190)	-0,0192	(0,125)	-0,105	(0,138)	0,0188	(0,131)	-0,0212	(0,134)
33	-0,227	(0,158)	-0,233**	(0,102)	-0,112	(0,117)	-0,213	(0,217)	-0,278	(0,219)
34	-0,229	(0,180)	-0,272**	(0,121)	-0,130	(0,143)	0,0483	(0,135)	-0,244*	(0,135)
35	0,0996	(0,106)	0,133*	(0,069)	0,212**	(0,084)	0,200**	(0,082)	0,227***	(0,076)
36	-0,106	(0,275)	-0,145	(0,174)	(.)	(.)	0,283	(0,186)	-0,0757	(0,198)

Una vez más se basará la evaluación en una medida del *output* fuera del control de los profesores que se pretende evaluar, en este caso las notas en el segundo cuatrimestre en las asignaturas comunes para los grados considerados: Estadística I, Introducción a la Contabilidad, Matemáticas II y Microeconomía. Considerar únicamente el resultado en las asignaturas del segundo cuatrimestre permitirá mitigar los posibles efectos derivados de la labor de otros profesores ya que los estudiantes tan solo habrán recibido un cuatrimestre de clases en la universidad. Una vez combinado el número de estudiantes y sus calificaciones en las distintas asignaturas del segundo cuatrimestre quedan un total de 14.910 observaciones.

Se dispone de datos para un total de 36 profesores de Matemáticas I. Se estimará el valor añadido de cada uno de ellos a partir de la metodología del Modelo 2 comentada en el apartado anterior, con la única diferencia de que, en este caso, se añadirá un grupo de variables ficticias que controlen los posibles efectos de cohortes.

Como podemos ver en la primera columna de la Tabla 3, se pueden encontrar profesores de Matemáticas I cuyos estudiantes obtienen resultados significativamente superiores y otros con resultados significativamente inferiores a la media, dentro de su grado y campus. Esta diferencia puede ser de algo más de tres décimas por encima de la media en el caso del profesor con mejores resultados (identificado con el número 27) y de hasta casi medio punto menos en el caso del profesor con peores resultados (identificado con el número 14). La desviación estándar del valor añadido del profesor es de 0,215. Por lo tanto, una desviación estándar en el valor añadido del profesor genera un aumento medio de 0,215 puntos adicionales en la

nota de las asignaturas posteriores consideradas. La segunda columna muestra que las estimaciones son parecidas si estimamos el modelo sin efectos aleatorios individuales.

Tal y como se ha comentado anteriormente, una de las principales dificultades al evaluar individualmente a los profesores en función de su *output* consistirá en distinguir correctamente el efecto del profesor de matemáticas del de otras disciplinas relacionadas. Para comprobar que estos efectos no constituyen un problema, las estimaciones de las columnas restantes introducen efectos fijos para controlar la posible influencia de los demás profesores del primer cuatrimestre (únicamente para aquellas asignaturas que son comunes para todos los estudiantes incluidos en la muestra). De este modo en las columnas restantes se controla por la posible influencia de los profesores de “Principios de Economía”, “Fundamentos de administración de empresas” e “Introducción al Derecho Civil Patrimonial”. Se puede observar que, en algunos casos, profesores que tenían un valor añadido significativamente distinto a la media dejan de serlo después de tener en cuenta estos efectos, este es el caso para los profesores identificados con los números 18 y 26. En cambio hay varios profesores cuyo valor añadido sigue siendo significativamente distinto a la media en todas las especificaciones, como es el caso de los profesores 3, 4 y 19. En la Tabla 3 se han resaltado aquellos profesores cuyos resultados difieren significativamente de la media en la mayoría de las especificaciones. Se observa que los profesores 1, 14, 17, 22, 27 y 35 tienen valores añadidos significativamente distintos de la media en todas las especificaciones excepto en una, por lo tanto, no podríamos descartar completamente que las estimaciones estén capturando únicamente el valor añadido del profesor de matemáticas.

TABLA 4. VALOR AÑADIDO DEL PROFESOR DE MATEMÁTICAS SOBRE LAS NOTAS DE LAS ASIGNATURAS DEL SEGUNDO CUATRIMESTRE (ESTADÍSTICA I, INTRODUCCIÓN A LA CONTABILIDAD, MATEMÁTICAS II Y MICROECONOMÍA). DESGLOSADO POR AÑOS.

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

PROFESOR	2008		2009	
	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR
1	-0,032	(0,213)	-0,112	(0,251)
2	-0,172	(0,257)	-1,907	(1,377)
3	-0,281	(0,249)	0,282	(1,382)
4	(.)	(.)	(.)	(.)
5	-0,531***	(0,198)	-0,342	(0,212)
6	(.)	(.)	(.)	(.)
7	0,307*	(0,174)	-3,686**	(1,525)
8	(.)	(.)	(.)	(.)
9	0,090	(0,266)	-0,245	(0,235)
10	(.)	(.)	0,103	(0,249)
11	-0,111	(0,189)	-0,136	(0,297)
12	(.)	(.)	-0,028	(0,271)
13	(.)	(.)	0,904	(1,285)
14	(.)	(.)	0,125	(0,254)
15	0,524**	(0,252)	0,227	(0,206)
16	-0,121	(0,169)	0,620**	(0,288)
17	(.)	(.)	(.)	(.)
18	0,383**	(0,182)	0,148	(0,206)
19	(.)	(.)	(.)	(.)
20	0,596*	(0,323)	-0,229	(0,203)
21	0,615***	(0,219)	-0,146	(0,250)
22	-1,082***	(0,172)	0,005	(0,198)
23	-0,358**	(0,173)	-0,091	(0,245)
24	(.)	(.)	(.)	(.)
25	(.)	(.)	-0,619	(1,312)
26	(.)	(.)	(.)	(.)
27	0,319	(0,202)	0,380	(0,232)
28	(.)	(.)	(.)	(.)
29	0,143	(0,166)	(.)	(.)
30	0,454*	(0,243)	(.)	(.)
31	0,341	(0,229)	-0,333	(0,233)
32	(.)	(.)	-0,429	(0,322)
33	(.)	(.)	(.)	(.)
34	(.)	(.)	0,241	(0,261)
35	-0,978	(0,126)	0,291	(0,204)
36	(.)	(.)	(.)	(.)

[CONTINÚA EN PÁGINA SIGUIENTE]

TABLA 4. VALOR AÑADIDO DEL PROFESOR DE MATEMÁTICAS SOBRE LAS NOTAS DE LAS ASIGNATURAS DEL SEGUNDO CUATRIMESTRE (ESTADÍSTICA I, INTRODUCCIÓN A LA CONTABILIDAD, MATEMÁTICAS II Y MICROECONOMÍA). DESGLOSADO POR AÑOS.

(Errores estándar entre paréntesis. ***, ** y * denotan significatividad al 1%, 5% y 10% respectivamente)

2010		2011		2012	
VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR	VA	ERROR ESTÁNDAR
0,303	(0,481)	0,664*	(0,383)	(.)	(.)
(.)	(.)	(.)	(.)	0,548**	(0,271)
-0,614*	(0,364)	-1,126	(1,351)	-0,356	(0,302)
-0,656***	(0,247)	(.)	(.)	(.)	(.)
0,346	(0,312)	0,262	(0,276)	(.)	(.)
0,137	(0,332)	0,337	(0,262)	(.)	(.)
0,601*	(0,312)	-0,525**	(0,254)	-0,229	(0,229)
(.)	(.)	(.)	(.)	-0,049	(0,225)
0,717**	(0,303)	-0,320	(0,256)	(.)	(.)
-0,0926	(0,314)	0,085	(0,347)	-0,040	(0,328)
(.)	(.)	0,222	(0,260)	-0,102	(0,316)
-0,472	(0,320)	0,047	(0,282)	0,139	(0,299)
-0,257	(0,284)	-0,004	(0,335)	0,040	(0,328)
-0,300	(0,440)	-0,390	(0,274)	(.)	(.)
(.)	(.)	(.)	(.)	-0,151	(0,301)
0,458	(0,291)	0,212	(0,310)	(.)	(.)
(.)	(.)	(.)	(.)	-0,159	(0,298)
(.)	(.)	(.)	(.)	(.)	(.)
(.)	(.)	-0,706***	(0,244)	-0,094	(0,450)
(.)	(.)	-0,269	(0,281)	0,131	(0,213)
-0,162	(0,277)	0,389*	(0,232)	-0,345	(0,310)
(.)	(.)	(.)	(.)	0,111	(0,232)
(.)	(.)	(.)	(.)	(.)	(.)
0,416	(0,354)	0,331	(0,438)	(.)	(.)
0,198	(0,368)	(.)	(.)	(.)	(.)
(.)	(.)	0,463	(0,367)	-0,006	(0,332)
0,100	(0,283)	-0,138	(0,270)	(.)	(.)
(.)	(.)	(.)	(.)	0,094	(0,237)
(.)	(.)	(.)	(.)	(.)	(.)
(.)	(.)	(.)	(.)	(.)	(.)
1,325	(1,137)	(.)	(.)	(.)	(.)
0,0992	(0,285)	(.)	(.)	(.)	(.)
-0,0364	(0,232)	(.)	(.)	(.)	(.)
(.)	(.)	-0,811**	(0,320)	(.)	(.)
-0,197	(0,192)	0,686***	(0,240)	(.)	(.)
(.)	(.)	(.)	(.)	0,206	(0,310)

A la luz de los resultados de la Tabla 3, la implantación de un sistema de evaluación de la docencia sugeriría recompensar a los profesores 1, 27 y 35. Al mismo tiempo se debería prestar atención al modo de mejorar las posibles deficiencias en la labor docente de los profesores 3, 4, 14, 17, 19 y 22. Aunque el valor añadido de estos profesores es significativamente inferior a la media en la mayoría de las especificaciones, se deberá tener cierta precaución a la hora de tomar medidas ya que, como hemos visto, algunos de estos profesores podrían estar capturando el efecto de otras asignaturas del primer cuatrimestre (profesores 14, 17 y 22). Aun así, este enfoque permitiría centrar la atención y los esfuerzos de evaluación y mejora en aquellos docentes para los cuales podría ser más necesario incrementar su efectividad docente.

Las regresiones de la Tabla 3 evalúan a los profesores basándose en los resultados de todos sus alumnos durante los cursos entre 2008 y 2012. Sin embargo, sería interesante llevar a cabo un seguimiento anual de los resultados. Esto proporcionaría dos ventajas. En primer lugar, reduciría las posibles diferencias en las medidas de *output*. Por ejemplo, podría ocurrir que la dificultad del examen de una asignatura del segundo cuatrimestre variara de un año a otro⁸. Evaluar a los profesores año a año permite controlar este tipo de efectos. En segundo lugar, mediante la comparación de la evaluación de los profesores en distintos cursos académicos, se podrá verificar la estabilidad de sus resultados, y hacer un seguimiento más preciso de su labor educativa. En este aspecto, los datos disponibles presentan ciertas limitaciones. Como se observa en la Tabla 4, tan solo 3 de los 36 profesores evaluados impartieron clase de Matemáticas I ininterrumpidamente entre los cursos 2008 y 2012. Ninguno de ellos está entre los que anteriormente habíamos identificado con efectos significativamente distintos de la media al considerar todos los cursos académicos de forma agregada. Por ejemplo, se puede ver cómo el profesor identificado con el número 21, para el que se dispone de datos para todos los años, obtiene resultados significativamente por encima de la media en los años 2008 y 2011 y resultados levemente y de manera no significativa por debajo de la media durante los otros tres cursos evaluados.

Fijando la atención en aquellos profesores que obtenían resultados significativamente distintos de la media sin separar por años académicos (Tabla 3), se observa cómo en el caso de los profesores 4, 29 y 33 únicamente se dispone de datos para un solo curso académico, lo que impediría comprobar la estabilidad temporal de sus evaluaciones. A pesar de no disponer de datos para todos los años académicos se puede observar cómo, por ejemplo, el profesor identificado con el número 27

8. En los resultados expuestos en la Tabla 3 este efecto se controlaba mediante la introducción de las variables ficticias de cohorte.

tiene valor añadido por encima de la media (con la única excepción del año 2011, cuando presenta una evaluación levemente y de manera no significativa por debajo de la media). Aunque el signo positivo del valor añadido se cumple para todos los años excepto 2011, se puede apreciar cómo las desviaciones no son significativas en cada uno de los años. Una de las consecuencias de evaluar separadamente los distintos cursos académicos es que se reduce notablemente el número de observaciones, lo que aumenta los errores estándar de las estimaciones y, en consecuencia, disminuyen los profesores con evaluaciones significativamente distintas de la media. Algo similar ocurre con los profesores 14, 19 y 22 pero con resultados por debajo de la media.

Finalmente, podremos usar la correlación entre las puntuaciones recibidas por los profesores en las encuestas docentes y el valor añadido para verificar la bondad del método de las encuestas a la hora de determinar la efectividad docente de un profesor. La Tabla 5 indica la correlación entre el grado de satisfacción de los estudiantes y el valor añadido de cada profesor calculado según cada una de las especificaciones de la Tabla 3. Como podemos ver, la correlación entre ambas medidas es muy cercana a cero en todas las especificaciones. Por lo tanto el método de las encuestas no parece adecuado para llevar a cabo las evaluaciones docentes. Este resultado es consistente con los encontrados en el trabajo de Braga *et al.* (2014), donde muestran que existe correlación positiva entre las notas que asigna el profesor y su valoración en las encuestas pero no entre el resultado en las encuestas y el valor añadido del profesor, medido como su impacto en las notas en cursos posteriores.

TABLA 5. CORRELACIÓN ENTRE CALIDAD DOCENTE Y PUNTUACIÓN EN LAS ENCUESTAS DE EVALUACIÓN DOCENTE

VALOR AÑADIDO DEL PROFESOR	PUNTUACIÓN EN LA ENCUESTA DE SATISFACCIÓN
VA1	0,0853
VA2	0,0902
VA3	0,0955
VA4	0,0554
VA5	0,1148

Este es un resultado que convendría estudiar con mayor profundidad. Cabe resaltar que en la Universidad Carlos III de Madrid la participación en estas encuestas es voluntaria y se lleva a cabo fuera de horas lectivas. Por un lado, esto supone una muy baja participación, además de un problema de autoselección; tan solo algunos estudiantes muy satisfechos o muy insatisfechos con la docencia podrían estar dispuestos a responder. Por lo tanto, los resultados de este estudio muestran poca relación entre el valor añadido de los profesores y sus evaluaciones en las encuestas.

CRÍTICAS Y DIFICULTADES DEL MÉTODO DEL VALOR AÑADIDO DEL PROFESOR

Como se ha comentado anteriormente, la importancia de la calidad docente junto a su no observabilidad aconsejan evaluar al profesorado mediante medidas basadas en sus resultados. Sin embargo, existen algunas críticas metodológicas y dificultades que se describen a continuación. Estas críticas deberán ser especialmente tenidas en cuenta ante la posibilidad de implementar las medidas de valor añadido como forma de evaluación y posterior compensación o promoción del profesorado.

Una de las principales críticas metodológicas gira alrededor de la fiabilidad de los tests. Como se ha descrito en la revisión de la literatura, la mayoría de estudios del valor añadido del profesor toman como medida del *output* la puntuación de los estudiantes en pruebas estandarizadas de conocimientos. Recientemente se ha cuestionado qué miden exactamente este tipo de pruebas, véase Jacob et al. (2008), Kane and Staiger (2008), Rothstein (2010) o Segal (2008) entre otros. Sin embargo, la literatura también ha aportado evidencia sustancial sobre la relación entre la puntuación en pruebas estandarizadas con la renta futura de los estudiantes y el nivel de educación alcanzado, véase Murnane et al. (1995). Por otro lado, la no observabilidad directa del conocimiento alerta sobre la posible presencia de errores de medición en las pruebas estandarizadas. En este sentido, la mayoría de estudios sobre valor añadido usan estimadores del valor añadido que corrigen estos posibles errores de medición. Aunque tener en cuenta esta posibilidad reduce las diferencias en el valor añadido de los profesores, éstas siguen siendo importantes, véase Hanushek y Rivkin (2010). Por otro lado, como muestran McCaffrey et al. (2009), el seguimiento de los profesores durante varios cursos permite minimizar los errores de medición.

Otra crítica habitual al método del valor añadido es el posible sesgo causado por la omisión de variables. Como se ha explicado en el apartado “La importancia de la calidad docente en el proceso educativo”, la oportunidad de elegir escuela o la posible formación de grupos basados en el resultado académico puede suponer un problema a la hora de llevar a cabo una evaluación basada en el valor añadido. Por esta razón, la metodología del valor añadido parece especialmente adecuada para evaluar al profesorado a nivel de escuela. Aun así, existe la posibilidad que la formación de grupos no sea aleatoria dentro de un mismo centro. Como se ha ilustrado en los ejemplos obtenidos a partir de los datos de la Universidad Carlos III de Madrid, la formación aleatoria de grupos permite la evaluación de la labor docente mediante el valor añadido. A la hora de llevar a cabo estudios similares a nivel de primaria o secundaria será necesario saber qué centros llevan a cabo una formación aleatoria

de sus grupos. Santín y Sicília analizan centros con asignación aleatoria de alumnos en el capítulo que coautorizan en este mismo proyecto de investigación.

El principal requerimiento a la hora de llevar a cabo una evaluación del profesorado basada en el enfoque del valor añadido será la existencia de datos longitudinales que hagan un seguimiento de estudiantes y profesores. En España no resulta habitual disponer de este tipo de datos en los niveles de educación primaria y secundaria. Facilitar y promover la disponibilidad de este tipo de datos será una medida clave y de escaso coste económico a la hora de mejorar la gestión educativa. En este sentido, la introducción de pruebas estandarizadas en la última reforma educativa debería ser vista como una oportunidad a partir de la cual empezar a publicar este tipo de datos que permitan facilitar la evaluación de resultados dentro del sistema educativo español. Un paso más allá, aunque ciertamente tendría un mayor coste, consistiría en extender las bases de datos con un seguimiento de la inserción laboral e indicadores de calidad de vida futuros de los estudiantes.

Finalmente, cabe destacar algunas de las limitaciones de estas medidas de evaluación. Obviamente, la efectividad docente sólo puede ser evaluada una vez se ha ejercido la labor educativa. Por lo tanto, medidas como el valor añadido del profesor no serán de utilidad a la hora de establecer qué criterios deben ser tenidos en cuenta para la contratación de nuevos profesores. Cabe recordar también que este tipo de medidas se basan habitualmente en la evaluación relativa de resultados. Por lo tanto, no permitirían diferenciar aquellas situaciones donde todos los profesores evaluados sean igualmente buenos de aquellas en las que todos sean igualmente malos. Por lo tanto, una evaluación del profesorado podría penalizar a los profesores de las escuelas con un personal docente por encima de la media.

CONCLUSIONES

La investigación en Economía de la Educación ofrece dos resultados referentes al profesorado con importantes implicaciones para la gestión educativa. En primer lugar, el papel del profesor resulta clave en el rendimiento de los estudiantes. En segundo lugar, la eficacia docente no se explica por las características observables del profesor como su nivel de estudios o la experiencia más allá del primer año de ejercicio.

A la vista de estos resultados, una buena gestión educativa requerirá identificar correctamente aquellas personas que puedan llevar a cabo la actividad docente de manera más eficaz. La imposibilidad de identificar aquellos profesores con unas mejores aptitudes docentes a partir de sus características observables (inputs) recomienda el uso de métodos de evaluación basados en sus resultados (outputs).

Como se ha explicado e ilustrado en el presente trabajo, uno de estos métodos consiste en el cálculo del valor añadido del profesor. Este método evalúa la labor docente basada en el rendimiento futuro de los estudiantes. Por lo tanto, convendría sustituir los viejos métodos de evaluación largamente usados en el sistema educativo español, como los sexenios o la asistencia a cursos, por métodos más avanzados, basados en los resultados como los que se han ilustrado en el presente trabajo.

La falta de datos educativos impide llevar a cabo este tipo de evaluación en gran parte del sistema educativo español. La introducción de pruebas estandarizadas de la última reforma educativa debería ser vista como una oportunidad para mejorar la información y la disponibilidad de bases de datos. Esto permitiría conseguir una mayor rendición de cuentas y evaluación en la tarea educativa. Cabe destacar que centralizar y publicar este tipo de datos no supondría un coste elevado y permitiría sentar las bases para una mejora del sistema educativo.

A pesar de su eficacia en la evaluación docente, el método del valor añadido no es aplicable a la hora de decidir la contratación de profesores noveles. Esta limitación unida al escaso poder explicativo de las características observables del profesor, dificulta el establecimiento de unos criterios para la incorporación de nuevo profesorado. Sin embargo, teniendo en cuenta la importancia de la experiencia durante el primer año en la mejora de la efectividad docente, parece lógico intentar establecer un sistema de entrada de nuevos profesores que facilite por un lado la adquisición de experiencia por parte de los nuevos profesores y que al mismo tiempo limite la exposición de los estudiantes a los profesores que todavía se encuentran en su primer año de ejercicio.

Finalmente, aplicar el método del valor añadido a la evaluación del profesorado presenta ciertas limitaciones como las críticas metodológicas y de aplicación. En este sentido, Hanushek y Rivkin (2010b) proponen una mayor autonomía de los centros como una alternativa o complemento a la evaluación basada en el valor añadido del profesorado. Esta propuesta se basa en la evidencia documentada por Jacob and Lefgren (2008) sobre la capacidad de los directores de los centros para identificar cuáles son los profesores con mayores y menores niveles de efectividad docente.

Los malos resultados educativos de España en pruebas internacionales de conocimiento, como el International Student Assessment Programme, advierten sobre la necesidad de mejoras urgentes del sistema educativo. La investigación en Economía de la Educación aporta algunas claves para acometer este tipo de mejoras, una de las más claras es el papel del profesor como elemento clave en el proceso educativo. La creación de un sistema de selección, evaluación y provisión de incentivos al profesorado deberá contribuir a la mejora que necesita el sistema educativo español.

REFERENCIAS

- Aaronson, D., Barrow, L. y Sander, W. (2007). "Teachers and student achievement in the Chicago public high schools". *Journal of Labor Economics*, 25(1), 95-135.
- Bagues, M., Berniell, I. y Cabrales, A. (2012) "The production function of College Education: Evidence from a Natural experiment". *Working Paper*.
- Braga, Michela, Marco Paccagnella y Michele Pellizzari. "Evaluating students' evaluations of professors". *Economics of Education Review* 41 (2014): 71-88.
- Chingos, M. M. y Peterson, P. E. (2011). "It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness". *Economics of Education Review*, 30(3), 449-465.
- Clotfelter, Charles T., Ladd, Helen F. y Vigdor, Jacob L. (2007). "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects". *Economics of Education Review*, Elsevier, vol. 26(6), pages 673-682, December.
- Cohen, D.K. y Murnane, R.J. (1985). "The merits of merit pay". *Public Interest* 80 (Summer), 3-30.
- Cohen, D.K. y Murnane, R.J. (1986). "Merit pay and the evaluation problem: Understanding why most merit pay plans fail and a few survive". *Harvard Educational Review* 56 (1), 1-17.
- Cohn, E. (1996). "Methods of teacher remuneration: Merit pay and career ladders". In: Becker, W.E., Baumol, W.J. (Eds.), *Assessing Educational Practices: The Contribution of Economics*. MIT Press, Cambridge, MA, pp. 209-238.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D. y York, R.L. (1966). *Equality of Educational Opportunity*. U.S. Government Printing Office, Washington, DC.
- Darling-Hammond, L., Berry, B. y Thoreson, A. (2001). "Does teacher certification matter? Evaluating the evidence". *Educational evaluation and policy analysis*, 23(1), 57-77.
- Goldhaber, D. D. y Brewer, D. J. (2000). "Does teacher certification matter? High school teacher certification status and student achievement". *Educational evaluation and policy analysis*, 22(2), 129-145.
- Hanushek, E. (1971). "Teacher characteristics and gains in student achievement: Estimation using micro data". *The American Economic Review*, 280-288.
- Hanushek, E. A. (1992). "The trade-off between child quantity and quality". *Journal of political economy*, 84-117.
- Hanushek, E. A. (1997). "Assessing the effects of school resources on student performance: An update". *Educational evaluation and policy analysis*, 19(2), 141-164.

- Hanushek, E. A. (2003). "The Failure of Input Based Schooling Policies". *The economic journal*, 113(485), F64-F98.
- Hanushek, E.A., Kain, J.F., O'Brien, D.M., Rivkin, S.G. (2005). "The market for teacher quality". NBER Working Paper 11154 (February).
- Hanushek, E. A. y Rivkin, S. G. (2006). "Teacher quality". *Handbook of the Economics of Education*, 2, 1051-1078.
- Hanushek, E. A. y Rivkin, S. G. (2010a). "Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools?" NBER Working Paper No. 15816.
- Hanushek, E. A. y Rivkin, S. G. (2010b). "Generalizations about using value-added measures of teacher quality". *The American Economic Review*, 267-271.
- Jacob, Brian A. y Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education". *Journal of Labor Economics*, 26(1): 101-36.
- Jepsen, C. y Rivkin, S.G. (2002). "What is the trade-off between smaller classes and teacher quality?" National Bureau of Economic Research.
- Kane, T. J., Rockoff, J. E. y Staiger, D. O. (2008). "What does certification tell us about teacher effectiveness? Evidence from New York City". *Economics of Education Review*, 27(6), 615-631.
- Kane, T. J. y Staiger, D. O. (2008). "Estimating teacher impacts on student achievement: An experimental evaluation" (No. w14607). National Bureau of Economic Research.
- McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood y Kata Mihaly. 2009. "The Inter-temporal Variability of Teacher Effect Estimates". *Education Finance and Policy*, 4(4): 572-606.
- Murnane, R.J. y Phillips, B.R. (1981). "Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance". *Economics of Education Review* 1 (4), 453-465.
- Murnane, Richard J., John B. Willett y Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination". *Review of Economics and Statistics*, 77(2): 251-66.
- Nye, B., Konstantopoulos, S. y Hedges, L. V. (2004). "How large are teacher effects?" *Educational evaluation and policy analysis*, 26(3), 237-257.
- Rivkin, S.G., Hanushek, E.A. y Kain, J.F. (2005). "Teachers, schools, and academic achievement". *Econometrica* 73 (2), 417-458.
- Rockoff, J.E. (2004). "The impact of individual teachers on student achievement: Evidence from panel data". *American Economic Review* 94 (2), 247-252.
- Rodríguez, R., Rubio, G. (2013). "Teaching Quality and Academic Research", Working Paper UAM.

- Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement". *Quarterly Journal of Economics*, 125(1): 175-214.
- Santín, D., Sicilia, G. "El impacto de la educación infantil en los resultados de primaria: evidencia para España a partir de un experimento natural". En *Reflexiones sobre el sistema educativo español*. Fundación Europea Sociedad y Educación. 2015.
- Sass, Tim R., Anastasia Semykina y Douglas N. Harris. "Value-added models and the measurement of teacher productivity". *Economics of Education Review*, 38 (2014): 9-23.
- Wachtel, H. K. (1998). "Student evaluation of college teaching effectiveness: A brief review". *Assessment & Evaluation in Higher Education*, 23(2), 191-212.

DIRECCIÓN DEL PROYECTO



José Abascal 57, 5º B
28003 Madrid
T 34 91 455 15 76
F 34 91 399 40 12

Rue de Trèves, 49 bte 3
B 1040 Bruxelles
T 32 2 280 6340
F 32 2 280 6338

Fundación Europea Sociedad y Educación
European Foundation Society and Education

www.sociedadyeducacion.org

PATROCINA

**FUNDACIÓN
RAMÓN ARECES**