

## **Determinantes de los resultados de las evaluaciones de profesores: el caso de la Universidad de los Andes\***

### ***Determinants of teacher evaluation: The case of Universidad de los Andes***

Alejandro Gaviria \*\*

Alejandro Hoyos \*\*\*

#### **Resumen**

Este trabajo estudia los determinantes de las evaluaciones realizadas por los estudiantes a los profesores y programas de la Universidad de los Andes. Se analiza cómo las características del curso, del profesor y el nivel de exigencia afectan el resultado de las evaluaciones. Los resultados muestran que las notas obtenidas por los estudiantes afectan de manera directa y positiva las evaluaciones de profesores. Así mismo, el tamaño de la clase afecta negativamente el resultado de las evaluaciones, y finalmente la experiencia de los maestros y su forma de vinculación laboral (planta o cátedra) no tienen un efecto significativo.

---

\* Los autores agradecen los valiosos comentarios de Hernando E. Mutis y Adriana Camacho.

\*\* Decano y profesor de la Facultad de Economía de la Universidad de los Andes e investigador del CEDE de la misma Universidad. Correo electrónico: [agaviria@uniandes.edu.co](mailto:agaviria@uniandes.edu.co).

\*\*\*Estudiante de Maestría en Economía de la Universidad de los Andes. Correo electrónico: [alhosu@gmail.com](mailto:alhosu@gmail.com), [ale-hoyo@uniandes.edu.co](mailto:ale-hoyo@uniandes.edu.co).

Este artículo fue recibido el 1° de mayo de 2008, modificado el siguiente 30 de mayo y aceptado el 7 de junio del mismo año.

*Palabras clave:* evaluaciones a profesores, incentivos en la educación, sistemas de pago por méritos.

*Clasificación JEL:* I23, J24, J33.

## **Abstract**

This paper examines the determinants of teacher and course evaluations at the Universidad de los Andes in Bogotá, Colombia. The paper explores how teacher and course characteristics affect evaluation results. Grades obtained by students directly and positively affect teacher evaluations. Class size negatively affects the results of the evaluations. Finally whether teachers are full or part-time as well as years of experience of teachers have a negligible impact on evaluations.

*Key words:* teacher evaluation, incentives on education, merit pay systems.

*JEL Classification:* I23, J24, J33.

## **Introducción**

El proceso educativo (la pedagogía, las técnicas de enseñanzas, las actitudes de los maestros, etc.) ha sido debatido intensamente por diversas disciplinas. El debate gravita alrededor de cuáles son los factores que contribuyen al aprendizaje, y cuáles las políticas que pueden incidir decididamente en estos factores. El debate es intenso e inconcluso, pues no existe un consenso, un acuerdo compartido sobre las políticas más eficaces en el campo educativo.

Hanushek, Kain y Rivkin (1998) muestran que la calidad del profesor (entendida en sentido amplio) es el factor que mejor explica las diferencias en los logros académicos de los estudiantes. Este hallazgo sugiere que la motivación, el esfuerzo, el grado de aprendizaje y (en últimas) la calidad de la educación impartida dependen principalmente de las características de los profesores. Por tanto, el diseño de sistemas de monitoreo y evaluación que permitan identificar y finalmente premiar a los buenos maestros puede considerarse

un instrumento esencial de cualquier política de mejoramiento de la calidad educativa. En suma, la existencia de un instrumento que permita clasificar a los profesores según su rendimiento contribuye al mejoramiento de la calidad de la educación.

Para poder diseñar un buen sistema de evaluaciones, se debe partir de una pregunta concreta: ¿quién es un buen profesor? Según Nerenz y Knop (1982), no existe un consenso sobre la respuesta a esta pregunta. Posiblemente por esta falta de consenso se han creado diferentes formas de evaluar a los profesores. O'Hanlon y Mortensen (1980) mencionan varias de ellas: la evaluación estandarizada de parte de los estudiantes, la evaluación de parte de los coordinadores y administradores, las opiniones de los colegas, la autoevaluación y los datos sobre el aprendizaje que utilizan pruebas estandarizadas de los estudiantes (Wright, Horn y Sanders, 1997; Rueda, 2006).

Este trabajo analiza el sistema de evaluación de profesores de la Universidad de los Andes. Este sistema está basado en evaluaciones estandarizadas realizadas por la totalidad los estudiantes de todos los cursos hacia el final de cada período académico. La evaluación es anónima y voluntaria, y está compuesta de veintisiete preguntas divididas en cuatro módulos: (i) el profesor, (ii) el curso, (iii) la metodología de evaluación y (iv) una autoevaluación del estudiante.

Se intenta establecer los determinantes de los resultados de las evaluaciones de profesores y programas. En particular se quiere establecer el efecto sobre resultados de las evaluaciones de tres grupos de variables: (i) las características del curso (tamaño de la clase, porcentaje de estudiantes que responden a la evaluación, nivel del curso dentro del pónsum, Facultad), (ii) del profesor (escolaridad, ordenamiento, profesor de planta) y (iii) del nivel de exigencia (calificación promedio de los estudiantes). También se quiere establecer si, en el caso específico de la Universidad de los Andes, se cumple el llamado efecto de reciprocidad entre calificaciones obtenidas por los estudiantes y la calificación asignada a los profesores (Clayson, Frost y Sheffet, 2006)<sup>1</sup>.

---

<sup>1</sup> El efecto de reciprocidad se refiere a que los estudiantes califican a sus profesores de acuerdo con la calificación que obtuvieron, sin hacer ningún tipo de consideración de su esfuerzo y rendimiento en la asignatura. Estudiantes con buenas calificaciones evalúan más positivamente a sus profesores, y estudiantes con malas calificaciones tienden a castigar a sus maestros en las evaluaciones.

La base de datos utilizada cuenta con información de las evaluaciones a profesores y programas de parte de los estudiantes al finalizar el período académico 2006 – 1. La base también cuenta con información sobre la calificación promedio que recibieron los estudiantes y algunas características del profesor, tales como la escolaridad, la formación académica y el ordenamiento. Es importante aclarar que sólo se tiene información para cursos de pregrado.

Uno de los aportes de este trabajo es la posibilidad de contar con una base de datos con información de sección<sup>2</sup>, lo cual permite construir un panel de datos para evaluar específicamente la hipótesis de reciprocidad y evitar problemas de endogeneidad al utilizar efectos fijos por profesor.

El análisis empírico permite entender, desde la perspectiva de los estudiantes, los atributos de un buen profesor, y permite, al mismo tiempo, determinar en qué medida los resultados de las evaluaciones están sesgados por las calificaciones de los estudiantes. A pesar de que en la Universidad de los Andes las evaluaciones a profesores no se tienen en cuenta, al menos de manera directa, en la remuneración, se encuentra una relación positiva entre la calificación que obtienen los estudiantes en el curso y la evaluación dada por éstos a los profesores; específicamente, se pudo establecer que la relación entre estas dos variables se debía a un efectos de reciprocidad. Este resultado sugiere que los profesores podrían utilizar el nivel de exigencia como un instrumento para *manipular* el resultado de las evaluaciones en busca de promoción y permanencia dentro de la institución.

Cabe señalar, finalmente, que el análisis muestra algunas diferencias en la calificación de los profesores y cursos según la Facultad. Algunas características del profesor y del curso también resultaron ser determinantes de la calificación de profesores.

Este documento está compuesto de varias secciones. La primera presenta una revisión de la literatura relevante; la segunda describe los datos; la tercera sección presenta la estrategia empírica; la cuarta

---

<sup>2</sup> Una asignatura está compuesta muchas veces por varias secciones, las cuales comparten un mismo horario y profesor. Algunos profesores pueden estar encargados de varias secciones de una misma asignatura.

sección expone los resultados y la última concluye y expone las implicaciones de los resultados.

## **I. Literatura sobre evaluaciones a profesores**

La literatura sobre las evaluaciones de profesores y programas puede dividirse en dos grupos: el primero está compuesto de las investigaciones que proponen diferentes sistemas de evaluación y plantean recomendaciones al respecto; el segundo está conformado por los trabajos que examinan cuantitativamente el efecto que tiene la implantación de las evaluaciones sobre los profesores y los estudiantes. La revisión de la literatura hace un repaso de los dos grupos mencionados, pero acentúa las investigaciones que analizan los sistemas de evaluaciones de forma cuantitativa y en particular la relación entre los resultados de las evaluaciones y las notas de los estudiantes.

### **A. Literatura sobre diseño y aplicación de las evaluaciones**

Las evaluaciones a profesores y programas surgieron como una herramienta que permitiría (i) determinar si los objetivos y las metas propuestas estaban siendo alcanzados, (ii) mejorar el rendimiento de los profesores y (iii) facilitar las decisiones sobre la promoción, remuneración y continuidad de los maestros (Wrightstone, 1943; Simpson, 1967; O'Hanlon y Mortensen, 1980). Pero diseñar un sistema de evaluación que cumpla con los objetivos señalados no ha sido tarea fácil, y ha tardado largo tiempo en consolidarse. Shinkfield y Stufflebeam (1995) señalan que fue en Inglaterra, hacia finales del siglo XIX, donde se estableció por primera vez un sistema de remuneración basado en resultados y donde se planteó inicialmente la necesidad de contar con una herramienta capaz de identificar los buenos maestros. Sin embargo, estos autores señalan que las primeras investigaciones académicas en este campo fueron posteriores a la Segunda Guerra Mundial, y que los hallazgos y las recomendaciones no serían recogidos hasta ya entrada la década de 1970, cuando se empezó a institucionalizar el uso de las evaluaciones a profesores<sup>3</sup>. Esto es, durante más de cien años se ha

---

<sup>3</sup> Para una revisión detallada sobre la evolución histórica de las evaluaciones a profesores, una buena referencia es Shinkfield y Stufflebeam (1995).

tratado (quizá sin mucho éxito) de diseñar un sistema formal de evaluaciones a profesores, capaz de identificar las características que hacen que un docente influya positivamente en el aprendizaje de sus pupilos.

En esa línea se han propuesto diferentes sistemas de evaluación: evaluaciones estandarizadas de parte de los estudiantes, evaluaciones de parte de los coordinadores y administradores, opiniones de los colegas, autoevaluación y datos sobre el aprendizaje, los cuales utilizan pruebas estandarizadas<sup>4</sup>. Del mismo modo se han hecho algunas recomendaciones importantes sobre la metodología que se debe seguir para su aplicación<sup>5</sup>.

La literatura presenta algunas recomendaciones que permitirían acopiar la mayor información posible sobre el rendimiento de los profesores y la calidad de los programas. Dentro de estas recomendaciones pueden mencionarse: las evaluaciones deben aplicarse con un propósito definido y la información obtenida no debe utilizarse con ningún otro fin; las evaluaciones no deben aplicarse después de un examen final y deben garantizar el anonimato de los estudiantes; los estudiantes deben estar convencidos de que las evaluaciones son relevantes para la toma de decisiones (Frey, 1976; Haider, 1976; O'Hanlon y Mortensen, 1980).

Actualmente, las evaluaciones realizadas por estudiantes son las más utilizadas (Giles y Pascoe, 2004). La razón tiene que ver posiblemente con el menor costo y la facilidad de recolección de los datos en este tipo de evaluaciones (O'Hanlon y Mortensen, 1980, Langbein, 2007). Sin embargo, según Clayson *et al.* (2006), la veracidad de este tipo de pruebas ha sido puesta en cuestión por la comunidad académica, especialmente cuando éstas son usadas como instrumento para la promoción y remuneración de los maestros.

---

<sup>4</sup> En la literatura el ejemplo más común sobre este tipo de pruebas es *The Tennessee Value – Added Assessment System (TVAAS)*. Se puede encontrar discusiones sobre la validez de esta prueba en Wright *et al.* (1997); Kupermintz, Shepard y Linn (2001) y Glass (2004).

<sup>5</sup> Hay una extensa literatura sobre las diferentes formas de evaluación y sobre su metodología de aplicación. Véanse, por ejemplo, Estrin (1958); Simpson (1967); O'Hanlon y Mortensen (1980); Shinkfield y Stufflebeam (1995) y Glass (2004).

## B. Literatura sobre la relación entre la calificación de los estudiantes y los profesores

Existen dos enfoques para explicar la relación que existe entre las evaluaciones de profesores y estudiantes. El primer enfoque defiende la validez de las evaluaciones realizadas por los estudiantes y argumenta que la relación positiva se explica principalmente por la motivación y efectividad del profesor. El segundo enfoque señala que existe un sesgo en este tipo de evaluaciones, como resultado de los llamados efectos de “clemencia” y “reciprocidad”. A continuación se explican en detalle cada uno de los dos enfoques.

Dentro del primer grupo de explicaciones, Marsh y Roche (1997) argumentan que la relación positiva que existe entre las calificaciones de los estudiantes y los resultados de las evaluaciones puede ser explicada porque los buenos profesores logran que sus estudiantes entiendan mejor los contenidos del curso, lo cual se ve reflejado en un mejor desempeño académico. Como los estudiantes son conscientes de su aprendizaje, califican mejor a sus maestros.

Howard y Maxwell (1982) ofrecen otra explicación. Estos autores proponen la hipótesis de la motivación, según la cual los profesores que *motivan* el interés por el contenido de sus cursos logran que sus estudiantes se esfuercen, obtengan buenas calificaciones y estén por tanto satisfechos con el curso y el profesor, por lo cual tienden a evaluar más positivamente a sus maestros. Cabe señalar, sin embargo, que quienes defienden<sup>6</sup> el uso de este tipo de evaluaciones argumentan que éstas no deben ser la única fuente de información, sino que deben ser complementadas con otros sistemas de evaluación (Cachin, 1995; d'Apollonia y Abrami, 1997).

Con todo, existe una serie de trabajos que cuestionan el uso de estas evaluaciones, con el argumento principal de que los resultados pueden ser determinados en cierta medida por la calificación que los estudiantes obtienen o esperan obtener. Si esto es así, los profesores tendrían

---

<sup>6</sup> Algunos de los autores que encuentran válido el uso de las evaluaciones de estudiantes a profesores son: Howard y Maxwell (1980); Howard y Maxwell (1982); Greenwald (1997); Marsh y Roche (1997); d'Apollonia y Abrami (1997); McKeachie (1997).

una herramienta poderosa (el nivel de exigencia) para incidir en el resultado de las evaluaciones (Bonesrønning, 2004). Los profesores, conscientes de que pueden obtener beneficios si logran una calificación más alta, simplemente inflarán deliberadamente las calificaciones de sus alumnos en espera de que éstos les correspondan con una buena evaluación (Langbein, 2007)<sup>7</sup>.

La primera hipótesis en este sentido plantea un relación dominada “clemencia”: los profesores que califican con un *nivel de exigencia* más bajo obtienen mejores calificaciones que aquellos que califican con un nivel más alto, pues los estudiantes, a la hora de evaluar, perciben la “clemencia” con la cual los profesores calificaron sus pruebas (Krautmann y Sander, 1999; Bacon y Novotny, 2002). Una segunda hipótesis sugiere que el efecto no es de “clemencia”, sino de reciprocidad: los estudiantes no se preocupan por qué tan estricto sea el profesor, sino por la calificación final, *independientemente del nivel de exigencia*. Es decir, estudiantes con buenas calificaciones evaluarán más positivamente a sus maestros, y quienes obtengan malas calificaciones tienden a castigarlos en las evaluaciones, sean o no estrictos los profesores (Clayson *et al.*, 2006).

Claramente no hay un consenso sobre la validez de las evaluaciones de los profesores realizadas por los estudiantes. Greenwald (1997) señala que las posiciones al respecto han estado divididas desde el decenio del setenta. El propósito de este trabajo no es demostrar si las evaluaciones realizadas por los estudiantes son un instrumento válido, pues parte de reconocer que las evaluaciones entregan información valiosa. En este sentido, analiza cuáles son las características del profesor y del programa que los estudiantes consideran importantes, e intenta, al mismo tiempo, cuantificar la importancia que éstos le dan a la calificación que ellos mismos recibieron.

En la mayoría de las instituciones en las cuales se ha encontrado una relación positiva entre calificaciones de profesores y estudiantes, se cuenta con un sistema de pago por méritos a profesores, el cual

---

<sup>7</sup> Jacob y Levitt (2002) presentan evidencia de cómo algunos profesores de educación básica primaria y bachillerato en Chicago realizan trampa en la presentación de pruebas estandarizadas de los estudiantes, las cuales muchas veces son tenidas en cuenta para su remuneración o la de los planteles educativos a los cuales pertenecen.

puede haber inducido a estos a bajar el nivel de exigencia, y puede haber generado, entonces, un fenómeno de inflación en las notas de los estudiantes. En la Universidad de los Andes no existe un esquema de remuneración docente basado en las evaluaciones de estudiantes, pero el fenómeno de inflación en las calificaciones sí parece evidente (Vallejo, 2004). Si efectivamente las evaluaciones a profesores son las causantes de la inflación de notas, no es porque éstos utilicen su nivel de exigencia para obtener mayores ingresos en el corto plazo, sino porque hacen lo propio con el fin de lograr la promoción y la permanencia dentro de la institución.

## **II. Datos**

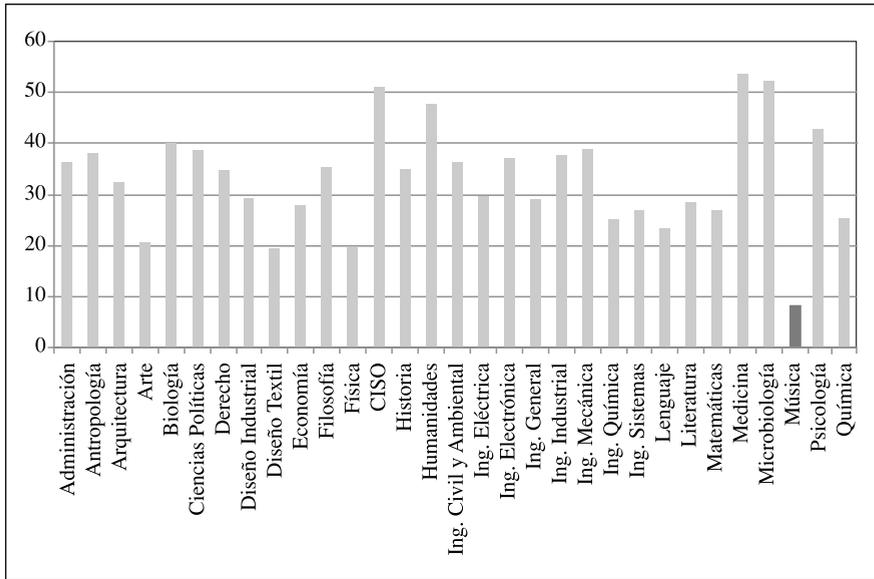
La base de datos usada en este trabajo incluye información de las evaluaciones a profesores y programas para los cursos de nivel de pregrado de la Universidad de los Andes para el primer semestre de 2006. Se cuenta con información de secciones y no de estudiantes, porque las evaluaciones son anónimas, lo cual hace imposible vincular la calificación de un estudiante con la evaluación que hizo de un profesor.

Las evaluaciones son respondidas por los estudiantes al finalizar el período académico. Las respuestas son anónimas y voluntarias. El cuestionario está compuesto de 27 preguntas que comprenden cuatro módulos (el profesor, el curso, las evaluaciones y una autoevaluación); el anexo 1 presenta el cuestionario completo de la evaluación. Además de las respuestas a los cuestionarios, la base de datos usada contiene información sobre el profesor: la escolaridad (pregrado, especialización, maestría, doctorado), el ordenamiento (instructor, profesor asistente, profesor asociado, profesor titular) y si es de planta o de cátedra. La base de datos también incluye información sobre la sección: el número de estudiantes inscritos y aquellos que respondieron a la evaluación, el promedio académico de las calificaciones obtenidas al finalizar el curso, el carácter (electivo o no electivo), el nivel dentro del pènsum (1 = introductorio, 2 = de mitad de carrera, 3 = énfasis), la Facultad y el Departamento que lo ofrece.

La base de datos se restringió a las secciones en las cuales más de 40% de los estudiantes habían contestado a la evaluación, con promedio

académico mayor a 2,0 y con al menos cinco estudiantes inscritos. Los cursos dictados por el Departamento de Música no se tomaron en consideración por su forma de evaluación y al pequeño número de estudiantes por clase<sup>8</sup>. El gráfico 1 presenta el tamaño promedio de las secciones para cada uno de los departamentos de la Universidad. Claramente las clases de música tienen un tamaño muy inferior al promedio y (por su naturaleza) usan metodologías diferentes. De allí su exclusión de la base de datos.

Gráfico 1. Tamaño promedio de las secciones por Departamento.



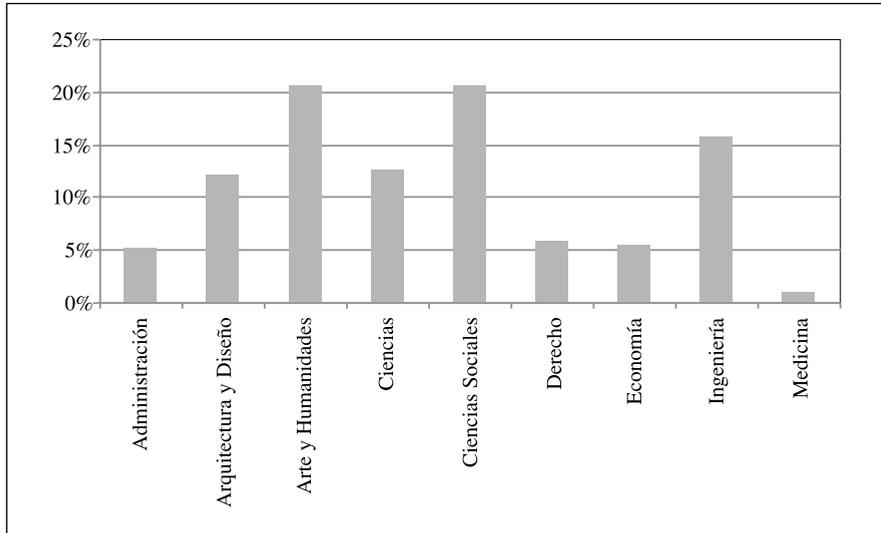
Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

Después de restringir la muestra, se tiene información para 1358 secciones, correspondientes a 770 cursos o asignaturas, de los cuales el 18% son cursos electivos; 38% son de nivel uno, es decir, cursos introductorios; 28% son cursos de nivel dos y 34% de nivel tres. La Facultad que más cursos dicta es Ciencias Sociales, con aproximadamente 22% de los cursos (Ciencias Sociales dicta 52% de los cursos electivos) y la que menos cursos dicta es Medicina, con 1,2% (la Fa-

<sup>8</sup> Los cursos de Música en su mayoría son personalizados. En ellos no se califica de acuerdo con la evaluación de pruebas escritas o presentación de trabajos, sino más bien, según la evolución en la interpretación musical mediante audiciones programadas.

cultad de Medicina fue abierta en 2004). En promedio, existen 31,4 estudiantes en cada sección. El gráfico 2 presenta la distribución de cursos dictados por Facultades.

**Gráfico 2.** Porcentaje de cursos dictados por cada Facultad.



*Fuente:* Oficina de Admisiones y Registro y cálculos de los autores.

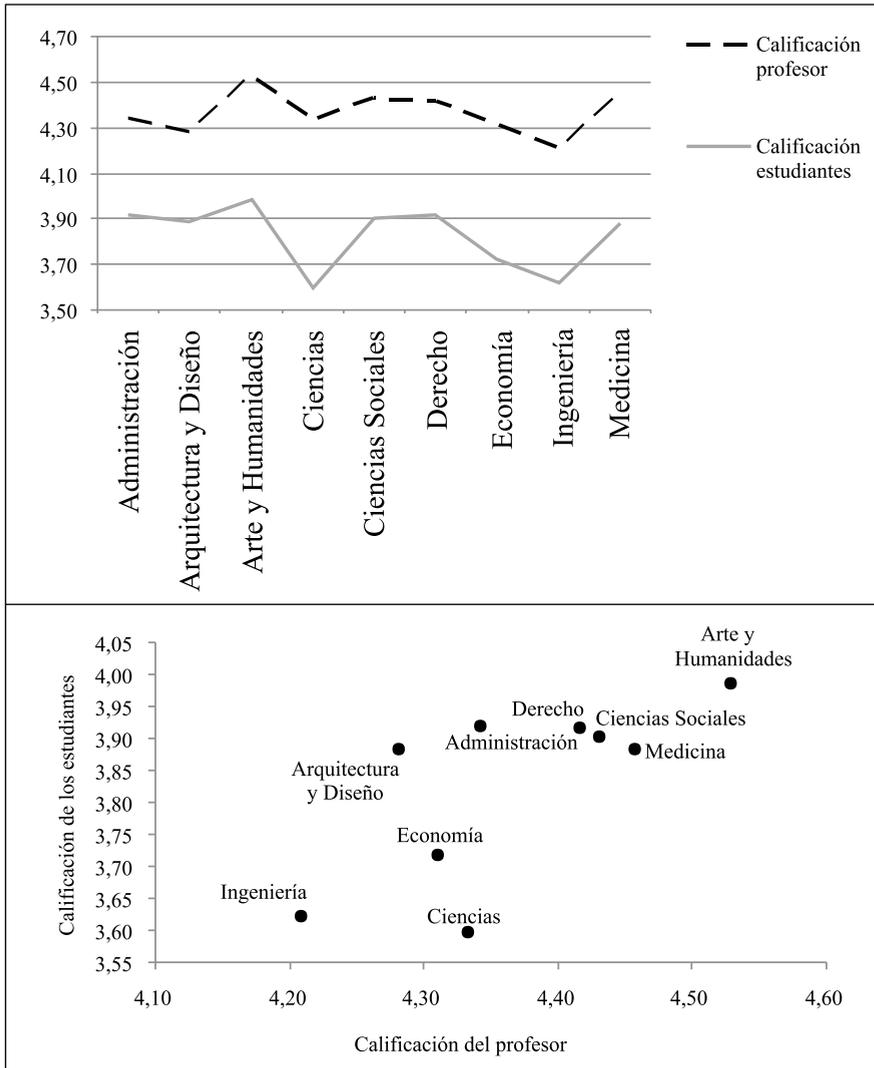
Los datos muestran que 825 profesores tienen a su cargo 1,6 secciones en promedio. La calificación global promedio de los profesores es 4,33 y la nota promedio de los estudiantes es 3,81. El gráfico 3 presenta el comportamiento promedio de cada una de estas calificaciones según la Facultad “dueña” del curso. Los datos muestran que en las Facultades de Ciencias e Ingeniería se encuentran (en promedio) las secciones con calificaciones de estudiantes y profesores más bajas. A primera vista se puede apreciar que las variables mostradas guardan estrecha relación entre sí.

El 56% de las secciones son dictadas por profesores de planta. Los profesores asistentes y asociados, con calificaciones promedio de 4,39 y 4,33, respectivamente, resultaron ligeramente mejor calificados que los instructores y profesores titulares, que obtuvieron calificaciones promedio de 4,23 y 4,30.

La correlación entre la calificación de los estudiantes y del profesor es de 0,24, y de 0,20 con la evaluación global del curso. Después de restringir la muestra, tres cuartas partes de los inscritos en una sección

respondieron a la encuesta. La calificación promedio de los estudiantes (3,81) es inferior al promedio simple de las preguntas del módulo de autoevaluación (4,20). La correlación entre estas dos calificaciones es 0,22. La correlación entre la calificación global del profesor ( $p11$ ) y el promedio del módulo del profesor ( $g1$ ) es de 0,97. El cuadro 1 resume las estadísticas descriptivas por sección de las principales variables después de restringir la muestra.

Gráfico 3. Calificación promedio del profesor y de los estudiantes por Facultad.



Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

Cuadro 1. Estadísticas descriptivas de las secciones.

Variable	Obs.	Media	Desv. estándar	Mín.	Máx.
<b>Evaluación</b>					
Calificación Global del Profesor ( <i>p11</i> )	1358	4,33	0,45	1,80	5,00
Calificación Global del Curso ( <i>p17</i> )	1358	4,30	0,40	2,40	5,00
Prom. Preguntas del Profesor ( <i>g1</i> )	1358	4,37	0,37	2,45	5,00
Prom. Preguntas del Curso ( <i>g2</i> )	1358	4,38	0,34	2,60	5,00
Prom. Preguntas de las Evaluaciones ( <i>g3</i> )	1358	4,26	0,40	2,23	5,00
Prom. Preguntas de Autoevaluación ( <i>g4</i> )	1358	4,20	0,26	3,16	4,96
Calificación Promedio de los Estudiantes ( <i>prom_estud</i> )	1358	3,81	0,42	2,00	5,00
Número de Estudiantes ( <i>num_est</i> )	1358	31,39	17,63	5,00	184,00
Porcentaje de Encuestados ( <i>porc_encues</i> )	1358	0,74	0,14	0,40	1,00
Curso Electivo ( <i>cbu</i> )	1358	0,16	0,36	0,00	1,00
Profesor de Planta ( <i>planta</i> )	1358	0,56	0,50	0,00	1,00
Nivel 1	1358	0,47	0,50	0,00	1,00
Nivel 2	1358	0,28	0,45	0,00	1,00
Nivel 3	1358	0,25	0,43	0,00	1,00

Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

Al comparar algunas de las variables más importantes por Facultad, se encontró que, en promedio, los cursos de Artes y Humanidades y Medicina fueron los que obtuvieron mejor calificación global de parte de los estudiantes. En cuanto a las preguntas sobre las evaluaciones y el monitor, Ingeniería obtuvo el promedio más bajo. La Facultad que tiene mayor proporción de secciones dictadas por profesores de planta es Ingeniería, con el 75,3%, y la que tiene proporción menor es Economía, con el 33,6%. El cuadro 2 presenta las estadísticas descriptivas de las secciones según la Facultad.

Cuadro 2. Estadísticas descriptivas de las secciones según la Facultad.

	<i>p11</i>	<i>p17</i>	<i>g1</i>	<i>g2</i>	<i>g3</i>	<i>g4</i>	<i>prom_estud</i>	<i>num_est</i>	<i>porc_encues</i> %	<i>cbu</i> %	<i>planta</i> %
Administración	4,34	4,28	4,36	4,39	4,21	4,28	3,92	36,30	74,07	9,38	64,58
Arquitectura y Diseño	4,28	4,26	4,31	4,33	4,21	4,31	3,89	29,63	73,28	4,00	37,00
Arte y Humanidades	4,43	4,43	4,47	4,48	4,46	4,25	4,10	25,13	74,36	28,29	52,63
Ciencias	4,33	4,23	4,37	4,32	4,23	4,15	3,60	28,94	75,25	20,00	53,95
Ciencias Sociales	4,43	4,39	4,46	4,47	4,42	4,13	3,90	35,38	74,27	38,07	69,54
Derecho	4,42	4,35	4,43	4,44	4,35	4,18	3,92	34,79	73,92	0,00	51,28
Economía	4,31	4,32	4,33	4,39	4,29	4,15	3,72	27,64	74,70	17,65	33,61
Ingeniería	4,21	4,21	4,25	4,30	4,03	4,18	3,62	32,88	72,68	5,98	75,30
Medicina	4,46	4,41	4,43	4,47	4,33	4,27	3,88	53,82	67,20	0,00	63,64

Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

### III. Metodología

Este trabajo utiliza modelos de regresión con efectos fijos para responder a las preguntas planteadas anteriormente. Se proponen varias especificaciones con el propósito de encontrar la que mejor se ajusta al problema planteado. Inicialmente se estima un modelo empírico que intenta establecer los determinantes de la evaluación global del profesor. Esta variable ( $p11_i$ ) está basada en la calificación dada por los estudiantes a la pregunta: ¿cuál es la calificación global del profesor de la  $i$ -ésima sección? El modelo utiliza las siguientes variables: el número de estudiantes ( $num\_est$ ), el promedio académico de los estudiantes ( $prom\_estud$ ), si los profesores son de planta ( $planta = 1$ ), si el curso es electivo ( $cbu = 1$ ) y el porcentaje de estudiantes que respondió a la evaluación ( $porc\_encues$ ). La ecuación (1) presenta la especificación utilizada.

$$p11_i = \beta_0 + \beta_1 \cdot num\_est_i + \beta_2 \cdot prom\_estud_i + \beta_3 \cdot porc\_encues_i + \beta_4 \cdot planta_i + \beta_5 \cdot cbu_i + \varepsilon_i \quad (1)$$

#### A. Diferencias entre Facultades

Es razonable esperar que los resultados de las evaluaciones y los criterios tenidos en cuenta por los estudiantes para evaluar a sus profesores

estén explicados por algunos factores propios de cada Facultad. Estos factores pueden incluir, entre otros, las diferencias en los contenidos, la dificultad de los cursos, las metodologías de enseñanza y evaluación y la “cultura” de la Facultad. Por tal razón, se estima un segundo modelo, en el cual se incluyen efectos fijos por Facultad. La ecuación (2) describe la especificación estimada, en la cual  $\theta_k$  representa el efecto sobre la calificación global del profesor de la  $i$ -ésima sección en la  $k$ -ésima Facultad.

$$p11_{ik} = \beta_0 + \beta_1 \cdot num\_est_{ik} + \beta_2 \cdot prom\_estud_{ik} + \beta_3 \cdot porc\_encues_{ik} + \beta_4 \cdot planta_{ik} + \beta_5 \cdot cbu_{ik} + \theta_k + \varepsilon_{ik} \quad (2)$$

## B. Ordenamiento del profesor y nivel del curso

Para comprobar si otras características, tales como el ordenamiento del profesor y el nivel del curso, son determinantes del resultado de las evaluaciones, se estimaron modelos adicionales. La ecuación (3) intenta establecer si el nivel del curso es una característica relevante en los resultados de las evaluaciones: el subíndice “ $l$ ” se refiere al nivel del curso,  $\lambda_l$  es el efecto sobre la calificación de los profesores del nivel  $l$ -ésimo. La ecuación (4) intenta busca identificar si el ordenamiento del profesor (éste puede ser un indicador de la experiencia docente) es determinante: la categoría omitida es la de profesores instructores.

$$p11_{il} = \beta_0 + \beta_1 \cdot num\_est_{il} + \beta_2 \cdot prom\_estud_{il} + \beta_3 \cdot porc\_encues_{il} + \beta_4 \cdot planta_{il} + \lambda_l + \varepsilon_{il} \quad (3)$$

$$p11_i = \beta_0 + \beta_1 \cdot num\_est_i + \beta_2 \cdot prom\_estud_i + \beta_3 \cdot porc\_encues_i + \beta_4 \cdot asociado_i + \beta_5 \cdot asistente_i + \beta_6 \cdot titular_i + \varepsilon_i \quad (4)$$

Las especificaciones anteriores permiten estimar los efectos de las características del profesor y del curso, pero no permiten distinguir el (los) mecanismo(s) que explican la existencia de una relación entre las calificaciones del profesor y de los estudiantes. En otras palabras: no permiten discernir entre los diferentes efectos mencionados anteriormente: motivación, clemencia, reciprocidad, etc. Por tal razón, se estimaron adicionalmente algunas especificaciones que incluyen efectos fijos por profesor y que permiten distinguir entre los efectos ya mencionados.

### C. Efectos fijos por profesor

Isely y Singh (2005) y McPherson (2006) utilizan esta estimación para prevenir dos posibles problemas: (i) un posible sesgo a causa de la omisión de variables no observadas del profesor determinantes en el resultado de las evaluaciones y (ii) la endogeneidad entre la calificación del profesor y del estudiante. La ecuación (5) presenta la especificación usada:  $\delta_j$  representa el efecto sobre la calificación global del j-ésimo profesor en la i-ésima sección.

$$p11_{ij} = \beta_0 + \beta_1 \cdot num\_est_{ij} + \beta_2 \cdot prom\_estud_{ij} + \beta_3 \cdot porc\_encues_{ij} + \beta_4 \cdot cbu_{ij} + \delta_j + \varepsilon_{ij} \quad (5)$$

Dentro de las variables del profesor que pueden ser “recogidas” por el efecto fijo figuran el ordenamiento (asociado, asistente, cátedra e instructor), la dedicación o diligencia del profesor (Teven y McCroskey, 1996), el nivel de exigencia en la calificación (Bacon y Novotny, 2002), la motivación y el interés despertado por el curso y el aspecto físico (Hamermesh y Parker, 2003). Todos estos aspectos pueden afectar los resultados de las evaluaciones y algunos de ellos (los no observados) podrían sesgar los estimadores.

En esta última especificación, si la variable *prom\_estud* resulta significativa y con un signo positivo, podemos estar seguros de que los estudiantes toman en cuenta la nota en sí misma (independientemente del nivel de exigencia) para evaluar a sus profesores, lo que constituye evidencia indirecta en favor del efecto de reciprocidad. Por ejemplo, si la relación fuese explicada por el efecto de motivación, propuesto por Howard y Maxwell (1982), la variable *prom\_estud* debería perder importancia al pasar de la primera a la quinta especificación, puesto que los efectos fijos estarían capturando la motivación inherente al estilo de cada profesor.

Pero, si la relación positiva entre las calificaciones de los estudiantes y la evaluación del profesor fuese explicada por el efecto de efectividad (eficacia pedagógica), este sería igualmente capturado por los efectos fijos, los cuales recogerían la habilidad intrínseca del profesor. Del mismo modo, esta especificación permitiría descartar el efecto de clemencia, pues el nivel de exigencia del profesor sería igualmente

capturado. En suma: los efectos fijos recogerían todos los aspectos “invariantes” del profesor: capacidad de motivación, eficacia pedagógica, nivel de exigencia, etc.

Es importante aclarar que todo lo anterior es posible, suponiendo que los profesores se desempeñan de forma similar en cada uno de los cursos; es decir, los profesores en todos sus cursos realizan sus exposiciones con la misma claridad, aplican más o menos las mismas técnicas y metodologías pedagógicas, califican con la misma rigurosidad y motivan por igual a los estudiantes.

Para aplicar esta metodología se toman únicamente los profesores que dictaban dos o más secciones. Si bien se pierden observaciones (se pasa de 1358 a 898 observaciones), la muestra continúa siendo suficientemente grande. Finalmente, cabe anotar que todas las especificaciones anteriores se estimaron alternativamente utilizando como variable dependiente el promedio de las preguntas del módulo del profesor, (preguntas 1 a 11), con el propósito de revisar si los resultados se mantenían después de redefinir la variable dependiente.

#### D. Efectos fijos por curso

Con el fin de identificar cuáles son los factores que determinan la calificación promedio de los cursos, se estiman varias especificaciones. Todas ellas usan como variable dependiente el promedio de las preguntas del módulo del curso (preguntas 12 a 17 del cuestionario). Las dos primeras especificaciones estimadas son idénticas a las descritas por las ecuaciones (1) y (2). La tercera especificación es similar a la ecuación (5), pero en lugar de los efectos fijos por profesor, contiene efectos fijos por curso (véase ecuación 6). La variable  $g2_{ij}$  es el promedio de la calificación de la  $i$ -ésima sección del  $j$ -ésimo curso y  $\phi_j$  es el efecto sobre la calificación del curso  $j$ -ésimo.

$$g2_{ij} = \beta_0 + \beta_1 \cdot num\_est_{ij} + \beta_2 \cdot prom\_estud_{ij} + \beta_3 \cdot porc\_encues_{ij} + \phi_j + \varepsilon_{ij} \quad (6)$$

#### IV. Resultados y análisis

El cuadro 3 presenta los resultados de las especificaciones descritas por las ecuaciones (1), (2) y (3) de la sección anterior.

Cuadro 3. Resultados de las regresiones para p11 por MCO.

Variable dependiente: Calificación Global del Profesor (P11).				
Ecuación	1	2	3	2-3
num_est	-0,002 (0,001) **	-0,002 (0,001) **	-0,002 (0,001) **	-0,002 (0,001) ***
prom_estud	0,263 (0,028) ***	0,251 (0,030) ***	0,274 (0,029) ***	0,257 (0,031) ***
porc_encues	0,228 (0,087) ***	0,212 (0,086) **	0,212 (0,087) ***	0,197 (0,086) **
Planta	0,027 (0,025)	0,029 (0,026)	0,039 (0,025)	0,039 (0,026)
Cbu	0,048 (0,028) *	0,016 (0,031)		
Nivel 2			-0,082 (0,030) ***	-0,080 (0,031) ***
Nivel 3			-0,087 (0,031) ***	-0,061 (0,031) **
Administración		0,063 (0,046)		0,067 (0,046)
Arquitectura y Diseño		0,009 (0,046)		-0,001 (0,046)
Arte y Humanidades		0,089 (0,048) *		0,069 (0,047)
Ciencias		0,123 (0,046) ***		0,100 (0,047) **
Ciencias Sociales		0,148 (0,042) ***		0,138 (0,040) ***
Derecho		0,142 (0,049) ***		0,147 (0,048) **
Economía		0,074 (0,049)		0,078 (0,049)
Medicina		0,233 (0,153)		0,187 (0,153)
Constante	3,191 (0,133) ***	3,175 (0,138) ***	3,209 (0,133) ***	3,212 (0,138) ***
R2 - Ajustado	0,069	0,085	0,0762	0,0904
Obs.	1358	1358	1358	1358

Errores estándar robusto en paréntesis<sup>9</sup>.

\*\*\* Significancia al 99%, \*\* significancia al 95%, \* significancia al 90%.

Ingeniería y el Nivel 1 son las categorías omitidas.

Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

Los resultados muestran que la variable *prom\_estud* es positiva y significativa en cada una de las especificaciones. Un aumento de una unidad en la calificación promedio de los estudiantes aumenta la calificación del profesor en 0,26. Este resultado es prácticamente idéntico al encontrado por Isely y Singh (2005). Estos autores estiman un modelo con efectos fijos por profesor y por curso en una universidad de los Estados Unidos (Grand Valley State University, Michigan), donde las

<sup>9</sup> Se encontró que algunos de los modelos presentan heteroscedasticidad, por lo cual se justifica el uso de errores estándar robustos. Específicamente, se encontró que la heteroscedasticidad era originada por la variable *num\_est*, lo cual no resulta sorprendente, dado que todas las variables son promedios de las observaciones por estudiante.

evaluaciones son utilizadas como parte del sistema de remuneración de los maestros. El efecto estimado está entre 0,21 y 0,33.

La variable *num\_est* es significativa en las tres especificaciones y tiene efecto negativo sobre la calificación del profesor. A medida que el tamaño de la clase aumenta, las interacciones alumno–profesor pueden verse afectadas tanto en cantidad como en calidad, pues conforme existan más alumnos, el tiempo disponible del profesor debe dividirse entre un número mayor de estudiantes<sup>10</sup>. La estimación muestra que por cada diez estudiantes adicionales, la calificación del profesor disminuye 0,02. Isely y Singh (2005) reportan que diez estudiantes adicionales reducen la calificación del profesor entre 0,03 y 0,04. McPherson (2006) reporta que el efecto en cuestión se sitúa entre 0,01 y 0,015. La magnitud de este coeficiente podría estar subestimada, porque algunas clases magistrales (se dictan en el mismo salón, a la misma hora, con el mismo profesor) podrían contar con varias secciones; esto es, la variable *num\_est* estaría midiendo el total de inscritos por sección y no por clase magistral. Sin embargo, no es posible hacer una corrección al omitir dichos cursos o controlar por el número de secciones, pues no es posible identificar si se trata de una clase magistral compuesta de varias secciones.

El porcentaje de encuestados (*porc\_encues*) es significativo en las tres especificaciones. Un aumento de 10 puntos porcentuales en la proporción de encuestados implica un aumento de 0,021 en la calificación global del profesor. Según McPherson (2006), esta variable debe ser incluida para prevenir un posible sesgo de selección. Este autor argumenta que el efecto en cuestión tiene un signo ambiguo: (i) puede ser un indicador del grado de interés de los alumnos en el curso y por tanto cabría esperar un efecto positivo sobre la calificación del profesor, o (ii) puede incidir en la proporción de estudiantes de bajo rendimiento académico que contestan a la evaluación y por tanto cabría esperar un efecto negativo. En este caso el efecto positivo parece primar.

---

<sup>10</sup> Esta idea es expuesta por Gaviria (2002), que se refiere a cómo un mayor número de hermanos en la familia afecta el rendimiento en la prueba del Icfes, como resultado de la división de los recursos financieros y tiempo dedicado por los padres para atender a cada hijo. Angrist y Lavy (1999) muestran evidencia de cómo el tamaño de la clase afecta el logro académico de los estudiantes y por tanto la calidad de la educación.

El carácter electivo de los cursos (*cbu*) no es significativo. Además, se realizó una prueba de Chow con el propósito de comprobar si la agrupación de los cursos electivos con los no electivos tenía validez. La prueba muestra que los coeficientes de ambos grupos son estadísticamente iguales y por tanto la agrupación es válida.

Se encontró, además, que los profesores de planta y de cátedra no presentaban diferencias estadísticamente significativas en el resultado de las evaluaciones. Los profesores de planta demuestran tener mejor conocimiento de los contenidos del curso, son más cumplidos con los horarios y atienden más consultas de los estudiantes por fuera de los horarios de clase; sin embargo, los profesores de cátedra motivan más el interés por el curso, realizan explicaciones más claras y fomentan más la participación de los estudiantes. Las fortalezas de unos y otros parecen contrarrestarse en la calificación final.

Además, profesores que dictan cursos del nivel uno (cursos introductorios) resultaron mejor calificados que aquellos que dictan cursos de mitad y de final de la carrera. Este resultado puede ser explicado si se supone que el nivel de exigencia de los estudiantes aumenta con el tiempo. A medida que los estudiantes avanzan en su carrera, no sólo adquieren conocimientos propios de su disciplina, los cuales les permiten evaluar la calidad de los cursos dictados, sino que tienen la posibilidad de responder a las evaluaciones teniendo como referente a otros profesores.

En cuanto a las diferencias entre Facultades, se encontró que los profesores de la Facultad de Ingeniería eran peor calificados que sus colegas de Artes y Humanidades, Ciencias, Ciencias Sociales y Derecho. Los profesores de la Facultad de Ingeniería tienen las menores calificaciones en cada una de las preguntas del módulo del profesor (preguntas 1 a 11).

El cuadro 4 presenta los resultados de la estimación para las especificaciones descritas por las ecuaciones (4) y (5). Este análisis permite especificar si la relación positiva entre calificaciones de los estudiantes y del profesor se debe al efecto de reciprocidad o a cualquier otro de los mecanismos descritos anteriormente. El análisis también permite determinar si existen diferencias en las evaluaciones relacionadas con

el ordenamiento de los maestros. La primera columna muestra los resultados de la primera especificación (ecuación 1) para la muestra restringida a los profesores que dictan dos o más cursos, con el propósito de hacer comparables los resultados con la especificación en la cual se incluyen los efectos fijos por profesor. Para la primera y tercera columnas se tienen 898 observaciones, porque se limitó la muestra a los profesores que dictaban dos o más secciones. En la columna 2 solamente se incluyeron los profesores de planta que tuvieran información sobre el ordenamiento, por lo cual se tienen 641 observaciones. Sin embargo, la muestra sigue siendo representativa para los cursos de la Universidad de los Andes.

**Cuadro 4.** Resultados de las regresiones para  $p11$  por MCO y efectos fijos.

Variable dependiente: calificación Global del Profesor (P11).										
Ecuación	1			4			5			
num_est	-0,001	-0,001		-0,001	-0,001		-0,002	-0,001		
prom_estud	0,27	-0,032	***	0,365	-0,046	***	0,270	-0,059	***	
porc_encues	0,213	-0,107	**	0,317	-0,123	***	-0,054	-0,123		
planta	-0,002	-0,034								
cbu	0,034	-0,032					-0,012	-0,048		
Profesor Asistente				0,081	-0,049	*				
Profesor Asociado				0,024	-0,051					
Profesor Titular				0,035	-0,067					
Constante	3,19	-0,151	***	2,789	-0,209	***	3,418	-0,259	***	
Efectos Fijos por Profesor		NO			NO			SI		***
R2	0,0729			0,1184			0,0642			
Obs	898			641			898			

Errores estándar robusto en paréntesis.

\*\*\* Significancia al 99%, \*\* significancia al 95%, \* significancia al 90%.

Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

Al pasar de la primera a la quinta especificación, en la cual se incluyen efectos fijos por profesor, el coeficiente asociado a la variable *prom\_estud* aumenta y sigue siendo significativo. Este resultado sugiere que el efecto de reciprocidad presentado propuesto por Clayson *et al.* (2006) explica la relación positiva entre la calificación de estudiantes y profesores. Al incluir efectos fijos por profesor, el porcentaje de encuestados (*porc\_encues*) deja de ser una variable significativa. Si

se supone que dentro de las características no observadas del profesor está la motivación y el interés que éste despierta en los estudiantes y se supone, además, que la variable *porc\_encues* es un indicador de la motivación, puede argumentarse que el efecto positivo de esta última variable está siendo capturado por los efectos fijos del profesor. Isely y Singh (2005) y McPherson (2006) encuentran que esta variable no es significativa.

En cuanto al ordenamiento del profesor, se encontró que los profesores asistentes obtenían una calificación ligeramente mayor que los instructores. Pero la diferencia no es significativa. Tampoco se encontraron diferencias significativas entre los coeficientes asociados a profesores asistentes, asociados y titulares. Los profesores con poca experiencia docente (instructores) resultaron peor calificados en cuanto a sus conocimientos y a las metodologías pedagógicas utilizadas en comparación con los más experimentados (asistentes, asociados e instructores). Pero estos últimos obtuvieron bajas calificaciones en cuestiones prácticas (respeto hacia los estudiantes, disponibilidad de tiempo para resolver dudas y cumplimiento de los horarios). Mediante un análisis de varianza, se estudió si el ordenamiento del profesor tenía algún efecto sobre el nivel de aprendizaje que los estudiantes reportaban haber logrado en un curso (pregunta 25). El análisis no mostró diferencias significativas.

Al estimar las especificaciones que usan como variable dependiente el promedio de todas las preguntas del módulo del profesor (*g1*), los resultados no presentan grandes cambios. La significancia de la variable *num\_est* aumenta y la variable *planta* resultan significativas y tener un efecto positivo sobre la evaluación. En promedio, la calificación de un profesor de planta es 0,043 unidades mayor que la de su colega de cátedra. El cuadro 5 presenta los resultados de las regresiones.

Las especificaciones 1, 5 y 6 se estimaron nuevamente con el promedio de las preguntas del curso (*g2*) como variable dependiente. El cuadro 6 presenta los resultados para cada uno de las tres especificaciones. La columna 3 incluye efectos fijos por curso. Con el propósito de tener una base de comparación, el cuadro también presenta los resultados para la calificación del profesor (*p11*) después de incluir efectos fijos por curso (columna 4). Las columnas tres y cuatro cuentan con 835 observaciones, dado que se limitó la muestra a cursos que tuvieran dos

o más secciones; sin embargo, la muestra sigue siendo representativa para los cursos de la Universidad.

**Cuadro 5. Resultados de las regresiones para g1 por MCO.**

Variable dependiente: promedio de la pregunta 1 a la 11 (G1).							
Ecuación	1			5			
num_est	-0,002	(0,001)	***	-0,002	(0,001)	**	
prom_estud	0,217	(0,023)	***	0,228	(0,050)	***	
porc_encues	0,196	(0,070)	***	-0,070	(0,096)		
Planta	0,043	(0,021)	**				
Cbu	0,035	(0,022)		-0,026	(0,038)		
Constante	3,427	(0,109)	***	3,637	(0,217)	***	
Efectos Fijos por Profesor	NO			SÍ			***
R2	0,075			0,0617			
Obs.	1.358			898			

Errores estándar robusto en paréntesis.

\*\*\* Significancia al 99%, \*\* significancia al 95%, \* significancia al 90%.

Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

**Cuadro 6. Resultados de las regresiones para g2 por MCO.**

Variable dependiente	1		2		3		4					
	g2		g2		g2		p11					
Ecuación	1		5		6		6					
num_est	-0,001	(0,001)	**	-0,002	(0,001)	**	-0,002	(0,002)	0,000	(0,002)		
prom_estud	0,147	(0,021)	***	0,146	(0,050)	***	0,049	(0,057)	0,212	(0,081)	***	
porc_encues	0,224	(0,065)	***	-0,019	(0,093)		-0,001	(0,084)	0,021	(0,138)		
Planta	0,043	(0,019)	**									
Cbu	0,060	(0,020)	***	0,016	(0,041)							
Constante	3.665	(0,094)	***	3.896	(0,208)	***	4.2387	(0,186)	***	3.497	(0,324)	***
Efectos fijos por Profesor	NO		SÍ		***		NO		NO			
Efectos fijos por Curso	NO		NO		SÍ		***		SÍ		***	
R2	0,051		0,0352		0,0210		0,0562					
Obs	1.358		898		835		835					

Errores estándar robusto en paréntesis.

\*\*\* Significancia al 99%, \*\* significancia al 95%, \* significancia al 90%.

Fuente: Oficina de Admisiones y Registro y cálculos de los autores.

Los resultados de las columnas 1 y 2 muestran que los efectos de todas las variables sobre la calificación del curso tienen el mismo signo y una magnitud similar a la mostrada en las tablas anteriores. El coeficiente de la variable *prom\_estud* es menor cuando se usa como variable dependiente la calificación del curso que cuando se usa la calificación del profesor. Este resultado tiene un significado intuitivo, dado que el efecto de reciprocidad aplica directamente sobre quien se califica: atañe más al profesor que al curso. Los cursos electivos son estadísticamente mejor calificados. En promedio, la calificación de un curso electivo es 0,06 unidades mayor que un curso no electivo con características similares.

La columna 3 presenta los resultados de la estimación, una vez incluidos los efectos fijos por curso. Ninguna de las variables es significativa; es decir, la calificación del curso parece estar determinada por características propias, tales como la dificultad, las metodologías de enseñanza y evaluación, entre otras, y no por la calificación que obtienen los estudiantes. Si se comparan los resultados de las columnas 3 y 4, se observa que, a diferencia de la calificación del curso, la del profesor sí está explicada por las calificaciones de los estudiantes. Una vez más se comprueba que los estudiantes tienen en consideración la calificación obtenida en la evaluación de los profesores pero no parecen tenerla en cuenta en la evaluación de los cursos.

## V. Conclusiones

Este trabajo presenta una serie de resultados que indican la existencia de una relación positiva entre las calificaciones de estudiantes y profesores. Los resultados sugieren, al mismo tiempo, que esta relación se explica por un efecto de reciprocidad. Los estudiantes parecen usar la calificación recibida, sin tener en cuenta su esfuerzo personal, como factor determinante de la evaluación de los profesores. Es decir, toman la calificación (independientemente de la dificultad del curso y del nivel del profesor) como un criterio importante al momento de evaluar a sus maestros.

Por la metodología empleada y la posibilidad de construir un panel de datos de profesor fue posible establecer que el efecto era de reciprocidad y, descartar otras hipótesis (clemencia, motivación y efectividad del

profesor) y permitir establecer que las evaluaciones de los profesores podrían estar sesgadas por la calificación de los estudiantes.

Este trabajo revela, entonces, que los profesores podrían utilizar el criterio de calificación como herramienta para obtener mejor evaluación de parte de los estudiantes. El trabajo no muestra que éste sea el caso. Pero sí presenta evidencia empírica que sugiere que la manipulación es posible. La relación positiva encontrada, similar a la reportada para el caso de otras instituciones, podría explicar parcialmente el fenómeno de inflación de notas documentado por Vallejo (2004). Sin embargo, para poder corroborar esta hipótesis de manera rigurosa, es necesario comparar los resultados antes y después de la introducción de la evaluación sistemática de los profesores de parte de los estudiantes.

El análisis muestra, además, que los estudiantes parecen no tener en cuenta la nota obtenida en su evaluación de los contenidos y la calidad de los cursos. El efecto mencionado parece estar circunscrito al caso de los profesores. Dicho análisis muestra que la calificación del profesor también está afectada por el tamaño de las clases, a pesar de que la magnitud del efecto es pequeña. No se encontró que el tipo de vinculación del profesor (cátedra o planta) fuera determinante de la calificación que asignan los estudiantes. Se encontró que cada grupo tenía fortalezas en diferentes aspectos. Estos resultados no sugieren que el efecto sobre el aprendizaje o la efectividad de profesores de planta o cátedra sea el mismo; simplemente muestra que la percepción promedio que tienen los estudiantes es similar. Del mismo modo, el ordenamiento del profesor resultó no ser determinante en la calificación asignada por los estudiantes.

También, se encontró que las características del curso, tales como el tamaño de la clase, el porcentaje de estudiantes que responde a la evaluación y el nivel (introductorio, de mitad de carrera o de énfasis), explican el resultado de las evaluaciones a profesores. El eventual carácter electivo de un curso parece no ser determinante. Al utilizar como variable dependiente la calificación global del profesor y el promedio de las preguntas del módulo de éste, los resultados no presentan cambios considerables.

Finalmente, este trabajo sintetizó una amplia discusión sobre las evaluaciones a profesores, la cual debería ser tenida en cuenta para futu-

ras revisiones del sistema. Se plantea que se deben imponer reglas y controles estrictos sobre la estrategia de aplicación de las evaluaciones (p. ej.: no permitir que sean llevadas a cabo durante o después de un examen); así mismo, se debe definir claramente el uso que se le va a dar a la información y se debe al mismo tiempo persuadir a la comunidad universitaria sobre la importancia que tienen las evaluaciones en el proceso de mejoramiento de la calidad de la educación.

## Referencias

- ANGRIST, J. and LAVY, V. (1999). "Using maimonides' rule to estimate the effect of class size on children's academic achievement", *Quarterly Journal of Economics*, 114(2):533-576.
- BACON, D. R. and NOVOTNY, J. (2002). "Exploring achievement striving as a moderator of the grading leniency effect", *Journal of Marketing Education*, 24(1):4-14.
- BONESRØNNING, H. (2004). "Can effective teacher behavior be identified?", *Economics of Education Review*, 23(3):237-247.
- CASHIN, W. E. (1995). "Student ratings of teaching: The research revisited", *IDEA Paper No.32*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- CLAYSON, D. E.; FROST, T. F. and SHEFFET, M. J. (2006). "Grades and the student evaluation of instruction: A test of the reciprocity effect", *Academy of Management Learning & Education*, 5(1):52-65.
- D'APOLLONIA, S. and ABRAMI, P. C. (1997). "Navigating student ratings of instruction", *American Psychologist*, 52(11):1198-1208.
- ESTRIN, H. A. (1958). "Measuring the quality of the teaching of english", *College Composition and Communication*, 9(2):111-113.

- FREY, P. W. (1976). "Validity of student instructional ratings: Does timing matter?", *The Journal of Higher Education*, 47(3):327-336.
- GAVIRIA, A. (2002). *Los que suben y los que bajan*. Bogotá, Fedesarrollo.
- GILES, O. and PASCOE, R. (2004). "Student evaluations; fact or fiction", in *Eastern Institute of Technology*. Consultado el 4 de septiembre de 2007 de: [http://www.naccq.ac.nz/conference05/proceedings\\_04/giles.pdf](http://www.naccq.ac.nz/conference05/proceedings_04/giles.pdf)
- GLASS, G. V. (2004). *Teacher evaluation. Policy brief*. Arizona State University College of Education, Education Policy Studies Laboratory, Tempe, AZ.
- GREENWALD, A. G. (1997). "Validity concerns and usefulness of student rating of instruction", *American Psychologist*, 52(11):1182-1186.
- HAIDER, G. (1976). "An evaluative-diagnostic model of education", *JAE*, 29(3):30-31.
- HAMERMESH, D. S. and PARKER, A. M. (2003). "Beauty in the classroom: Professor's pulchritude and putative pedagogical productivity", *National Bureau Economic Research. Working Paper 9853*.
- HANUSHEK, E. A.; KAIN, J. F. and RIVKIN, S. G. (1998). "Teachers, schools, and academic achievement", *National Bureau Economic Research. Working Paper 6691*.
- HOWARD, G. S. and MAXWELL, S. E. (1980). "The correlation between grades and student satisfaction: a case of mistaken causation?", *Journal of Educational Psychology*, 72(6):810-820.
- HOWARD, G. S. and MAXWELL, S. E. (1982). "Do grades contaminate student evaluations of instruction?", *Research in Higher Education*, 16(2):175-188.

- ISELY, P. and SINGH, H. (2005). "Do higher grades lead to favorable student evaluations?", *Journal of Economic Education*, 36(1):29-42.
- JACOB, B. A. and LEVITT, S. D. (2002). "Rotten apples: An investigation of the prevalence and predictors of teacher cheating", *National Bureau Economic Research. Working Paper 9413*.
- KRAUTMANN, A. C. and SANDER, W. (1999). "Grades and student evaluations of teachers", *Economics of Education Review*, 18(1):59-63.
- KUPERMINTZ, H.; SHEPARD, L. and LINN, R. (2001). *Teacher effects as a measure of teacher effectiveness: Construct validity considerations in TVAAS (Tennessee Value Added Assessment System)*. Seattle, WA., University of Colorado at Boulder. Symposium at the National Council on Measurement in Education (NCME) Annual Meeting.
- LANGBEIN, L. (2007). "Management by results: Student evaluation of faculty teaching and the mis-measurement of performance", *Economics of Education Review*, doi:10.1016/j.econedurev.2006.12.003.
- MARSH, H. W. and ROCHE, L. A. (1997). "Making student's evaluations of teaching effectiveness effective", *American Psychologist*, 52(11):1187-1197.
- MCKEACHIE, W. J. (1997). "Student ratings. The validity of use", *American Psychologist*, 52(11):1218-1225.
- MCPHERSON, M. A. (2006). "Determinants of how students evaluate teachers", *Journal of Economic Education*, 37(1):3-20.
- NERENZ, A. G. and KNOP, C. K. (1982). "A time-based approach to the study of teacher effectiveness", *The Modern Language Journal*, 66(3):243-254.

- O'HANLON, J. and MORTENSEN, L. (1980). "Making teacher evaluation work", *The Journal of Higher Education*, 51(6):664-672.
- RUEDA, M. R. (2006). "Efecto de la exigencia en calificación sobre el desempeño académico. Estudio de los cursos de Matemáticas básicos en la Universidad de los Andes", *Documento CEDE 39*, Centro de Estudios sobre Desarrollo Económico.
- SHINKFIELD, A. J. and STUFFLEBEAM, D. L. (1995). *Teacher evaluation: Guide to effective practice. Evaluation in education and human services*. Norwell, MA., Kluwer Academic Publishers.
- SIMPSON, R. H. (1967). "Evaluation of college teachers and teaching", *Journal of Farm Economics*, 49(1):286-298.
- TEVEN, J. J. and MCCROSKEY, J. C. (1996). "The relationship of perceived teacher caring with student learning and teacher evaluation", *Education Resources Information Center*. Consultado el 28 de mayo de 2007 de: [http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/16/92/f8.pdf](http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/16/92/f8.pdf)
- VALLEJO, H. (2004). "Evolución reciente del sistema de notas de pregrado en la Universidad de los Andes y la Facultad de Economía, con una propuesta para el futuro", *Documento CEDE 04*. Centro de Estudios sobre Desarrollo Económico.
- WRIGHT, S. P.; HORN, S. P. and SANDERS, W. L. (1997). "Teacher and classroom context effects on student achievement: Implications for teacher evaluation", *Journal of Personnel Evaluation in Education*, 11(1): 57-67.
- WRIGHTSTONE, J. W. (1943). "Evaluation of an activity program", *Journal of Educational Sociology*, 17(2):114-124.

## Anexo 1. Cuestionario de la evaluación a profesores y programas.

Cada una de las preguntas es calificada por los estudiantes en una escala de 1 a 5.

1. Total desacuerdo; 2. En desacuerdo; 3. Ni de acuerdo ni en desacuerdo; 4. De acuerdo; 5. Total acuerdo; NA. No aplica.

	#	Pregunta	1	2	3	4	5	NA
Profesor	1	El profesor demuestra conocimiento adecuado de los temas del curso						
	2	Las presentaciones y explicaciones del profesor son claras						
	3	El profesor motiva el interés por el curso						
	4	Las metodologías pedagógicas usadas por el profesor son adecuadas para los objetivos del curso						
	5	Las metodologías pedagógicas usadas por el profesor fomentan una participación activa del estudiante						
	6	La relación del profesor con los estudiantes es de mutuo respeto						
	7	El profesor cumple adecuadamente los horarios y calendario del curso						
	8	El profesor atiende consultas del curso en horarios de atención						
	9	El sistema de calificaciones ha sido claramente establecido desde el inicio del curso	SÍ		NO			
Curso	10	Las calificaciones son entregadas oportunamente						
	11	La calificación global del profesor es:						
	12	Los objetivos del curso han sido claramente establecidos						
	13	El programa desarrolla adecuadamente los objetivos del curso						
	14	El profesor cumple con el programa del curso						
	15	El curso aporta de manera significativa al desarrollo de su propia capacidad crítica y/o de análisis						
	16	El curso aporta de manera significativa al desarrollo de su propia creatividad y/o capacidad de síntesis						
	17	La calificación global del curso es:						



Evaluaciones	18	Las evaluaciones son consistentes con los objetivos del curso						
	19	Las evaluaciones fomentan el aprendizaje						
	20	El material docente es adecuado para el programa del curso						
	21	El material docente está disponible para el estudiante						
	22	El monitor tiene la preparación adecuada para las funciones que se le asignan en el curso						
Autoevaluación	23	El monitor está disponible para las funciones que se le han asignado						
	24	Califique el esfuerzo que usted ha dedicado a este curso						
	25	Califique el nivel de aprendizaje que usted ha logrado en este curso						
	26	Califique su participación en clase						
	27	Califique su asistencia a clase						

28. A partir de los aspectos calificados *muy positivamente*, comente los factores y criterios tomados en cuenta.

29. A partir de los aspectos calificados con 1 ó 2, comente los factores y criterios tomados en cuenta.

