



ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

XLIX Reunión Anual

Noviembre de 2014

ISSN 1852-0022

ISBN 978-987-28590-2-2

DETERMINANTES DE CÓMO LOS
ESTUDIANTES EVALÚAN A LOS DOCENTES: EL
CASO DE LA FACULTAD DE CIENCIAS
ECONÓMICAS DE LA UNIVERSIDAD CATÓLICA
ARGENTINA

Ciocchini, Francisco
Rabassa, Mariano
Sabbioni, Guillermo

Determinantes de Cómo los Estudiantes Evalúan a los Docentes: El Caso de la Facultad de Ciencias Económicas de la Universidad Católica Argentina¹

Francisco J. Ciocchini – Facultad de Ciencias Económicas, UCA

Mariano J. Rabassa – Facultad de Ciencias Económicas, UCA y CONICET

Guillermo S. Sabbioni – Facultad de Ciencias Económicas, UCA

Abstract: *Students' evaluations of courses and their teachers have become an important tool used by university administrators to assess the effectiveness of teachers. This paper analyzes the determinants of students' evaluations for the undergraduate courses taught in the School of Economics at Universidad Católica Argentina. Using more than 47 thousand evaluations, corresponding to 1668 courses, we find that class size, the presence of more than one professor by course, teachers' experience, and its educational attainment are all relevant characteristics to explain students' evaluations.*

Resumen: *Las evaluaciones que los estudiantes realizan sobre los cursos y los profesores que los dictan son una herramienta importante para evaluar el desempeño de los docentes. En este trabajo analizamos los determinantes de dichas evaluaciones para el caso de los cursos de grado ofrecidos por la Facultad de Ciencias Económicas de la Universidad Católica Argentina. En base a más de 47 mil evaluaciones correspondientes a 1668 cursos, encontramos que el tamaño del curso, la cantidad de docentes por curso, la experiencia del profesor y el nivel educativo del docente aparecen como características relevantes a la hora de explicar las evaluaciones.*

JEL: I23, J24, J33.

¹ Los autores agradecen la colaboración de Mario Karg. Información de contacto: Dr. Guillermo S. Sabbioni, Director de la Licenciatura en Economía, UCA. E-mail: gsabbioni@uca.edu.ar.

1. Introducción

Las instituciones universitarias necesitan evaluar el desempeño de sus docentes. Para ello suelen utilizar evaluaciones estandarizadas en las que los alumnos se expiden acerca de las materias que toman y los docentes que las dictan. Dichas evaluaciones son una herramienta importante para los administradores universitarios, no sólo a la hora de proveer *feedback* a los docentes, sino también al momento de asignar docentes a determinados cursos, decidir promociones u otorgar aumentos salariales (Feldman, 2007).

La validez de este instrumento de evaluación ha sido largamente discutida en la literatura, sobre todo teniendo en cuenta que existen otros mecanismos para evaluar a los docentes, como las pruebas estandarizadas a los alumnos, evaluaciones de expertos en enseñanza, y la evaluación por parte de pares, administradores universitarios o expertos en educación y pedagogía. Y si bien es claro que cada una de estas alternativas tiene sus ventajas y desventajas, existe consenso respecto de que las evaluaciones estandarizadas que los alumnos completan sobre los cursos y sus profesores son un instrumento válido para la gestión universitaria (Marsh, 2007).

Hay muchos estudios, tanto en la literatura de ciencias de la educación como del área de economía, sobre la utilidad de las evaluaciones y, especialmente, sobre los determinantes de una buena evaluación. Sin embargo, existe muy poca evidencia --nula a nuestro entender-- para la Argentina. En este estudio utilizamos más de 47 mil encuestas correspondientes a cursos de grado ofrecidos por la Facultad de Ciencias Económicas de la Pontificia Universidad Católica Argentina (UCA) durante un período de cinco semestres con el fin de analizar los determinantes de una buena evaluación. En particular, el estudio se concentra tanto en las características del docente (experiencia, género, nivel educativo alcanzado, etc.) como en las características de los cursos (tamaño, número de docentes, carácter electivo, etc.).

Nuestro trabajo se destaca en al menos dos puntos con respecto a la literatura existente. En primer lugar, el diseño de la encuesta que responde el alumnado de la UCA permite reducir varios sesgos presentes en la inmensa mayoría de los estudios. En segundo lugar, las características curriculares de la Universidad permiten también reducir errores de medición en variables claves, como por ejemplo, el tamaño del curso.

Los resultados muestran un efecto negativo y significativo, aunque de baja magnitud, del tamaño del curso sobre las evaluaciones docentes. También se observa una diferencia en cómo los alumnos perciben a las distintas carreras de la Facultad, y un impacto significativamente negativo cuando el curso es dictado por más de un docente. No se aprecian diferencias en cuanto el género del docente, aunque sí en cuanto a su nivel educativo máximo alcanzado, dado que tener un doctorado aumenta significativamente las evaluaciones obtenidas.²

En la siguiente sección se revisa la literatura referida a las evaluaciones docentes por parte del alumnado. La tercera sección presenta detalladamente la base de datos y provee estadísticas descriptivas básicas que sirven para motivar los modelos econométricos de la cuarta sección. La

² Claramente el título de doctor en sí mismo no es el responsable del impacto positivo en las evaluaciones, sino más bien sirve como un indicador de los conocimientos y motivación del docente.

quinta sección presenta los principales resultados. El trabajo concluye con una discusión y recomendaciones de gestión universitaria.

2. Literatura relacionada

En general, la literatura referida a las evaluaciones docentes por parte de los alumnos se centra en dos cuestiones. En primer lugar, una vasta fracción, especialmente en el área de ciencias de la educación, discute la validez de este instrumento como medida del desempeño docente y como instrumento de gestión educativa. En segundo lugar, una serie de estudios empíricos, más cercanos al espíritu de este trabajo, estudia los determinantes de una buena evaluación docente. A continuación describimos en líneas generales ambos tipos de trabajos.

2.1 Validez de las encuestas como mecanismo de evaluación docente e instrumento de gestión educativa

La utilización de encuestas estandarizadas a los alumnos como mecanismo de evaluación docente e instrumento de gestión educativa ha despertado polémicas desde sus orígenes.³ En esta sección discutimos brevemente esta cuestión siguiendo a Aleamoni (1999), que resume la discusión a través de una lista de dieciséis “mitos” relacionados con este tipo de encuestas y los contrapone con los “hechos” reportados en la literatura empírica sobre el tema.

Una primera crítica es que, debido a su inmadurez y falta de experiencia, los estudiantes no son capaces de hacer juicios consistentes sobre sus docentes (mito 1); o que solamente podrían emitir juicios razonables después de cierto tiempo -probablemente años- de finalizado el curso (mito 4). Adicionalmente, se suele argumentar que las encuestas a los estudiantes no son más que concursos de popularidad entre los docentes, en los que los profesores más “amigables” resultan siempre ganadores (mito 3). Peor aún, es común escuchar que los docentes tienden a “sobornar” a los alumnos con mayores notas a la espera de cierta reciprocidad en las encuestas (mito 13). Como consecuencia, algunos creen que los docentes solamente pueden ser evaluados por pares con vasta experiencia en el aula y excelentes publicaciones académicas (mito 2).

Dos características deseables de cualquier método de evaluación son su fiabilidad y validez. En el caso de las encuestas docentes, una alta (baja) fiabilidad correspondería a una baja (alta) disparidad de opiniones sobre un mismo profesor entre los estudiantes de un mismo curso. Por su parte, un alto (bajo) grado de validez se correspondería con una alta (baja) correlación entre los resultados de las encuestas y los de métodos alternativos de evaluación de la efectividad docente (por ejemplo, exámenes estandarizados a los alumnos al finalizar el curso, evaluación por observadores especialmente entrenados, etc.). Una de las críticas más importantes a las encuestas estandarizadas es que no son un instrumento fiable y válido para evaluar la efectividad docente (mito 5).

³ Uno de los primeros trabajos sobre el tema fue el de Heilman y Armentrout en 1936.

También se suele argüir que ciertos factores afectan “injustamente”, o en forma “sesgada”, las evaluaciones recibidas por los docentes.⁴ Entre ellos se cuentan: el tamaño del curso –a mayor tamaño, peores evaluaciones- (mito 6);⁵ el género del profesor y/o del alumno (mito 7); la hora del día en que se dicta el curso (mito 8); el año de la carrera en que se encuentra el alumno (mito 11); si la materia es obligatoria u optativa (mito 9); la disciplina -matemática, arte, negocios, etc.- a la que corresponde la materia (mito 14); si la materia corresponde al campo principal (*major*) o a un campo complementario (*minor*) de estudio (mito 10); la categoría del docente –asistente, pro-titular, titular- (mito 12).

Finalmente, se suele afirmar que darle información (*feedback*) a los docentes sobre los resultados de las encuestas no tiene efectos sobre la calidad de la instrucción (mito 16). En particular, la utilidad de las encuestas es menor cuando dichos resultados se basan en respuestas a preguntas de carácter global (sobre el curso o sobre el docente), en lugar de provenir de encuestas con múltiples ítems - claridad del instructor, interacción con los estudiantes, organización del curso, etc.- (mito 15).

Después de revisar la literatura empírica producida entre 1924 y 1998, Aleamoni (1999) concluye que los mencionados “mitos” son efectivamente “mitos”, no respaldados por los “hechos”. Según este autor, las investigaciones más recientes respaldan el uso de encuestas estandarizadas, en combinación con un *feedback* apropiado hacia los docentes, como mecanismo útil tanto para el mejoramiento de la calidad educativa como para la toma de decisiones administrativas.⁶

Otros autores, como Marsh y Roche (1997) y Marsh (2007) obtienen conclusiones similares a las de Aleamoni (1999).⁷ En particular, Marsh y Roche (1997) resumen sus conclusiones de la siguiente manera: “[B]ajo condiciones apropiadas, las encuestas docentes estandarizadas son (a) multidimensionales;⁸ (b) fiables y estables; (c) dependen fundamentalmente de las características del docente y no del curso dictado; (d) son relativamente válidas en comparación con una variedad de indicadores de efectividad docente; (e) no se ven demasiado afectadas por variables que podrían, hipotéticamente, introducir sesgos (por ejemplo, lenidad en las notas, tamaño del curso, carga de trabajo durante el curso, interés previo en la materia); y (f) útiles para mejorar la efectividad educativa cuando se las complementa con *feedback* apropiado al docente.”⁹

⁴ En esta literatura se dice que hay sesgo (*bias*) cuando una característica del estudiante, del docente o del curso afecta las evaluaciones recibidas por el docente (ya sea positiva o negativamente) sin estar relacionada con los determinantes de la buena docencia, como el aprendizaje logrado por los alumnos (Centra, 2003).

⁵ Nótese que la existencia de correlación entre el tamaño del curso y las evaluaciones docentes no refleja necesariamente un sesgo, en el sentido definido anteriormente. En particular, si una clase de menor tamaño facilita el aprendizaje a través de una relación más personal entre el docente y los alumnos, entonces no diríamos que el tamaño de la clase está introduciendo un sesgo.

⁶ Recientemente, Taylor y Tyler (2012) muestran que en el caso de docentes del nivel inicial, las evaluaciones combinadas con *feedback* pueden mejorar el esfuerzo y las capacidades del docente, no solo en el corto plazo sino también en el largo plazo.

⁷ Para una visión discordante, véase Becker, Bosshardt y Watts (2011).

⁸ Los cuestionarios incluyen múltiples ítems sobre el docente y el curso.

⁹ Traducción de los autores.

2.2 Determinantes de la evaluación docente

Más allá de su validez como instrumento de gestión universitaria, lo cierto es que este tipo de evaluaciones estandarizadas son intensivamente utilizadas por la gran mayoría de los establecimientos educativos. Por ejemplo, Becker y Watts (1999) reportan datos de una encuesta a los directores de los departamentos de economía de universidades de los Estados Unidos donde se muestra que aproximadamente el 83% de los departamentos evalúa a sus docentes utilizando—casi en forma exclusiva— encuestas estandarizadas a los alumnos.¹⁰ Al respecto, parece apropiado preguntarse cuáles son los determinantes de una buena evaluación docente.

Los estudios empíricos suelen focalizarse en dos grupos de factores: aquellos relacionados con el tipo de curso y aquellos relacionados con las características docentes. Dentro del primer grupo el mayor esfuerzo se observa en inferir el impacto del tamaño de la clase sobre la percepción que los alumnos tienen sobre el desempeño del docente. En general, se concluye que los estudiantes en cursos más grandes tienden a calificar menos satisfactoriamente a los docentes. La interpretación que se desprende es que en los grupos más grandes no es posible una adecuada interacción docente-alumno. Por ejemplo, Gaviria y Hoyos (2008) encuentran que para el caso de la Universidad de Los Andes, Colombia, el aumento del tamaño del curso repercute negativamente sobre las calificaciones que reciben los docentes, aunque dicho efecto es de magnitud reducida. Un resultado similar es reportado por McPherson (2006) e Illanes y Sapelli (2012), para el caso de la Universidad de North Texas, EEUU, y la Universidad Católica de Chile, respectivamente. El trabajo de Illanes y Sapelli es particularmente interesante porque cuantifican un potencial *trade-off* entre reducir el tamaño de los cursos -abriendo nuevos cursos- pero colocando a docentes con menos experiencia al frente de ellos. Su conclusión es que el beneficio de reducir el tamaño del curso (de baja magnitud) no compensaría a los costos de utilizar docentes más inexpertos.

El tratamiento del tamaño del curso en los estudios empíricos es de por sí complejo. En primer lugar, la existencia de “clases magistrales” -donde un docente experimentado dicta clases a un grupo grande de alumnos, los cuales luego se separan en grupos más pequeños con docentes de menor rango- necesariamente sesga el efecto del tamaño del curso, ya que el tamaño oficialmente registrado es el de los grupos pequeños. En segundo lugar, la posibilidad que tienen los alumnos de elegir qué cursos tomar, tal es el caso en las universidades en los Estados Unidos, introduce endogeneidad entre las evaluaciones y el tamaño de los cursos.

Además de estimar el impacto del tamaño, otras características de los cursos han recibido atención en la literatura, por ejemplo, la diferentes calificaciones que supuestamente reciben los cursos de diferentes carreras (mito 14). La evidencia no es concluyente.

Por otra parte, entre las características de los docentes se destaca entre los estudios empíricos aquellos que buscan mostrar que los docentes pueden influir sobre las evaluaciones que reciben de sus alumnos reduciendo la exigencia del curso u otorgando notas más altas (fenómeno denominado *reciprocidad*). La evidencia se basa generalmente en correlaciones observadas entre las notas de los

¹⁰ En un trabajo reciente que pretende actualizar estos resultados, Becker, Bosshardt y Watts (2012) reportan que dichos porcentajes continúan siendo extremadamente elevados. Además reportan que el peso que los administradores le otorgan a este tipo de evaluaciones al momento de evaluar a los docentes es de aproximadamente el 50%, en promedio, pero que en algunos casos supera el 75%.

alumnos y las evaluaciones (véase como ejemplo Spooren y Mortelmans, 2006). Una explicación alternativa a esta correlación positiva es que los mejores docentes motivan más a sus alumnos. Un alumno más motivado es un mejor estudiante que obtiene notas más altas, quien en última instancia refleja su satisfacción otorgando calificaciones más altas a su docente. Lejos se encuentra esta situación del comportamiento estratégico –casi de soborno- que se plantea en el mito 13. El desafío es poder diferenciar ambas situaciones empíricamente. Algunos estudios econométricos, por ejemplo Isely y Singh (2005), McPherson (2008) y Gaviria y Hoyos (2008) incluyen una serie de efectos fijos de modo de tratar de aislar la respuesta correcta. Los estudios demuestran que la correlación positiva se debe a la existencia de reciprocidad.

Finalmente, la inmensa mayoría de los estudios empíricos de los determinantes de las evaluaciones deben superar el problema de sesgo de selección producto de que las encuestas son de carácter voluntario. Por lo tanto, solo aquellos alumnos con determinados incentivos completaran la encuesta. Generalmente, los estudios controlan por la proporción de alumnos que responden la encuesta o restringiendo la muestra a solo aquellos cursos con altos porcentajes de respuestas.

3. Datos

Para este estudio utilizamos 47881 encuestas completadas por los alumnos de la Facultad de Ciencias Económicas (de ahora en más, “la Facultad”) de la UCA.¹¹ Dichas encuestas corresponden a 1668 cursos ofrecidos durante cinco semestres, comprendidos entre el principio de 2011 y mediados de 2013, para los cuales la Facultad utilizó a 462 docentes.¹²

Nuestra observación está compuesta por la dupla docente-curso, ya que un profesor que dicta más de un curso (sea de la misma materia o no) recibe de parte de los directivos de la Facultad los resultados de las evaluaciones para cada uno de los cursos que dictó.

A continuación se describe la encuesta utilizada, junto con otra información proveniente de los registros oficiales de la Universidad. Posteriormente se presentan estadísticas descriptivas.

¹¹ La base de datos con la que contamos corresponde a las carreras de: Licenciatura en Administración de Empresas, Contador Público, Licenciatura en Economía y Organización Bancaria. Las primeras tres carreras corresponden a un plan de estudios de cuatro años mientras que en el caso de Organización Bancaria, los datos corresponden tanto a la Tecnicatura (carrera de tres años) como a la Licenciatura (carrera de cuatro años).

¹² Se eliminaron todas las observaciones correspondientes a los Talleres de las carreras de Contador Público y Licenciatura en Administración de Empresas (por ejemplo, Taller de Desarrollo de Habilidades Interpersonales, Taller de Presentaciones Efectivas, entre otros). Dichas observaciones no fueron analizadas en este trabajo debido a que los Talleres no corresponden a materias en el sentido tradicional de la palabra. En particular, los Talleres constan de sólo dos clases y no requieren de la aprobación de un examen final, sino que se aprueban sólo con la asistencia. Luego de eliminar las encuestas correspondientes a estos Talleres, la base de datos cuenta con 47881 observaciones.

3.1 Descripción de la encuesta

Las encuestas que realizan los alumnos son de carácter obligatorio y anónimo, mediante un formulario electrónico que debe completarse una vez finalizado el curso, pero antes de rendir el examen final de la materia.¹³

En dicha encuesta el estudiante evalúa –con un puntaje entre 1 y 5, siendo 5 el más alto- a cada uno de los docentes del curso en forma individual¹⁴, respecto a siete campos relacionados con las cualidades del docente: (i) claridad en la exposición de los contenidos, (ii) claridad en el empleo del lenguaje técnico, (iii) utilización de los ejemplos en función de la comprensión, (iv) clima de trabajo en el aula, (v) disposición ante las preguntas, (vi) relación entre el tiempo dedicado a la explicación y la complejidad del tema, y (vii) asistencia y puntualidad.¹⁵

El carácter obligatorio de la encuesta presenta una gran ventaja sobre los trabajos existentes, los cuales casi en su totalidad utilizan encuestas voluntarias. La obligatoriedad permite eliminar el sesgo de selección. En simples palabras, si las encuestas fuesen voluntarias, solamente aquellos alumnos con determinados incentivos –ya sea recompensar al buen docente o castigar un pobre desempeño- completarían la evaluación. Si bien no es posible determinar a priori la dirección y magnitud del sesgo cuando las encuestas son voluntarias, es claro que no es posible descartar su existencia.

Adicionalmente, el hecho de que las encuestas sean un procedimiento centralizado desde la administración de la Facultad, y que estas sean solicitadas al alumno luego de terminado el curso pero antes de la evaluación final, permite a su vez eliminar cualquier tipo de comportamiento estratégico por parte del profesor. En otras palabras, el profesor no tiene la capacidad de influir sobre quiénes contestan la encuesta, ya que no es él quien determina en qué momento se realiza. Asimismo, la incertidumbre existente por el lado del alumno, ya que al momento de completar la encuesta todavía no rindió el examen final de la materia, reduciría otro tipo de comportamiento estratégico usualmente reportado en la literatura, el cual consiste en que el profesor otorga buenas calificaciones en busca de reciprocidad por parte del alumnado.¹⁶ Es más, el docente tampoco conoce el resultado de las evaluaciones al momento de tomar el examen final.¹⁷ Una alternativa usualmente utilizada en los análisis econométricos consiste en controlar por las notas que los alumnos obtuvieron en el curso, lo

¹³ La obligatoriedad está implícita en el hecho de que el alumno debe contestar la evaluación docente antes de poder inscribirse para rendir exámenes finales o para cursar las materias subsiguientes.

¹⁴ El alumno evalúa a todos los profesores que tuvo en el semestre, en cada una de las materias que cursó, haya o no aprobado la cursada. La única excepción es que si en una materia determinada el alumno no tuvo al menos un 60% de asistencia, no se le pide que complete la encuesta. En el caso de los cursos correspondientes al primer semestre, dicho porcentaje mínimo de asistencia se calcula para las clases transcurridas hasta fines de mayo inclusive, mientras que para los cursos correspondientes al segundo semestre, se calcula para las clases transcurridas hasta fines de octubre.

¹⁵ Lamentablemente la encuesta no consulta a los estudiantes sobre la percepción de lo aprendido en clase ni permite realizar una autoevaluación del desempeño de los alumnos en el curso. Y si bien los alumnos tienen un campo libre para realizar comentarios, sólo en contadas ocasiones lo utilizan.

¹⁶ Claramente el alumno puede inferir qué tan exigente ha sido un docente mediante las notas que ha obtenido a lo largo del curso (por ejemplo, a través de exámenes parciales, trabajos prácticos, etc.). Sin embargo, cierto grado de incertidumbre persiste al completar la evaluación con antelación a rendir el examen final de la materia.

¹⁷ Generalmente la administración de la Facultad distribuye los resultados de las encuestas unos meses después de realizarlas.

cual en este caso equivaldría a controlar por la nota obtenida durante la cursada de la materia. Sin embargo, no contamos aún con esta información.

La misma base de encuestas permite a su vez inferir el tamaño del curso, es decir el número de alumnos que compartieron al mismo profesor en una misma aula, así como el número de profesores que dictaron el curso. Cabe notar que el tamaño del curso se determina en base a los alumnos que responden en el cuestionario, que como se mencionó es obligatorio. Puede haber, sin embargo, algunas diferencias respecto del número de inscriptos en el curso al comienzo de la clase debido a que algunos alumnos deciden abandonar el curso o no cumplen con las asistencias necesarias para poder evaluarlo. En estas circunstancias Becker y Powers (2001) sugieren que utilizar el tamaño al inicio del curso es preferible ya que así se reduce la potencial endogeneidad del tamaño del curso. Si bien es un punto válido, no creemos que sea tan relevante para nuestro caso en particular ya que la currícula en la Facultad es esencialmente fija para todos los estudiantes. En consecuencia, es poco probable que un estudiante pueda hacer *shopping* de cursos, ya que más tarde o más temprano va a tener que tomar determinado curso con determinado profesor. Igualmente, en última instancia, lo relevante para la experiencia educativa (interacción con el docente) no necesariamente es la cantidad de alumnos inscriptos sino los que realmente acudieron al aula.

Asimismo, dicha base de datos también contiene la carrera, la materia, el año de la materia en el plan de estudios, el semestre y el turno en el que se ofreció el curso que el alumno evaluó. La información permite también distinguir entre materias electivas u obligatorias. Por último, la base de datos contiene información sobre el promedio de notas obtenido por el alumno que está evaluando (previo al momento de tomar el curso que evalúa).¹⁸

Cabe destacar que en la Facultad existen muy pocas clases que se dictan bajo la modalidad de clase magistral, donde un docente concentra a varias comisiones que luego se separan para tener clases con un docente de menor rango. Esto permite eliminar el error de medición en el tamaño del curso.¹⁹

Toda esta información fue complementada con registros oficiales de la Universidad concernientes a los docentes. En particular, se logró recolectar información sobre el género del docente, su fecha de nacimiento, su titulación (grado, magister o doctorado), el cargo docente (asistente, adjunto, pro-titular y titular), y si el docente tiene además un cargo adicional en la Facultad (por ejemplo, dirección de una carrera, investigación, etc.). Si bien sería bueno contar con información acerca de la producción científica de los docentes, lamentablemente la Universidad no mantiene un registro perfectamente actualizado de dicha información.

¹⁸ Dado que las encuestas son anónimas, contar con el promedio de notas del alumno podría servir (a priori) para inferir qué tan buen estudiante es el alumno que está completando la evaluación docente.

¹⁹ La determinación exacta del tamaño del curso requirió de mucho esfuerzo, ya que los cursos que alumnos de una carrera toman en alguna otra carrera aparecen como distintos cursos en la base de datos de las encuestas. Para poder asignar correctamente a los alumnos que compartieron un mismo curso, es decir, un mismo profesor en una misma aula en el mismo momento del tiempo, se utilizó información adicional acerca de los nombramientos de profesores, que facilitaron los administradores de la Facultad.

3.2 Estadísticas descriptivas

Nuestra muestra consta de 2806 duplas curso-profesor. Para cada uno de estos pares calculamos el promedio de las evaluaciones sobre todos los alumnos del curso, para cada uno de los siete campos. La Tabla 1 muestra la media y la mediana para cada campo entre las 2806 observaciones.

Dada la altísima correlación existente entre los distintos campos de evaluación, con la excepción de “asistencia y puntualidad” (ver Tabla 2), se optó por calcular el promedio de las evaluaciones entre los siete campos. De hecho, este promedio, y no el desagregado, es el que los directivos de la Facultad envían habitualmente a los docentes.²⁰ Esta alta correlación entre los campos sugiere, en parte, el fracaso del diseño de la encuesta para capturar distintas facetas del desempeño docente.²¹ Una explicación posible indicaría que los alumnos tienen en su mente una evaluación global de la duplacurso-docente, y que la calificación que otorgan en determinado campo es una leve variación sobre dicha calificación global.

La Figura 1 muestra la distribución de la calificación total, es decir, el promedio de las calificaciones de los siete campos evaluados. La distribución está claramente sesgada hacia la izquierda, donde relativamente pocos docentes obtienen calificaciones muy bajas. El promedio de la evaluación total para los 2806 pares curso-docente es de 3.98. Solamente 6 cursos-docentes obtuvieron una calificación perfecta de parte de sus alumnos, varios de ellos en cursos de tamaño reducido.

De las 2806 duplas curso-docente, un 56% corresponde a la Licenciatura en Administración de Empresas, un 27% a la carrera de Contador Público, un 14% a Economía, y el 3% restante a Organización Bancaria. Por otra parte, el 62% de los cursos tuvieron a dos docentes a cargo y el 22% tuvo a tres docentes asignados. Solo el 16% tuvo a un sólo docente frente al curso. La cantidad de docentes en un curso ha sido largamente ignorada en la literatura como un posible factor determinante de las calificaciones que asignan los estudiantes.

Aproximadamente en el 43% de los casos el docente fue una mujer, aunque dicho porcentaje varía considerablemente según el año en que se dicte la materia en el plan de estudio. En el primer año de la carrera predominan las docentes mujeres, pero a medida que avanzan las carreras predominan los docentes hombres, llegando a representar casi un 80% de las observaciones correspondientes al cuarto y último año.²²

²⁰ La Facultad le informa al docente los resultados de sus evaluaciones como un modo de brindar *feedback* acerca de su desempeño—al menos, desde el punto de vista de los alumnos.

²¹ Alternativamente uno podría inferir que si un docente es efectivo, lo será en cada una de la facetas de la enseñanza educativa, aun en lo referido a la asistencia y puntualidad.

²² En realidad la proporción de mujeres en la plantilla de profesores de la Facultad es aún menor al 43%. De los 462 docentes que dictaron cursos en los cinco semestres bajo análisis tan solo un tercio son mujeres. La razón de esta diferencia se debe a que las mujeres tienden a dar más cursos que los hombres, especialmente, como se mencionó, en las materias de primer año.

Sobre las 2806 duplas de curso-docente, apenas en un cuarto de ellas el docente posee un título de posgrado. Al igual que en el caso de la distribución por género, los cursos en los últimos años de los planes de estudio presentan un porcentaje más alto de docentes más calificados.²³

En el presente análisis calculamos la edad del docente al inicio del curso en base a la fecha de nacimiento. El docente promedio a cargo de un curso tiene 49.3 años de edad. El panel superior de la Figura 2 muestra la distribución de edades. Se observa en el gráfico una clara disminución de docentes cuyas edades superen los 65 años.

El tamaño del curso es sin lugar a duda una de las variables más importantes en este tipo de análisis. El panel inferior de la Figura 2 muestra la distribución del tamaño del curso para el caso de los cursos ofrecidos por la Facultad. El curso promedio tiene 30 alumnos y raramente supera los 60 alumnos.²⁴ La Tabla 3 muestra el promedio de alumnos por curso discriminando por carreras. Dentro de las carreras más populares de la Facultad, los cursos de la Licenciatura en Administración de Empresas son los más grandes, seguidos por los de la carrera de Contador Público y luego por los de la Licenciatura en Economía.

¿Cómo influyen estas características en las evaluaciones que realizan los estudiantes? Antes de proceder al análisis econométrico conviene analizar las calificaciones promedio para dichas características, lo cual se muestra en la Tabla 4. En general no se observan grandes diferencias en la evaluación promedio que recibe cada característica. Dentro de la poca variabilidad quizás haya que notar las mejores evaluaciones que reciben aquellos docentes con título de doctorado en comparación con el resto de los docentes. Dicha diferencia es además estadísticamente significativa según un test de diferencia de medias. También se aprecia una diferencia en las evaluaciones que reciben los cursos de las distintas carreras. En particular, los cursos de la Licenciatura en Administración de Empresas parecen ser los más castigados por los estudiantes.

La cantidad de docentes asignados al curso también parece ser una variable para destacar. Los docentes de cursos con un solo profesor obtienen calificaciones más altas que los docentes de cursos con múltiples profesores. Esta diferencia también es estadísticamente significativa.

Para finalizar con esta sección presentamos en la Figura 3 estimaciones no-paramétricas de la relación entre la edad del docente y el tamaño del curso respecto a las evaluaciones del alumnado. El panel superior muestra una relación interesante entre la edad del docente y las calificaciones. Se aprecia una relación con forma de “u invertida” que alcanza un máximo alrededor de los 45 años. Las calificaciones para los docentes de 30 años son prácticamente iguales que las evaluaciones que obtiene un docente de 65 años. Sin embargo, es interesante notar que después de los 65 años las evaluaciones suben. Obviamente esto no se debe a una mejora en todos los docentes una vez que superan dicha edad, sino más bien a un cambio en la composición de la planta de docentes mayores.

²³ En este contexto el término “calificado” se entiende como el grado académico logrado por el docente y no por la efectividad al frente de una clase.

²⁴ Los picos en el histograma podrían estar relacionados con el hecho de que la Facultad no tiene una política establecida respecto al tamaño máximo de un curso (es decir, el tamaño a partir del cual es conveniente dividir en dos a un curso determinado). En consecuencia, al aplicarse ciertos criterios en forma errática, o con mayor o menor rigurosidad dependiendo del semestre, la distribución del tamaño del curso no se asemeja a una normal.

En particular, lo que sucede es que luego de cumplir la edad de jubilación, sólo los mejores docentes continúan dando clases.²⁵

El panel inferior de la Figura 3 muestra la relación no-paramétrica entre el tamaño del curso y las calificaciones. La relación es claramente negativa hasta los 30 alumnos por curso -exactamente el tamaño promedio de un curso ofrecido por la Facultad- aunque para tamaños de cursos mayores parece no haber una penalidad adicional.

4. Modelos econométricos

En la sección previa mostramos algunas relaciones no-condicionadas entre las calificaciones que los alumnos le asignan a la dupla curso-profesor. Claramente estas relaciones pueden estar “contaminadas” por el impacto de otras variables. Por ejemplo, los resultados sugerían que no hay una diferencia significativa entre la calificación promedio de los docentes hombres y las docentes mujeres, aunque sí existía una diferencia importante entre aquellos profesores que han alcanzado el título de doctor y aquellos que no lo han alcanzado. Sin embargo, como la inmensa mayoría de los docentes con doctorado son hombres, es probable que la relación no-condicionada entre la evaluación obtenida y el género esté sesgada en contra de las mujeres.

En esta sección desarrollamos los modelos econométricos que utilizaremos para determinar el impacto que cada una de las características del curso y las características del docente tienen sobre la evaluación docente. El punto de partida es el siguiente modelo,

$$Y_{i,c} = \alpha + \mathbf{X}\beta + \mathbf{Z}\gamma + \omega_i + \varepsilon_{i,c} \quad (1)$$

donde $Y_{i,c}$ es la calificación total que obtuvo el docente i , es decir, el promedio de los siete campos de evaluación para todos los alumnos de un curso, c . El curso es el conjunto de alumnos que estuvieron en una misma aula, el cual se determina a partir de las siguientes variables: año, semestre, carrera, materia, comisión y turno. Introducir a la carrera en la definición del curso es necesario ya que la misma materia puede ser ofrecida por más de una carrera. Claramente, un docente puede dictar (sólo o en conjunto con otro docente) más de una materia o más de una comisión de una misma materia durante un mismo semestre. En estos casos, el docente obtiene de parte de los directivos de la Facultad tantas evaluaciones totales como materias o comisiones tenga a cargo. Este hecho es crucial para nuestro estudio, ya que permite determinar con precisión el tamaño del curso y asignar la calificación correspondiente. Por el contrario, si las evaluaciones se procesaran conjuntamente en el caso que el docente dictara clases en dos comisiones, sería imposible distinguir a qué curso corresponde cada evaluación, lo cual impediría calcular el tamaño de cada uno de los cursos.

²⁵En este aspecto la Universidad debe aplicar estrictamente la legislación laboral argentina. Esto implica que una vez cumplida la edad mínima exigida por ley (diferente para hombres y mujeres) el docente tiene que jubilarse. No obstante este requisito legal, la Facultad puede, si así lo desea, invitar al docente a continuar dictando clases aun luego de la jubilación. Lógicamente, la Facultad sólo invita a continuar dando clases a sus docentes más destacados.

Los determinantes de la evaluación total $Y_{i,c}$ pueden separarse en aquellos relacionados con el curso (\mathbf{X}) y aquellos relacionados con el profesor (\mathbf{Z}). En referencia al curso, contamos con el tamaño del mismo, el cual incluimos en forma lineal y no lineal, y un conjunto de variables binarias por: carrera, el año de la materia en el programa de estudios, año calendario y semestre, el turno, si la materia es optativa u obligatoria y si el curso estuvo dictado por más de un docente. Extrañamente, el impacto que la cantidad de docentes en el curso tiene sobre las calificaciones de los estudiantes ha sido olvidado en este tipo de estudios. Cursos con más de un docente suelen tener problemas de coordinación o los docentes pueden tener estilos de enseñanza y/o evaluación muy disímiles, afectando negativamente a la experiencia educativa del alumno.

En cuanto a las características del docente, contamos con su edad al comienzo del curso, la cual también incluimos en forma lineal y no-lineal, su titulación (grado, maestría o doctorado), su rango en la escala docente de la UCA (asistente, adjunto, pro-titular o titular) y el género del docente.

Finalmente incluimos variables binarias temporales por año calendario o por semestre-año calendario, ω_t , de forma de capturar cualquier cambio en la política de la Facultad que haya afectado a todos los cursos simultáneamente, como por ejemplo la determinación del presupuesto para cada semestre.

El modelo descrito no permite controlar por factores no-observables relacionados con la habilidad intrínseca de cada docente, o su capacidad para dar clases sobre una materia en especial en los casos en que el docente dicte más de un curso. Para controlar por estos factores estimamos el siguiente modelo.

$$Y_{i,c} = \alpha + \mathbf{X}\beta + \mathbf{Z}\gamma + \omega_t + \delta_i + \varepsilon_{i,c} \quad (2)$$

donde δ_i representa en forma alternativa un efecto fijo por docente o un efecto fijo para la dupla materia-docente. La ventaja de esta especificación respecto de la anterior es obvia. Sin embargo, el efecto fijo no sólo captura la habilidad/motivación no-observada, sino que absorbe también todas aquellas características del docente o de la dupla materia-docente que no varían en el tiempo—en particular, el género del docente y su titulación.

Alternativamente se podrían utilizar estimadores de panel para controlar por factores no-observables. Sin embargo, seguir esta alternativa no sería apropiado en este contexto por la siguiente razón: el panel se define sobre la dupla curso-docente, el cual contempla entre otras cosas la comisión del curso que dictó el docente. Parece entonces poco intuitivo asumir que hay alguna característica no-observable que hace que un mismo docente dicte clases en forma diferente entre dos comisiones de una misma materia en un mismo turno. Quizás se podría redefinir la variable dependiente eliminando la comisión en su construcción, pero esto generaría un problema en cuanto a la definición del tamaño de curso para aquellos docentes que hayan dictado la misma materia en más de una comisión durante el mismo turno.

5. Resultados

La Tabla 5 muestra distintas especificaciones relacionadas con el modelo presentado en la ecuación (1). En la primera columna, tanto el tamaño del curso como la edad del docente se introducen en forma lineal. El tamaño del curso tiene un impacto negativo y altamente significativo sobre la calificación total que recibe el docente. Cada alumno adicional en el aula reduce la evaluación docente en 0.0044 (o equivalente, un aumento de un desvío estándar en el tamaño del curso reduce la calificación en una décima parte de un desvío estándar). Este resultado está en línea con la opinión habitual de los docentes, quienes suelen expresar su descontento cuando deben dictar clases a un grupo numeroso de alumnos. De todos modos, y no obstante la significancia estadística encontrada, el impacto del tamaño del curso es de baja magnitud, lo cual es consistente con lo reportado en la literatura.

Un resultado interesante es que los cursos que son dictados por más de un profesor obtienen calificaciones peores por parte de los alumnos. El coeficiente indica que poner a más de un docente en un curso reduce las evaluaciones en 0.16 (aproximadamente un tercio de un desvío estándar). Este impacto no es alto en magnitud pero resulta elevado en comparación con el efecto del tamaño del curso. Este resultado merece ser destacado, debido a que no había sido estudiado previamente en la literatura, y también debido a que podría tener implicancias desde el punto de vista de la gestión. Se presume que la penalidad en la evaluación de los docentes que comparten un curso puede estar asociada a dificultades de coordinación entre ellos, que son percibidas en forma negativa por los estudiantes. De ser así, y a menos que dicha falta de coordinación pudiera ser corregida, quizás sería deseable evitar la presencia de más de un profesor en el curso, en virtud del impacto negativo que esto tiene en la percepción de los alumnos. Sin embargo, y no obstante la mejora que esto significaría en términos de la evaluación, esto podría convertirse en un problema para aquellas universidades que utilizan masivamente profesores a tiempo parcial.²⁶

La Tabla 5 también muestra que si bien las materias optativas tienen un impacto positivo sobre la calificación, tal cual es de esperar dado que los alumnos revelan sus preferencias al elegir dichos cursos, ese impacto no es estadísticamente significativo. A su vez, los resultados indican que los cursos ofrecidos en el segundo semestre reciben peores evaluaciones, en promedio, respecto a los cursos que se dictan durante la primera parte del año. Esto puede deberse al cansancio de los alumnos hacia el final del año lectivo, especialmente cuando responden la encuesta de evaluación (generalmente en diciembre), o a las características especiales que puedan tener los cursos en dicho semestre.

No parece haber una diferencia en las evaluaciones que reciben los cursos en los tres turnos que tiene la Facultad: mañana, tarde y noche, aunque sí se observan diferencias por carrera. Respecto a la carrera omitida, Licenciatura en Administración de Empresas, las carreras de Licenciatura en Economía y Contador Público reciben evaluaciones más bajas por parte de los alumnos.²⁷ Este resultado sugiere que hay factores referidos a la dupla curso-docente (no incluidos en esta regresión)

²⁶ En estos casos, contar con dos o más profesores en el curso atenúa el riesgo de que se pierdan clases por la ausencia imprevista de uno de ellos (debido a viajes laborales, por ejemplo), pero a un costo en términos de evaluaciones.

²⁷ De por sí este resultado es interesante ya que la media no-condicionada de las evaluaciones por carrera era precisamente más baja para la Licenciatura en Administración de Empresas que para las otras carreras ofrecidas por la Facultad. El resultado de la Tabla 5 muestra entonces que las otras carreras tienen atributos, del curso y del profesor, más deseables para los alumnos.

que están distribuidos en forma diferente entre las carreras, o simplemente a que los alumnos de las carreras peor evaluadas son más exigentes o están menos satisfechos.

Con respecto a las características de los docentes, los resultados en la primera columna de la Tabla 5 sugieren que los docentes de mayor edad obtienen menores calificaciones, aunque esto puede enmascarar cualquier tipo de relación no-lineal entre la edad del docente y su desempeño, tal cual lo evalúan los alumnos. Cada año adicional del docente reduce sus calificaciones en 0.003; una magnitud de por sí muy baja.

En esta primera especificación, los docentes hombres obtienen calificaciones más bajas, aunque la significatividad aparece recién al 10%. Este resultado puede deberse a algún sesgo (discriminación) de parte de los alumnos (ver Sección 2) o a alguna variable omitida que se encuentre correlacionada con el género del docente.²⁸ Por último, aquellos docentes con título de doctorado reciben consistentemente mejores evaluaciones de parte de sus alumnos. El coeficiente es similar en magnitud al impacto de tener un sólo docente en el curso. El efecto positivo de tener un doctorado puede estar relacionado con mayores conocimientos por parte del docente como así también con una mayor motivación para las tareas académicas.²⁹

La segunda especificación de la Tabla 5 introduce no-linealidades en el tamaño del curso y en la edad del docente. El tamaño del curso sigue presentando un efecto negativo y es además casi lineal. La edad del docente presenta una forma de “u-invertida” con un máximo alrededor de los 50 años. Ambos resultados confirman las formas funcionales presentadas en la Figura 3, aunque el pico de evaluaciones se da aproximadamente 5 años antes en la estimación no-paramétrica.³⁰ El resto de los coeficientes, con excepción del género del docente, mantienen su significatividad y su magnitud.

Las últimas tres columnas de la Tabla 5 muestran tres especificaciones para distintas sub-muestras. En la primera se eliminan los cursos con menos de cinco alumnos, en la segunda se eliminan las materias optativas y en la última se eliminan los cursos correspondientes a la Licenciatura y la Tecnicatura en Organización Bancaria. Los resultados no varían significativamente respecto a los reportados en la columna 2.

Finalmente, de la Tabla 5 queda por destacar el bajo poder predictivo de estos modelos. Todas las variables incluidas apenas explican conjuntamente el 10% de la variabilidad de las evaluaciones. Esto es de esperarse ya que en estas especificaciones no se controla por la habilidad, motivación o conocimientos del docente para dar la clase, ni por otras características no observables de la materia o del docente.

²⁸ Algunas variables omitidas que pueden estar correlacionadas con el género del docente pueden ser: la severidad con la que los docentes evalúan a sus alumnos, la amabilidad en el trato, la motivación, etc.

²⁹ Para analizar con más profundidad el impacto de tener un doctorado incluimos interacciones entre esa variable y las distintas carreras, en cada una de las especificaciones de la Tabla 5. El argumento usualmente escuchado es que el doctorado solamente es útil en la Licenciatura en Economía, pero no en las restantes carreras de la Facultad -de hecho la mayor parte de los profesores con doctorado dictan cursos de economía. Los resultados (no reportados) no sugieren que tener un doctorado tenga un impacto diferencial por carrera.

³⁰También se probó con una *spline* para ambas variables y se obtuvieron resultados similares (no reportados).

La Tabla 6 presenta los resultados del modelo presentado en la ecuación (2), el cual incluye efectos fijos por docente, materia o ambos. La primera columna muestra los resultados de incluir *dummies* por docente. La capacidad predictiva de este modelo aumenta considerablemente respecto de la reportada en la Tabla 5. Las variables de tamaño del curso prácticamente no se alteran, aunque el efecto negativo de tener más de un profesor se atenúa en un 40%.

La segunda columna reporta el modelo con efectos fijos por materia. Vale la pena notar que la inclusión de no-observables por materia mejora relativamente poco la bondad de ajuste del modelo. Gran parte de la variabilidad en las evaluaciones permanece sin ser explicada. Una vez que se incluyen *dummies* por materia el efecto negativo del segundo semestre desaparece, sugiriendo que la penalización que se observaba en la Tabla 5 y en la primera columna de la Tabla 6 se debía al tipo de cursos dictados en cada semestre y no al efecto del cansancio de los alumnos.

Resulta interesante notar que cuando se incluyen los efectos fijos por materia, el género del docente vuelve a ser significativo, pero esta vez al 5%. Nuevamente, los hombres obtienen calificaciones más bajas que las mujeres. Finalmente, la última columna de la Tabla 6 introduce efectos fijos por materia-docente. Los resultados no varían sustancialmente de los reportados en las especificaciones anteriores.

Debe reiterarse que en los resultados de este estudio no está presente un sesgo que muchas otras investigaciones han tenido que enfrentar: las encuestas que dan lugar a nuestros resultados son obligatorias, de manera que las conclusiones en este estudio no están influidas por la opinión mayoritaria de los alumnos que tienen una percepción predominantemente negativa o positiva del docente.

6. Discusión final

Optimizar la gestión de los recursos es clave en cualquier organización, pero más aún en las instituciones educativas. Esto es así debido a que los colegios y universidades contribuyen a la formación del capital humano que será, en parte, responsable de nuestro crecimiento económico en el futuro. En virtud de esto, la opinión de los estudiantes acerca del desempeño de los docentes en el proceso de enseñanza debe ser tenida en cuenta, a la vez que entendemos cuáles son los factores que influyen en la percepción que tienen los alumnos.

En este contexto, y si bien puede ser riesgoso utilizar exclusivamente las evaluaciones estandarizadas de parte de los alumnos -tal como advierten Becker y Watts (1999), Becker (2000) y Becker, Bosshardt y Watts (2011)- la validez de este instrumento no puede ser ignorada. Los departamentos de economía y de muchas otras disciplinas alrededor del mundo continúan consultando a sus alumnos acerca de su percepción sobre el desempeño de sus docentes. En consecuencia, y dado que probablemente éste siga siendo un método ampliamente utilizado a nivel universitario, e incluso secundario, comprender cuáles son los determinantes de dichas evaluaciones es importante.

En este sentido, los resultados del presente estudio aportan información útil para la gestión de los recursos académicos. En primer lugar, una sugerencia que se desprende de nuestros resultados es que para mejorar la experiencia del estudiante respecto de su proceso de aprendizaje, reducir la

cantidad de docentes a sólo uno por curso equivale a reducir el tamaño del curso en aproximadamente 12 alumnos. En otras palabras, la eliminación de los eventuales problemas de coordinación entre docentes compensaría las dificultades asociadas a tener esa cantidad de alumnos más en el aula. En consecuencia, dado que reducir el número de docentes en un mismo curso no requeriría de una inversión adicional de recursos económicos -a diferencia de la reducción en el tamaño de los cursos, lo cual requiere abrir más cursos- nuestros resultados sugieren la existencia de oportunidades de mejorar la eficiencia.³¹

Sin embargo, este *trade-off* entre el número de docentes y el tamaño del curso podría ser costoso en universidades que se apoyan fuertemente en el uso de profesores *part-time*, o cuya actividad principal se encuentra por fuera de la docencia. En estos casos, si la presencia de un sólo profesor por curso incrementa el riesgo de ausentismo por motivos inesperados (por ejemplo, un viaje de trabajo), esto podría anular el efecto positivo derivado de la ausencia de problemas de coordinación. En este sentido, recordemos que el campo Asistencia y Puntualidad, dentro de los siete campos evaluados por los alumnos de la Facultad, era el que presentaba una menor correlación con los demás, lo cual sugiere que la evaluación global del profesor se ve afectada por la percepción de los alumnos en términos de asistencia, con lo cual incrementar el riesgo de ausencias no sería recomendable.

Para finalizar, quizás es valioso relacionar nuestros resultados con los de Becker y Watts (2001), quienes analizan cuáles son los métodos de enseñanza comúnmente utilizados en la ciencia económica. Estos autores se refieren a un estudio previo (Becker y Watts, 1998), en el cual mostraban que los profesores de economía son menos propensos a utilizar métodos alternativos a la tradicional exposición (o *lecture*), a la vez que señalan que los profesores de economía obtienen peores evaluaciones que los profesores en otras disciplinas. Según Becker y Watts (2001), en años recientes surgieron varios estudios acerca del problema de la enseñanza de la economía y sobre cómo utilizar métodos pedagógicos alternativos. Sin embargo, estos autores muestran cierto escepticismo, preguntándose si el interés adicional acerca de cómo se enseña la economía ha ocasionado un cambio efectivo a la hora de dictar clases. Con este objetivo, realizaron una encuesta para determinar si los profesores de economía habrían modificado sus hábitos a la hora de enseñar. Para bien o para mal, sus resultados muestran que el método de enseñanza predominante sigue siendo el tradicional "*chalk-and-talk*" (o sea, exponer y explicar al tiempo que se escribe en el pizarrón). En virtud de esto, los autores sugieren que los departamentos de economía estarían corriendo un gran riesgo a largo plazo, frente a otras disciplinas en las cuales se utilizan con mayor frecuencia otros métodos alternativos (y más innovadores) de enseñanza, como los experimentos en clase y el uso intensivo de Internet. En conclusión, en la medida que no haya incentivos apropiados para la adopción de estos métodos novedosos de enseñanza, probablemente continuemos buscando oportunidades de eficiencia en el tamaño del curso, la cantidad de docentes por curso, y otras variables similares.

³¹Asimismo, recordando lo expuesto por Illanes y Sapelli (2012), reducir el número de alumnos en cada aula requeriría abrir nuevos cursos, los cuales posiblemente haya que cubrirlos con docentes menos experimentados.

7. Bibliografía

- Aleamoni, Lawrence M. (1999). "Student rating myths research facts from 1924 to 1998," *Journal of Personnel Evaluation in Education*, 13(2): 153-166.
- Becker, William (2000). "Teaching economics in the 21st century", *Journal of Economic Perspectives*, 14(1): 109-119.
- Becker, William E. y Michael Watts (1999). "How departments of economics evaluate teaching," *Journal of Economic Education*, 43(3): 325-333. *American Economic Review*, 89(2): 344-349.
- Becker, William E., William Bosshardt y Michael Watts (2012). "How departments of economics evaluate teaching," *Journal of Economic Education*, 43(3): 325-333.
- Becker, William. E. y John R. Powers (2001). "Student performance, attrition, and class size given missing student data," *Economics of Education Review*, 20(4): 377-88.
- Centra, John A. (2003). "Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work?" *Research in Higher Education*, 44(5): 495-518.
- Feldman, Kenneth A. (2007). "Identifying exemplary teachers and teaching: Evidence from students ratings," in *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, R.P. Perry and J.C. Smart (eds.), 93-147. Springer Netherlands.
- Gaviria, Andrés y Alejandro Hoyos (2008). "Determinantes de los resultados de las evaluaciones de profesores: el caso de la Universidad de los Andes," *Revista Desarrollo y Sociedad*.
- Heilman, J. D., y W. D. Armentrout (1936). "The rating of college teachers on ten traits by their students," *Journal of Educational Psychology*, 27(3): 197-216.
- Illanes, Gastón y Claudio Sapelli (2012). "Class size and teacher effects in higher education," Documento de Trabajo 418, Instituto de Economía, Pontificia Universidad Católica de Chile.
- Isely, Paul y Harinder Singh (2005). "Do higher grades lead to favorable student evaluations?" *Journal of Economic Education*, 36(1): 29-42.
- McPherson, Michael A. (2006). "Determinants of how students evaluate teachers," *Journal of Economic Education*, 37(1):3-20.
- Marsh, Herbert W. (2007). "Students evaluation of university teaching: Dimensionality, reliability, validity, potential biases and usefulness," in *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, R.P. Perry and J.C. Smart (eds.), 319-383. Springer Netherlands.
- Marsh, Herbert W. y Lawrence A. Roche (1997). "Making students' evaluations of teaching effectiveness effective the critical issues of validity, bias, and utility," *American Psychologist*, 52(11): 1187-1197.

Taylor, Eric S. y John H. Tyler (2012). "The effect of evaluation on teacher performance," *American Economic Review*, 102(7): 3628–3651.

Spooren, Pieter y Dimitri Mortelmans (2006). "Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students give higher ratings?" *Educational Studies*, 32(2): 201–214.

TABLA 1. Evaluación promedio por campo para la dupla curso-docente

	media	desvio est.	mínimo	máximo
contenidos	3.899	0.5707	1.00	5
lenguaje técnico	3.955	0.5246	1.00	5
ejemplos	3.944	0.5402	1.00	5
trabajo en aula	3.989	0.5455	1.00	5
preguntas	4.099	0.5376	1.33	5
explicaciones/complejidad	3.860	0.5462	1.00	5
asistencia & puntualidad	4.082	0.5654	1.00	5

Nota: número de observaciones 2806

TABLA 2. Correlación entre campos de evaluación

	contenidos	lenguaje técnico	ejemplos	trabajo en aula	preguntas	explicaciones/complejidad	asistencia & puntualidad
contenidos	1						
lenguaje técnico	0.9625	1					
ejemplos	0.9495	0.9515	1				
trabajo en aula	0.8896	0.8983	0.8979	1			
preguntas	0.8544	0.8624	0.8711	0.8941	1		
explicaciones/complejidad	0.9242	0.9213	0.9236	0.8798	0.8808	1	
asistencia & puntualidad	0.6017	0.6251	0.6168	0.6162	0.5998	0.6284	1

Nota: número de observaciones 2806

TABLA 3. Distribución de obs. por carrera

	frecuencia	porcentaje
Cont. Público	751	26.76
Lic. en Adm. en Empresas	1,566	55.81
Lic. en Economía	388	13.83
Org. Bancaria	75	2.67
Tec. Org. Bancaria	26	0.93

Nota: número de observaciones 2806

TABLA 4. Evaluación total de acuerdo a las características

	media	desvio est.	min.	max.
<i>A. Curso</i>				
Mat. Optativa: no	3.97	0.506	1.05	5.00
Mat. Optativa: si	4.06	0.418	2.89	5.00
Multiples docentes: no	4.13	0.460	2.43	5.00
Multiples docentes: si	3.95	0.506	1.05	5.00
Carrera 1: Lic. Adm.	3.88	0.476	1.85	5.00
Carrera 2: Cont. Pub.	4.10	0.478	1.79	5.00
Carrera 3: Lic. Econ.	4.08	0.564	1.05	5.00
Carrera 4 : Org. Banc.	4.10	0.556	2.43	4.87
Carrera 5: Tec. Org. Banc.	4.26	0.376	3.58	4.83
Programa 1: primer año	3.95	0.493	1.05	5.00
Programa 2: segundo año	3.98	0.524	1.79	5.00
Programa 3: tercer año	3.94	0.529	2.12	5.00
Programa 4: cuarto año	4.03	0.463	1.85	5.00
Turno 1: mañana	3.93	0.513	1.79	4.88
Turno 2: tarde	3.99	0.494	1.05	5.00
Turno 3: noche	4.04	0.483	2.19	5.00
Semestre1: primero	4.00	0.499	1.05	5.00
Semestre 2: segundo	3.94	0.508	1.85	5.00
Año 1: 2011	3.94	0.510	1.79	5.00
Año 2: 2012	3.97	0.499	1.05	5.00
Año 3: 2013	4.05	0.486	2.09	4.92
<i>B. Docente</i>				
Genero: hombre	3.98	0.524	1.05	5.00
Genero: mujer	3.97	0.474	2.00	5.00
Educación 1: grado	3.96	0.493	1.05	5.00
Educación 2: magister	3.96	0.527	2.00	5.00
Educación 3: doctor	4.17	0.498	2.46	5.00

Nota: número de observaciones 2806

TABLA 5. Resultados

	(1)	(2)	(3)	(4)	(5)
tamaño	-0.00440*** (0.000776)	-0.00994*** (0.00310)	-0.0108*** (0.00302)	-0.0106*** (0.00322)	-0.0103*** (0.00312)
tamaño^2		8.55e-05* (4.37e-05)	9.66e-05** (4.27e-05)	9.26e-05** (4.52e-05)	8.98e-05** (4.39e-05)
mat. optativa (1= electiva)	0.0371 (0.0426)	0.0348 (0.0426)	0.0178 (0.0427)		0.0355 (0.0425)
+1 docente	-0.163*** (0.0243)	-0.156*** (0.0245)	-0.147*** (0.0245)	-0.171*** (0.0258)	-0.147*** (0.0250)
edad	-0.00281*** (0.000825)	0.0120** (0.00604)	0.0111* (0.00598)	0.0110* (0.00615)	0.0152** (0.00619)
edad^2		-0.000152** (6.13e-05)	-0.000141** (6.04e-05)	-0.000141** (6.24e-05)	-0.000186*** (6.30e-05)
genero (1= hombre)	-0.0361* (0.0204)	-0.0294 (0.0207)	-0.0334 (0.0205)	-0.0268 (0.0213)	-0.0252 (0.0212)
educación3: doctor	0.160*** (0.0403)	0.154*** (0.0409)	0.145*** (0.0416)	0.174*** (0.0427)	0.150*** (0.0418)
educacion2: magister	-0.00570 (0.0236)	-0.0128 (0.0238)	-0.0106 (0.0235)	-0.0136 (0.0247)	-0.00780 (0.0241)
segundo año	-0.0192 (0.0270)	-0.0186 (0.0272)	-0.0143 (0.0274)	-0.0243 (0.0272)	-0.0114 (0.0278)
tercer año	-0.0656** (0.0287)	-0.0688** (0.0289)	-0.0631** (0.0290)	-0.0730** (0.0292)	-0.0615** (0.0293)
cuarto año	0.0303 (0.0304)	0.0295 (0.0306)	0.0445 (0.0304)	0.0175 (0.0311)	0.0342 (0.0308)
carrera2: Cont. Pub.	-0.172*** (0.0228)	-0.167*** (0.0229)	-0.165*** (0.0229)	-0.173*** (0.0237)	-0.164*** (0.0229)
carrera3: Lic. Econ.	-0.0954*** (0.0347)	-0.0945*** (0.0346)	-0.0945*** (0.0345)	-0.115*** (0.0355)	-0.0920*** (0.0348)
carrera4: OrgBanc.	-0.0467 (0.0662)	-0.0410 (0.0670)	-0.0354 (0.0671)	-0.0622 (0.0671)	
carrera5: Tec. Org. Banc.	0.0260 (0.0775)	0.0159 (0.0786)	0.0211 (0.0787)	-0.00703 (0.0786)	
año2: 2012	0.0246 (0.0206)	0.0218 (0.0206)	0.0151 (0.0206)	0.0240 (0.0212)	0.0179 (0.0210)
año3: 2013	0.0846*** (0.0264)	0.0823*** (0.0264)	0.0756*** (0.0262)	0.0791*** (0.0271)	0.0833*** (0.0266)
semestre (1= segundo)	-0.0432** (0.0206)	-0.0448** (0.0206)	-0.0495** (0.0205)	-0.0509** (0.0212)	-0.0446** (0.0209)
turno1: mañana	-0.0476 (0.0316)	-0.0342 (0.0315)	-0.0209 (0.0313)	-0.0350 (0.0316)	-0.0340 (0.0315)
turno2: tarde	0.00900 (0.0353)	0.0208 (0.0350)	0.0259 (0.0351)	0.0297 (0.0353)	0.0193 (0.0350)
constante	4.531*** (0.0685)	4.241*** (0.163)	4.256*** (0.162)	4.298*** (0.167)	4.161*** (0.166)
observaciones	2,806	2,806	2,781	2,684	2,705
adj. R2	0.097	0.101	0.098	0.103	0.098

Nota: Variable dependiente es la evaluación total (promedio de los 7 campos de evaluación). Errores estándares robustos en paréntesis, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

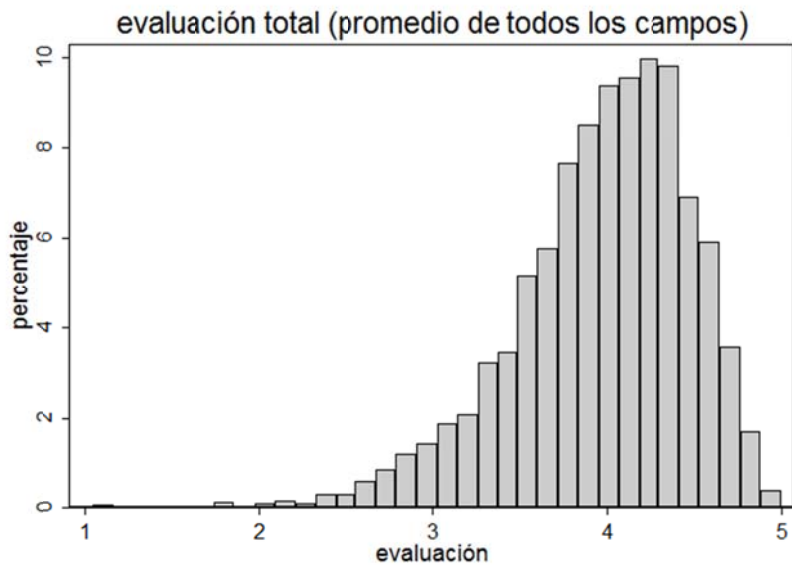
TABLA 6. Resultados (cont.)

	(1) Materia FE	(2) Docente FE	(3) Mat.+Doc. FE
tamaño	-0.00958*** (0.00305)	-0.0109*** (0.00268)	-0.0108*** (0.00351)
tamaño^2	9.15e-05** (4.33e-05)	9.55e-05** (3.73e-05)	9.57e-05** (4.76e-05)
mat. Optative (1= elective)	0.122 (0.212)	0.0516 (0.0487)	0.208 (0.152)
+1 docente	-0.141*** (0.0296)	-0.0813*** (0.0256)	-0.0836** (0.0383)
edad	0.00985 (0.00663)	-0.123** (0.0538)	-0.128** (0.0635)
edad^2	-0.000123* (6.69e-05)	0.000759* (0.000449)	0.000923* (0.000534)
segundo año		-0.0145 (0.0312)	
tercer año		0.0165 (0.0381)	
cuarto año		0.156*** (0.0444)	
carrera2: Cont. Pub.		-0.161*** (0.0244)	
carrera3: Lic. Econ.		-0.0988** (0.0491)	
carrera4: Org. Banc.		0.0620 (0.0922)	
carrera5: Tec. Org. Banc.		0.0978 (0.103)	
año2: 2012	0.0284 (0.0201)	0.0143 (0.0213)	0.00872 (0.0241)
año3: 2013	0.106*** (0.0261)	0.102* (0.0569)	0.0928 (0.0639)
semestre (1= segundo)	-0.0120 (0.0224)	-0.0287* (0.0164)	-0.0341 (0.0215)
turno: mañana	-0.0644* (0.0330)	-0.0554* (0.0301)	-0.0406 (0.0394)
turno: noche	-0.0134 (0.0367)	0.0435 (0.0350)	0.0368 (0.0450)
genero (1=hombre)	-0.0525** (0.0242)		
educacion3:: doctor	0.110** (0.0527)		
educacion2: magister	-0.0280 (0.0257)		
constante	4.148*** (0.171)	8.451*** (1.824)	8.225*** (2.110)
observaciones	2,806	2,806	2,806
Adj. R2	0.258	0.609	0.734

Nota: Variable dependiente es la evaluación total (promedio de los 7 campos de evaluación).

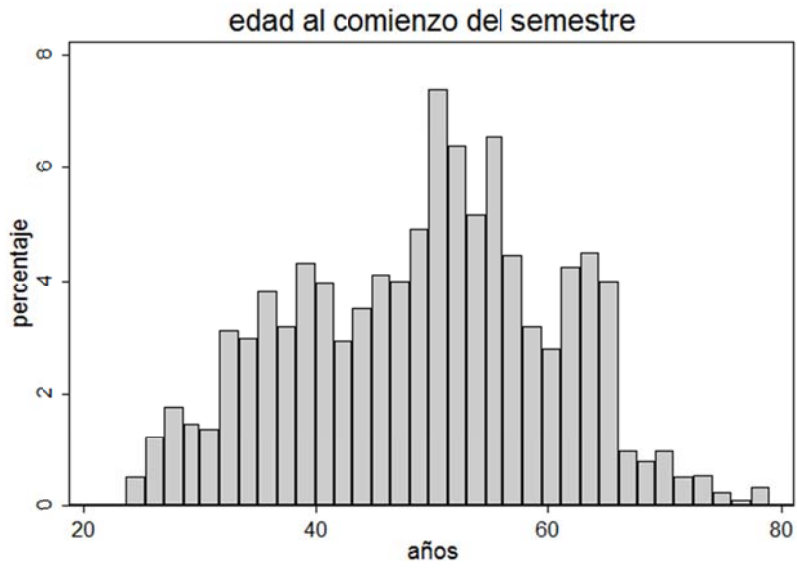
*Errores estándares robustos en paréntesis, *** p<0.01, ** p<0.05, * p<0.1*

FIGURA 1. Distribución de calificaciones

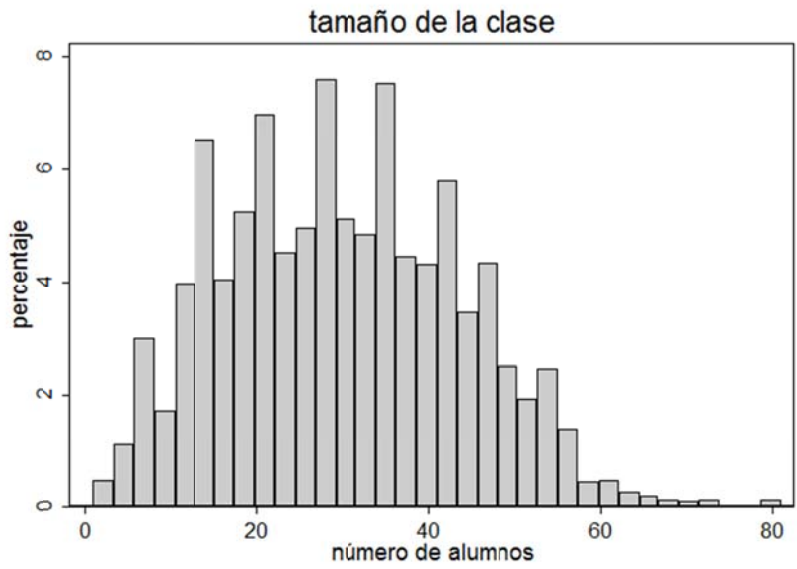


Nota: observaciones curso-profesor (n=2806)

FIGURA 2. Distribución de la edad del docente y el tamaño del curso



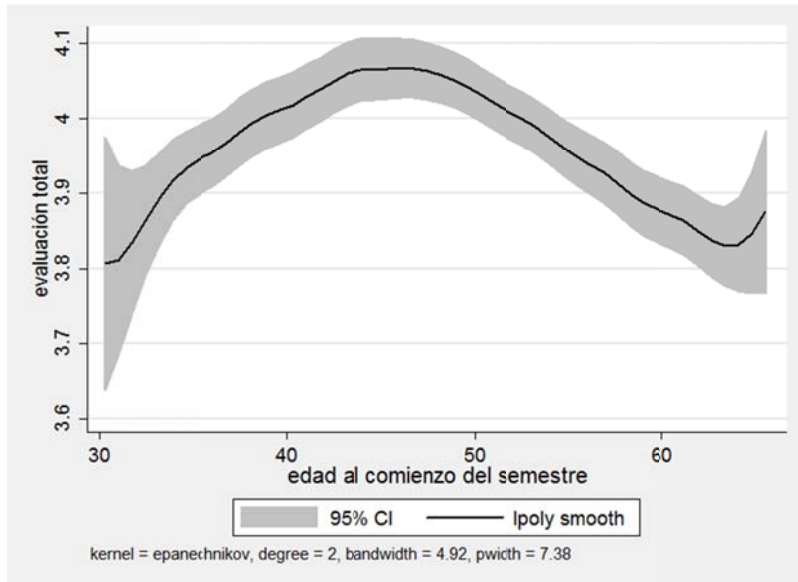
(a) edad del docente al comienzo del curso



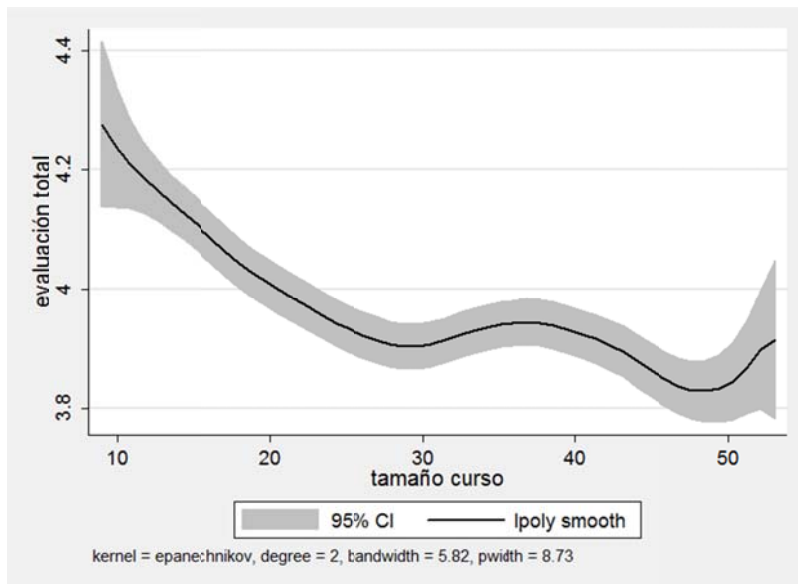
(b) tamaño de curso

Nota: observaciones curso-profesor (n=2806)

FIGURA 3. Estimaciones no-paramétricas sobre la evaluación total



(a) edad del docente al comienzo del curso



(b) tamaño del curso

Nota: observaciones curso-profesor (n=2806)