

# ¿Hombres “cracks” y mujeres “amables”? Sesgos de género en encuestas de profesores

---

Nicolás Urdaneta Andrade

Agosto de 2021

Documento CEDE

Estudiantes

# 46

© 2021, Universidad de los Andes, Facultad de Economía, CEDE. Calle 19A No. 1 – 37 Este, Bloque W. Bogotá, D. C., Colombia Teléfonos: 3394949- 3394999, extensiones 2400, 2049, 2467

[infocede@uniandes.edu.co](mailto:infocede@uniandes.edu.co)  
<http://economia.uniandes.edu.co>

Impreso en Colombia – Printed in Colombia

La serie de Documentos de Trabajo CEDE se circula con propósitos de discusión y divulgación. Los artículos no han sido evaluados por pares ni sujetos a ningún tipo de evaluación formal por parte del equipo de trabajo del CEDE. El contenido de la presente publicación se encuentra protegido por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por tanto su utilización, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso, digital o en cualquier formato conocido o por conocer, se encuentran prohibidos, y sólo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito del autor o titular. Las limitaciones y excepciones al Derecho de Autor, sólo serán aplicables en la medida en que se den dentro de los denominados Usos Honrados (Fair use), estén previa y expresamente establecidas, no causen un grave e injustificado perjuicio a los intereses legítimos del autor o titular, y no atenten contra la normal explotación de la obra.

Universidad de los Andes | Vigilada Mineducación Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

## Documento CEDE

### Estudiantes

Los documentos CEDE-Estudiantes contienen los resultados de las tesis de maestría (PEG o MEcA) de los estudiantes de la Facultad que se consideran merecedoras de este honor. Extraordinariamente por postulación de al menos un profesor de planta de la Facultad; se pueden considerar trabajos de estudiantes de pregrado.

# ¿Hombres “cracks” y mujeres “amables”? Sesgos de género en encuestas de profesores

Nicolás Urdaneta Andrade \*

## Resumen

Este artículo estudia la existencia de sesgos de género en las evaluaciones de docencia de una universidad en Colombia. Para esto, se utiliza un modelo de regresión lineal con efectos fijos de estudiante y que controla por características de los estudiantes, profesores y los cursos. Se encuentra que, en promedio, las profesoras de planta reciben evaluaciones 0.07 desviaciones estándar menores que sus pares masculinos. Mientras que, no hay diferencias por género entre asistentes graduados ni profesores de cátedra. Sin embargo, las heterogeneidades son amplias. En cursos con profesores de 35-55 años, con más de 50 estudiantes, o entre menor sea la fracción de estudiantes mujeres, las profesoras son penalizadas más fuertemente. En estos casos se encuentran sesgos entre 0.1-0.2 desviaciones estándar y representan el 40 % de las evaluaciones. Adicionalmente, las mujeres tienden a ser descritas con palabras relacionadas a características personales (“amable” o “comprensiva”) y los hombres con temas asociados al curso o con adjetivos positivos (“tema”, “parcial”, “crack”). A partir de lo anterior, se proponen medidas que disminuyan los sesgos en los estudiantes y un análisis de las evaluaciones consciente de las diferencias por género.

*Palabras Clave:* género; profesor; evaluación docente; sesgos.

*Códigos JEL:* J16; I21; I29; C52.

---

\*Tesis de grado de Maestría en Economía, Universidad de Los Andes, email: n.urdaneta@uniandes.edu.co. Agradezco todos los comentarios y el acompañamiento de Adriana Camacho y los comentarios de Diana Pérez, Lewis Polo, Pedro Cabra, Germán Orbegozo, Mateo Arbeláez, Jimena Hurtado, Ignacio Sarmiento, Julián Mariño y los asistentes del seminario ADES en la Universidad de Los Andes. También agradezco el acceso a la información por parte de Julián Mariño, Cristhian Calderón y Mayra Neisa del Centro de Evaluación de la Universidad de Los Andes por los datos de las encuestas de evaluación; a Fabio Sánchez, Ángela Patiño y Tatiana Mojica por la información del SPADIES Uniandes; a Sonia Murillo y Sandra Acosta por la información anonimizada de los profesores de la universidad; y a Yenni Amón y Claudia Meza por la información de Admisiones y Registro. Esta tesis tuvo aprobación del Comité de ética de la Facultad de Economía y fue revisada por el Centro de Evaluación de la Universidad de los Andes previo a su publicación. Toda la programación de este trabajo está disponible en el siguiente repositorio: <https://github.com/Nurdaneta/Tesis-PEG-Public>.

# ¿“Smart” men and “kind” women? Gender bias in student evaluations of teaching

Nicolás Urdaneta Andrade \*\*

## Abstract

This paper studies the existence of gender bias in the student evaluations of teaching at a university in Colombia. For this, I use a linear regression with student fixed effects that also accounts student, instructor and course characteristics. I find that, on average, full-time female professors receive evaluations lower in 0.07 standard deviation than male professors. Meanwhile, there are no gender differences among adjunct professors or teacher assistants. Nevertheless, there is ample heterogeneity. In course with professors with 35-55 years old, with more than 50 students, or with a smaller share of female students, female professors are penalized the most. In such cases I find that biases have a magnitude about 0.1-0.2 standard deviations and represent 40% of all evaluations. Additionally, women are mostly described with words related to personal characteristics (“kind” or “comprehensive”) and men are associated with course-related topics or positive adjectives (“topic”, “exam”, “smart”). Based on the results found, I propose measures to reduce bias in students and that allow an analysis on the evaluations to take into account the gender differences.

*Keywords:* gender; professor; student evaluation of teaching; bias.

*JEL codes:* J16; I21; I29; C52.

---

\*\* Master's thesis in economics, University of Los Andes, email: n.urdaneta@uniandes.edu.co. I am grateful for all the comments and companionship of Adriana Camacho as well as comments from Diana Pérez, Lewis Polo, Pedro Cabra, Germán Orbegozo, Mateo Arbeláez, Jimena Hurtado, Ignacio Sarmiento, Julián Mariño, and ADES seminar participants at University of Los Andes. I am also grateful to the following people for granting me access to the information of this project: Julián Mariño, Cristhian Calderón, and Mayra Neisa from the Evaluation Center at University of Los Andes for the evaluations data; Fabio Sánchez, Ángela Patiño, and Tatiana Mojica for information on the SPADIES Uniandes; Sonia Murillo and Sandra Acosta for the anonymized information from university instructors; and Yenni Amón and Claudia Meza for the information from Registry and Admissions. This thesis has been approved by the Economics Faculty Ethics Committee and the document was seen by the Centro de Evaluación of the University of los Andes prior to publication. All the code replication material of this work is available at: <https://github.com/Nurdaneta/Tesis-PEG-Public>.

# 1 Introducción

¿Son las profesoras universitarias evaluadas de manera sistemáticamente diferente a sus pares masculinos en las encuestas de docencia, únicamente como resultado de su género? Las evaluaciones de profesores que responden los estudiantes en las universidades están siendo revaluadas debido a que pueden reflejar otros factores más allá de las habilidades docentes. Un motivo de particular preocupación es la existencia de un diferencial en la forma en que los estudiantes evalúan a sus profesores en función de su género. Para determinar si existen dichos sesgos y bajo qué casos, se utiliza información de las encuestas de profesores en una universidad en Colombia. A partir de un mejor entendimiento sobre en qué contextos hay sesgos contra las mujeres, se proponen posibles soluciones para evaluar de manera más equitativa a todos los profesores.

Una calificación diferenciada por género puede tener repercusiones en la carrera profesional y en el mercado laboral para las mujeres en la academia. Muchas universidades, incluida la universidad de estudio<sup>1</sup>, utilizan las encuestas docentes como insumo en sus procesos de promoción de profesores. Además, es posible que ante una discriminación las mujeres afronten la necesidad de mejorar en las evaluaciones. Por este motivo, pueden realizar un esfuerzo adicional en sus clases, lo que limita su tiempo a otros ámbitos del trabajo como, por ejemplo, la investigación (Mengel et al., 2019). También es posible que al buscar formas de mejorar su rendimiento en las evaluaciones, estos métodos resulten no ser favorables para el aprendizaje de los estudiantes (Gaviria y Hoyos, 2008). Entre estas posibilidades está dar calificaciones sistemáticamente más elevadas o enseñar conocimientos específicos para los trabajos asignados, lo que puede traducirse en un menor nivel de aprendizaje (Carrell y West, 2010). Probablemente debido a estos efectos discriminatorios, las mujeres parecen ser más receptivas a la información de las evaluaciones (Buurman et al., 2020).

La posible discriminación en las evaluaciones de docencia contrasta con la forma como principalmente se entiende la discriminación en el mercado laboral; a partir de menores salarios y mayor nivel de desempleo. Un sesgo en las evaluaciones tiene repercusiones directas sobre la calidad del trabajo de los profesores y su probabilidad de promoción, lo que plausiblemente resulta en menores salarios eventualmente. De manera que, las discriminaciones que pueden afrontar las mujeres en este contexto también pueden estar presentes en otros ámbitos profesionales. Por ejemplo, es común que los servidores públicos tengan evaluaciones sobre su desempeño<sup>2</sup>. Por lo anterior, las evaluaciones de docencia representan uno de los múltiples casos en los que puede haber discriminación por una percepción subjetiva y que se traslada a una brecha en la carrera profesional de las mujeres respecto a los hombres.

---

<sup>1</sup>El estatuto profesoral de la universidad hace explícito que estas encuestas hacen parte del proceso de promoción de profesores (Universidad de Los Andes, 2020).

<sup>2</sup>En Colombia, el Gobierno Nacional tiene una guía completa para la evaluación del desempeño laboral de los servidores públicos (Presidencia de la República de Colombia, 2020).

Este trabajo es la primera evaluación que, a conocimiento del autor, aborda los sesgos de género en una universidad en América Latina, a diferencia de la mayoría de los trabajos sobre este tema que han sido realizados en Estados Unidos o en Europa<sup>3</sup>. En este sentido, América Latina es una región de particular interés en el estudio de discriminación de género puesto que ya se ha documentado ampliamente que hay una mayor discriminación contra las mujeres en ámbitos profesionales en la región en comparación con otros países occidentales<sup>4</sup>.

Para estudiar las evaluaciones de profesores se hace uso de información detallada de 352,654 encuestas que respondieron más de 17 mil estudiantes acerca de 3 mil profesores (35.5% mujeres) en 5,800 cursos distintos durante cuatro semestres. En esta universidad, el 91% de los profesores consulta las encuestas todos los semestres. En adición, el 63% ha ajustado las actividades del curso debido a los resultados de la encuesta y el 51% reporta haber ajustado la metodología<sup>5</sup>.

La mayoría de la literatura previa ha estado considerablemente limitada por la cantidad de evaluaciones que pueden observar y no cuentan con tanta riqueza de información. En particular, debido a que cuentan con información de pocas facultades dentro de un mismo estudio o con estudiantes únicamente de primeros semestres<sup>6</sup>. La diferencia es que en esta universidad se cuenta con información de 38 programas académicos en distintas áreas.

Con las información disponible es posible identificar diferencias en dos elementos de las evaluaciones. En primer lugar, se determina qué diferencias hay en variables sobre los puntajes que reciben los profesores en tanto a su docencia o al curso, y en un indicador de si el estudiante recomienda tomar la clase con ese profesor. En segundo lugar, se determinan diferencias en qué tan positivos o negativos son los comentarios que escriben los estudiantes a partir de un análisis de sentimientos. Este componente de análisis de texto es un elemento adicional que se ha dejado de lado en la literatura y que también es relevante debido a los efectos sobre la docencia de los profesores<sup>7</sup>.

La principal estrategia para identificar los sesgos de género consiste en un modelo de regresión lineal con efectos fijos de estudiante como propone [Boring \(2017\)](#) que adicionalmente incluye controles de características del estudiante, del curso y del profesor. Bajo esta metodología se define un sesgo de género como una diferencia en el puntaje de los profesores a causa de su género que no se puede explicar por características del estudiante, curso o del profesor. En este sentido, un

<sup>3</sup>Trabajos realizados en EEUU: [Basow y Silberg \(1987\)](#); [Boring et al. \(2016\)](#); [Macnell et al. \(2014\)](#); [Schmidt \(2015\)](#). Trabajos realizados en Europa: [Boring \(2017\)](#); [Boring y Philippe \(2021\)](#); [Boring et al. \(2016\)](#); [Mengel et al. \(2019\)](#). El único trabajo conocido fuera de estas regiones es [Keng \(2020\)](#), que utiliza información de una universidad en Taiwan.

<sup>4</sup>Algunos ejemplos que documentan estas brechas de género son [Dulce et al. \(2019\)](#); [Iregui-Bohórquez et al. \(2020\)](#); [Conceição et al. \(2002\)](#); [Muller y Casabonne \(2020\)](#)

<sup>5</sup>Información obtenida de una encuesta realizada por la Vicerrectoría Académica de la universidad a profesores de cátedra y planta de los cuales respondieron 700 profesores.

<sup>6</sup>Los trabajos que abarcan la muestra más amplia que se conoce son los de [Boring \(2017\)](#) con 4,362 estudiantes y 20,197 evaluaciones y el de [Mengel et al. \(2019\)](#) con 4,848 estudiantes y 19,952 observaciones.

<sup>7</sup>Algunos estudios que analizan el texto, pero de manera limitada son [Mitchell y Martin \(2018\)](#) y [Storage et al. \(2016\)](#). También está el portal de [Schmidt \(2015\)](#) que utiliza los comentarios de *Rate My Professor*.

elemento fundamental es evitar que se documenten sesgos que sean resultado de diferencias en habilidades u otras características de los docentes. Para evitar este problema de variable omitida se utiliza un modelo de valor agregado, de acuerdo a Chetty et al. (2014), para medir las características de un profesor que determinen las notas de sus estudiantes como la generosidad en las calificaciones y sus habilidades.

Por su parte, los efectos fijos de estudiante alivian la preocupación relacionada con que los estudiantes puedan escoger los cursos que toman. Los trabajos de Boring (2017), Boring et al. (2016), Macnell et al. (2014) y Mengel et al. (2019) se esfuerzan en tener evaluaciones en las que los estudiantes son asignados de manera aleatoria a sus cursos. Sin embargo, se puede perder generalidad al restringir estudios a unas ciertas carreras o estudiantes de ciertas características (ej. de primeros semestres). Con los efectos fijos se espera que las características del estudiante que podrían determinar su selección de cursos no afecte la identificación de los sesgos. En particular porque, en promedio cada estudiante se observa en 20.2 evaluaciones distintas y en varios semestres, a diferencia de otros trabajos en los que cada estudiante se observa menos de 5 veces en promedio. Con esto, se pueden capturar características intrínsecas de un estudiante por las cuales podrían escoger ciertos cursos o profesores. Adicionalmente, se corroboran los resultados al estudiar casos puntuales en los que la facilidad con la que los estudiantes escogen cursos es limitada.

Con esta estimación se encuentra que las profesoras de planta son calificadas 0.07 desviaciones estándar peor que los hombres en los puntajes relacionados con su docencia. Las profesoras de planta son recomendadas con una probabilidad menor en 1.4 puntos porcentuales y los comentarios en sus evaluaciones son 2 puntos porcentuales menos positivos y más negativos. Entre todas las medidas, los sesgos tienen magnitudes similares entre 0.05-0.07 desviaciones estándar. Dentro de estas diferencias, los sesgos son ligeramente más grandes entre estudiantes hombres y no siempre son significativos entre estudiantes mujeres. Sin embargo, entre los profesores de cátedra y los asistentes graduados no hay diferencias en las evaluaciones de hombres y mujeres.

Estos resultados contrastan con los trabajos recientes de Boring (2017), Mengel et al. (2019) y Keng (2020) que son los más similares en metodología y datos. Los resultados de esos trabajos apuntan a sesgos entre 2 y 4 veces más grandes con respecto al sesgo documentado para profesores de planta. Algunas razones plausibles que pueden explicar esta diferencia son: 1) el contexto de la universidad. Por ejemplo, la universidad de estudio parece tener profesoras con más experiencia y está más cerca de balance de género en el cuerpo estudiantil que en otros estudios<sup>8</sup>; 2) No controlar explícitamente por las habilidades de los profesores; 3) Un caso de sesgo de publicación en el que los trabajos tienen mayor facilidad de publicarse en caso de demostrar una mayor discriminación.

No obstante, hay heterogeneidades en las evaluaciones y en muchos casos sí hay un sesgo importante contra las mujeres. A partir de un modelo de “árboles causales honestos” desarrollado por

---

<sup>8</sup>Boring (2017) reconoce que los sesgos evidentemente dependen del contexto particular de cada universidad.

[Athey y Imbens \(2016\)](#) se caracterizan las heterogeneidades detrás del resultado principal. Con el “árbol” se encuentra que, entre los profesores que tienen experiencia pero todavía están avanzando en su carrera (35 a 55 años), cuando los cursos no tienen tantos estudiantes y los profesores tienen una menor influencia en las notas de los estudiantes, las mujeres se ven ampliamente desfavorecidas. Estos casos equivalen al 40 % de las evaluaciones y los sesgos van desde 0.07 desviaciones estándar hasta 0.2 desviaciones estándar. A pesar de esto, principalmente entre profesoras muy jóvenes (menos de 4 años de experiencia), cuando hay una amplia mayoría de estudiantes mujeres (mayor al 70 %) o cuando hay muy pocas estudiantes los sesgos son positivos. Pero el sesgo positivo está apenas en el 15 % de las evaluaciones y no supera diferencias de 0.1 desviaciones estándar.

Finalmente, para profundizar el análisis de los comentarios se estudian las diferencias en el lenguaje utilizado por género del profesor. Con un modelo de regularización Lasso con distribución logística se determina qué palabras, de las 5,000 más frecuentes, predicen mejor el género del profesor sobre el que se escribió el comentario. Este método se inspira en el ejercicio realizado por [Wu \(2018\)](#) que encuentra discriminación en el lenguaje utilizado para describir a las mujeres en el portal *Econ Job Rumors*.

Entre los comentarios de los profesores hombres predomina un lenguaje sobre el curso y la docencia, mientras que para las mujeres es más común el uso de palabras asociadas a estereotipos femeninos y no necesariamente relacionadas con el curso. Los cinco adjetivos que más se destacan en los comentarios acerca de las mujeres son “querido”, “amable”, “dulce”, “paciente” y “comprendativo”. Mientras que los cinco adjetivos principales para los hombres son “chistoso”, “crack”, “gracioso”, “interesante” y “entretenido”<sup>9</sup>. Además, para los hombres es más común encontrar palabras relacionadas al curso como “profesor”, “complementario”, “tema” y “parcial”. Estos resultados son similares a lo que muestra [Schmidt \(2015\)](#) de las evaluaciones del portal *Rate My Professor* para universidades en EEUU y Canadá<sup>10</sup>.

Ante esto, se proponen algunas recomendaciones con el fin de disminuir los sesgos y evitar que las evaluaciones perjudiquen a las mujeres en su carrera profesional. La primera está en buscar un balance de género en la composición de género de los estudiantes, pues de mantenerse una mayoría de estudiantes hombres en algunos cursos los sesgos seguramente seguirán presentes en esos casos. Segundo, recordarle a los estudiantes antes de responder las encuestas la importancia que tienen las evaluaciones para los profesores e informar sobre los sesgos que se han documentado contra las mujeres. Sobre esto, los trabajos de [Boring y Philippe \(2021\)](#) y [Peterson et al. \(2019\)](#) muestran

---

<sup>9</sup>En la sección 4 se detalla cómo se evita el problema de encontrar palabras asociadas a un género debido a que la palabra de entrada implica el género del sujeto del que se habla (como “profesor” y “profesora”).

<sup>10</sup>El trabajo de [Schmidt \(2015\)](#) analiza el lenguaje de todos los comentarios en *Rate My Professor*. El portal del autor está diseñado para que el usuario explore distintas palabras y no determina cuáles son las que se utilizan con más frecuencia por género del profesor.

evidencia que estas intervenciones pueden ser exitosas. En tercer lugar, se pueden incrementar los insumos que se utilizan para las evaluaciones de desempeño docente por parte de la universidad. Por ejemplo, un modelo de valor agregado permite tener una medida objetiva sobre qué tanto aporta un profesor al aprendizaje de sus estudiantes. Y por último, se pueden presentar los puntajes agregados de las encuestas para cada profesor después de regresar esta variable contra el género del profesor (y otras variables relevantes). Esto evita que los resultados que observan los profesores para mejorar su docencia y la universidad para evaluarlos tenga estos sesgos en promedio. Infortunadamente, la eficacia de esta última propuesta se ve afectada por el grado de heterogeneidad en los sesgos. Con estas medidas concretas se pueden disminuir los sesgos en las encuestas y se puede contar con mejores herramientas para evaluar a los profesores en su trabajo docente.

La presente investigación se nutre de una extensa literatura sobre sesgos de género en evaluaciones de profesores. Los trabajos iniciales sobre este tema contaban con una estrategia de identificación a partir de una diferencia incondicional, al igual que con una cantidad de evaluaciones muy limitada ([Basow 1998](#); [Basow y Silberg 1987](#); [Johnson et al. 2013](#); [Centra y Gaubatz 2000](#); [Das y Das 2001](#); [Burke et al. 2017](#); [Nargundkar y Shrikhande 2014](#) entre otros). En ellos, la evidencia sobre la existencia de sesgos es poco concluyente pues no siempre se encuentra un sesgo contra las mujeres ([Centra y Gaubatz, 2000](#)). Más recientemente, los trabajos de [Boring \(2017\)](#), [Macnall et al. \(2014\)](#), [Mengel et al. \(2019\)](#) y [Keng \(2020\)](#) dieron un vuelco hacia una identificación más limpia para atribuir los sesgos a efectos causales y en estos encuentran que los sesgos son considerablemente altos. Así, este trabajo trae cosas importantes de las aproximaciones de ambas ramas de la literatura. Por un lado, es fundamental identificar correctamente los sesgos. Y, por otra parte, se estudian los mecanismos por los cuales también puede haber sesgos a través de efectos heterogéneos y el contenido de las evaluaciones con los comentarios.

El resto de este trabajo se organiza de la siguiente manera. La sección 2 detalla el contexto de los estudiantes y las evaluaciones en la universidad. En la sección 3 de datos se describe el contenido de las encuestas y cómo se calculan las variables de resultado. Luego, se encuentran las secciones de estrategia empírica y resultados. Más adelante, en la sección 6 se proponen las recomendaciones para disminuir los sesgos y sus efectos. Finalmente, la sección 7 concluye.

## 2 Contexto del proceso de encuestas

Es de preocupación de los trabajos más recientes si los estudiantes pueden escoger o no a sus profesores y cursos. En caso de no ser aleatorio, es posible que esto resulte determinando los puntajes de los profesores. En el caso de esta universidad, al comienzo de cada semestre, los estudiantes escogen las clases que desean tomar<sup>11</sup> En este proceso cada estudiante tiene un turno

<sup>11</sup>En particular, el reglamento de estudiantes de la universidad hace obligatorio que los estudiantes puedan saber quién es el profesor responsable de cada curso al momento de escoger materias.

de inscripción que varía según la cantidad de créditos aprobados. El primer turno lo tienen los estudiantes de primer semestre y después de ellos los turnos se asignan comenzando por quienes tengan más créditos aprobados hasta llegar a los estudiantes de segundo semestre que toman los últimos turnos. Estas limitaciones en la facilidad para inscribir cursos pueden dar luces sobre cómo cambian los sesgos a medida que un estudiante puede escoger a sus profesores.

Los estudiantes deben responder las encuestas de evaluación docente al final del semestre antes de comenzar el periodo de exámenes finales, por lo que para ese momento los estudiantes no conocen su nota final en la materia<sup>12</sup>. En consecuencia, la nota que recibe un estudiante está fuertemente asociada a una mejor evaluación docente (como se ve en la Tabla A1)<sup>13</sup>. En adición, los profesores únicamente conocen el resultado de las evaluaciones después del final del semestre, lo cual disminuye la preocupación que los profesores asignen notas debido a los resultados de sus evaluaciones.

Finalmente, responder las evaluaciones es opcional, pero en caso de no hacerlo, los estudiantes son penalizados con ser asignados el último turno de inscripción de materias el semestre siguiente. De esta forma, en el periodo estudiado, entre el 91 y 94 % de las evaluaciones fueron respondidas. Con esto, hay una menor preocupación de una posible selección sobre quiénes responden la encuesta y quiénes no. Por su parte, es completamente opcional hacer comentarios en el espacio previsto en la encuesta sobre el curso y el profesor. De las encuestas que se respondieron, en el 66.1 % también se respondió al menos una de las cuatro preguntas abiertas de comentarios. Aunque es probable que haya alguna selección, con un ejercicio de robustez parece que esta selección no tiene influencia en los resultados, de igual forma a lo que encuentran [Mengel et al. \(2019\)](#).

### 3 Datos

Este trabajo utiliza una amplia información obtenida de distintas dependencias de la universidad sobre los estudiantes de pregrado en cuatro semestres a partir del segundo semestre del 2017. Se cuenta con 17,400 estudiantes y 3,013 profesores de los cuales el 44.9 % de los estudiantes y el 35.5 % de los profesores son mujeres. Entre los datos disponibles está la información de las encuestas docentes e información académica y socio demográfica de los estudiantes y características de los profesores. Para este análisis, se utilizan únicamente los cursos con aprobación numérica<sup>14</sup>. De esta forma, se estudian 352,654 evaluaciones que representan el 85.9 % del total de las evaluaciones entre todos los cursos de los 38 programas de pregrado que ofrece la universidad.

En primer lugar, en las evaluaciones, cada facultad determina un grupo de preguntas para ca-

<sup>12</sup>Por reglamento, los profesores deben entregar al menos el 30 % de la nota a mitad del semestre. Al momento de responder la encuesta un estudiante puede conocer un porcentaje incluso mayor de su calificación final.

<sup>13</sup>Esta correlación también se ha documentado esto en otros trabajos. Ver [Johnson et al., 2013; Mengel et al., 2019; Boring, 2017](#) en otros casos y [Gaviria y Hoyos \(2008\)](#) para esta universidad.

<sup>14</sup>Los cursos se califican de 1 a 5, con dos decimales, y se aprueba con un puntaje de 3.

da curso en función de los objetivos de aprendizaje del mismo y del tipo y metodología docente. El banco de preguntas de toda la universidad contiene 316 preguntas que están categorizadas por su enfoque sobre el curso, el profesor y un indicador de si se recomienda tomar el curso con ese profesor<sup>15</sup>. Dado que cada curso tiene preguntas distintas, se define la calificación agregada por categoría como un promedio de las preguntas estandarizadas, de forma similar a Mengel et al. (2019). En particular, la ecuación que define las calificaciones agregadas está descrita por la ecuación (1). En esta, el estudiante  $i$  responde  $N_{ipct}$  preguntas sobre el profesor  $p$  en el curso  $c$  en el semestre  $t$ . La respuesta a la pregunta  $q$  corresponde a la variable  $y_{ipct}^q$ . Para estandarizar  $y_{ipct}^q$  se utilizan el promedio ( $\bar{y}_t^q$ ) y su desviación estándar ( $\sigma_t^q$ ) por semestre. Con este procedimiento se tiene una medida razonable por categoría cuyos efectos se miden en desviaciones estándar<sup>16</sup>.

$$Y_{ipct} = \frac{1}{N_{ipct}} \sum_{q=1}^{N_{ipct}} \frac{(y_{ipct}^q - \bar{y}_t^q)}{\sigma_t^q} \quad (1)$$

En segundo lugar, las evaluaciones también incluyen los comentarios que los estudiantes escriben sobre sus profesores y cursos. Todas las encuestas tienen espacio para cuatro comentarios en los que se les pide a los estudiantes que describan los aspectos positivos tanto del curso como del profesor y que den recomendaciones sobre cómo podría mejorar la clase o la práctica docente del profesor. En este trabajo se le da un manejo a los comentarios de dos formas. La primera consiste en calcular una probabilidad que el comentario sea positivo o negativo. Para esto se utiliza la librería de python *pysentimiento* que está diseñada para realizar análisis de sentimiento en español (Pérez et al., 2021). Este algoritmo toma como insumo el texto y retorna la probabilidad que ese sea positivo, neutro o negativo. En segundo lugar, se tabula la frecuencia con la que se utiliza cada palabra en los comentarios. Esto permite descubrir si el lenguaje es diferente por género del profesor. Sobre esto, la aproximación empírica a nivel de palabras se detalla en la sección 5.4.

Estas medidas de análisis de sentimientos parecen capturar de manera adecuada qué tan positivo o negativo es un comentario. En la Tabla 1 se encuentra que, entre los comentarios que piden resaltar los aspectos positivos del curso o del profesor, la probabilidad promedio que este sea positivo es 0.63 y 0.66 puntos porcentuales mayor a la probabilidad de que sea negativo, respectivamente. De manera análoga, entre los aspectos por mejorar, la probabilidad promedio que un comentario sea positivo respecto a la probabilidad de ser negativo es 0.4 puntos porcentuales menor entre comentarios del curso y 0.27 puntos porcentuales entre comentarios del profesor.

También se cuenta con la información del SPADIES para esta universidad<sup>17</sup>, una base de datos

---

<sup>15</sup>En los casos que hay más de un profesor para una sección de un curso, todos los profesores reciben el mismo puntaje en las preguntas relacionadas al curso.

<sup>16</sup>La variable descrita en la ecuación (1) se vuelve a estandarizar por semestre para garantizar que tenga una desviación estándar igual a 1.

<sup>17</sup>Esta es la información que la universidad le debe reportar al Ministerio de Educación Nacional. En particular, estos

que se utiliza con el fin de hacer seguimiento a los estudiantes de pregrado y prevenir la deserción. De esta fuente se tienen resultados académicos y características socio-demográficas de los estudiantes, incluidas sus notas, estrato socio-económico, educación y ocupación de sus padres, entre otra información. Finalmente, se dispone de datos acerca de los profesores en el último semestre en el que estuvieron activos en la universidad. Dicha información corresponde a su edad, el cargo docente que tienen, los años de antigüedad en su contrato y si en el primer semestre del 2020 estaban activos en la universidad.

**Tabla 1: Probabilidad promedio de clasificar un comentario como positivo, neutro o negativo, por categorías de comentario**

	Aspectos positivos del curso	Aspectos por mejorar del curso	Aspectos positivos del profesor	Aspectos por mejorar del profesor
Prob. positivo	0.68	0.08	0.72	0.13
Prob. neutro	0.27	0.44	0.22	0.47
Prob. negativo	0.05	0.48	0.06	0.40

Promedios estadísticamente distintos entre sí.

En la Tabla 2 se presenta una diferencia de medias de las principales variables por género del profesor. En esta tabla se puede ver que tan solo algunas variables tienen diferencias significativas entre profesores hombres y mujeres con una magnitud relevante. Por un lado, las evaluaciones de profesores hombres y mujeres tienen estudiantes de la misma edad, están cursando el mismo semestre, tienen el mismo promedio en el semestre, se demoran lo mismo respondiendo la encuesta y es igual de probable que cursen una carrera STEM. Además, no hay una diferencia relevante en tanto a la cantidad de créditos inscritos, la nota obtenida en el curso o si el estudiante decide comentar. Sin embargo, en promedio las profesoras son más jóvenes y con menos años de experiencia pues es más probable que sean profesores asistentes o asistentes graduadas que asociadas o titulares. Además, hay más estudiantes mujeres en cursos de profesoras, si bien la diferencia es pequeña.

datos corresponden a una versión más amplia de información reportada al Ministerio.

**Tabla 2: Diferencia de medias de características observadas de estudiantes, profesores y curso por género del profesor**

	Prof. Hombres	Prof. Mujeres	Diferencia		Prof. Hombres	Prof. Mujeres	Diferencia
Estudiante mujer	0.44	0.47	0.031***	Edad profesor	41.50	38.58	-2.920***
Edad estudiante	20.49	20.51	0.014	Influencia sobre notas	-0.03	0.07	0.107***
Semestres cursados	5.19	5.23	0.038	Profesor activo	0.78	0.74	-0.031***
Carrera STEM	0.48	0.49	0.006	Asistente graduado	0.21	0.25	0.045***
Créditos inscritos	18.50	18.58	0.088***	Profesor de cátedra	0.21	0.21	-0.005
Promedio del semestre	3.81	3.81	-0.005	Profesor asistente	0.11	0.13	0.021**
Nota del curso	3.87	3.90	0.035***	Profesor asociado	0.29	0.25	-0.042***
Créditos del curso	2.42	2.30	-0.124***	Profesor titular	0.07	0.04	-0.027***
Tiempo de respuesta (mins)	7.08	7.12	0.041	N. de estudiantes	43.60	40.75	-2.851***
Escribe comentario	0.65	0.64	-0.011**	Observaciones (Totales)	238,691	110,208	348,899

Diferencia de medias obtenida de una regresión lineal con efectos fijos de facultad del curso. Pruebas de hipótesis realizadas con errores estándar con cluster a nivel de curso. La diferencia de medias del tiempo de respuesta restringe las evaluaciones que se respondieron en menos de 3 horas debido a que es común que muchos estudiantes comiencen a responder un día y terminen en otro. Aunque el promedio de tiempo de respuesta es de 16 horas, el 85% de las encuestas se responden en menos de media hora y el 87.5% en menos de 3 horas. La influencia sobre las notas de un profesor corresponde a cuánto aporta un profesor a la nota de sus estudiantes. Para ver más detalles de esta medida ver la sección 4.1 \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

## 4 Estrategia empírica

### 4.1 Estrategia empírica principal

En este trabajo se argumenta que la percepción de un estudiante  $i$  sobre el profesor  $p$  en el curso  $c$  y periodo  $t$ ,  $Y_{ipct}^*$ , está determinada por cuatro elementos como se observa en la ecuación (2). Dicha percepción general es una variable latente y no equivale perfectamente a la evaluación que se observa en las encuestas principalmente debido a que las evaluaciones utilizan un puntaje en una escala de ordenada de 1 a 5 sobre múltiples aspectos diferentes.

$$Y_{ipct}^* = f \left( \begin{array}{cccc} \phi_{it} & , & \psi_{ct} & , & v_{pct} & , & \tau_{ipct} \\ \text{Características} & & \text{Características} & & \text{Características} & & \text{Interacción} \\ \text{profesor} & & \text{curso} & & \text{estudiante} & & \text{estudiante} \\ \text{Percepción del} & & & & & & & \text{y profesor} \\ \text{estudiante sobre} & & & & & & & \end{array} \right) \quad (2)$$

En esta estructura, que se asemeja a la de Mengel et al. (2019)<sup>18</sup>, se tiene un componente de características de los profesores para enseñar cada curso ( $\phi_{pct}$ ), que es seguido por las características

<sup>18</sup>La principal diferencia con esa estructura es que estos autores consideran que la evaluación de un estudiante es función únicamente de la nota del estudiante, el esfuerzo que le dedica a la clase y la experiencia en la clase (los “factores blandos”).

intrínsecas de cada individuo por las cuales este podría calificar más con puntajes más altos o bajos a sus profesores ( $v_{it}$ ). En tercer lugar están los factores del curso que pueden afectar el ambiente de aprendizaje pero son exógenas a las características del profesor y del estudiante ( $\Psi_{ct}$ ). Estos factores pueden ser la cantidad de estudiantes en la clase, la composición por género de la clase, entre otros. Finalmente, hay un componente de relacionamiento entre estudiantes y profesores, que incorpora lo que Mengel et al. (2019) consideran como “factores blandos” de la experiencia que un estudiante tuvo en la clase con un cierto profesor ( $\tau_{ipct}$ ). A partir de esta estructura, un sesgo de género se define como una diferencia en las evaluaciones que no está explicada por ninguno de los factores mencionados puesto que el género no debería determinar la percepción que un estudiante tiene sobre su profesor.

Con el fin de llevar esta estructura conceptual a una estrategia empírica se toma una aproximación de forma reducida en la siguiente ecuación (3):

$$Y_{ipct} = \beta_1 \text{Prof M}_p + \beta_2 \text{Prof M}_p \times \text{Est M}_i + \gamma X_{ict} + \tau W_{ct} + \delta Z_{pt} + \theta_i + \delta_t + \varepsilon_{ipct} \quad (3)$$

La variable dependiente de esta ecuación,  $Y_{ipct}$ , corresponde a una las medidas de las evaluaciones, sea el puntaje agregado del profesor o del curso, un indicador si el estudiante recomienda al profesor o la probabilidad de que el comentario en la encuesta sea positivo o negativo. Este modelo incluye un indicador del sexo del profesor junto con un diferencial por el género del estudiante. Además, se controla por características del estudiante en el vector  $X_{ict}$ , características del curso en el vector  $W_{ct}$ , características del profesor en  $Z_{pt}$  y efectos fijos de estudiante ( $\theta_i$ ) y semestre ( $\delta_t$ ). Este modelo se asemeja a la metodología principal de Boring (2017) que también incluye efectos fijos de estudiante<sup>19</sup>, aunque no modela directamente las habilidades de los profesores como sí hace este trabajo.

En particular, un modelo de regresión simple con el género del profesor resulta en resultados incorrectos debido a los factores mencionados que pueden determinar el puntaje de evaluaciones y que se pueden correlacionar con el género del profesor. En esencia, esta situación corresponde a un problema de variable omitida y por el cual la ecuación (3) busca controlar por los cuatro componentes de la estructura en la ecuación (2).

Los controles de estudiante incluidos son la edad, el promedio del semestre, el promedio acumulado, su situación académica<sup>20</sup> y el semestre cursado. Entre estos controles y los efectos fijos de estudiante se da una aproximación al componente de características de los estudiantes,  $\phi_{it}$ , del

---

<sup>19</sup>Boring también tiene una especificación de efectos fijos de profesor en vez de estudiante. Sin embargo esta especificación no permite observar diferencias por género de profesor sino de estudiante.

<sup>20</sup>La situación académica de un estudiante consiste en si este se encuentra en un periodo de prueba académica por tener un promedio acumulado inferior a 3.25, o si está en periodo de reintegro por haber sido suspendido tras durar dos periodos consecutivos en prueba académica. Esta variable no captura casos de suspensión disciplinaria.

marco conceptual<sup>21</sup>. Estos controles se incluyen con el fin de evitar una posible variable omitida de los estudiantes que pueda afectar los resultados de la evaluación de los profesores en función de su género.

Luego, se incluyen como controles a nivel del curso, en el vector  $W_{ct}$ , la facultad del curso, créditos del curso, la cantidad de estudiantes, fracción de estudiantes mujeres en el curso, cantidad de profesores que enseñan en la clase y un indicador de si el curso es magistral o complementario. Estos componentes se relacionan al factor  $\Psi_{ct}$  del marco conceptual que sugiere que hay elementos del curso que son independientes del profesor y del estudiante que pueden afectar cómo un estudiante percibe el desempeño de sus profesores. Por ejemplo, el tamaño del curso suele disminuir el puntaje de las evaluaciones de los profesores (Johnson et al., 2013; Gaviria y Hoyos, 2008) y en la Tabla 2 se evidencia que el tamaño de las clases es en promedio menor entre profesoras. Esto podría llevar a un coeficiente sesgado que indica que a los hombres les va peor en las evaluaciones.

En tercer lugar, están las características del profesor que incluyen su edad, nivel de ordenamiento (profesor de cátedra, asistente, asociado, titular o asistente graduado), antigüedad del contrato y su influencia en la nota de sus estudiantes. En conjunto, estos términos buscan medir las habilidades y experiencia de los profesores del marco conceptual ( $\theta_{pct}$ ), de los cuales, la influencia de los profesores en las notas, como aproximación a las habilidades y la generosidad al calificar, es un elemento central puesto que las mujeres parecen tener una mayor influencia que los hombres (Tabla 2 y Figura A2) y una mayor influencia está correlacionada con mejores puntajes en las evaluaciones (Tabla A1).

El último elemento es la relación entre el estudiante y profesor en la clase ( $\tau_{ipct}$ ). A partir de los datos disponibles, esta relación se aproxima a partir de la nota del estudiante y el tiempo de respuesta en la evaluación. La nota que recibe el estudiante es particularmente relevante puesto que, en los cursos con mujeres los estudiantes reciben notas ligeramente más altas (Tabla 2), pero una nota mayor está ampliamente correlacionada con una evaluación mejor (Tabla A1). La nota captura elementos del aprendizaje del estudiante, su motivación en el curso, entre otras cosas. No obstante, es improbable que estas medidas controlen todos los elementos relevantes que pueden determinar la relación entre estudiante y profesor como la capacidad que tiene un profesor para conectarse con cada estudiante.

Más allá de las características observadas de los profesores, hay una serie de características no observadas que son fundamentales. Sus habilidades y la generosidad que tienen en las evaluaciones son dos potenciales variables omitidas relevantes. En este sentido, la influencia que un profesor tiene sobre la nota de sus estudiantes captura parte de las habilidades del profesor en mejorar el rendimiento de sus estudiantes como de su generosidad en las calificaciones. Para capturar la

---

<sup>21</sup>Dada la cantidad de veces que se ve a los estudiantes, es posible incluir efectos fijos de estudiante  $\times$  semestre para ser flexible en eventos por semestre para cada estudiante como estar en prueba académica. Incluir este nivel de efectos fijos no tiene un efecto importante sobre los resultados.

influencia de un profesor sobre las notas de los estudiantes se utiliza un modelo de valor agregado a partir de estimar la ecuación (4), que sigue de cerca a Chetty et al. (2014)<sup>22</sup>. En esta ecuación, la variable dependiente es la nota que un estudiante  $i$  obtuvo en la clase  $c$  con el profesor  $p$  en el semestre  $t$ . Mientras que al lado derecho están las características del estudiante y del curso junto con efectos fijos de profesor-semestre ( $\tilde{\phi}_{pt}$ ).

$$Nota_{ipct} = \alpha \tilde{X}_{ipct} + \kappa \tilde{W}_{ct} + \tilde{\theta}_i + \tilde{\phi}_{pt} + v_{ipct} \quad (4)$$

En la ecuación (4) incluye controles a nivel del estudiante ( $\tilde{X}_{ipct}$ ) que pueden medir sus habilidades y desempeño académico como su edad, promedio acumulado, créditos inscritos ese semestre, el promedio para las demás materias en el semestre, semestre cursado y su situación académica. Adicionalmente, se incluyen efectos fijos de estudiante para controlar por características intrínsecas no observadas ( $\tilde{\theta}_i$ ). También se incluyen controles a nivel de curso ( $\tilde{W}_{ct}$ ) que pueden afectar el desempeño del estudiante como la cantidad de profesores que enseñan la clase, si es un curso magistral o complementario y la cantidad de estudiantes en el curso<sup>23</sup>.

Después de estimar este modelo, se construye la variable  $m_{pt} = \frac{\tilde{\phi}_{pt}}{\sigma_\phi}$ , donde  $\sigma_\phi$  es la desviación estándar de los coeficientes de efectos fijos. Este término  $m_{pt}$  se incluye como un polinomio de grado tres entre el vector de controles  $\theta_{pct}$  de la ecuación (3) para incorporar estas características no observadas de los profesores de manera flexible<sup>24</sup>. Al controlar por esta medida, que busca capturar en parte las habilidades de los profesores, se evita que haya un problema de variable omitida. En la Tabla A5 se muestra el ajuste de este modelo y de sus componentes y se evidencia que el ajuste del modelo cae considerablemente si se excluyen los efectos fijos del profesor ( $\tilde{\phi}_{pt}$ ).

Debido a que los profesores asignan las notas esto implica que la influencia de los profesores sobre la nota de los estudiantes contiene principalmente dos elementos: las habilidades y la generosidad al calificar. Puesto que no es una medida exclusiva de las habilidades puede ser de

---

<sup>22</sup>Esta estimación es una simplificación de la que se detalla en el apéndice 1 de Chetty et al. (2014). Estos autores buscan un predictor del valor agregado del profesor  $p$  en el semestre  $t$  a partir de información de otros semestres que el profesor enseña. En este contexto eso implica perder información del 25% de los profesores por la alta rotación de asistentes graduados, aunque solamente correspondan al 6% de las evaluaciones, y por eso no se sigue tan de cerca la metodología. Adicionalmente, el motivo del ajuste de estos autores se debe a un problema que únicamente aparece cuando la regresión de interés tiene como variable dependiente la nota del estudiante. De igual forma, los resultados se mantienen si se mide el valor agregado exactamente con la metodología de los autores (resultados disponibles bajo pedido).

<sup>23</sup>Chetty et al. (2014) no incluye efectos fijos de estudiante ni características del curso. En particular estos controles ayudan a aliviar un problema por posible selección de los estudiantes a las clases que toman. Para una revisión de modelos de valor agregado, referirse a Koedel et al. (2015).

<sup>24</sup>Bjerk y Ozbeklik (2018) también utilizan un modelo de valor agregado para estudiar sesgos de género en un contexto de evaluaciones de rectores de colegio sobre sus profesores y tienen medidas para habilidades en lectura y matemáticas de sus estudiantes. Mengel et al. (2019) usa una medida similar para caracterizar posibles efectos heterogéneos, pero no lo utilizan como un control en sus estimaciones. Keng (2020) tiene un enfoque muy similar para entender los mecanismos de los sesgos.

preocupación, como sugiere [Keng \(2020\)](#) y, por eso, ese trabajo propone utilizar las notas de cursos siguientes. Por ejemplo, se estima cuánto aportó al aprendizaje un profesor de cálculo I cuando sus estudiantes toman cálculo II. Esto elimina cualquier componente de calificación y evaluaciones puesto que el profesor del primer curso no influye en la calificación del siguiente curso, pero sí en qué tan bien están preparados sus estudiantes. Debido a esto, en este trabajo se evita utilizar el término “valor agregado” para definir la variable  $m_{pr}$ . De contar con información de datos no anonimizados de cursos se podrían calcular exclusivamente las habilidades. Sin embargo, los resultados de sesgos de género de [Keng \(2020\)](#) no varían considerablemente con ambas aproximaciones, lo que sugiere que utilizar la medida de influencia en las notas puede capturar de manera apropiada la varianza en habilidades de los profesores.

En el modelo principal (3), el coeficiente  $\beta_1$  indica cómo un estudiante hombre evalúa en promedio a una profesora respecto a sus pares masculinos, mientras el término  $\beta_1 + \beta_2$  indica cómo una estudiante mujer evalúa en promedio a una profesora respecto a sus pares masculinos. De tal forma que, en promedio, no hay sesgos si  $\beta_1 = \beta_2 = 0$ . Es de particular interés descubrir si hay diferencias en los sesgos por género del estudiante dado los resultados de [Boring \(2017\)](#) y [Mengel et al. \(2019\)](#) en los que los sesgos provienen principalmente de los estudiantes hombres. El fundamento de estas diferencias proviene de trabajos previos en los que es probable que un estudiante tenga una mayor afinidad con un profesor con el que sienta una mayor cercanía y que resulta ser un profesor de su mismo género ([Crosby y Reinardy, 1993](#); [Das y Das, 2001](#)).

Con esta metodología, el supuesto de identificación de los sesgos es que, condicional en las características de los estudiantes, cursos y profesores, y los efectos fijos de estudiante, el género del profesor no se correlaciona con alguna variable omitida. En otras palabras, estas características modelan correctamente los componentes descritos en el marco teórico que pueden afectar tanto el género del profesor como las evaluaciones. Con esto, una probable amenaza a esta identificación es que los estudiantes tienen cierta capacidad de elección sobre sus clases. Esto se hace improbable con los efectos fijos de estudiante pues se logra observar a los estudiantes en muchos cursos por lo que se espera capturar apropiadamente la motivación intrínseca de cada estudiante a favorecer cursos con ciertos profesores. Como medida de robustez, en la sección de resultados se muestra evidencia de los cursos de la facultad de medicina y de los estudiantes de primeros semestres en los que la posibilidad de elección es limitada y en los cuales no hay diferencias importantes en los resultados.

Determinar la existencia de sesgos en un contexto en el que los estudiantes no puedan escoger sus cursos (como en [Boring \(2017\)](#) y [Mengel et al. \(2019\)](#)) implica estudiar pocos cursos de ciertas áreas, sin tener un resultado generalizable, como sugiere [Keng \(2020\)](#). Además, estos casos tampoco garantizan resultados insesgados. Sin incorporar medidas de características no observadas de los profesores. Por ejemplo, puede haber aleatoriedad en los estudiantes asignados, pero no en

las habilidades de los profesores (u otras características) con respecto a su género, lo que lleva a resultados incorrectos.

## 4.2 Efectos heterogéneos

Este trabajo utiliza “árboles causales honestos” de [Athey y Imbens \(2016\)](#) para determinar en cuáles contextos hay sesgos de género y en cuáles no. Este “árbol” categoriza los efectos heterogéneos a partir de las covariables disponibles de manera no lineal. Con base en un árbol de regresión, se utilizan los controles disponibles para predecir la variable dependiente partiendo la muestra a partir de cortes en las covariables tal que se maximice la diferencia en la variable dependiente entre profesores hombres y mujeres. Al final, cada nodo final del árbol caracteriza una sub-muestra que corresponde a un efecto heterogéneo. Esta metodología permite descubrir efectos heterogéneos, a partir de interacciones y no linealidades entre las variables, en las evaluaciones de profesores que se beneficia de la riqueza en la información disponible en este trabajo.

El procedimiento de implementación es el siguiente: I) se divide la muestra de datos en dos partes de manera aleatoria; II) en una de las muestras se estima un árbol de regresión cuya variable a predecir es  $\hat{y}_{ipct}$  y los predictores son todas las características disponibles del estudiante, el curso y el profesor, donde  $\hat{y}_{ipct}$  es la variable  $y_{ipct}$  después de ser regresada contra los controles que se incluyen en la regresión principal junto con los efectos fijos de estudiante y semestre. Este paso es fundamental porque evita que el árbol de predicción encuentre relaciones en la variable dependiente provenientes de covariables incluidas en los controles. Por último, III), se estima una regresión con efectos heterogéneos como se especifica en la ecuación (5) en la segunda muestra aleatoria<sup>25</sup>. Para estimar apropiadamente la regresión, la variable del sexo del profesor también es regresada contra los controles y efectos fijos<sup>26</sup>. En esta ecuación,  $h$  representa la sub-muestra caracterizada por cada nodo final del árbol de regresión. Con esto, se estiman  $H$  coeficientes de interés que caracterizan los efectos heterogéneos.

$$\hat{y}_{ipct} = \sum_{h=1}^H \beta_0^h + \sum_{h=1}^H \beta_1^h \widehat{\text{Prof M}_p} + \varepsilon_{ipct} \quad (5)$$

En un ejercicio adicional se repite el paso II) un total de 1,000 veces, cambiando las variables utilizadas y en sub muestras, con el fin de estimar una gran cantidad de árboles distintos y tener una medida más robusta de la distribución de los efectos heterogéneos. Esto consiste en un “bosque causal honesto” que se inspira en el “random forest” ([Wager y Athey, 2018](#)). Esta estrategia permite calcular un coeficiente para cada evaluación como el promedio de los coeficientes estimados.

<sup>25</sup>Se considera que el árbol causal es “honesto” al utilizar una muestra para estimar el árbol y otra muestra para estimar los coeficientes de la regresión. Esto evita un posible sobre ajuste.

<sup>26</sup>Por simplicidad, se muestra únicamente un efecto para todos los estudiantes, en vez de uno diferencial por género del estudiante.

dos a través de los distintos árboles. Con el bosque causal se tiene una medida más robusta de la distribución de los efectos heterogéneos que con un único árbol, pero se pierde la interpretación de las particiones.

## 5 Resultados

### 5.1 Resultados principales

La Tabla 3 presenta los resultados de la regresión principal para las tres variables de resultado de los puntajes y las medidas de qué tan positivos o negativos son los comentarios. Con esto se puede afirmar que el sesgo promedio enfrentado por una profesora al ser mujer existe, pero es pequeño. Entre las preguntas acerca del profesor la diferencia es de 0.03-0.04 desviaciones estándar. Mientras que para el puntaje del curso el efecto es menor y no alcanza a ser significativo, algo razonable dado que los puntajes del curso son iguales para profesores distintos que enseñan un mismo curso. La diferencia en la probabilidad de que una mujer sea recomendada para una clase es de un punto porcentual menor, sin embargo este efecto es muy pequeño dado que el 90 % de los estudiantes recomiendan tomar clase con sus profesores. Finalmente, también hay un sesgo de magnitud similar al de la primera columna contra las mujeres en la probabilidad que los comentarios sean positivos o negativos. Las profesoras enfrentan una probabilidad ligeramente menor de tener comentarios positivos, y mayor de tener comentarios negativos, equivalentes a 0.03-0.04 desviaciones estándar. Además, la Tabla A12 indica que en las preguntas que se refieren a las habilidades docentes tienen el mayor sesgo mientras que no hay sesgos en las preguntas asociadas a retroalimentación y trato a los estudiantes.

**Tabla 3: Diferencias en las evaluaciones por género del profesor**

	Puntaje del profesor (1)	Puntaje del curso (2)	Recomienda al profesor (3)	Comentario positivo (4)	Comentario negativo (5)
Profesor mujer ( $\beta_1$ )	-0.035*** (0.011)	-0.016 (0.010)	-0.012*** (0.004)	-0.012*** (0.005)	0.014*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.009 (0.010)	0.015 (0.009)	0.005 (0.003)	0.007* (0.004)	-0.004 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
Obs	349,541	350,214	302,087	226,619	226,619
R <sup>2</sup>	0.25	0.28	0.14	0.20	0.19
Prom. var. dep.	-0.01	-0.01	0.89	0.52	0.21
$\beta_1 + \beta_2$	-0.044***	-0.001	-0.007*	-0.005	0.010***

Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Las observaciones varían por columna porque no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.

Por su parte, en todas las variables de resultado, a excepción de las preguntas del profesor, el sesgo es menor para las estudiantes mujeres que para los estudiantes hombres. Únicamente es significativa la diferencia entre estudiantes hombres y mujeres en la probabilidad de tener un comentario positivo, pero en tres casos el sesgo de las estudiantes mujeres no es significativo al 5 %. Este es un indicador sugestivo de mayores sesgos por parte de los hombres que de las mujeres, alineado con los resultados de [Basow y Silberg \(1987\)](#), [Boring \(2017\)](#), [Centra y Gaubatz \(2000\)](#), [Mengel et al. \(2019\)](#) y [Keng \(2020\)](#) donde la mayoría de los sesgos provienen de estudiantes hombres.

Al omitir los controles y los efectos fijos de estudiante se obtienen resultados que sugieren la ausencia de sesgos. Una regresión simple como la del panel A de la Tabla A6 sugiere erróneamente que, en promedio, no existe ningún sesgo. Una vez se incluyen los controles se encuentran sesgos (panel B). Específicamente, esta diferencia proviene de variables como la nota del estudiante obtenida en el curso, las características del profesor como su nivel de ordenamiento y la influencia sobre las notas del estudiante (ver Tabla A11), ampliamente correlacionadas con las evaluaciones de los profesores y mayores entre las profesoras (ver Tablas 2 y A1 y las Figuras A2 y A1).

Al incluir efectos fijos de curso en la especificación principal (ecuación 3) se capturan diferencias provenientes de los distintos tipos de curso pero, únicamente se aprovecha la variación de cursos enseñados por profesores de ambos géneros, correspondientes al 40 % de las evaluaciones. La Tabla A7 muestra que incluir estos efectos fijos no tiene prácticamente ninguna diferencia sobre

<sup>26</sup>El experimento aleatorio hipotético en este caso consiste en asignar un profesor a muchos cursos distintos pero intercambiando su género. Con lo que el tratamiento es a nivel de curso, y por lo tanto los clusters son a este nivel.

los sesgos estimados. Más allá de esto, incluir estos efectos fijos genera una selección no aleatoria en los cursos que puede afectar los sesgos calculados<sup>27</sup>.

En general, estos resultados contrastan con varios trabajos recientes que sugieren amplias brechas en las evaluaciones entre hombres y mujeres. Por ejemplo, [Centra y Gaubatz \(2000\)](#) documenta seis trabajos realizados en los años 80 y 90 en los cuales no hay diferencias por género del profesor y cinco trabajos en los que sí hay diferencias. No obstante, el problema de estos trabajos es la falta de una identificación limpia en la que se pueda afirmar que las diferencias son a causa del género del profesor y que cuenten con una amplia cantidad de evaluaciones.

Los trabajos más recientes, con una mejor identificación, han documentado diferencias importantes. [Mengel et al. \(2019\)](#) encuentra que las evaluaciones son 0.21 desviaciones estándar más bajas para las mujeres entre los estudiantes hombres y 0.08 para las estudiantes mujeres. Por su parte, [Boring \(2017\)](#) documenta diferencias en puntajes equivalentes a 0.2 desviaciones estándar. [Y Keng \(2020\)](#), con una identificación más cercana a la del presente trabajo, encuentra que las evaluaciones de las mujeres son entre 0.11 y 0.16 desviaciones estándar más bajas.

¿Por qué los resultados de este trabajo sugieren un sesgo menor? Puede que las diferencias en los sesgos entre trabajos correspondan parcialmente a heterogeneidad en habilidades no incorporada en estos trabajos. Asimismo, las diferencias en los sesgos pueden estar asociadas al contexto de cada estudio. Por ejemplo, en [Boring \(2017\)](#), las profesoras tienen una experiencia mucho menor que los hombres. En específico, el 43 % de las mujeres son estudiantes de doctorado y el 17 % son de planta, mientras que el 26 % de los hombres son estudiantes de doctorado y el 41 % son de planta, a diferencia de esta universidad en Colombia en la cual hay mayor similitud en la composición de categorías por género (ver Tabla 2). Otros factores como la composición de género del cuerpo estudiantil pueden ser relevantes, como se muestra en la siguiente sección. La universidad del presente trabajo está más cerca de tener balance de género entre los estudiantes que la universidad estudiada por [Mengel et al. \(2019\)](#) en la que apenas el 38 % son mujeres.

Sumado a lo anterior, estos trabajos no encuentran de manera inequívoca sesgos en todos los contextos en esas universidades. [Mengel et al. \(2019\)](#) encuentra que los efectos son negativos únicamente para profesoras que son estudiantes de doctorado. Mientras que entre profesoras de planta y cátedra los efectos son nulos entre estudiantes hombres pero positivos (y de hasta 0.26 desviaciones estándar) entre estudiantes mujeres. Y más allá de esto, existe una gran posibilidad que esta literatura sufra de sesgo de publicación. Los trabajos que tienen mayor probabilidad de ser publicados son aquellos que encuentran sesgos contra las mujeres, por lo cual se desconocen resultados en contextos en los cuales no hay diferencias por género o en los que las mujeres aventajan a sus pares.

---

<sup>27</sup>[Keng \(2020\)](#) y [Mengel et al. \(2019\)](#) incluyen efectos fijos de curso en sus especificaciones principales.

## 5.2 Robustez de resultados principales

Los resultados de la Tabla 3 no parecen ser explicado por la posibilidad que tienen los estudiantes de escoger a sus profesores. Por una parte, al incluir efectos fijos de estudiante disminuye la preocupación por la elección de los estudiantes de ciertos profesores debido a sus características no observadas. Por otra parte, entre los estudiantes de la facultad de medicina, donde no es posible escoger los cursos, los sesgos son similares. La Tabla A3 indica que, entre estos estudiantes, únicamente entre las estudiantes mujeres hay un sesgo un poco mayor con respecto al del resultado principal.

Además del caso anterior, se puede aprovechar que la facilidad de inscribir cursos depende de la cantidad de créditos aprobados. Al estimar los efectos por semestre cursado del estudiante, tampoco se ve una diferencia sustancial entre los coeficientes (Figura A3). Es más, en el primer semestre, en el cual los estudiantes tienen el primer turno de elección, en promedio no hay diferencias. Este hallazgo va contra la idea que los sesgos son más bajos en esta universidad debido a los estudiantes escogen a sus profesores.

Así como los sesgos no dependen de poder escoger a los profesores, tampoco dependen del interés en responder la encuesta. Por un lado, las estimaciones principales controlan por el tiempo de respuesta de los estudiantes como proxy de su interés en la encuesta. Por otro lado, en la Tabla A8 del apéndice se comprueba que los coeficientes son prácticamente iguales cuando se excluyen las observaciones de los estudiantes que parecen menos interesados en la encuesta. Lo cual se realiza excluyendo dos grupos de observaciones: 1) los que no escriben comentarios para esa evaluación y 2) los estudiantes marcan exactamente la misma respuesta para todas las preguntas.

En otros aspectos, puede ser problemático que sea opcional escribir comentarios en las evaluaciones debido a que los estudiantes que comentan quizás son los más interesados y que podrían ser menos propensos a penalizar a un profesor por su género en las encuestas. Esto afectaría la validez de los resultados de las columnas 4 y 5 de la Tabla 3. Para aliviar esta preocupación, se utiliza un modelo de selección de Heckman en el que la restricción de exclusión es la fracción de encuestas en las que cada estudiante comenta en ese semestre, excluyendo la observación analizada<sup>28</sup>. Por lo que, condicional en las características observadas del estudiante, si este comenta más evaluaciones en un semestre es más probable que también comente en la evaluación *ipct*<sup>29</sup>. Esto corresponde a un caso en el que escribir menos comentarios está asociado a una menor disponibilidad de tiempo y en el que la motivación a comentar varía por semestre. Parte de esto parece ser válido debido a que, en la muestra observada, el 95.5 % de los estudiantes comenta en al menos una encuesta en cada

<sup>28</sup>Esta es una variación de la restricción de exclusión que utiliza Mengel et al. (2019) para modelar la decisión de responder la encuesta. Ellos utilizan la probabilidad de respuesta de encuestas en el pasado para el estudiante.

<sup>29</sup>Se realiza la estimación sin efectos fijos de estudiante para evitar un problema de parámetros incidentales y por limitaciones computacionales.

semestre. Lo que sugiere que, comentar no parece estar asociado a características no observadas de algunos estudiantes. En la Tabla A4 se puede ver que los coeficientes cambian de manera marginal respecto a la estimación principal, a pesar que hay evidencia de algo de selección.

Finalmente, para comprobar la robustez de los resultados de las medidas agregadas de curso y profesor se utiliza una medida calculada con un modelo de escala de respuesta (RSM)<sup>30</sup>. Esta forma de calcular puntajes considera que la percepción latente del estudiante sobre su profesor se puede medir incorrectamente si no se consideran las diferencias entre las preguntas y cómo se decide responder cada una. Por ejemplo, la diferencia entre estar “completamente de acuerdo” y “de acuerdo” con que el profesor es puntual en clase es distinta a considerar si este explica bien los temas de clase. Con este modelo RSM se construyen puntajes latentes para cada evaluación y se estandarizan. Los coeficientes encontrados con los puntajes de este modelo, ver Tabla A10, sugieren que los sesgos son muy similares a los presentados en la Tabla 3.

### 5.3 Efectos heterogéneos

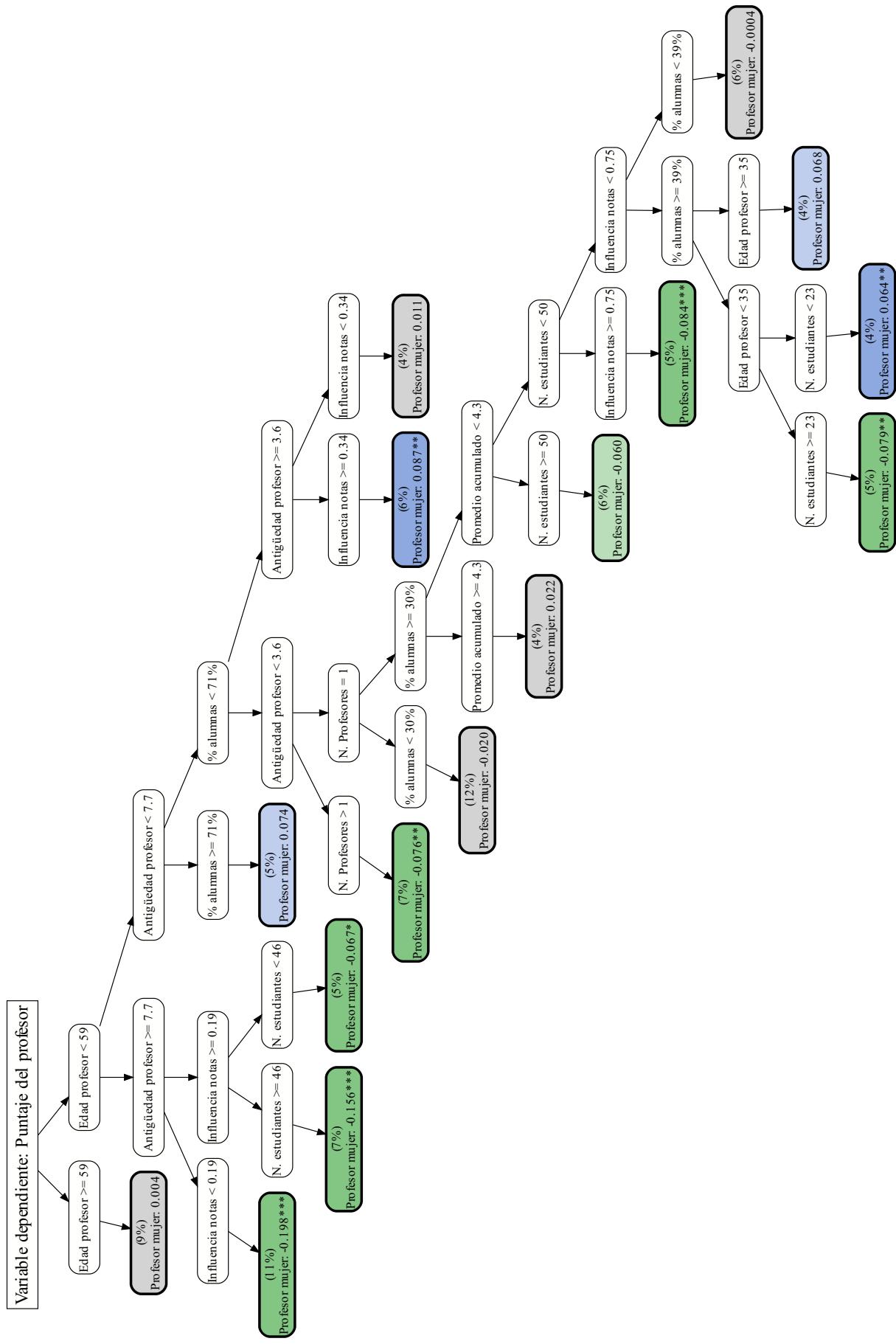
Los resultados del árbol causal sugieren que, detrás de los efectos promedio que parecen pequeños, en realidad hay una amplia heterogeneidad en los sesgos de género. En la Figura 1 se puede ver las predicciones de los sub-grupos en los que las ramificaciones más altas son las de mayor importancia y los nodos finales indican el coeficiente para el género del profesor entre las evaluaciones caracterizadas. Se observa que los sesgos se comportan considerablemente de manera no lineal en las variables observadas. No obstante, resaltan algunas generalidades: 1) los sesgos son mayores en las evaluaciones de profesores con menor experiencia (menos de 59 años) y 2) cuando la cantidad de estudiantes es mayor (más de 45-50 estudiantes). En estos casos las profesoras pueden percibir una evaluación que alcanza a estar entre 0.08 y 0.20 desviaciones estándar por debajo de los hombres. Además, estos casos corresponden al 40% de todas las evaluaciones respondidas (azul oscuro en la figura).

Por otro lado, los casos en los que no hay diferencias o estas favorecen a las mujeres se destacan tres caracterizaciones: 1) cuando los profesores tienen más experiencia (más de 59 años), 2) cuando hay muchas estudiantes mujeres en el curso (más del 20%) y 3) cuando tienen un rango de experiencia entre 3.5 y 8 años. Sin embargo, los casos con sesgos positivos tienen diferencias menores, que rondan entre 0.06 y 0.13 desviaciones estándar, y representan apenas el 20% de las evaluaciones (verde oscuro en la figura).

Al realizar este mismo ejercicio para las demás variables de resultado se encuentran patrones muy similares. Por ejemplo, en el caso de las preguntas asociadas al curso se encuentra que las evaluaciones en las que las mujeres son penalizadas se caracterizan por variables y particiones

<sup>30</sup>El Centro de Evaluación de esta universidad utiliza una versión de este modelo conocida como modelo de respuesta graduada. Para más información sobre teoría de respuesta al ítem ver ([Engelhard, 2014](#); [Tyek Han, 2013](#))

**Figura 1: Efectos heterogéneos del género del profesor en los puntajes del profesor**



Estimación para el árbol causal honesto. El árbol se estima en una muestra aleatoria y luego se estiman las regresiones en la muestra restante, cada una cuenta con 174,731 observaciones. Orden de particiones del árbol van de arriba del nodo inicial hacia los nodos finales. En la estimación se dejan al menos 2,000 observaciones de cada género de profesor en un nodo final antes de podar el árbol. El árbol se poda bajo el criterio de menor error de validación cruzada, en un proceso de 5 iteraciones de validación cruzada en la muestra aleatoria de estimación del árbol. Nodos finales muestran en paréntesis el porcentaje de las observaciones de la muestra de estimación de regresiones. Colores de los nodos finales indican signo y significancia de los coeficientes: tonos verdes implican coeficientes positivos y gris para coeficientes cercanos a cero. Tonos oscuros representan casos de significancia estadística, tonos claros son casos de coeficientes que no son cercanos a cero pero no significativos. La variable dependiente es regresada contra todos los controles de la estimación principal y efectos fijos de estudiante y semestre antes de la estimación del árbol causal. Errores estándar con cluster a nivel de curso. \*p<0.10, \*\*p<0.05, \*\*\*p<0.01.

cercanas al árbol de la Figura 1. En esa variable hay grupos de evaluaciones en las que los sesgos contra las mujeres alcanzan 0.19 desviaciones estándar. A su vez, hay algunos contextos en los que las mujeres son evaluadas de manera más favorable en 0.10 desviaciones estándar. Nuevamente, los contextos en los que hay sesgos negativos ocurren en casi el doble de evaluaciones que en los que hay sesgos positivos<sup>31</sup>.

Con el bosque causal se encuentra que los efectos heterogéneos pueden ser incluso más grandes que lo sugerido por el árbol. La Tabla 4 muestra que al rededor del 40 % de las evaluaciones tienen una penalización hacia las mujeres de al menos 0.1 desviaciones estándar, mientras que menos del 10 % de las evaluaciones obtienen el resultado opuesto. Sin embargo, esta tabla muestra que, apenas el 10 % de las evaluaciones tienen un sesgo de una magnitud similar al resultado promedio encontrado por Mengel et al. (2019) y el 30 % de las evaluaciones penalizan a las profesoras de manera semejante a lo sugerido por Keng (2020).

**Tabla 4: Deciles de la distribución de los efectos del género del profesor en el puntaje sobre el profesor**

	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %
Estudiante mujer	-0,229	-0,169	-0,130	-0,096	-0,065	-0,035	-0,003	0,035	0,089
Estudiante hombre	-0,244	-0,182	-0,140	-0,106	-0,075	-0,045	-0,013	0,024	0,076

Estimación para el bosque causal honesto. Se utilizan 1,000 árboles para estimar los coeficientes. A cada observación se le asigna como coeficiente el promedio de los coeficientes a través de los distintos árboles estimados. En cada árbol se utiliza una parte de las covariables para evitar una alta correlación entre los resultados de los distintos árboles. Se calibran los parámetros de mínima cantidad de observaciones en cada nodo, el máximo desbalance del tratamiento en cada nodo, la penalización de dicho desbalance, la fracción de la muestra utilizada para las particiones honestas, la fracción de la muestra utilizada para crear cada árbol y la cantidad de variables utilizadas en cada árbol en un proceso de validación cruzada. La variable dependiente e independiente es regresada primero contra todos los controles de la estimación principal y los efectos fijos de estudiante y semestre antes de la estimación del bosque para encontrar únicamente efectos heterogéneos.

Ahora, más allá de los resultados del árbol causal hay algunos factores que pueden ser causantes de diferencias y por eso se exploran directamente. Por un lado, existe un debate sobre si en los cursos más intensivos en matemáticas y en las ciencias naturales las mujeres son penalizadas fuertemente. Una hipótesis que surge debido al contexto particularmente discriminatorio afrontado por las mujeres en carreras STEM donde la fracción de mujeres es considerablemente baja<sup>32</sup>. En segundo lugar, qué relevante es el efecto encontrado en el árbol acerca de la representación de estudiantes mujeres en las clases. Tercero, si el sesgo disminuye con la experiencia de los profesores. Así como lo sugiere el árbol causal, las mujeres con mayor experiencia tienen puntajes menos desfavorables que sus pares masculinos. Por lo que algo que esta heterogeneidad puede provenir de un efecto de modelo a seguir en el que las mujeres resaltan sobre sus pares debido a que tienen un cargo alto siendo la minoría en dichos cargos más altos. En este caso, uno esperaría que si hay un efecto de modelo a seguir, el sesgo disminuye entre profesores con más experiencia. Y, finalmente,

<sup>31</sup>Estos árboles se encuentran disponibles bajo pedido.

<sup>32</sup>Siglas en inglés para ciencia, tecnología, ingeniería y matemáticas.

se considera qué tan generalizables son los resultados encontrados para esta universidad para el resto de los estudiantes universitarios del país.

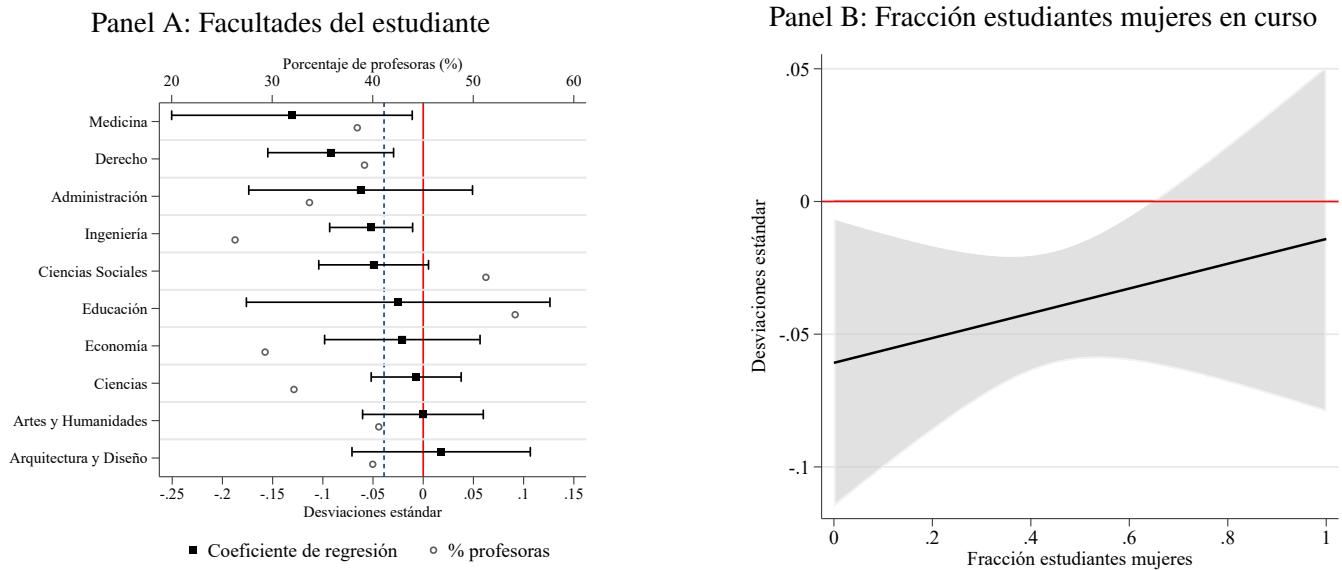
Sobre las diferencias por facultades, el Panel A de la Figura 2 muestra que no hay grandes diferencias. La hipótesis que las carreras STEM tienen un contexto en el que las mujeres son más desfavorecidas que en las demás carreras no parece tener lugar. Incluso al nivel de la carrera del estudiante como en la Figura A4 se puede ver que entre las carreras con mayor penalización están tanto algunas ingenierías como carreras de humanidades y ciencias sociales. Además, en matemáticas e ingeniería eléctrica las mujeres son penalizadas menos que en la mayoría de las carreras. Esto contrasta con la evidencia de Mengel et al. (2019) que señala que en los cursos más intensivos en matemáticas es donde los estudiantes están más sesgados contra las mujeres<sup>33</sup>.

Aunque no parece haber relación entre la cantidad de profesoras mujeres en cada facultad, sí hay un rol importante en la cantidad de estudiantes mujeres en cada curso. El panel B muestra que la representación femenina entre estudiantes es determinante: entre menos mujeres haya mayores son los sesgos . En particular, los efectos son negativos hasta que se alcanza un 65 % de mujeres en la clase. En un curso mediano, con 43.5 % de mujeres, el sesgo promedio es de 0.05 desviaciones estándar. Además, entre los profesores de planta, el sesgo en el curso mediano es más pronunciado con 0.08 desviaciones estándar.

---

<sup>33</sup>Estos autores pueden comparar únicamente cursos dentro de una escuela de negocios, a diferencia de comparar programas académicos.

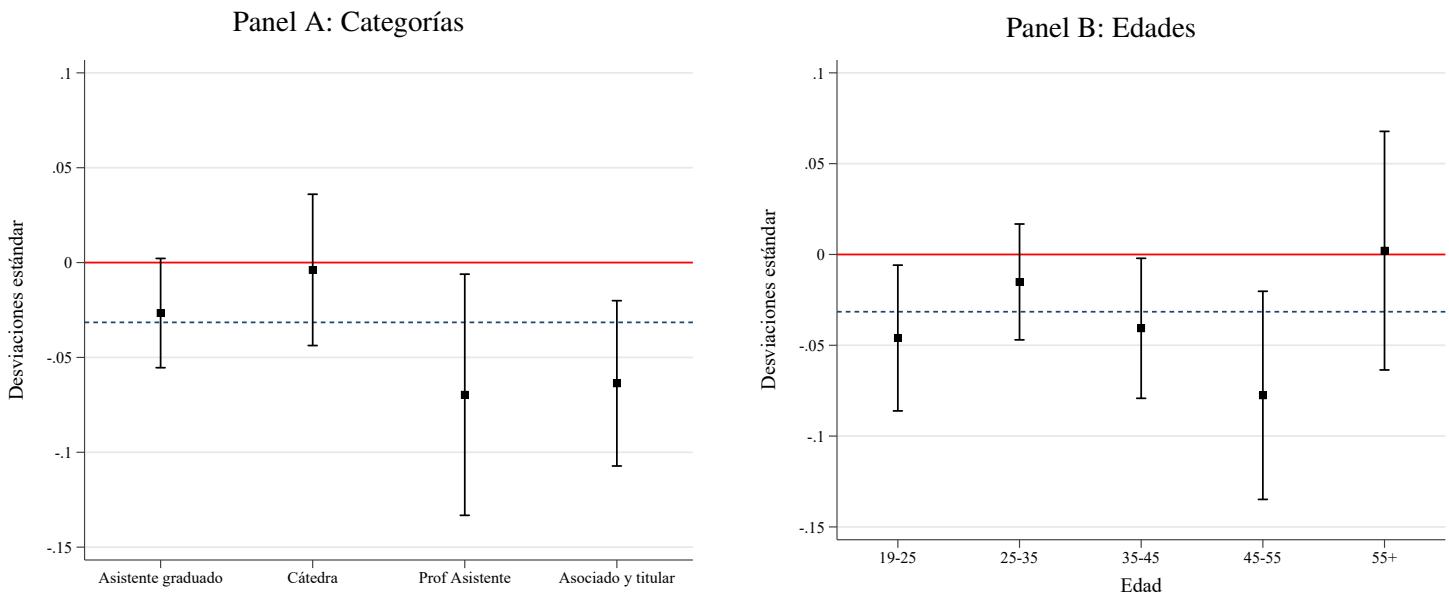
**Figura 2: Efecto del género del profesor en sus puntajes. Por facultades del curso y fracción de estudiantes mujeres en el curso**



Errores estándar con cluster a nivel de curso, intervalos de confianza al 95 %. Panel A: Línea punteada representa el sesgo promedio, equivalente al de la Tabla 3. Panel B: Línea sólida corresponde a coeficiente de sexo del profesor e interacción con la cantidad de mujeres estudiantes. Líneas punteadas: intervalos de confianza de 95 %. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: categorías de edad, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres en el curso, cantidad de profesores que dictan el curso, créditos del curso y un indicador de si es curso complementario. Se incluyen efectos fijos de semestre. Los profesores de cátedra menores de 30 años se categorizan como asistentes graduados.

En tanto a la experiencia, el Panel A de la Figura 3 muestra que las profesoras penalizadas son de planta (asociadas, asistentes y titulares) y no las asistentes graduadas o profesoras de cátedra. Las diferencias entre los profesores de planta son más notorias, con un sesgo de 0.07 desviaciones estándar. Lo que coincide con el Panel B que indica que estos sesgos se presentan entre profesores en un rango de edad entre 35 y 55 años. Una posible conjectura es que esas evaluaciones reflejan discriminaciones en el mercado laboral así como que a estas edades las mujeres tienen hijos y pueden estar dedicando un mayor tiempo a su crianza. Esto resulta preocupante en cuanto a que en estas edades los profesores están avanzando en sus carreras profesionales. No obstante, así como lo sugería el árbol causal, cuando aumenta la experiencia incluso más, no hay diferencias. Una explicación plausible es que solamente el 25 % de los profesores con más de 55 años son mujeres por lo que podrían tener un mayor reconocimiento.

**Figura 3: Efecto del género del profesor en sus puntajes y fracción de profesoras por grupos de edad y categorías del profesor**



Errores estándar con cluster a nivel de curso, intervalos de confianza al 95 %. Línea punteada representa el sesgo promedio, equivalente al de la Tabla 3. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: categorías de edad, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres en el curso, cantidad de profesores que dictan el curso, créditos del curso y un indicador de si es curso complementario. Se incluyen efectos fijos de semestre. Los profesores de cátedra menores de 30 años se categorizan como asistentes graduados.

En línea con lo anterior, resulta preocupante que los sesgos estén presentes entre estudiantes de semestres 6 a 9, como indica la Figura A3. Esto sugiere que en los cursos electivos (que usualmente comienzan en el 5to o 6to semestre) hay mayores sesgos. Un resultado sorprendente puesto que en estos cursos los profesores deben tener más experiencia en los temas del curso y es justo en los cursos en los que las mujeres al parecer resultan desfavorecidas.

Finalmente, estos resultados no parecen provenir de la composición particular de esta universidad. La mayoría de los estudiantes provienen de contextos muy favorables económicamente y no representan al estudiante universitario promedio del país, como se puede ver en la Tabla A2. Sin embargo, no hay diferencias en las evaluaciones dependiendo del estrato socio-económico del estudiante, el nivel educativo o la ocupación de los padres del estudiante. Por una parte, en el árbol causal estas variables no son relevantes en ninguna partición. Y además, al hacer explícita la heterogeneidad en un modelo lineal, los sesgos de género son iguales en todos los estratos, niveles educativos y ocupaciones de la madre del estudiante. En particular, no se puede rechazar la hipótesis que los coeficientes de la diferencia entre hombres y mujeres son iguales para ninguna de estas tres variables al 10 % de significancia<sup>34</sup>.

<sup>34</sup>Resultados disponible bajo pedido.

## 5.4 Análisis de palabras de los comentarios

La información disponible sobre los comentarios es particularmente útil para dar evidencia sobre cómo los estudiantes evalúan a sus profesores puesto que tienen libertad para escribir lo que deseen. Además del análisis de sentimientos para descubrir si los comentarios son positivos o negativos, también se puede ver si las palabras utilizadas son distintas entre hombres o mujeres. Con un modelo de penalización Lasso con distribución logística se busca predecir el género del profesor correspondiente al comentario a partir de las palabras utilizadas así como Wu (2018). Para este modelo de predicción se usan las 5,000 palabras que aparecen con mayor frecuencia entre todos los comentarios, las cuales representan el 98.3 % de todas las palabras<sup>35</sup>.

El modelo Lasso identifica qué palabras predicen el género del profesor asociado al comentario a través de una penalización que asegura que si una palabra tiene una capacidad de predicción muy baja esta se lleve hasta cero. Los coeficientes de la regresión Lasso se obtienen de resolver el problema de minimización de la ecuación (6) en donde el parámetro de penalización es  $\lambda$ . Además, se incluyen efectos fijos de facultad ( $\gamma_f$ ), sin tener penalización del Lasso, para evitar encontrar correlaciones entre una palabra y el género del profesor que sean explicadas por el contexto de las facultades.

$$\beta^{Lasso} = \arg. \min_{\beta} \sum_i \left[ y_i \times (\beta_0 + \mathbf{x}'_i \beta + \gamma_f) - \log \left( 1 + e^{(\beta_0 + \mathbf{x}'_i \beta + \gamma_f)} \right) \right] + \lambda \sum_{j \geq 1} |\beta^j| \quad (6)$$

Con el fin de evitar que este modelo seleccione palabras de manera trivial pues en español las palabras pueden tener género femenino o masculino, todas las palabras se reemplazan por su raíz que es neutra de género. Por ejemplo, sería sencillo identificar que un estudiante se refiere a su “profesor” o “profesora” y así conocer el género del profesor. Por lo que llevarlas a su raíz sería convertirlas en “profesor” en ambos casos. Para esto, primero, se eliminan de la estimación las palabras vacías (*stopwords* en inglés) como “él” y “ella”. Y luego, se busca la raíz de todas las palabras que aparecen al menos 4 veces entre todos los comentarios haciendo uso del diccionario de la Real Academia Española (RAE)<sup>36</sup>. Adicionalmente, se reemplazaron los nombres y apellidos que aparecen en los comentarios por los términos “NOMBRE” y “APELIDO”.

Los resultados de la selección Lasso se pueden ver en la Figura (4), una nube de las palabras con las 100 palabras que mejor predicen que el comentario sea de una mujer o un hombre res-

<sup>35</sup>Las principales 5,035 palabras equivalen al 98.3 % de todas las palabras utilizadas en los comentarios después de ser reemplazadas por su raíz. Estas son palabras utilizadas al menos 41 veces entre todos los comentarios.

<sup>36</sup>Este procedimiento es imperfecto puesto que se utiliza *webscraping* con el diccionario en línea de la RAE. Este puede encontrar palabras equivocadas, especialmente cuando hay errores gramaticales o palabras en inglés. Para corregir posibles errores se hizo una revisión manual en la que se ajustó apenas el 4.85 % de una lista de 13,637 palabras que corresponden a las palabras utilizadas al menos 20 veces entre todos los comentarios y abarcan el 97.8 % de todas las palabras antes de procesar con su raíz.

pectivamente. Entre más grande sea una palabra en la nube esto indica que dicha palabra es un mejor predictor del género. Las palabras que mejor predicen que el comentario sea de un hombre y no de una mujer se refieren a temas de docencia: “profesor”, “complementario”, “tema”, “parcial” y “tablero”. Al igual que con varios adjetivos que se pueden considerar como positivos: “crack”, “chistoso”, “gracioso”, “grandioso”, “bacán” y “genio”. En los comentarios a las mujeres se encuentran menos referencias sobre las clases y se resalta mucho un uso de palabras asociadas a estereotipos femeninos como “amable”, “comprensivo”, “alegre”, “lindo”, “dulce”, “cariñoso”, “hermoso”, “dulce” y “amigo”. Adicionalmente, hay palabras negativas en la nube para las mujeres que son un reflejo de estereotipos negativos. A ellas se les refieren por ser “escueleras” o “bravas”. Sin embargo, hay hombres que son catalogados por ser “machistas” o “sexistas”, un indicador que en los comentarios se reconocen aspectos negativos de los hombres y asociados a un problema de género.

**Figura 4: Nube de palabras que mejor predicen el género del profesor en un comentario a partir de regresión Lasso**



Resultados del modelo de regularización Lasso con distribución logística con efectos fijos de facultad del curso. Las palabras “NOMBRE” y “APELLIDO” corresponden a nombres y apellidos que fueron reemplazados por estos términos para ser considerados iguales. Cada nube muestra las 100 palabras que mejor predicen cada género. Modelo cuenta con 5,035 palabras, de las cuales 1,784 tienen un coeficiente distinto de cero. Modelo estimado con 5 iteraciones de validación cruzada para obtener un parámetro de penalización. El AUC de validación cruzada óptimo fue de 0.67. Número de observaciones en la estimación: 166,297. Se incluyen datos de comentarios del semestre 2019-2 a diferencia del resto de estimaciones debido a que no hay información disponible sobre la nota de los estudiantes para este último semestre. Debido a limitaciones computacionales únicamente se realiza la estimación en el 50% de las observaciones. El pre-procesamiento de palabras quita las tildes de la figura presentada.

En adición a lo anterior, con las palabras en la nube se nota que los estudiantes reflejan una mayor cercanía hacia las mujeres al ser descritas como “mamá” y “amor” al igual que por su nombre (“NOMBRE”). Los hombres, por su parte, parecen ser relacionados de manera mucho más profe-

sional. A ellos se les referencia más como “profesor” y por su apellido. Incluso las referencias al humor en los profesores masculinos con palabras como “chistoso”, “chiste”, “gracioso” y “humor” está asociado a un mayor estatus y como un resultado positivo ([Evans et al., 2019](#)). Finalmente, como complemento a la nube de palabras, en la Tabla [A9](#) se muestran los resultados de la penalización para las 50 palabras que mejor predicen cada género con sus respectivos coeficientes.

Estos resultados concuerdan con [Schmidt \(2015\)](#) que recopiló los comentarios del portal RateMyProfessor y muestra que a las mujeres se les refiere por ciertos estereotipos y con frecuencia con un lenguaje negativo mientras que a los hombres se les destacan sus habilidades al igual que se les reconocen más cuestiones profesionales que personales. En general, esta discriminación hacia las mujeres en los comentarios puede tener implicaciones tan relevantes como los puntajes de las evaluaciones. Con comentarios desmotivantes y menos relevantes al curso las mujeres pueden ver perjudicada su docencia y avance en términos profesionales.

## 6 Discusión y recomendaciones: ¿cómo disminuir los sesgos de género y sus implicaciones?

A pesar que en este caso los sesgos parecen menores en promedio que en otros trabajos, estos no necesariamente son bajos al considerar ciertos grupos de profesores y cursos. Lo que lleva a buscar soluciones para disminuir los sesgos a su vez que se logra aliviar las implicaciones que pueden tener sobre las profesoras. Lo primero es que una gran parte de los sesgos están asociados a una menor representación femenina en los cursos. Esto sugiere que mientras haya una cantidad cursos con muchos más estudiantes hombres que mujeres, es más probable que los sesgos contra las profesoras sean mayores.

Segundo, para evitar que los estudiantes den evaluaciones sesgadas es útil darles información antes de responder las encuestas con la importancia que tienen para los profesores e ilustrar hechos específicos de cómo las mujeres han sido perjudicadas en ellas. Estas intervenciones parecen ser exitosas de acuerdo a las investigaciones de [Boring y Philippe \(2021\)](#) y [Peterson et al. \(2019\)](#). Esta parte es fundamental por dos razones. Por un lado, evita que las evaluaciones estén sesgadas desde el principio más allá de cualquier heterogeneidad. Y por otra parte, puede hacer que el lenguaje en los comentarios no sea discriminatorio<sup>37</sup>. No obstante, esta medida puede disminuir los sesgos observados pero no necesariamente elimina los sesgos implícitos de los estudiantes.

Tercero, una opción para que los resultados de las evaluaciones que reciben los profesores y que utiliza la universidad para evaluarlos es regresar los puntajes de las evaluaciones contra el género del profesor. Esto evita que haya diferencias en promedio por género del profesor. Como

<sup>37</sup>La evidencia de estas intervenciones únicamente ha explorado efectos sobre los puntajes de las evaluaciones pero plausiblemente puede disminuir los sesgos en los comentarios también.

las diferencias suelen ser condicionales en otras variables también es relevante controlar por otros factores como la nota que recibe el estudiante, el tamaño de la clase, la categoría del profesor, la edad del profesor, la facultad del curso, entre otros<sup>38</sup>. Esto elimina la variación promedio por género pero es insuficiente para corregir los sesgos presentes en complicadas heterogeneidades. Para esto se pueden eliminar las variaciones por género en un sub grupo. Por ejemplo, como se evidenció que los sesgos varían considerablemente por grupos etarios de los profesores, se puede controlar por categorías de edad del profesor con una interacción del sexo del profesor. Esto garantiza que las evaluaciones de los profesores no dependan del género de los profesores de las mismas categorías.

Y por último, como un objetivo principal es evaluar correctamente a los profesores, las universidades deben ampliar la cantidad de medidas que utilizan para evaluar a sus profesores. Con esto en mente, el valor agregado de los profesores puede ser una estadística más objetiva de las habilidades de los profesores para mejorar el desempeño de sus estudiantes (cuando los profesores no determinan la nota que reciben sus estudiantes). En particular, puede ser útil para los profesores conocer su lugar en la distribución de las evaluaciones al igual que en el valor agregado.

## 7 Conclusiones

Este trabajo realiza una evaluación de la existencia de sesgos de género en las encuestas de profesores de una universidad en Colombia. Se aprovecha la amplia disponibilidad de información de estudiantes, cursos y profesores. Se utilizó un modelo de efectos fijos de estudiantes junto con controles de estudiante, curso y profesor para determinar la magnitud de los sesgos. Con esto, se encuentra que en promedio las mujeres son calificadas igual a sus pares masculinos entre asistentes graduados y de cátedra, pero en caso de ser profesoras de plantas reciben evaluaciones menores en 0.07-0.08 desviaciones estándar.

Además, hay una importante heterogeneidad que depende de las características del curso y del profesor, principalmente. Estas heterogeneidades reflejan que cuando el profesor tiene entre 35 y 55 años, la clase tiene más estudiantes y entre menos estudiantes sean mujeres en el curso las profesoras se ven más penalizadas. Una gran parte del sesgo se explica por la falta de representación femenina entre los estudiantes en los cursos. Esas situaciones, en las que las mujeres son penalizadas en al menos 0.1 desviaciones estándar, corresponden al 40 % de las evaluaciones. Por el contrario, los casos en los que el sesgo es positivo para las mujeres con la misma magnitud corresponde al 10 % de las evaluaciones.

Por su parte, con un estudio del lenguaje en los comentarios de las evaluaciones se puede determinar que es más común referirse a un profesor por cuestiones académicas y describirlo como

---

<sup>38</sup>El Centro de Evaluación de la universidad actualmente le entrega a los profesores el resultado de sus encuestas controlando por la nota que recibieron sus estudiantes junto con una medida incondicional.

“profesor”, “crack”, “chistoso”. Mientras que a las mujeres se les referencia por ser “amable”, “querida”, “dulce”, que son términos que no están asociados a su práctica docente. Esto sugiere que los sesgos también están presentes en aspectos más sutiles como la forma en la que los estudiantes se refieren a sus profesores.

En conjunto, estos resultados demuestran que las mujeres siguen afrontándose a situaciones de discriminación. Esta forma de sesgos no está directamente asociada al mercado laboral pero sí tiene potenciales efectos en una menor probabilidad de promoción y menor desempeño en su carrera académica al buscar dedicarle más tiempo a sus clases que a otros aspectos como la investigación. Pero también puede ser un reflejo de discriminaciones en otros aspectos. Así como las mujeres afrontan una mayor discriminación al tener hijos, justamente en ese rango de edad (35-55) es cuando más penalizadas se ven en las evaluaciones.

Para que las universidades eviten evaluar a sus profesoras injustamente se pueden utilizar insu-  
mos adicionales en dicho proceso, como el valor agregado. También pueden ajustar los puntajes de  
las evaluaciones por género junto con las demás características que influyen en las evaluaciones.  
Esta, sin embargo, puede ser insuficiente debido a la enorme heterogeneidad en los sesgos. Final-  
mente, hay evidencia en otros trabajos que muestran que es posible disminuir los sesgos si se le  
recuerda a los estudiantes la importancia de las evaluaciones y se les presenta de manera explícita  
los sesgos en las evaluaciones contra las mujeres.

## Referencias

- Athey, S. y Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.
- Basow, S. A. (1998). Student evaluations: The role of gender bias and teaching styles. In Collins, L. H., Chrisler, J. C., y Quina, K., editors, *Career strategies for women in academe: Arming Athena*, pages 195–156. Sage Publications, Inc.
- Basow, S. A. y Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Education Psychology*, 79(3):308–314.
- Bjerk, D. y Ozbeklik, S. (2018). Using Samples-of-Opportunity to Assess Gender Bias in Principal Evaluations of Teachers: A Cautionary Tale. *Journal of Labor Research*, 39(3):235–258.
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145:27 – 41.
- Boring, A., Ottoboni, K., y Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.

- Boring, A. y Philippe, A. (2021). Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics*, 193:104323.
- Burke, A. S., Head-Burgess, W., y Siders, M. (2017). He's smart and she's nice: Student perceptions of male and female faculty.
- Buurman, M., Delfgaauw, J., Dur, R., y Zoutenbier, R. (2020). When Do Teachers Respond to Student Feedback? Evidence from a Field Experiment. *IZA Discussion Paper Series.*, (12907).
- Carrell, S. y West, J. (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432.
- Centra, J. A. y Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71(1):17–33.
- Chetty, R., Friedman, J. N., y Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.
- Conceição, P., Hall, J., Hsu, Y.-C., Jahic, A., Kovacevic, M., Mukhopadhyay, T., Ortubia, A., Rivera, C., y Tapia, H. (2002). Tackling social norms. a game changer for gender inequalities. Working paper, UNDP.
- Crosby, F. J. y Reinardy, L. M. (1993). Closeness between female college students and their professors: A novel assessment technique. *Sex Roles*, 28:477–483.
- Das, M. y Das, H. (2001). Business students' perceptions of best university professors: Does gender role matter? *Sex Roles*, 45:665–676.
- Dulce, O. V., Maldonado, D., y Sánchez, F. (2019). ¿influyencian mujeres a otras mujeres? el caso de las docentes en áreas stem en bogotá. *Documentos de Trabajo - Escuela de Gobierno Alberto Lleras Camargo*.
- Engelhard, G. (2014). *Item Response Theory Models for Rating Scale Data*. American Cancer Society.
- Evans, J. B., Slaughter, J. E., Ellis, A. P., y Rivin, J. M. (2019). Gender and the evaluation of humor at work. *Journal of Applied Psychology*, 104(8):1077–1087.
- Gaviria, A. y Hoyos, A. (2008). Determinantes de los resultados de las evaluaciones de profesores: el caso de la universidad de los andes. *Desarrollo y Sociedad*, pages 87 – 118.

Iregui-Bohórquez, A. M., Melo-Becerra, L. A., Ramírez-Giraldo, M. T., y Tribín-Uribe, A. M. (2020). The path to gender equality in Colombia: Are we there yet? *Borradores de Economía* 1131, Banco de la Republica de Colombia.

Johnson, M. D., Narayanan, A., y Sawaya, W. J. (2013). Effects of course and instructor characteristics on student evaluation of teaching across a college of engineering. *Journal of Engineering Education*, 102(2):289–318.

Keng, S.-H. (2020). Gender bias and statistical discrimination against female instructors in student evaluations of teaching. *Labour Economics*, 66:101889.

Koedel, C., Mihaly, K., y Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47:180 – 195.

Macnell, L., Driscoll, A., y Hunt, A. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*.

Mengel, F., Sauermann, J., y Zölitz, U. (2019). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association*, 17(2):535–566.

Mitchell, K. M. W. y Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3):648–652.

Muller, M. y Casabonne, U. (2020). Closing gender gaps in latin america and the caribbean. *World Bank Group. Working paper*, 128525.

Nargundkar, S. y Shrikhande, M. (2014). Norming of student evaluations of instruction: Impact of noninstructional factors. *Decision Sciences Journal of Innovative Education*, 12(1):55–72.

Peterson, D., Biederman, L., Andersen, D., Ditonto, T., y Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLOS ONE*, 14(5).

Presidencia de la República de Colombia (2020). Guía para la evaluación del desempeño laboral.

Pérez, J. M., Giudici, J. C., y Luque, F. (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks.

Schmidt, B. (2015). Gendered language in teacher reviews. How it was published.

Storage, D., Horne, Z., Cimpian, A., y Leslie, S.-J. (2016). The frequency of “brilliant” and “genius” in teaching evaluations predicts the representation of women and african americans across fields. *PLOS ONE*, 11(3):1–17.

Tyek Han, K. (2013). Item response models used within wingen.

Universidad de Los Andes (2020). Estatuto profesoral.

Wager, S. y Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wu, A. H. (2018). Gendered language on the economics job market rumors forum. *AEA Papers and Proceedings*, 108:175–79.

## Apéndices

**Tabla A1: Correlación entre variables de las encuestas y características del estudiante, curso y profesor**

Variable	Puntaje del profesor	Puntaje del curso	Recomienda al profesor	Prob. comentario positivo	Prob. comentario negativo
Nota del curso	0.236***	0.245***	0.048***	0.084***	-0.053***
Créditos del curso	-0.014***	-0.014**	-0.001	-0.009***	0.009***
Tiempo de respuesta (mins)	-0.002***	-0.004***	-0.001***	-0.001***	0.001***
Influencia sobre notas	0.113***	0.119***	0.029***	0.054***	-0.034***
Edad profesor	-0.005***	-0.004***	-0.001***	-0.001***	0.001***
Antigüedad del profesor	-0.006***	-0.004***	-0.001***	-0.001**	0.001***
N. de estudiantes	-0.001***	-0.001***	-0.000	-0.000***	0.000***
N. de profesores	-0.044***	-0.101***	-0.003	-0.034***	0.032***

Correlaciones entre variables de resultado y características del estudiante, del curso y del profesor. Coeficientes obtenidos de una regresión entre la variable de resultado correspondiente, la característica y efectos fijos de facultad del curso y de semestre. Pruebas de hipótesis con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

**Tabla A2: Características demográficas de bachilleres y profesionales en Colombia comparados con los estudiantes de la universidad estudiada**

	Bachilleres	Estudiantes universitarios	Estudiantes universidad estudio		Bachilleres	Estudiantes universitarios	Estudiantes universidad estudio
<i>Distribución por género y por estrato socioeconómico</i>							
Fracción mujeres	0.54	0.60	0.45	Ninguno	0.18	0.18	0.00
Estrato 1	0.33	0.15	0.04	Primaria	0.27	0.21	0.03
Estrato 2	0.37	0.34	0.10	Bachillerato	0.32	0.30	0.11
Estrato 3	0.21	0.34	0.16	Técnico	0.10	0.12	0.10
Estrato 4	0.05	0.11	0.26	Profesional	0.11	0.13	0.44
Estrato 5	0.02	0.04	0.21	Posgrado	0.02	0.07	0.33
Estrato 6	0.01	0.02	0.23	Total estudiantes	562,461	237,124	2,634

Columna 2 y 6: estudiantes de bachillerato que presentaron el Saber 11 en alguno de los dos semestres del 2018. Columna 3 y 7: estudiantes universitarios que presentaron el Saber Pro en el 2018. Columna 4 y 8: estudiantes de la universidad de estudio que ingresaron en el segundo semestre del 2018 o primer semestre del 2019 (el ingreso es un semestre después de presentar el Saber 11). El Saber 11 es un requisito para ingresar a la educación superior y el Saber Pro es un requisito para el grado de una institución de educación superior. La extensa mayoría de los bachilleres a nivel nacional presenta el Saber 11.

**Tabla A3: Diferencias en las evaluaciones por género del profesor en cursos de medicina**

	Puntaje del profesor (1)	Puntaje del curso (2)	Recomienda al profesor (3)	Comentario positivo (4)	Comentario negativo (5)
Profesor mujer ( $\beta_1$ )	-0.026 (0.052)	-0.040 (0.041)	0.002 (0.011)	-0.026 (0.018)	0.015 (0.013)
Profesor mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.066* (0.035)	-0.016 (0.031)	-0.023*** (0.008)	0.017 (0.017)	-0.007 (0.015)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
Obs	18,845	19,372	17,348	12,477	12,477
R <sup>2</sup>	0.23	0.35	0.16	0.24	0.25
Prom. var. dep.	-0.05	-0.13	0.90	0.46	0.27
$\beta_1 + \beta_2$	-0.092**	-0.056*	-0.021**	-0.009	0.008

Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en scriptsize semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando y nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Las observaciones varían por columna porque no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.

**Tabla A4: Estimaciones de comentarios con corrección de Heckman**

	Comentario positivo			Comentario negativo		
	EF (1)	MCO (2)	Heckman (3)	EF (4)	MCO (5)	Heckman (6)
Estudiante mujer	0.000 (.)	-0.005** (0.002)	-0.007*** (0.003)	0.000 (.)	0.010*** (0.002)	0.012*** (0.002)
Profesor mujer ( $\beta_1$ )	-0.012** (0.006)	-0.012* (0.006)	-0.012** (0.006)	0.014*** (0.004)	0.013*** (0.004)	0.013*** (0.004)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	0.007 (0.005)	0.010** (0.005)	0.010** (0.005)	-0.004 (0.004)	-0.007* (0.004)	-0.007* (0.004)
$\rho$			-0.118*** (0.008)			0.148*** (0.010)
$\ln(\sigma)$			-0.868*** (0.001)			-1.067*** (0.005)
E.F. estudiante	Sí	No	No	Sí	No	No
Obs	226,619	227,451	350,831	226,619	227,451	350,831
Prom. var. dep.	0.52	0.52	0.52	0.21	0.21	0.21
$\beta_1 + \beta_2$	-0.005	-0.001		0.010**	0.006*	

Errores estándar con cluster a nivel de profesor. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. En la estimación de efectos fijos de estudiante el coeficiente de sexo del estudiante es igual a cero por colinealidad con los efectos fijos. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando y nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre.

**Tabla A5: Ajustes del modelo de valor agregado y sus componentes**

	Modelo principal completo	Sin EF profesor	Sin controles estudiante	Sin controles curso	Solo EF profesor
	(1)	(2)	(3)	(4)	(5)
R <sup>2</sup>	0.652	0.478	0.350	0.649	0.344
R <sup>2</sup> ajustado	0.625	0.452	0.334	0.622	0.328
RMSE	0.397	0.481	0.530	0.400	0.532
Controles estudiante	Sí	Sí	No	Sí	No
Controles curso	Sí	Sí	Sí	No	No
E.F. profesor-semestre	Sí	No	Sí	Sí	Sí
Obs	335,517	335,517	335,517	335,517	335,517

Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, créditos inscritos en el semestre, promedio del semestre, promedio acumulado, promedio en otros cursos en el semestre, situación académica, semestre que está cursando, programa académico, efectos fijos de estudiante. A nivel de curso: facultad del curso, cantidad de estudiantes, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario.

**Tabla A6: Diferencias en las evaluaciones por género del profesor - Con controles y sin controles**

	Puntaje del profesor (1)	Puntaje del curso (2)	Recomienda al profesor (3)	Comentario positivo (4)	Comentario negativo (5)
Panel A: Sin controles					
Estudiante mujer	0.027*** (0.006)	-0.025*** (0.007)	-0.005** (0.002)	0.009*** (0.003)	0.005** (0.002)
Profesor mujer ( $\beta_1$ )	0.007 (0.012)	0.021* (0.012)	-0.006 (0.005)	0.009* (0.005)	-0.000 (0.004)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.015 (0.010)	0.019* (0.010)	0.005 (0.003)	0.009* (0.005)	-0.006 (0.004)
E.F. estudiante	No	No	No	No	No
E.F. curso	No	No	No	No	No
R <sup>2</sup>	0.00	0.00	0.01	0.01	0.00
$\beta_1 + \beta_2$	-0.008	0.040***	-0.001	0.017***	-0.006*
Panel B: Con controles					
Estudiante mujer	0.006 (0.005)	-0.042*** (0.005)	-0.010*** (0.002)	-0.005** (0.002)	0.010*** (0.002)
Profesor mujer ( $\beta_1$ )	-0.031*** (0.011)	-0.015 (0.011)	-0.011*** (0.004)	-0.012** (0.005)	0.013*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.013 (0.010)	0.019** (0.010)	0.005 (0.003)	0.010** (0.004)	-0.007* (0.004)
E.F. estudiante	No	No	No	No	No
E.F. curso	No	No	No	No	No
R <sup>2</sup>	0.03	0.04	0.05	0.05	0.04
$\beta_1 + \beta_2$	-0.043***	0.004	-0.006*	-0.001	0.006*

Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. En la estimación de efectos fijos de estudiante el coeficiente de sexo del estudiante es igual a cero por colinealidad con los efectos fijos. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando y nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Las observaciones varían por columna porque no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.

**Tabla A7: Diferencias en las evaluaciones por género del profesor - Con y sin efectos fijos de curso**

	Puntaje del profesor	Puntaje del curso	Recomienda al profesor	Comentario positivo	Comentario negativo
Panel A: Modelo principal en cursos con profesores hombres y mujeres					
Profesor mujer ( $\beta_1$ )	-0.021 (0.014)	-0.004 (0.012)	-0.009 (0.006)	-0.005 (0.006)	0.005 (0.005)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.014 (0.013)	0.021* (0.012)	0.004 (0.005)	0.012* (0.006)	-0.004 (0.005)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
E.F. curso	No	No	No	No	No
Obs	139,246	139,818	115,013	86,243	86,243
R <sup>2</sup>	0.30	0.35	0.23	0.29	0.29
Prom. var. dep.	-0.03	-0.04	0.88	0.50	0.22
$\beta_1 + \beta_2$	-0.035**	0.017	-0.004	0.007	0.001
Panel B: Modelo principal con efectos fijos de curso					
Profesor mujer ( $\beta_1$ )	-0.028* (0.014)	-0.012 (0.012)	-0.010 (0.006)	-0.011* (0.006)	0.010** (0.005)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.016 (0.012)	0.019* (0.011)	0.000 (0.004)	0.012** (0.006)	-0.003 (0.005)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
E.F. curso	Sí	Sí	Sí	Sí	Sí
Obs	139,246	139,818	115,013	86,243	86,243
R <sup>2</sup>	0.36	0.42	0.31	0.34	0.34
Prom. var. dep.	-0.03	-0.04	0.88	0.50	0.22
$\beta_1 + \beta_2$	-0.044***	0.007	-0.009*	0.001	0.007*

Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. Panel A es igual a la Tabla 3 pero se limita a las observaciones con variación por género del profesor intra-curso para ser comparable con el panel B. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando y nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Las observaciones varían por columna porque no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.

**Tabla A8: Diferencias en las evaluaciones por género del profesor dependiendo del interés del estudiante en la encuesta**

	Puntaje del profesor (1)	Puntaje del curso (2)	Recomienda al profesor (3)	Comentario positivo (4)	Comentario negativo (5)
Panel A: Muestra completa					
Profesor mujer ( $\beta_1$ )	-0.035*** (0.011)	-0.016 (0.010)	-0.012*** (0.004)	-0.012*** (0.005)	0.014*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.009 (0.010)	0.015 (0.009)	0.005 (0.003)	0.007* (0.004)	-0.004 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
Obs	349,541	350,214	302,087	226,619	226,619
R <sup>2</sup>	0.25	0.28	0.14	0.20	0.19
Prom. var. dep.	-0.01	-0.01	0.89	0.52	0.21
$\beta_1 + \beta_2$	-0.044***	-0.001	-0.007*	-0.005	0.010***
Panel B: Estudiantes que también comentan					
Profesor mujer ( $\beta_1$ )	-0.038*** (0.014)	-0.020 (0.012)	-0.012** (0.005)	-0.012*** (0.005)	0.014*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.005 (0.012)	0.024** (0.011)	0.004 (0.004)	0.007* (0.004)	-0.004 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
Obs	226,068	226,578	221,027	226,619	226,619
R <sup>2</sup>	0.26	0.31	0.16	0.20	0.19
Prom. var. dep.	-0.01	-0.04	0.88	0.52	0.21
$\beta_1 + \beta_2$	-0.043***	0.004	-0.008*	-0.005	0.010***
Panel C: Evaluaciones sin respuestas sistemáticas					
Profesor mujer ( $\beta_1$ )	-0.029** (0.013)	-0.005 (0.011)	-0.013*** (0.005)	-0.014*** (0.005)	0.015*** (0.004)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.021* (0.011)	0.010 (0.010)	0.003 (0.004)	0.007 (0.005)	-0.002 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
Obs	249,500	250,011	227,625	175,285	175,285
R <sup>2</sup>	0.24	0.28	0.17	0.21	0.21
Prom. var. dep.	-0.17	-0.27	0.86	0.47	0.24
$\beta_1 + \beta_2$	-0.050***	0.005	-0.011**	-0.007*	0.013***

Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando y nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Las observaciones varían por columna porque no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.

**Tabla A9: Principales palabras que predicen el género del profesor - Modelo LASSO**

	Mujeres				Hombres			
1	querido	0,073	amor	0,027	profesor	-0,131	civil	-0,023
2	actividad	0,070	moda	0,027	chiste	-0,058	arquitectura	-0,022
3	amable	0,067	cerámica	0,027	chistoso	-0,055	tema	-0,022
4	mujer	0,055	paciente	0,026	APELLIDO	-0,046	plataforma	-0,022
5	laboratorio	0,052	creativo	0,026	contrato	-0,046	ciencia	-0,022
6	invitado	0,051	tributario	0,025	física	-0,045	parcial	-0,022
7	italiano	0,050	comprendensivo	0,025	crack	-0,040	más	-0,021
8	hongo	0,050	Francia	0,025	cirugía	-0,036	man	-0,021
9	lamina	0,047	muy	0,025	hombre	-0,036	audiencia	-0,020
10	idioma	0,046	Corbusier	0,023	presto	-0,035	machista	-0,020
11	francés	0,045	autoconocimiento	0,023	ejemplo	-0,034	rhino	-0,020
12	ilustración	0,040	textil	0,023	gracioso	-0,033	duro	-0,020
13	color	0,038	cariñoso	0,023	tablero	-0,033	filosofía	-0,019
14	NOMBRE	0,036	hermoso	0,023	saber	-0,031	diferencial	-0,019
15	biología	0,034	joyería	0,023	gran	-0,028	conocimiento	-0,019
16	tejido	0,034	lindo	0,023	interesante	-0,028	evaluacion	-0,018
17	mamá	0,034	analogo	0,022	patrimonio	-0,028	him	-0,018
18	portugués	0,032	geociencias	0,022	jazz	-0,027	computacional	-0,018
19	dulce	0,032	feminismo	0,022	videojuego	-0,027	bomba	-0,018
20	entregado	0,032	alegre	0,022	entretenido	-0,026	simulacion	-0,018
21	she	0,031	bravo	0,022	humor	-0,026	clase	-0,017
22	actitud	0,031	paciencia	0,021	desviar	-0,026	letra	-0,017
23	arte	0,030	ayuda	0,021	programa	-0,025	finanza	-0,017
24	traje	0,028	técnicas	0,021	divagar	-0,024	materno-infantil	-0,017
25	diapositiva	0,028	escuelero	0,021	emprendimiento	-0,024	bacán	-0,017

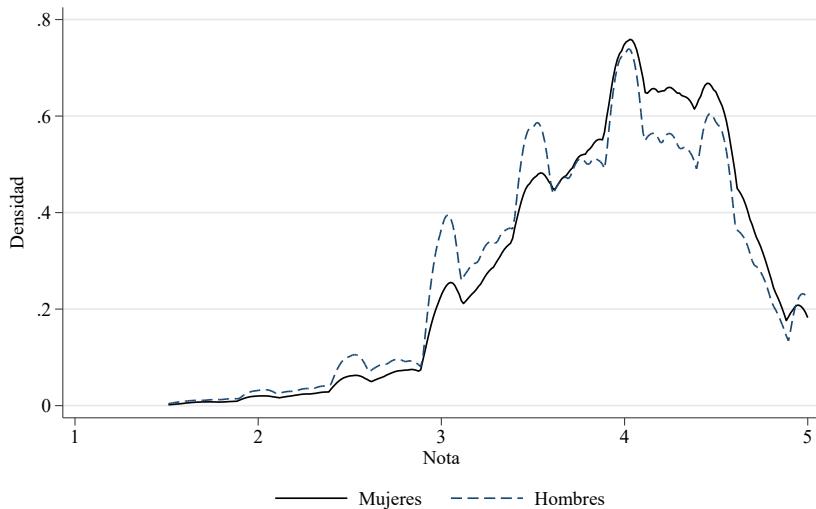
Resultados del modelo de regularización Lasso con distribución logística con efectos fijos de facultad del curso. Cada valor representa el coeficiente de Lasso asociado a esa palabra. Palabras ordenadas de mayor a menor, la primera columna representa los principales 25 predictores y la segunda columna los siguientes 25, tanto para hombres como para mujeres. El modelo cuenta con 5,035 palabras, de las cuales 1,784 tienen un coeficiente distinto de cero. Modelo estimado con 5 iteraciones de validación cruzada para obtener un parámetro de penalización. El AUC de validación cruzada óptimo fue de 0.67. Número de observaciones en el modelo: 166,297. Se incluyen datos de comentarios del semestre 2019-2 a diferencia del resto de estimaciones debido a que no hay información disponible sobre la nota de los estudiantes para este último semestre. Debido a limitaciones computacionales únicamente se realiza la estimación en el 50% de las observaciones.

**Tabla A10: Diferencias en las evaluaciones por género del profesor con modelo de respuesta al ítem**

	Puntaje del profesor	Puntaje del curso
	(1)	(2)
Profesor mujer ( $\beta_1$ )	-0.027** (0.011)	-0.015 (0.010)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	0.007 (0.010)	0.017* (0.009)
E.F. estudiante	Sí	Sí
Obs	350,205	350,211
R <sup>2</sup>	0.29	0.31
Prom. var. dep.	-0.02	-0.02
$\beta_1 + \beta_2$	-0.020*	0.002

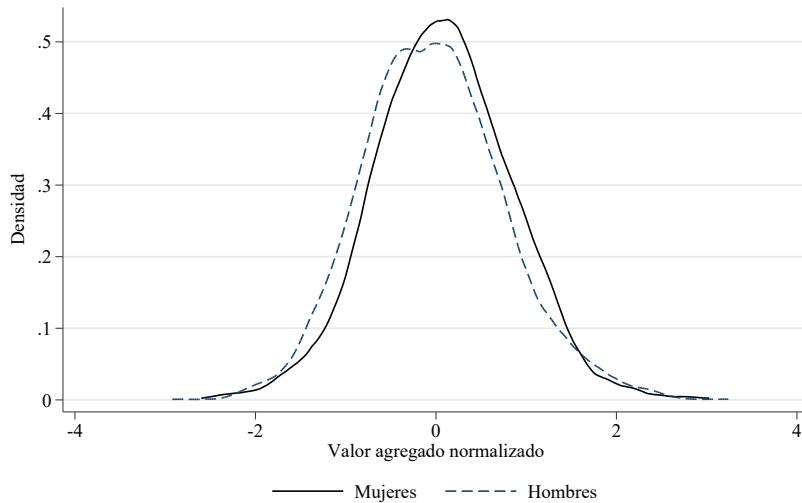
Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Observaciones varían por columna debido a que no todos los estudiantes deben responder las mismas preguntas. Puntajes calculados con un modelo de respuesta al ítem de escala de calificación (RSM). Variable dependiente estandarizada.

**Figura A1: Distribución de las notas de los estudiantes por género del profesor**



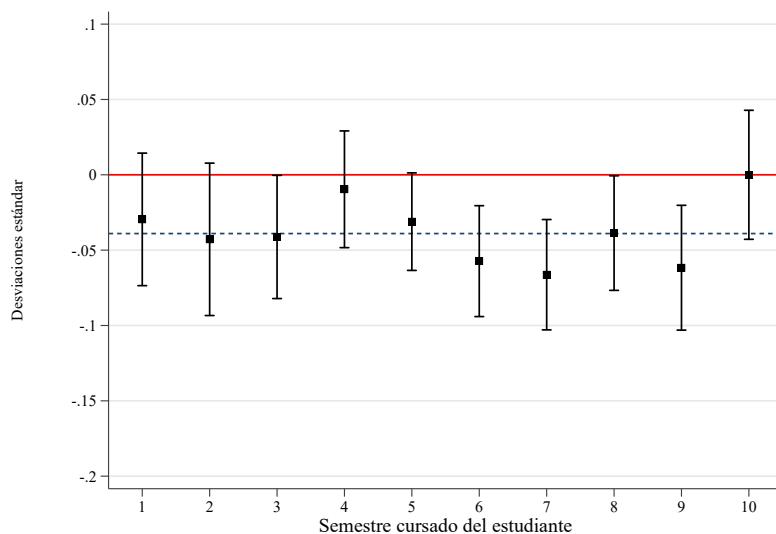
Promedio mujeres: 3.96. Promedio hombres 3.87. Diferencia significativa al 1 %. Diferencia de las distribuciones significativa al 1 % de acuerdo al test de Kolmogorov-Smirnov. Aglomeración en ciertos puntos corresponde a que con frecuencia las notas definitivas de los estudiantes se redondean a múltiplos de 0.5.

**Figura A2: Distribución de la influencia sobre las notas por género del profesor**



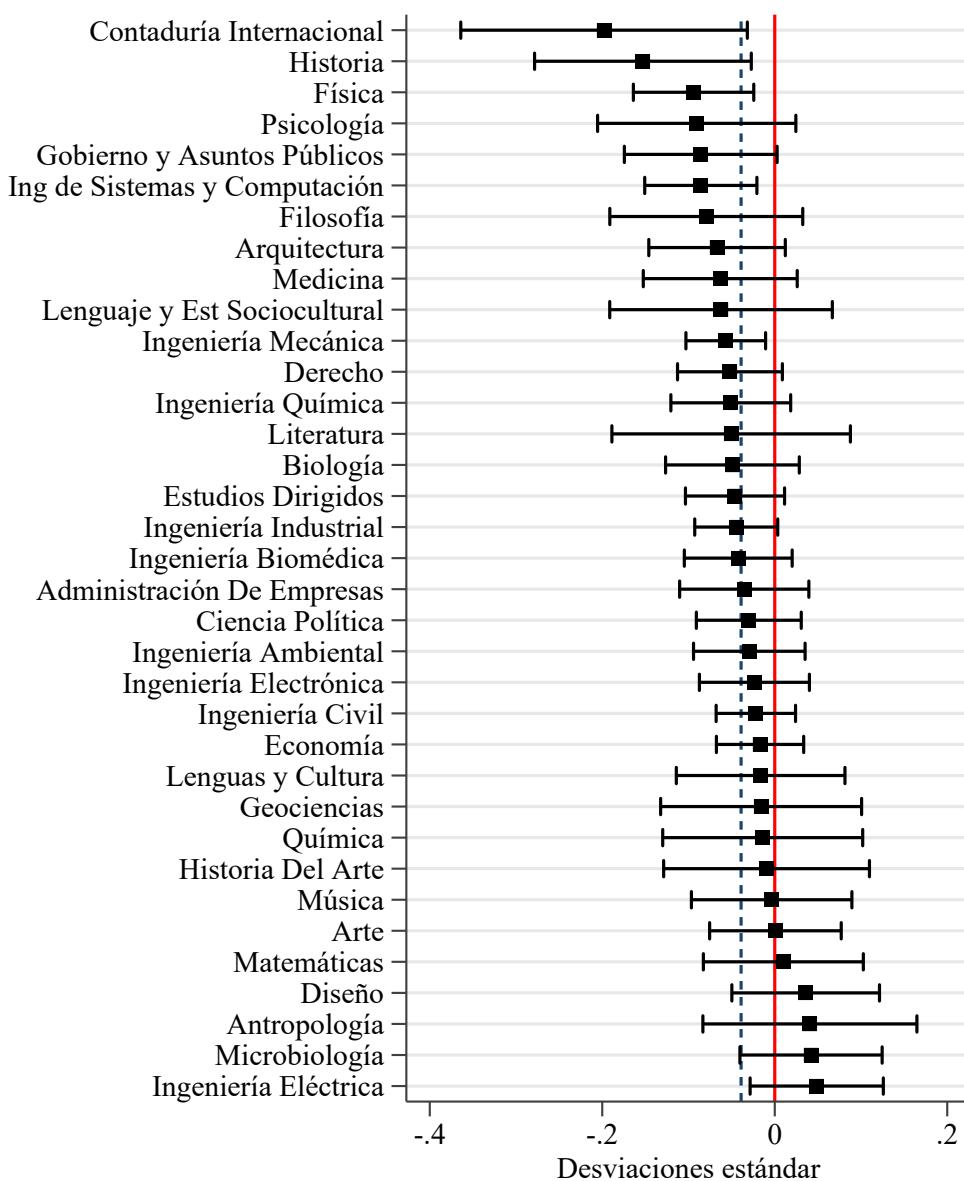
Promedio mujeres: 0.082. Promedio hombres -0.043. Diferencia de medias significativa al 1%. Diferencia de las distribuciones significativa al 1% de acuerdo al test de Kolmogorov-Smirnov. Distribuciones calculadas con la influencia al nivel de profesor-semestre y se muestran las distribuciones después de regresarlas contra efectos fijos de facultad del curso y semestre y luego de sumar la media de la variable para mayor comparabilidad. Influencia sobre las notas se calcula como efectos fijos de profesor-semestre en una regresión de la nota obtenida por un estudiante contra efectos fijos de estudiante, edad del estudiante, edad al cuadrado, promedio acumulado, créditos inscritos, promedio del semestre de las demás materias ese semestre, semestres cursados, si el estudiante ya se graduó o se va a graduar en el periodo de muestra, si desertó la universidad durante la muestra, situación académica del estudiante, cantidad de profesores que enseñan la clase, cantidad de estudiantes en el curso, indicador de magistral/complementaria, créditos del curso y facultad del curso.

**Figura A3: Diferencias en las evaluaciones por género del profesor dependiendo del semestre cursado del estudiante**



Errores estándar con cluster a nivel de curso, intervalos de confianza al 95 %. Línea punteada representa el sesgo promedio, equivalente al de la Tabla 3. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres en el curso, cantidad de profesores que dictan el curso, créditos del curso y un indicador de si es curso complementario. Se incluyen efectos fijos de semestre.

**Figura A4: Efectos heterogéneos en la diferencia de puntaje del profesor por programa del estudiante**



Errores estándar con cluster a nivel de curso, intervalos de confianza al 95 %. Línea punteada representa el sesgo promedio, equivalente al de la Tabla 3. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres en el curso, cantidad de profesores que dictan el curso, créditos del curso y un indicador de si es curso complementario. Se incluyen efectos fijos de semestre. En la figura se excluyen los programas de Ingeniería General, Licenciaturas y Narrativas Digitales por tener una baja cantidad de evaluaciones. El programa del estudiante corresponde al programa en que fue matriculado a la universidad y no necesariamente el que está cursando en el semestre que es observado.

**Tabla A11: Diferencias en las evaluaciones por género del profesor - Sin notas e influencia del profesor sobre las notas**

	Puntaje del profesor (1)	Puntaje del curso (2)	Recomienda al profesor (3)	Comentario positivo (4)	Comentario negativo (5)
Panel A: Resultados principales					
Profesor mujer ( $\beta_1$ )	-0.035*** (0.011)	-0.016 (0.010)	-0.012*** (0.004)	-0.012*** (0.005)	0.014*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.009 (0.010)	0.015 (0.009)	0.005 (0.003)	0.007* (0.004)	-0.004 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
R <sup>2</sup>	0.25	0.28	0.14	0.20	0.19
$\beta_1 + \beta_2$	-0.044***	-0.001	-0.007*	-0.005	0.010***
Panel B: Sin influencia sobre las notas					
Profesor mujer ( $\beta_1$ )	-0.031*** (0.011)	-0.014 (0.010)	-0.010** (0.004)	-0.010** (0.005)	0.012*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.011 (0.010)	0.014 (0.009)	0.004 (0.003)	0.006 (0.004)	-0.003 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
R <sup>2</sup>	0.25	0.28	0.14	0.20	0.19
$\beta_1 + \beta_2$	-0.041***	0.000	-0.006*	-0.004	0.009***
Panel C: Sin nota del estudiante					
Profesor mujer ( $\beta_1$ )	-0.037*** (0.011)	-0.019* (0.010)	-0.013*** (0.004)	-0.013*** (0.005)	0.014*** (0.003)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.005 (0.010)	0.021** (0.009)	0.006* (0.003)	0.009** (0.004)	-0.005 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
R <sup>2</sup>	0.23	0.26	0.13	0.20	0.18
$\beta_1 + \beta_2$	-0.042***	0.002	-0.007*	-0.004	0.009***
Panel D: Sin nota del estudiante ni influencia sobre las notas					
Profesor mujer ( $\beta_1$ )	-0.021* (0.012)	-0.002 (0.011)	-0.009* (0.005)	-0.007 (0.005)	0.010*** (0.004)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.009 (0.010)	0.016* (0.010)	0.005 (0.003)	0.007 (0.005)	-0.004 (0.004)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí
R <sup>2</sup>	0.22	0.25	0.12	0.18	0.17
$\beta_1 + \beta_2$	-0.031***	0.014	-0.004	-0.000	0.007**
Panel E: Sin ningún control					
Profesor mujer ( $\beta_1$ )	0.007 (0.012)	0.021* (0.012)	-0.006 (0.005)	0.009* (0.005)	-0.000 (0.004)
Est. mujer $\times$ Prof. mujer ( $\beta_2$ )	-0.015 (0.010)	0.019* (0.010)	0.005 (0.003)	0.009* (0.005)	-0.006 (0.004)
E.F. estudiante	No	No	No	No	No
R <sup>2</sup>	0.00	0.00	0.01	0.01	0.00
$\beta_1 + \beta_2$	-0.008	0.040***	-0.001	0.017***	-0.006*

Errores estándar con cluster a nivel de curso. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. En la estimación de efectos fijos de estudiante el coeficiente de sexo del estudiante es igual a cero por colinealidad con los efectos fijos. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre, promedio acumulado, situación académica, semestre que está cursando y nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del curso, cantidad de estudiantes, fracción de estudiantes mujeres, cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Observaciones varían por columna debido a que no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.

**Tabla A12: Diferencias en las evaluaciones por género del profesor - Por temas de preguntas**

	Fomento de autonomía	Fomento pensamiento crítico	Ambiente de confianza	Habilidades de presentación	Actividades orientadas a objetivos	Retroalimentación y criterios de calificación
Profesor mujer ( $\beta_1$ )	(1) -0.105*** (0.023)	(2) -0.073*** (0.018)	(3) -0.054*** (0.015)	(4) -0.033* (0.018)	(5) -0.019* (0.011)	(6) -0.017 (0.012)
Est. mujer × Prof. mujer ( $\beta_2$ )	-0.008 (0.022)	0.019 (0.017)	-0.009 (0.012)	0.006 (0.017)	0.020** (0.010)	0.022** (0.011)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí	Sí
Obs	65,917	171,545	222,748	125,981	258,891	251,317
R <sup>2</sup>	0.32	0.25	0.22	0.25	0.25	0.26
Prom. var. dep.	-0.01	0.00	-0.01	-0.01	-0.01	-0.01
$\beta_1 + \beta_2$	-0.112***	-0.054***	-0.063***	-0.027	0.001	0.005
Trato a estudiantes	Actividades consistentes con evaluaciones	Comunica claramente los objetivos	Retratalimentación constructiva	Retratalimentación es oportuna	Retratalimentación es	Espacios para reflexionar sobre desempeño
Profesor mujer ( $\beta_1$ )	(1) -0.016 (0.010)	(2) -0.015 (0.013)	(3) -0.006 (0.017)	(4) -0.002 (0.011)	(5) 0.004 (0.012)	(6) 0.024** (0.011)
Est. mujer × Prof. mujer ( $\beta_2$ )	-0.014 (0.009)	0.021* (0.012)	0.000 (0.017)	0.022** (0.011)	0.019 (0.011)	0.010 (0.011)
E.F. estudiante	Sí	Sí	Sí	Sí	Sí	Sí
Obs	321,350	199,489	86,028	222,993	202,714	172,797
R <sup>2</sup>	0.23	0.26	0.29	0.26	0.26	0.30
Prom. var. dep.	-0.00	-0.01	-0.01	-0.02	-0.02	-0.02
$\beta_1 + \beta_2$	-0.029***	0.005	-0.006	0.020*	0.023**	0.034***

Errores estándar con cluster a nivel de curso. \* p<0.10. \*\* p<0.05. \*\*\* p<0.01. Controles incluidos en la regresión a nivel de estudiante: edad, edad<sup>2</sup>, número de materias inscritas en el semestre, promedio del semestre acumulado, situación académica, semestre que está cursando, nota obtenida en el curso. A nivel de profesor: edad, edad<sup>2</sup>, antigüedad del contrato del profesor, categoría del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de curso: facultad del profesor, polinomio grado 3 de la influencia sobre las notas. A nivel de estudiante: cantidad de profesores que dictan el curso, créditos y un indicador de curso magistral/complementario. Se incluyen efectos fijos de semestre. Observaciones varían por columna debido a que no todos los estudiantes deben responder las mismas preguntas y no todos deciden comentar.