

Número 615

**Gender Stereotypes:
The Case of Mis Profesores.com in Mexico**

**EVA O. ARCEO GÓMEZ
RAYMUNDO M. CAMPOS VÁZQUEZ**

ENERO 2019

CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONONÓMICAS



Advertencia

Los Documentos de Trabajo del CIDE son una herramienta para fomentar la discusión entre las comunidades académicas. A partir de la difusión, en este formato, de los avances de investigación se busca que los autores puedan recibir comentarios y retroalimentación de sus pares nacionales e internacionales en un estado aún temprano de la investigación.

De acuerdo con esta práctica internacional congruente con el trabajo académico contemporáneo, muchos de estos documentos buscan convertirse posteriormente en una publicación formal, como libro, capítulo de libro o artículo en revista especializada.

D.R. © 2019, Centro de Investigación y Docencia Económicas A.C.
Carretera México Toluca 3655, Col. Lomas de Santa Fe, 01210,
Álvaro Obregón, Ciudad de México, México.
www.cide.edu

www.LibreriaCide.com

Oficina de Coordinación Editorial
editorial@cide.edu
Tel. 5081 4003

Agradecimientos

Agradecemos la asistencia de investigación de Raquel Yunoen Badillo, Sergio López y Raúl Murga.

Abstract

In this study we analyze the extent of gender stereotypes in student evaluations of college professors on the internet site MisProfesores.com in Mexico. We download more than 600,000 evaluations for the period 2008-2018. The evaluations include three scores on a scale of 0 to 10: how easy it is to obtain a good grade, how much the professor helps his or her students obtain good grades, and how clearly the professor presents the concepts of the course. The site also allows students to comment on the professor and the class, and we performed a quantitative text analysis of these comments. We found that women receive lower scores than their male counterparts, although the difference is relatively small: 1-2% of a standard deviation. Students refer more to the appearance and personality of female professors, and describe them more often as “bad” or “strict.” They also refer to women in less respectful terms, calling them “maestra” (“teacher”), but calling men “profesor” or “licenciado” (the title corresponding to their academic degree), and they use less positive language for women (“good” vs. “great” or “excellent” for men). Finally, words associated with qualities of service (traditionally stereotyped as feminine) favor women more than men; whereas, words with traditionally masculine associations have a negative impact on women’s evaluations.

Keywords: Gender; Stereotypes; Big Data; Teaching evaluations; Mexico

Resumen

En esta investigación analizamos la existencia de estereotipos de género y su magnitud en las evaluaciones realizadas por estudiantes a profesores en el sitio de internet MisProfesores.com. Para ello, descargamos poco más de 600,000 evaluaciones de estudiantes en el período de 2008-2018. Las evaluaciones incluyen tres calificaciones con escala del 0 al 10 relacionadas con la facilidad de obtener una buena calificación, la ayuda que el profesor da a sus alumnos para obtener una buena calificación y la claridad con la que el profesor enseña los conceptos en la clase. Además, el sitio permite incluir comentarios sobre el profesor y la clase, con los cuales hicimos análisis de texto para su análisis estadístico. Encontramos que las profesoras reciben peores calificaciones que sus contrapartes hombres, aunque esta magnitud es relativamente pequeña de 1-2% de una desviación estándar. Los estudiantes se refieren con mayor frecuencia a la apariencia y personalidad de las profesoras y las describen como “malas” o “estrictas” con mayor frecuencia. Igualmente, se refieren a las mujeres con palabras menos deferentes (“maestra”) que a los hombres (“profesor” o “licenciado”), o menos positivas (“buena” vs. “gran”/“excelente”). Finalmente, las palabras asociadas a cualidades de servicio (con estereotipos femeninos) tienen un efecto positivo en la puntuación de las profesoras en comparación con los profesores, mientras que palabras con asociaciones masculinas tienen un efecto negativo en las profesoras en comparación con los profesores.

Palabras claves: Género; Estereotipos; Grandes Datos; Evaluaciones docentes; México

Gender Stereotypes: The Case of MisProfesores.com in Mexico*

Eva O. Arceo Gómez[♦]

Raymundo M. Campos Vázquez[♠]

Abstract

In this study we analyze the extent of gender stereotypes in student evaluations of college professors on the internet site MisProfesores.com in Mexico. We download more than 600,000 evaluations for the period 2008-2018. The evaluations include three scores on a scale of 0 to 10: how easy it is to obtain a good grade, how much the professor helps his or her students obtain good grades, and how clearly the professor presents the concepts of the course. The site also allows students to comment on the professor and the class, and we performed a quantitative text analysis of these comments. We found that women receive lower scores than their male counterparts, although the difference is relatively small: 1-2% of a standard deviation. Students refer more to the appearance and personality of female professors, and describe them more often as “bad” or “strict.” They also refer to women in less respectful terms, calling them “*maestra*” (“teacher”), but calling men “*profesor*” or “*licenciado*” (the title corresponding to their academic degree), and they use less positive language for women (“good” vs. “great” or “excellent” for men). Finally, words associated with qualities of service (traditionally stereotyped as feminine) favor women more than men; whereas, words with traditionally masculine associations have a negative impact on women’s evaluations.

Keywords: Gender; Stereotypes; Big Data; Teaching evaluations; Mexico.

JEL: D82; D83; I20; J16; O54.

This version: January 2019

* We are grateful for research assistance from Raquel Yunoen Badillo, Sergio López, and Raúl Murga. Any errors or omissions are the responsibility of the authors.

[♦] Centro de Investigación y Docencia Económicas, Carretera Mexico-Toluca 3655, Col. Lomas de Santa Fe, Mexico City, C.P. 01210. Phone: +52 (55) 5727-9800, ext. 2759. Fax: +52 (55) 5727-9878. Email: eva.arceo@cide.edu.

[♠] El Colegio de México, Centro de Estudios Económicos, Camino al Ajusco 20, Col. Pedregal de Santa Teresa, Mexico City, C.P. 10740, Tel.: +52 (55) 5449-3000, ext. 4153. Email: rmcampos@colmex.mx.

1. Introduction

Recent years have seen an interest in the study of women's low representation in different areas, of which academia is one. In the Mexican National System of Researchers (Sistema Nacional de Investigadores, SNI), for example, women make up only one-third of the total. This low representation has important consequences. The lack of diversity in the sciences leads to the neglect of issues of vital importance for the welfare of women, such as the treatment of medical conditions that affect only them,¹ or the impact of public policy on their lives.² Another consequence of this lack of diversity, or lack of role models, is that fewer women choose academic careers, perpetuating the idea that academia is a masculine field.

One of the reasons given for this low representation is that the evaluation of academic work is subjective and can be guided by gender stereotypes. Men and women can be evaluated based on different standards of academic work and behavior. Such stereotypes can therefore bias tenure decisions against women. In this article, we present evidence of such bias and stereotypes in teaching evaluations, using big data from the internet site MisProfesores.com in Mexico. We download all the evaluations submitted to the site from 2008 to 2018, a total of slightly more than 600,000 student evaluations.

This data source has advantages and disadvantages. One advantage is that the evaluations are completely anonymous, so they closely reflect students' opinions. The site includes a section for narrative comments, on which we perform a quantitative text analysis. It also includes information about location, institution, and area of study (or of the subject of

¹ For instance, a quick search on Google Academic© for "male breast cancer" has 20,600 hits, whereas that of "female breast cancer" has 47,400 hits, even though female breast cancer has 3.8 times the incidence of male breast cancer.

² It took 17 years after the release of Viagra for women to have an FDA approved pill to help with their sexual dysfunction issues. The FDA rejected two pills before finally approving the pink pill, because it had secondary effects which, incidentally, were like those of other approved medications such as antifungals, and the approval came with many restrictions on its use.

the course), which we also use in the analysis. Among the disadvantages are that we do not know whether the evaluations represent a random sample of the students taking the course, because the students submitting the evaluations are self-selecting.³ Also, given the complete anonymity of the process, we do not know the gender or any other demographic variable of the student submitting an evaluation. We therefore cannot analyze who is most likely to discriminate or express their opinions in different ways.

This type of study is pioneering in Mexico for its scientific analysis of big data generated on the internet. In the U.S., this type of data has been used extensively to analyze student's opinions about their university courses. In one of the first examples in the literature, Kindred and Mohammed (2005) analyze why students evaluate their professors or use information from RateMyProfessors.com (RMP), the U.S. equivalent of MisProfesores.com. Felton et al. (2008) use the data from RMP to analyze the correlations between the quality of a course and the perceptions of the students. Studies have also analyzed the validity of this type of online evaluation, including what is really evaluated on the site (Bleske-Rechek and Michels 2010; Gregory 2012; Otto, Sanford, and Ross 2008) and comparing it with institutional teaching evaluations (Brown, Baillie, and Fraser 2009; Clayson 2014; Coladarci and Kornfield 2007; Sonntag, Bassett, and Snyder 2009; Timmerman 2008). In general, such studies have found that this type of voluntary and informal evaluation is correlated with the quality of the course and does not necessarily offer an extreme perspective on teaching performance. The findings indicate that the evaluations in RMP are positively correlated with formal institutional evaluations, although some note differences between evaluations

³In the U.S., Bleske-Rechek and Michels (2010) find few differences between the students who use RateMyProfessors.com and those who do not. It would be necessary to perform a similar analysis for Mexico to be certain of the validity of the sample from MisProfesores.com.

submitted to RMP before and after institutional evaluations (Legg and Wilson 2012), as well as halo or contagion effects in positive evaluations (Clayson 2014; Lewandoski, Higgins and Nardone 2012), suggesting possible biases.

Closer to the intent of our study, Stuber et al. (2009) investigate possible gender biases in the RMP evaluations and find no statistically significant differences between men and women. Reid (2010) looks for evidence of racial or gender biases in the evaluations, and finds that African-American and Asian-American professors receive lower scores, but finds little gender bias. More recent studies have attempted to control for differences between male and female professors, for example in teaching style or in the type of courses taught. MacNell, Driscoll, and Hunt (2015) artificially vary the gender of professors of online courses and find that women receive lower evaluation scores than men. Boring (2017) and Mengel, Sauermann, and Zölitz (2018) take advantage of a random assignment of male and female professors to courses and find biases against the women. Finally, Mitchell and Martin (2018) find gender biases in evaluation of online courses. They also analyze the language students use in answers to open-ended questions and find that they describe male professors in terms of their competencies and female professors in terms of their appearance and personality. Our own study applies the methods of this latter study to analyze the evaluations in MisProfesores.com.

This topic is not only important in itself, but may also have broader implications. The majority of Mexican universities have a system of student evaluations of professors. In some institutions this evaluation is an important factor in tenure or promotion decisions. At the same time, we know that there is a gap in the representation of male and female professors in Mexico. According to INEGI (2015), 56% of college and university professors are men and 44% are women. Among researchers enrolled in the National System of Researchers

(SNI), the difference is even greater: 66% men and 34% women. If student evaluations reflect gender stereotypes and biases, and are used in tenure and promotion decisions, that could help to explain the gender gap in teaching positions. It is not possible to prove such a link with the data from our study, but we hope that our results encourage further reflection and empirical analysis of the validity of these evaluations.

Our principal findings are as follows. First, there is evidence that women receive lower scores and a lesser perception of their performance than men, although the difference is relatively small: 1-2% of a standard deviation. Second, female professors receive far more comments than men about their appearance and personality, as found also by Mitchell and Martin (2018). Third, evaluations commonly refer to men as “*profesor*” or “*licenciado*” (the title corresponding to their academic degree), but to women as “*maestra*” (“teacher”), showing greater respect to men than to women. This difference is pronounced: women receive twice as many comments with the word “*maestra*” as men. Fourth, there is great heterogeneity with respect to field of study, type of institution, and geographic region. Fifth, the word used most often to describe a class taught by a man is “excellent,” while that for a class taught by a woman is “good.” Sixth, the proportion of words referring to the demanding nature of the class or to being a “bad” professor is greater for women. Finally, words associated with qualities of service (traditionally stereotyped as feminine) favor women more than men; whereas, words with traditionally masculine associations have a negative impact on women’s evaluations.

This article is organized as follows. Section 2 reviews the literature on the topic of gender stereotypes and biases. In Section 3 we describe the process of acquiring and coding the data. Descriptive results are presented in Section 4. Section 5 provides a discussion of the empirical results and our conclusions.

2. The Literature on Gender Stereotypes

Stereotypes are defined as characteristics or behaviors expected of a particular social group (Ellemers 2018; Koch, D’Mello, and Sackett 2015). These stereotypes lead us to categorize people by general group attributes rather than by their individual characteristics. Stereotypes can thus have damaging consequences, including conscious explicit discrimination (such as not being hired or denied promotion for a job) or unconscious biases (such as automatically assigning certain types of work to women and not to men).⁴

Gender stereotypes include those of men and women. Men are more commonly identified with individualist stereotypes, such as domination, ambition, and independence. Women are most often identified with communal stereotypes, including concern for others, emotional expressiveness, and other characteristics related to caring (Koch, D’Mello, and Sackett 2015). Given this dichotomy, women face a tradeoff between being agreeable toward others and being considered competent (Bohnet 2016). If women do not conform to the social norm of communal behavior (the expected stereotype), they may face negative consequences (they might, for example, be considered less competent in positions of authority). Many professional women thus face daily decisions about whether to go along with or against stereotypes about their behavior.

One of the professions most studied by social scientists is that of college and university professors. It is possible that teaching in higher education has stereotypes that favor men, because being a professor is considered a position of authority or because

⁴ For a fuller analysis of these phenomena, see Bayer and Rouse (2016), and Bertrand, Chung, and Mullainathan (2005). For an analysis of the Mexican context and qualitative evidence of gender stereotypes, see Rocha-Sánchez and Díaz-Loving (2005).

academia is a masculine environment. If, as the theory suggests, a lack of adherence to the stereotype is punished in teaching evaluations and these are used for decisions about hiring or promoting academic personnel, the mechanism of the teaching evaluation could introduce a gender bias into the representation of women on faculties and in institutional hierarchies.

Numerous studies have in fact shown that student evaluations are biased against women because of these stereotypes. Competence and teaching methods are not factors. For example, Mitchell and Martin (2018) analyze their own teaching evaluations (one is a man and the other a woman) for in-person and online classes. The online classes were identical, as was their interaction with students. They analyze not only official university evaluations, but also those from the website RateMyProfessors.com. In both sets of evaluations, students show a bias against the female professor. The bias is the same for the online and in-person classes, which indicate the effect of a stereotype and not the way in which the class is taught. Moreover, the final grades given by each professor were similar: if anything, students' grades were higher in the female professor's class.

In a similar study, MacNell, Driscoll, and Hunt (2015) perform an experiment to assess students' gender bias. There are two professors and four online classes. For two classes they identify the professors' true gender; for the other two they publicize the course as being taught by a man when it is in fact a woman, or vice versa. They find that students favor the man over the woman. Researchers have also taken advantage of the fact that in some universities students are randomly assigned to classes (so there is no selection bias for or against a professor), and they have obtained similar results, showing a consistent bias against female professors from both male and female students (Boring 2017; Mengel, Sauermann, and Zölitz 2018).

Gender stereotypes not only affect teaching evaluations, but also the demands of students on professors. El-Alayli, Hansen-Brown, and Ceynar (2018) interview professors about the requests they receive from students. They find that students ask for more favors, special treatment, and acts of friendship from female than from male professors. They also interview students about their expectations from professors, and find that they expect motherly treatment, greater attention to their emotions, and other demonstrations of communal behavior from female professors. Students also say they are more likely to ask for favors and special treatment from female professors, to insist more to get what they want, and to react much more negatively when female professors refuse to give them special treatment. These demands add to the emotional stress and the uncompensated workload women already have, undermining their possibilities for promotion and tenure (Babcock et al. 2017; Guarino and Borden 2017).

Evidence has also been found in other contexts for the importance of gender stereotypes. Goldin and Rouse (2000) find that when a curtain is used in orchestral auditions in the U.S., so that evaluators do not know if the musician is male or female, more women are hired. Bohnet (2016) describes different national studies of how stereotypes affect the promotion of female attorneys. Miller, Eagly, and Linn (2015) find on a national level that gender stereotypes in scientific fields are related to the number of women in a university, as well as the number of female researchers. Sarsons (2018) finds that female doctors are seen in evaluations and recommendations as more affected by the death of a patient than male doctors. This not only affects the doctor who suffers the death of a patient, but it extends to other women as well, showing the reach of negative stereotypes of gender. In Mexico, Arceo and Campos (2014) find that employers offer more jobs to light-skinned women than to dark-skinned ones, in line with a stereotype of beauty that is imposed on women but not on men.

Conformity with stereotypes takes place socially in childhood and adolescence. This can be seen in the calculation of risk preferences (adult women are more risk-averse than adult men) and of competition (men have greater preferences for competition). Experiments with children and adolescents have shown that there are no gender differences in these preferences in early childhood (before the age of six), but that the differences begin before adolescence and then solidify (Andersen et al. 2013; Dreber, von Essen, and Ranehill 2014; Sutter and Glätzle-Rützler 2015). Results from matriarchal cultures show no gender differences in adult preferences (Gneezy, Leonard, and List 2009). These results point to the important role of social and cultural norms in strengthening and maintaining certain stereotypes.

Stereotypes can be countered with a greater presence of women in stereotypically male roles. If, for example, being a college professor is biased in favor of men, one way of combating that stereotype is to increase the representation of women in that area. Studies have shown that this approach can have positive results in reducing gender stereotypes. Beaman et al. (2012), for example, show that local governments headed by women changed people's perception of women and what girls could accomplish in the future. In a laboratory experiment, Dasgupta and Asgari (2004) find similar results, which show the importance of having leaders of one's own gender. Affirmative action and quotas are mechanisms that could help achieve the goal of diminishing or eliminating stereotypes more quickly. The introduction of these mechanisms in different countries and contexts has not had significant negative effects, and in most case has had positive effects that help to eliminate gender stereotypes.⁵

⁵ For an overview of this topic, see Bohnet (2016).

3. Data

In this study we use data taken from the website MisProfesores.com, from which we download all student evaluations submitted from 2008 to 2018, a total of approximately 600,000 individual evaluations of 64,577 professors and teachers.⁶ Students submitting evaluations have the option of searching for the professor or teacher in the existing record or adding one manually. The student then assigns three scores: (1) “Difficulty: How easy is it to get a good grade, assuming that the student completes all of the assigned work?” (2) “Help: How much does the professor/teacher help students with extra credit or individual help?” and (3) “Clarity: How clearly does the professor/teacher explain the concepts?”⁷ The student can then include optional comments about the course and the professor. Although the site includes the instructions “Please include only constructive comments. Any offensive or inappropriate comment will be deleted,” in our text analysis we found many comments that did not follow this rule.

We use the three scores as they are provided by the website. Following previous studies, we perform a text analysis of the written comments. We first analyze the words used most often in all the comments. Then, using the 800 most used (51% of the comment text), we divide them into positive and negative, in five areas that refer specifically to the professor/teacher: performance, assertiveness, appearance, personality, and recognition. This division corresponds to that used by Mitchell and Martin (2018). Table 1 shows the list of words used in each area, and whether they are positive or negative. If a comment is both

⁶ It is not possible to know the precise number of professors because the entries depend on how students enter their names on the website. If the name is entered in exactly the same way, the website assumes they are the same person, but if there are slight differences, they are listed separately.

⁷ These are the descriptions as presented on the website.

positive and negative in the same category, it is considered to be neutral. In addition, we look at whether students refer to the instructor using terms such as “professor,” “doctor,” “teacher” (*profesor/a, doctor/a, maestro/a*), and consider this variable as a measure of formality or respect. Use of “*profesor/a,*” “*doctor/a,*” “*licenciado/a,*” “*ingeniero/a,*” or “*catedrático/a*” was considered positive, and “*maestro/a*” or “*miss*” (an English word sometimes applied to female teachers in primary or secondary schools) was considered negative.

Table 1. Words Included in Opinions Analyzed

Opinion	Positive Words	Negative Words
Performance	good, excellent, committed, super, well, learn, teach, better/best, recommendable, recommended, recommend, clear, interesting, knows well, intelligent, warmth, motivates, great, perfect, ability, flexible	awful, bad, badly, worse, confuses, confusing, boring, terrible, doesn't know
Assertiveness	easy, exam, grades, homework, help, gives, pass, accessible, get, exercise, assignment, readings, topic, practice, understand, project, explains, attendance, turn in/return, activity, enjoyable, prepare, help, willing, willingness, fair, on time, flexible, pass	doesn't show up, difficult, easy, demanding, demands, strict, complicated, late, boring, unfair, fail
Appearance	tall, handsome, pretty, young	old, ugly, disgusting
Personality	person, cool, nice, concerned, love, patient, human, sweet, nice, intelligent, attitude, patience, humor, fun, I love, great, believes, feels, on time, interested, entertaining fun, passion, easygoing	boring, disgusting, asshole, dictator, terrible, lazy, tedious, rude, dominating, angry, arrogant, late
Recognition	professor/teacher, teacher, doctor, engineer, prof	idiot, bastard, damn, faggot, bitch, shit

Notes: All text in Spanish was changed to lower case, without accent marks, in order to avoid errors. The five divisions of opinions as well as some of the words were suggested by Mitchell and Martin (2018).

When students enter the name of a professor, the website associates it with a university and department, and the students then choose the class they took. We can therefore see the name of the institution and classify it according to type (public, private, or technical school, high school, normal school), geographical location (in most cases the state; in many cases the city or town), and area of study (according to the professor/teacher's department).

For convenience, we divide the latter into law; business administration and accounting; economics (including business and finance); sociology (including education, demography, urban studies, and related fields); political science and international studies (including communication and government); arts and humanities; exact sciences (including engineering, mathematics, chemistry, and related fields); biology and medicine (including health sciences and psychology); tourism, agriculture and architecture (including design, veterinary sciences, and related fields); and computer and information science. Using these categories, 64% of the evaluations can be assigned to a field; most of the rest cannot be assigned because the field is not defined in the evaluation.

Finally, the most important variable in the study is gender. The website does not automatically identify the gender of the professor/teacher, and unfortunately their names are not written in a standard format: some are last name first and some are not. In order to identify gender we therefore use text analysis. We first use databases of common surnames to remove these from the name field, leaving only first and middle names. We then use lists of names associated with men and women to identify their gender.⁸ This procedure works for Mexican names, but not if the professor/teacher has a foreign name. In some cases we manually identify gender. We perform random reviews in rounds of 1000 names until we find no errors in the identification of gender. In the end, only 0.6% of the evaluations were not identified (the majority of these because the name given was not on the list, and a minority because it was a foreign name whose gender was difficult to identify).

Table 2 shows the number of observations and statistics for gender, fields of study, type of institution, and geographical locations. They are divided into those that can be

⁸ Surnames were obtained from http://www.mapadeapellidos.eu/apellidos_clasificacion.php, and men's and women's names from <https://www.guiainfantil.com/>.

identified as pertaining to universities and those that cannot. There is a total of slightly more than 600,000 observations, almost 566,000 of which can be identified with universities.

Table 2. Descriptive Statistics of the Aggregate Data (percentages)

	Total	University	
		No	Yes
Number of observations	602,919	37,111	565,808
A. Gender			
Female	34.2	46.2	33.4
Male	65.2	52.8	66
Not identified	0.6	1.1	0.6
B. Area of Study			
Law			10.6
Business Administration			6.4
Economics			7.5
Sociology/Demography			2.4
International Studies/Poli. Sci.			4.2
Arts and Humanities			7.1
Mathematics/Engineering			21.7
Medicine/Biology			7.5
Tourism/Architecture/Design			3.6
Computer/Information Science			4.1
Not identified			24.8
C. Public, Private, or Technical			
Public			67.1
Private			21.5
Technical			6.5
Not identified			4.9
D. Region			
Mexico City	54.6	52.3	54.7
North	10.9	21.9	10.2
Central	25.8	19.5	25.1
South	9.7	6.3	9.9

Notes: Authors' calculations. Gender was identified by the professor's/teacher's name. Area of study: law (criminal defense, forensic, or tax law), business administration (including accounting), economics (business and finance), sociology/demography (including education, urban studies, sustainability studies, social sciences), international studies/political science (including communication and government), arts/humanities (including history, literature, cinema, photography, music), mathematics and engineering (including physics, chemistry, statistics, actuarial sciences), biology/medicine (including health sciences and psychology), tourism/design (including architecture, agriculture, veterinary sciences, and ecology), and computer/information sciences (systems, programming, and software). Type of university was assigned according to the text identifying the institution. Regions: Mexico City, north (all the northern border states plus Baja California Sur), south (Guerrero, Puebla, Morelos, Tlaxcala, Veracruz, Oaxaca, Chiapas, Campeche, Tabasco, Yucatán, and Quintana Roo), Central (all of the remaining states). There are very few missing observations of region; these are not included.

We can see in Table 2 that there are fewer women teaching at the university level than at the high school level. For universities, 33.4% of the evaluations are of women and 66% of men. This difference is similar to the gender distribution in the National System of Researchers (SNI). Nearly one-fourth of the university-level evaluations cannot be assigned to a specific area of study, given that students did not provide this information to the website. The area that received the greatest number of evaluations is the exact sciences (mathematics, engineering, and similar fields), with 21.7% of the responses, followed by law, with 10.6%. The rest of the fields each had between 2.4 and 7.5%. Given the small number of observations in each area, we combine the groups further. The final groups are: law (10.6%); exact sciences plus computer and information science (25.8%); business administration and economics (13.9%); sociology, arts, humanities, and international studies (13.7%); and biology, medicine, tourism, and architecture (11.1%).

Table 2 also shows whether the university is public (research), private, or technical (technical schools and polytechnic universities, not including the Instituto Politécnico Nacional). Students from public universities submitted 67.1% of the evaluations and those from polytechnic universities 6.5%, totaling 73.6% from public universities. Those from private universities submitted 21.5%, and less than 5% came from students whose schools could not be identified. These figures are close to the figures of the Secretary of Public Education (SEP 2017), which finds 70.6% of students in public education.

Universities in Mexico City account for more than half of the evaluations, followed by the central-western region (which includes all of the states north of Mexico City except the northern border states and Baja California Sur), with 25.1% of the total. The northern border region and Baja California Sur account for 10.2%, and the southern region (all of the

states south of Mexico City, plus Veracruz) for 9.9%. There is a general bias in the evaluations in favor of locations with several universities: Mexico City, Jalisco, and Puebla.

The results presented below are based only on the evaluations from college and university students; they do not include those from students at high schools. With these evaluations we can identify the academic department where the course is given, and also whether the institution is public or private.

4. Results

The results shown below are analyzed for the different groupings described above. First we present the aggregate results, followed by analyses by area of study, type of institution, and geographic region. They focus on the criteria of difficulty, help, and clarity, as well as the analysis of opinions based on students' textual comments.

4.1 Aggregate Results

Table 3 shows the average scores as well as the percentage of positive and negative opinions from different groups with respect to male and female professors, along with the differences, standard errors, and p -values.

Table 3: Average Scores and Opinions from Comments

Variable	Women	Men	Difference	Standard Error	<i>p</i> -value
A. Scores					
Difficulty	6.61	6.58	0.03	[.008]	0.001
Help	7.03	7.09	-0.06	[.009]	0.000
Clarity	7.11	7.17	-0.06	[.009]	0.000
Average	6.92	6.95	-0.03	[.008]	0.000
B. Textual Comment					
Performance					
Positive	57.6	57.2	0.41	[.150]	0.007
Negative	17.6	15.8	1.8	[.114]	0.000
Difference	40.1	41.4	-1.4	[.231]	0.000
Assertiveness					
Positive	39.2	37.3	1.9	[.147]	0.000
Negative	21.8	19.9	1.9	[.124]	0.000
Difference	17.4	17.4	0.00	[.228]	0.995
Appearance					
Positive	1.8	1.6	0.20	[.039]	0.000
Negative	0.89	0.92	-0.03	[.029]	0.344
Difference	0.87	0.64	0.23	[.049]	0.000
Personality					
Positive	22.3	18.7	3.53	[.123]	0.000
Negative	8.6	8.2	0.43	[.084]	0.000
Difference	13.6	10.5	3.1	[.160]	0.000
Recognition					
Positive	48.6	50.8	-2.2	[.151]	0.000
Negative	0.56	0.87	-0.32	[.025]	0.000
Difference	48	49.9	-1.9	[.155]	0.000
Formality					
Positive	24.6	34.2	-9.6	[.135]	0.000
Negative	24.8	12.8	12.0	[.122]	0.000

Notes: Authors' calculations. Panel A includes scores from universities, excluding observations without identifiable gender, for a total of 562,599 observations. Numbers shown are averages, with a maximum score of 10. Panel B additionally excludes observations without textual comments, for a total of 488,589 observations. Numbers shown are percentages. Definitions of positive and negative opinions are shown in Table 1. Positive formality refers to a reference in the textual comments to "professor" or terms that refer to the instructor's degree (licenciado/a, ingeniero/a, doctor/a) and negative formality refers to a reference to "teacher" or "miss."

Female professors score better than their male counterparts only in difficulty. Women score lower in help, clarity, and in the average of the three scores. All of these differences are statistically significant at the 1% level, with standard deviations close to 3, so that the

difference between men and women is approximately .01-.02 standard deviations. It is not a very large difference, but it is statistically significant, and it is consistent with the results of prior studies in which female professors scored lower than male professors, although the difference is smaller than in those studies (Boring 2017; Mengel, Sauermann, and Zölitz 2018; Mitchell and Martin 2018).

The text analysis in different categories shows evidence consistent with the existence of gender stereotypes. Comments about the performance of male and female professors are the most common: 57% of the comments include remarks about professors' performance, with a greater number of positive comments ("good" and similar terms) about women. However, women also receive more negative comments ("bad," "terrible," and similar terms) about their performance than men. In the aggregate, women have a lower average score for performance than men: their average score is 1.4 percentage points lower than that of men, a difference of approximately 3.5%.

Positive terms about assertiveness include words like "easy," "accessible," "assignment," and "fair," while negative terms include such words as "difficult," "fail," "unfair." Female professors receive more comments about assertiveness, both positive and negative, than their male counterparts. Overall, male and female professors receive approximately the same number of comments (as shown in the difference row).

Previous studies have shown that student comments refer to the appearance and personality of female professors more than those of men. Comments about the appearance of men, either positive or negative, are not very common: a little more than 2% of the total. Negative comments about appearance are approximately equal between women and men. However, women receive more positive comments about their appearance. If the productivity of the professor (knowledge absorbed by the student) is all that matters, the physical

appearance of the professor should be irrelevant. Although the percentage of comments mentioning appearance is small, female professors do receive more of them.

Comments about personality show a similar pattern. Women receive more positive comments about their personality than men (almost 3.5 percentage points, or 20% more). But at the same time they receive slightly more negative comments about their personality. Overall, women receive more positive than negative comments, and the difference in this category is substantial (3.1 percentage points, a little less than 30% of the difference in men).

Women receive fewer comments about recognition than men, either positive or negative. This is also consistent with previous studies (Mitchell and Martin 2018) in which students recognize female professors less than their male counterparts.

Finally, we analyze the extent to which men and women are referred to as “professor” or a word referring to their academic degree, as opposed to “teacher” or “miss” (an English word sometimes used to refer to female primary or secondary school teachers). The former show respect and are classified as positive; the latter imply lesser authority and are classified as negative (Mitchell and Martin 2018). The differences in the terms used are substantial. Men are referred to using the positive terms in approximately 34% of the comments, but they are applied to women in only 25%. That is, the use of positive terms for women is 9 percentage points, or 30% less than that for men. At the same time, women are referred to as “teachers” twice as often as men.

These results are consistent with those of the previous studies discussed in the literature review. Women are scored lower than men, although the magnitude of the difference is relatively small. Women also receive comments consistent with gender stereotyping. They receive more negative comments about their performance, more comments about their personality and appearance, less recognition, and they are referred to

using less formal language than men. In the following sections we analyze these differences for subgroups defined by area of study, type of institution, and geographic region. We include only the gender differences in the following variables: average score, performance, assertiveness, appearance, personality, and recognition. Given that the differences in scores and opinions have different scales, to facilitate comparison we present the results in terms of standard deviations, except for those concerning formality as defined by use of words like “professor” versus words like “teacher” (positive versus negative opinions as defined in Table 2).

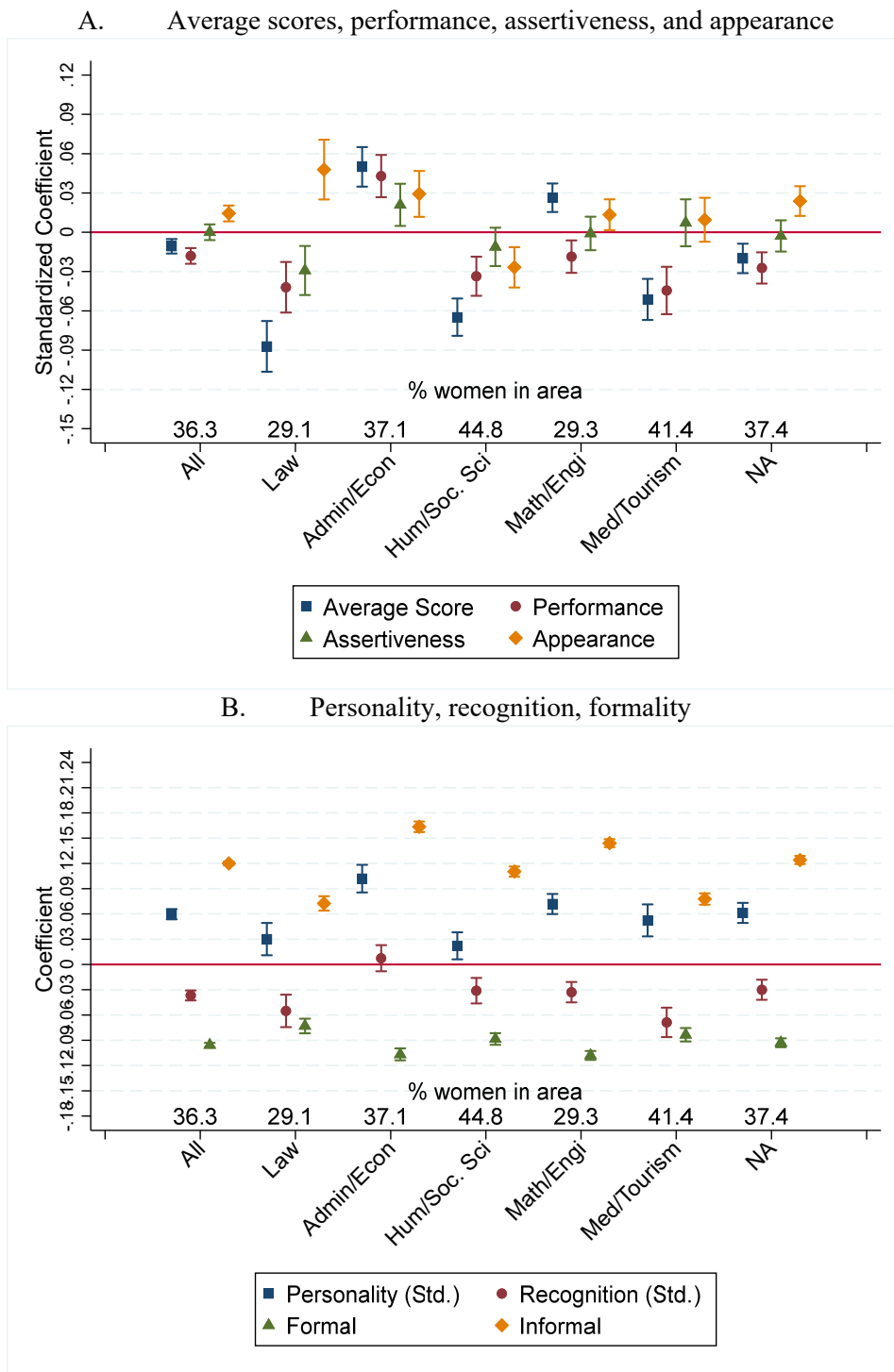
4.2 By Area of Study

For simplicity we divide the areas of study into six groups: law, administration, and economics (includes accounting, business, and finance), humanities and social sciences (sociology, demography, arts, and others), mathematics and engineering (includes computer and information science and related fields), medicine and tourism (includes biology, architecture, and design), and finally those that cannot be identified. Panel A includes the average scores for difficulty, help, and clarity, and for opinions about performance, assertiveness, and appearance (including only observations that include comments). Panel B includes opinions about personality, recognition, and formality. A positive effect indicates that women receive a higher score or greater proportion of comments; a negative effect indicates the opposite.

Figure 1 shows effects in terms of standard deviations (with the exception of those regarding formality) by area of study. We include the results for all observations (those that appear in Table 2). The figure shows a great degree of heterogeneity among fields of study. The field with the lowest average score and opinion regarding performance is law, indicating

that female professors receive substantially less value for their work in this area. Negative effects are also seen in the category including humanities and social sciences, and in the one including medicine, tourism, and architecture. Women are not evaluated more negatively in all areas: in business administration and economics they are generally evaluated more positively than men. In mathematics the results are mixed: although women have a better average score than men, they are evaluated more negatively on performance. The figure also includes the proportion of female professors evaluated in each area of study (the proportion of distinct individuals, not the proportion of evaluations) to analyze its relationship with the scores; no correlation is observed. Although there are few female law professors, there are likewise few in mathematics and engineering, but in the latter fields women do not receive such negative scores. At the same time, there is a greater presence of female professors in the humanities and social sciences, but women in these fields have the second lowest score of all the groups.

Figure 1: Effects by Area of Study



Note: Authors' elaboration. 95% confidence intervals. All variables are standardized except for formality (comments referring to "professor" versus "teacher"). Observations are included only where comments are included in the evaluation. Admin./econ. includes administration, accounting, economics, business, finance, and similar fields; hum./soc. sci. includes sociology, demography, urban studies, arts, humanities, and similar fields; math/eng. include mathematics, engineering, physics, computer and information science, and similar fields; med./tourism includes medicine, biology, tourism, architecture, design, and similar fields; NA indicates that the evaluation could not be assigned to an area of study.

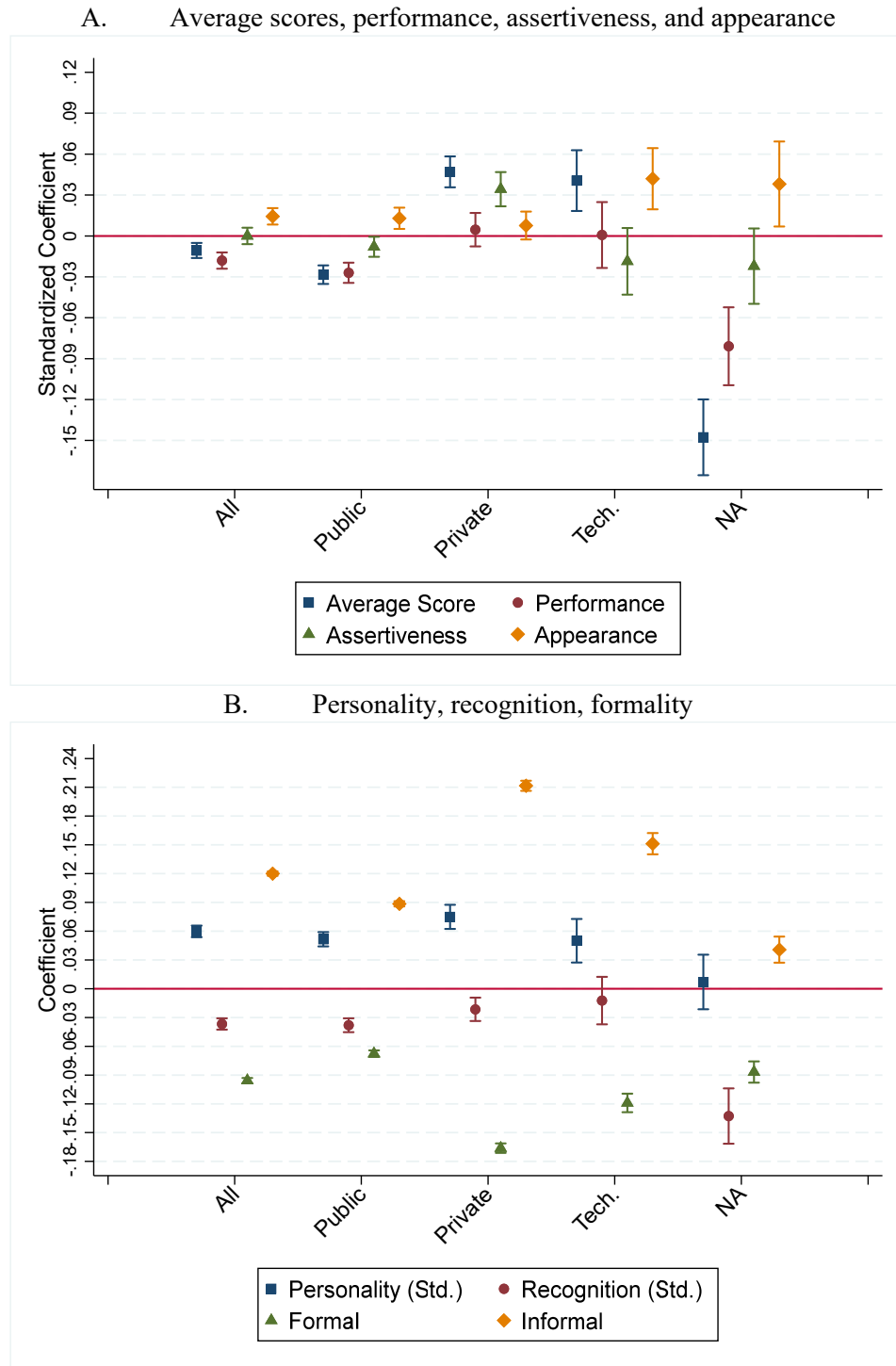
Panel B shows that in general, student comments about personality refer more often to women and that female professors receive lower scores for recognition. The greatest differences in language used by students are found in the words used to identify professors: men are far more commonly referred to as "professor" or "*licenciado*" (referring to their academic degree) and women as "teacher" or "miss" (an English word sometimes used to refer to female primary and secondary school teachers). There are large differences between areas of study, most notably in administration and economics, as well as medicine, architecture, biology, and other fields.

4.3 By Type of Institution

Figure 2 shows differences by type of institution: public, private, technical, and not identifiable. In public universities there are more comments related to negative gender stereotypes than in private ones, especially in the average score and opinions about performance and assertiveness. Technical schools show no very large bias. In universities that could not be identified (generally because information was not supplied), comments about female professors are more negative than in other institutions. However, panel B shows that it is much more common in private than in public universities to refer to female professors as "teacher" and to male professors as "professor," "*licenciado*," or a similar term. The differences are great: in private universities, women are referred to with the formal terms

17 percentage points less than men, and they are referred to with the informal terms 21 percentage points more.

Figure 2: Effects by type of institution

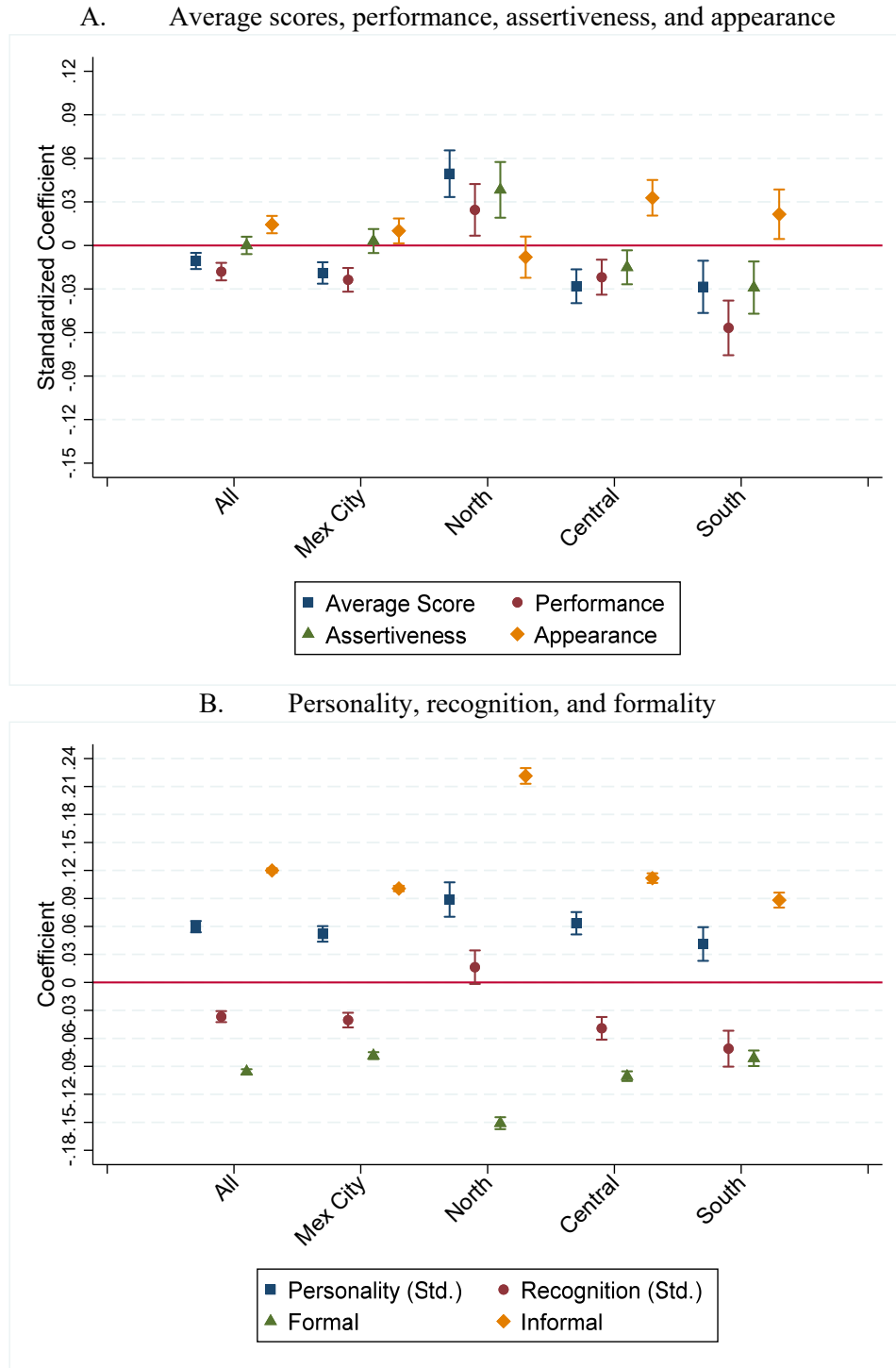


Note: Authors' elaboration. 95% confidence intervals. All variables are standardized except for formality (comments referring to "professor" versus "teacher"). Observations are included only where comments are included in the evaluation. Tech. refers to technical schools.

4.4 By Geographic Region

The region that gives the worst scores and the worst evaluation of performance to women is the south, although the confidence intervals are relatively large. It is followed by the center and Mexico City, which have a more precise estimation. Only in the north are formal terms used more often for men and informal terms more often for women. The southern region appears to show bias in scoring and commenting on performance, while using informal terms for women and formal terms for men less often than the other regions.

Figure 3: Effects by geographic region



Note: Authors' elaboration. 95% confidence intervals. All variables are standardized except for formality (comments referring to “professor” versus “teacher”). Observations are included only where comments are included in the evaluation. “North” includes all northern border states plus Baja California Sur; “south” includes all states south of Mexico City: Guerrero, Puebla, Tlaxcala, Morelos, Veracruz, Chiapas, Oaxaca, Campeche, Yucatán, and Quintana Roo; “central” refers to all remaining states except Mexico City.

4.5 Most Used Words

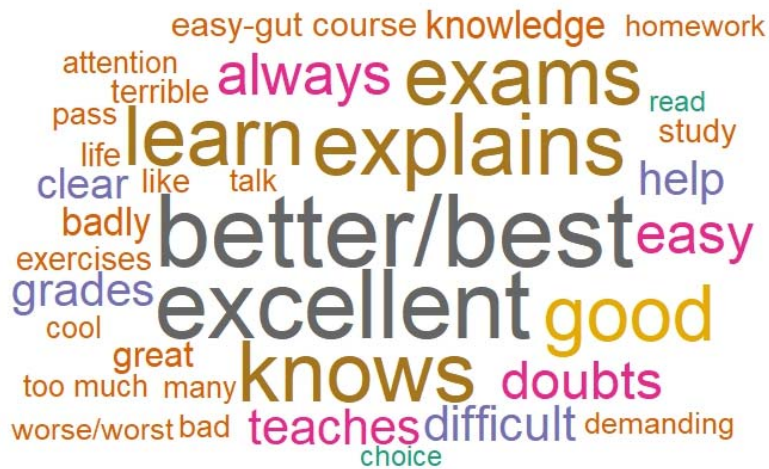
Finally, to find the differences in the words used in comments, we determine the words used most often in the comments. Figure 4 shows word clouds for male and female professors, excluding articles, simple verbs, pronouns, and other words known as “stop words” (e.g., “the,” “is,” “there is”). Diacritical marks are removed. The word used most often for female professors is “good”; for their male counterparts it is “excellent.” This observation confirms what has already been shown: that in general, men receive better evaluations for performance than women. Apart from this clear difference, other words are in general used equally for men and women.

Figure 4: Most used words by gender

A. Women



B. Men



Note: Authors' elaboration. The most used words are shown for each sex, not including stop words (pronouns, articles, etc.). We omit the word "buen" in Spanish because the translation is similar to "Bueno"/ "Buena". In Spanish, the most frequent words for men are "Excellent" and "Bueno".

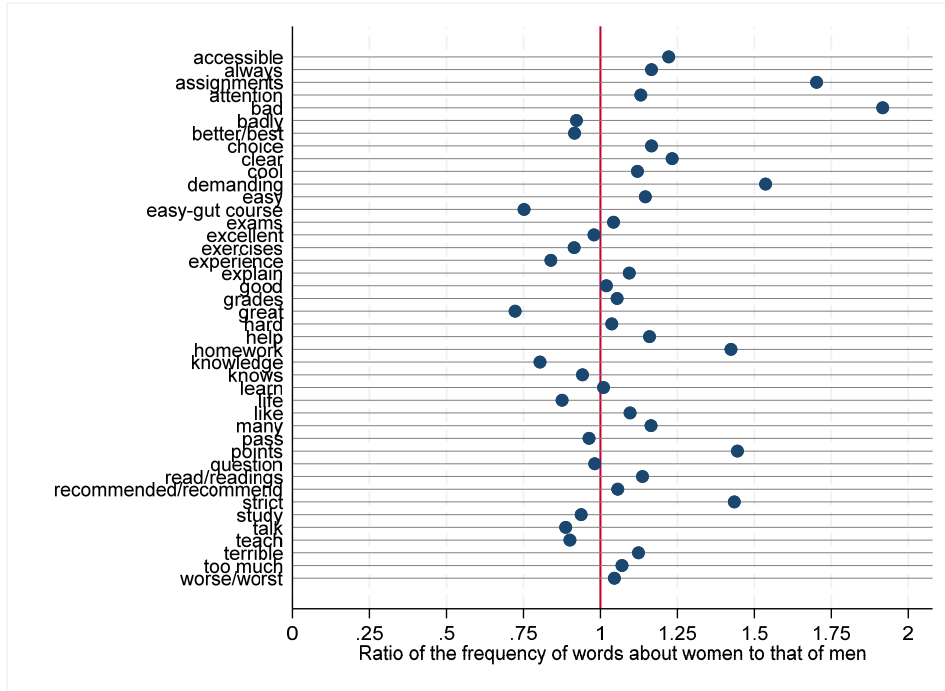
Figure 4 does not allow the visualization of the percentage difference in words for men and women. To address this problem and simplify the visualization, we use the 50 words shown in the figure, counting together words that differ only in their gendered endings ("buen," "buena," and "bueno," for example, all meaning "good," are counted together). Also counted together are minor variations like single and plural forms, as well as different conjugations of the same verb. We then add the words that appear only in the top 50 for both men and women. This leaves us with just 42 distinct words to analyze, 0.03% of the total number of unique words mentioned.

Given that a larger number of words appear in the comments about men (a total of 3,098,664 words, 102,907 distinct words) than about women (a total of 1,591,149 words, 66,849 distinct words), we calculate the proportion of occurrences for each gender. That is, for women, we count the number of times a word was mentioned and divide it by the total number of words mentioned in comments about women, and similarly for men. If the same words are used about women and men, this rate should be the same for both sexes.

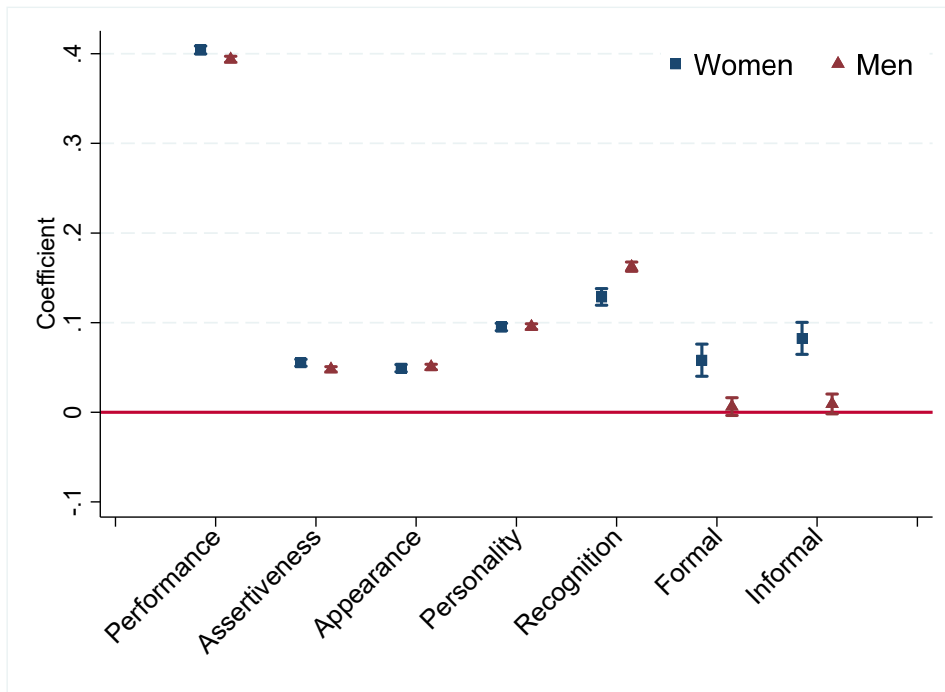
Figure 5 shows the ratio of frequencies for words used about women to those used about men (panel A). The five words used much more often about men than women (those with ratios closest to zero) are “great,” “easy” (“*barco*”), “knowledge,” “experience,” and “life.” The words used much more often about women than men (with a greater value on the x-axis) are “bad,” “assignment,” “demanding,” “points,” and “strict.” Notably, the word “bad” is used almost twice as often about women than about “men.”

Figure 5: Ratios of frequencies of most used words

A. Ratio of the frequency of most used words about women to that of men



B. Determinants of scores



Note: Authors' elaboration. Words that differ by gendered or plural endings or that are conjugations of the same verb are grouped together. Panel A is the ratio of occurrences referring to women to occurrences referring to men. Panel B shows the coefficients of a regression by gender with the average score as dependent variable, and the standardized measures of performance, assertiveness, appearance, personality, recognition, and the variables formal and informal, as explanatory variables.

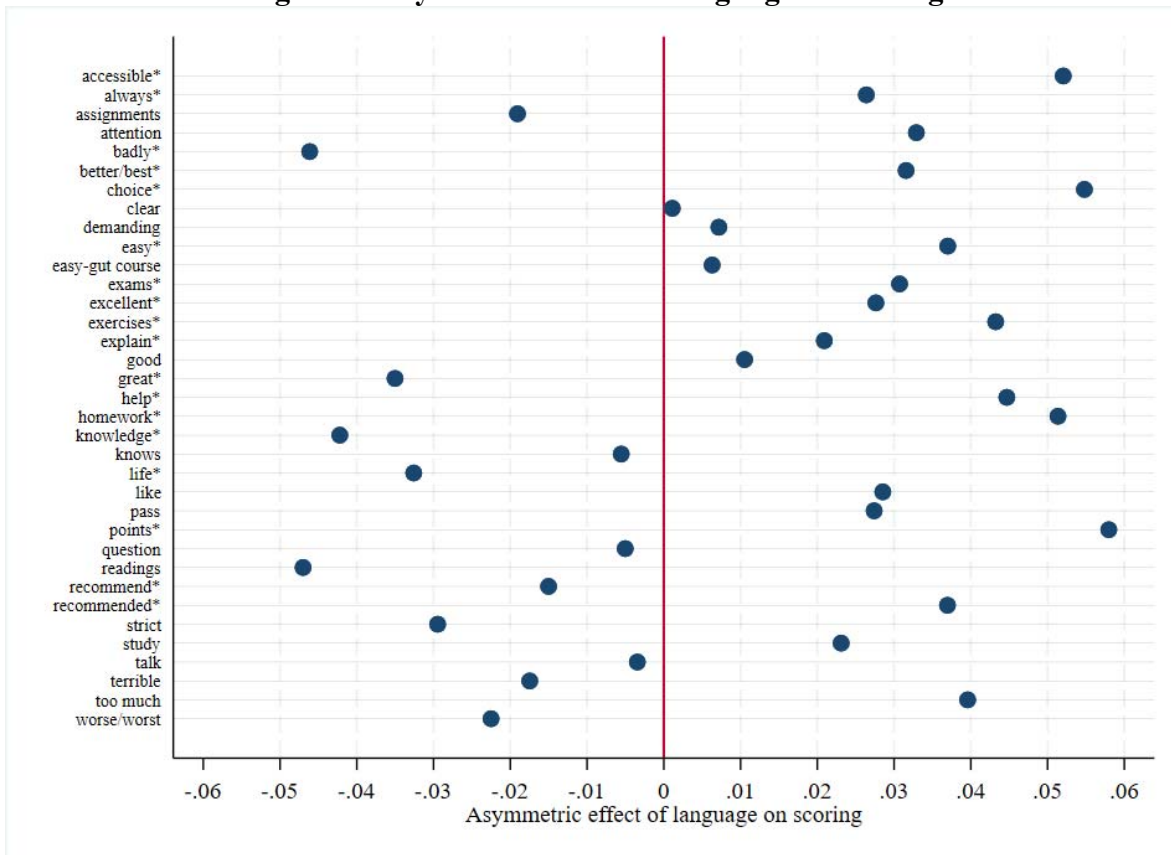
Panel B includes the determinants of the scores by gender. That is, a regression is estimated with each student's average score for a professor as the dependent variable and performance, assertiveness, personality, appearance, recognition, formality, and informality as independent variables. All of the variables have been standardized (except, as before, the opinions of formality and informality). These results are presented to determine whether the correlations are distinct within gender: whether there is an opinion associated more with a score for one gender than for the other. In general, the results are similar for men and women. The opinion most correlated with the average score is performance, followed by recognition. The coefficients are all similar, with the exception of recognition, which is slightly higher for men than for women. The coefficients that are not correlated with the scores are those for formality and informality for men; however, these are correlated for women. Their magnitude is statistically significant but small: less than 0.1, and less for other opinions.

Finally, we analyze how the language used in the evaluations affects the score given to male and female professors. We have seen that personality characteristics considered masculine are rated lower for women; that is, there is a penalty for not conforming to the gender stereotype. Figure 6 shows the results of this analysis. Each point is the coefficient of a regression of the total score on the interaction of a dummy variable for women with a dummy variable for the word used in the evaluation.⁹ Words that describe better service to students (a stereotypically female quality) by female professors—such terms as “accessible,” “help,” “easy,” and “points”—favor women more than men. However, some words related more to stereotypically male qualities—like “knowledge,” “strict,” and “experience”—favor women less than men. The fact that these asymmetries are observed precisely for words

⁹ The regression is controlled for the dummy variable for the word, the dummy for women, the interaction of these dummies, and the sentiment of the opinion.

expressing masculine or feminine stereotypes provides further evidence of the existence of gender stereotypes that affect female professors.

Figure 6: Asymmetric effect of language on scoring



Note: Authors' elaboration. Words that differ by gendered or plural endings or that are conjugations of the same verb are grouped together. Each point represents the coefficient of the total score in the interaction of a dummy variable for women with a dummy for the presence of a word in the evaluation. The regression controls for gender, the presence of the word, and the general score for the opinion in the comment. *signifies that the coefficient is statistically significant at the 5% level.

5. Discussion and Conclusions

The existence and use of gender stereotypes can have consequences for women's representation in certain spaces where it does not conform to stereotypes. Such is the case for teaching in higher education. Universities have traditionally been male spaces, and the entrance of women into academia has been met with both conscious and unconscious reactions, including gender stereotypes. Previous studies have found that female professors

are evaluated by different standards than their male counterparts, which translates into lower scores in evaluations of their teaching. Many universities use such evaluations in assessing the productivity and performance of their faculty. Thus, gender stereotypes among the student population can adversely affect the tenure and promotion of female professors.

In this study we investigate the existence of gender stereotypes in teaching evaluations on the website MisProfesores.com. Our data consist of all submissions to the site from 2008 to 2018, a total of 600,000 evaluations from all of Mexico. We make direct use of the scores assigned and also perform a text analysis of opinions in the “comments” section of each evaluation. Consistent with recent studies, we find that female professors do indeed receive worse evaluations than their male counterparts, although the magnitude of the difference is relatively small: 1-2% of a standard deviation. We also find that women are evaluated by different standards, as both their personality and appearance play a greater role in their evaluation than in that of men. Students likewise describe the work of their female professors using more adjectives that are negative or less positive than those they use for male professors: women are described as “bad” twice as often as men, and, unlike men, as “good” much more often than “excellent.” Finally, students refer to the female professors with less respectful words like “teacher” or “miss,” in contrast to the greater use of “professor” or terms corresponding to the academic degree of male professors.

A curious finding is that students make greater use of the words “strict,” “demanding,” “homework,” and “assignments” in evaluating courses taught by women. A woman who demands more and assigns more homework is not part of the female stereotype, so it is possible that students are punishing female professors for not conforming to the stereotype. Likewise, the scores for “help” are lower for women. Previous studies suggest that students expect much more help from women than from men, and that when they do not

receive it, they react much more negatively against women than against men. All of these stereotypes, which help to form certain kinds of expectations, end up affecting teaching evaluations.

Finally, we find that words associated with qualities of service (traditionally stereotyped as feminine), like “accessible,” “help,” “easy,” and “points,” favor women more than men. However, words with traditionally masculine associations, like “knowledge,” “experience,” and “strict,” have a negative impact on women’s evaluations. These asymmetries reflect the biases present in teaching evaluations that result from gender stereotypes.

Given these results and the fact that it will take time to change gender stereotypes, it has been suggested that teaching evaluations not be used as a mechanism for measuring the productivity of female professors, especially when teaching is considered in tenure and promotion decisions. Our own results for the Mexican case add to this concern, and we recommend that such instruments not be used to evaluate women in processes of recruitment, promotion, or tenure in academic institutions. Continuing to do so could affect the representation of women in academia, with all of the consequences that implies.

References

- Andersen, Steffen, Seda Ertac, Uri Gneezy, John A. List, and Sandra Maximiano (2013). "Gender, Competitiveness, and Socialization at a Young Age: Evidence from a Matrilineal and a Patriarchal Society." *Review of Economics and Statistics*, 95(4): 1438-43.
- Arceo, Eva, and Raymundo Campos (2014). "Race and marriage in the labor market: a discrimination correspondence study in a developing country." *American Economic Review*, 104(5): 376-280.
- Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart (2017). "Gender differences in accepting and receiving requests for tasks with low promotability." *American Economic Review*, 107(3): 714-747.
- Bayer, Amanda, and Cecilia Rouse (2016). "Diversity in the economics profession: A new attack on an old problem." *The Journal of Economic Perspectives*, 30(4): 221-242.
- Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova (2012). "Female leadership raises aspirations and educational attainment for girls: a policy experiment in India." *Science*, 335(6068): 582-586.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan (2005). "Implicit discrimination." *American Economic Review*, 95(2): 94-98.
- Bleske-Rechek, April, and Kelsey Michels (2010). "RateMyProfessors.com: Testing assumptions about students use and misuse." *Practical Assessment, Research & Evaluation*. 15(5): 1-12.
- Bohnet, Iris. (2016). *What works: Gender equality by design*. Harvard University Press, Cambridge, MA.

- Boring, Anne. (2017). "Gender biases in student evaluations of teaching." *Journal of Public Economics*, 145: 27-41.
- Brown, Michael J., Michelle Baillie, and Shawndel Fraser (2009). "Rating Ratemyprofessors.com: A comparison of online and official student evaluations of teaching." *College Teaching*, 57(2): 89-92.
- Clayson, Dennis E. (2014). "What does ratemyprofessors.com actually rate?" *Assessment & Evaluation in Higher Education*, 39(6): 678-698.
- Coladarci, Theodore e Irv Kornfield (2007). "RateMyProfessors.com versus formal in-class student evaluations of teaching." *Practical Assessment, Research & Evaluation*, 12(6): 1-15.
- Dasgupta, Nilanjana, and Shaki Asgari (2004). "Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping." *Journal of Experimental Social Psychology*, 40(5): 642-658.
- Dreber, Anna, Emma von Essen, and Eva Ranehill (2014). "Gender and Competition in Adolescence: Task Matters." *Experimental Economics*, 17(1): 154-72.
- El-Alayli, Amani, Ashley A. Hansen-Brown, and Michelle Ceynar (2018). "Dacing backwards in high heels: Female professors experience more work demand and special favor requests, particularly from academically entitled students." *Sex Roles*, 79(3-4): 136-150.
- Ellemers, Naomi. (2018). "Gender Stereotypes." *Annual Review of Psychology*, 69: 275-298.
- Felton, James, Peter T. Koper, John Mitchell and Michael Stinson (2008). "Attractiveness, easiness and other issues: student evaluations of professors on Ratemyprofessors.com." *Assessment & Evaluation of Higher Education*, 33(1): 45-61.

- Gneezy, Uri; Kenneth L Leonard, and John A. List (2009). "Gender differences in competition: Evidence from a matrilineal and a patriarchal Society." *Econometrica*, 77(5): 1637-64.
- Goldin, Claudia and Cecilia Rouse (2000). "Orchestrating impartiality: The impact of "blind" auditions on female musicians." *American Economic Review*, 90(4): 715-741.
- Gregory, Karen M. (2012). "How undergraduates perceive their professors: A corpus analysis of Rate My Professor." *Journal of Educational Technology Systems*, 40(2): 169-193.
- Guarino, Cassandra M. and Victor M. H. Borden (2017). "Faculty service loads and gender: Are women taking care of the academic family?" *Research in Higher Education*, 58(6): 672-694.
- INEGI (2015). "Estadísticas a propósito del día del maestro", Instituto Nacional de Estadística and Geografía.
<http://www.inegi.org.mx/saladeprensa/aproposito/2015/maestro0.pdf>. Accessed July 2018.
- Kindred, Jeannette and Shaheed N. Mohammed (2005). "“He will crush you like an academic ninja!”: Exploring teacher ratings on Ratemyprofessors.com." *Journal of Computer-Mediated Communication*, 10(3). <https://doi.org/10.1111/j.1083-6101.2005.tb00257.x>
- Koch, Amanda J., Susan D. D’Mello, and Paul R. Sackett (2015). "A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making." *Journal of Applied Psychology*, 100(1): 128-161.
- MacNeill, Lillian, Adam Driscoll, and Andrea N. Hunt (2015). "What’s in a name: Exposing gender bias in student ratings of teaching." *Innovative Higher Education*, 40(4): 291-303.

- Mengel, Friederike, Jan Sauermann, and Ulf Zölitz (2018). "Gender bias in teaching evaluations." *Journal of the European Economic Association*, forthcoming. <https://doi.org/10.1093/jeea/jvx057>.
- Miller, David, Alice Eagly, and Marcia Linn. (2015). "Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations." *Journal of Educational Psychology*, 107(3): 631-644.
- Mitchell, Kristina M.W., and Jonathan Martin (2018). "Gender bias in student evaluations." *PS: Political Science & Politics*, 51(3): 648-652.
- Legg, Angela M., and Janie H. Wilson (2012). "RateMyProfessors.com offers biased evaluations." *Assessment & Evaluation in Higher Education*, 37(1): 89-97.
- Lewandoski Jr., Gary W., Emma Higgins, and Natalie N. Nardone (2012). "Just a harmless website? An experimental examination of RateMyProfessors.com's effect on student evaluations." *Assessment & Evaluation in Higher Education*, 37(8): 987-1002.
- Otto, James, Douglas A. Sanford, Jr., and Douglas N. Ross (2008). "Does ratemyprofessor.com really rate my professor?" *Assessment & Evaluation in Higher Education*, 33(4): 355-368.
- Reid, Landon D. (2010). "The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com." *Journal of Diversity in Higher Education*, 3(3): 137-152.
- Rocha-Sánchez, Tania E., and Rolando Díaz-Loving (2005). "Cultura de género: la brecha ideológica entre hombres y mujeres." *Anales de Psicología*, 21(1): 42-49.
- Sarsons, Heather (2018). "Interpreting signals in the labor market: evidence from medical referrals." Working paper, Harvard University. <http://bit.ly/2zuENYB>. Accessed July 2018.

SEP (2017). "Sistema Educativo de los Estados Unidos Mexicanos." Secretaría de Educación Pública, México.

http://www.planeacion.sep.gob.mx/Doc/estadistica_e_indicadores/principales_cifras/principales_cifras_2016_2017.pdf. Accessed July 2018.

Sonntag, Michael E., Jonathan F. Bassett, and Timothy Snyder (2009). "An empirical test of the validity of student evaluations of teaching made on RateMyProfessors.com." *Assessment & Evaluation in Higher Education*. 34(5): 499-504.

Sutter, Matthias and Daniela Glätzle-Rützler (2015). "Gender Differences in the Willingness to Compete Emerge Early in Life and Persist." *Management Science*, 61(10): 2339-2354.

Stuber, Jenny M., Amanda Watson, Adam Clarke, and Kristin Staggs (2009). "Gender expectations and on-line evaluations of teaching: evidence from RateMyProfessors.com." *Teaching in Higher Education*, 14(4): 387-399.

Timmerman, Thomas (2008). "On the validity of RateMyProfessors.com." *Journal of Education for Business*, 84(1): 55-61.

Supplementary Materials

Table S1. Words in Spanish included in the analysis

Sentimiento	Palabras Positivas	Palabras Negativas
Desempeño	buena, bueno, buen, excelente, excel, comprometido, comprometida, super, bien, aprend, ensen, mejor, recomendable, recomendado, recomendada, clara, claro, recomiendo, interesante, domina, inteligente, calidad, motiva, genial, perfecto, capacidad, flexible	pesimo, pesima, mal, mala, malo, peor, confunde, confuso, tediosa, tedioso, horrible, no sabe
Asertividad	facil, examen, califica, tareas, ayuda, pone, pasar, pasas, pasa, accesible, sacar, ejercicio, trabajo, lecturas, tema, practica, entiend, proyecto, explica, asistencia, entrega, actividad, amena, prepar, apoya, apoyo, dispuesto, dispuesta, disposicion, justo, justa, puntual, flexible, aprueba, aprob	falta, dificil, barco, exigente, exige, estricto, estricta, complicad, impuntual, pesado, pesada, injusto, injusta, reprueba, reprob
Apariencia	alto, guapo, guapa, joven	viejo, vieja, feo, fea, asco
Personalidad	persona, onda, amable, preocup, amor, paciente, humano, humana, linda, agradable, inteligente, actitud, paciencia, humor, divertida, amo , chido, chida, cree, siente, puntual, atento, atenta, entretenida, divertido, divertida, pasion, tranquil	aburrida, aburrido, asco, mamon, despota, nefasto, mamona, flojo, floja, tediosa, tedioso, grosero, grosera, prepotente, enoja, arrogante, impuntual
Reconocimiento	profesor, maestro, maestra, profesora, doctor, doctora, ingeniero, ingeniera, prof, catedratic	pendej, cabron, pinche, puto, puta, mierda

Table S2. English-Spanish dictionary

Spanish	English	Spanish	English
accesible	accesible	explica	explains
aprender	learn	explicar	explains
aprendes	learn	fácil	easy
atención	attention	fáciles	easy
ayuda	help	gran	great
barco	easy-gut course	gusta	like
buena	good	habla	talk
buena onda	cool	leer	read
bueno	good	mal	badly
califica	grades	mala	bad
calificacion	grades	malo	bad
clara	clear	mejor	better/best
claro	clear	mejores	better/best
conocimiento	knowledge	muchas	many
conocimientos	knowledge	opción	choice
demasiado	too much	pasas	pass
difícil	difficult	peor	worse/worst
difíciles	difficult	pesima	terrible
duda	doubts	pesimo	terrible
dudas	doubts	puntos	points
ejercicios	exercises	recomendable	recommended/recommend
ensena	teaches	recomiendo	recommended/recommend
estricta	strict	sabe	knows
estudiar	study	siempre	always
exámenes	exams	tarea	homework
excelente	excellent	trabajos	assignments
exigente	demanding	vida	life
experiencia	experience		

Documentos
de trabajos
eBooks Novedades
Fondo
editorial
Revistas
eLIBROS LIBROS

www.LibreriaCide.com