FACULTAD ECONOMÍA Y NEGOCIOS
UNIVERSIDAD DE CHILE
DEPARTAMENTO DE ECONOMÍA

SERIE DE DOCUMENTOS DE TRABAJO

SDT 483

Does Facial Structure Explain Differences in Students Evaluations of Teaching? The Role of Perceived Dominance

Autores:
Valentina Paredes
Francisco Pino
David Díaz

Santiago, Mayo de 2019

sdt@econ.uchile.cl
econ.uchile.cl/publicaciones

Does Facial Structure Explain Differences in Students Evaluations of Teaching? The Role of Perceived Dominance*

Valentina Paredes† Francisco Pino‡ David Díaz§

May 2, 2019

Abstract

Dominance is usually viewed as a positive male attribute, but this is not typically the case for women. Using a novel dataset of teacher evaluations in a school of Business and Economics of a selective university, we construct the face width-to-height ratio (fWHR) as a proxy for dominance to assess whether individuals with a higher ratio obtain better student evaluations of teaching. Our results suggest that a higher fWHR is associated with a better evaluation for male faculty, while the opposite is the case for females. These results are not due to differences in teachers' productivity. Because teacher evaluations are relevant for pay and promotion, this might contribute to the underrepresentation of women in economics.

JEL Codes: J16, I23, M51, A22.

Keywords: Student Evaluations of Teaching, Facial Structure, Dominance, Gender biases.

*We thank Osea Giuntella, Ana Nuevo-Chiquero and Esteban Puentes for their helpful comments. Funding: This work was supported by Fondecyt Inicio [Project 11170968], and the Centre for Social Conflict and Cohesion Studies (COES) [CONICYT/FONDAP/15130009]. Correspondence should be addressed to Valentina Paredes, Department of Economics, University of Chile, Diagonal Paraguay 257, Santiago, Chile. E-mail: vparedes@fen.uchile.cl.

†Department of Economics, University of Chile.

‡Department of Economics, University of Chile.

§Department of Management, University of Chile.

1 Introduction

Some characteristics are not equally valued in women as in men. For example, several studies have found that women who behave in a dominant, assertive or agentic way are not as well received as men who engage in the same behaviors (Williams and Tiedens, 2016; Rudman, 1998; Butler and Geis, 1990; Bertrand, 2011). This phenomenon is known as the backlash effect (Rudman and Glick, 2001) and can affect women's evaluations (Eagly, Makhijani, and Klonsky, 1992) or likeability (Costrich, Feinstein, Kidder, Marecek, and Pascale, 1975; Powers and Zuroff, 1988).

One area where women might be encountering backslash is academia. Recent studies show that women are underrepresented not only in engineering and physical sciences, but also in social sciences (Ginther and Kahn, 2009; Ceci, Ginther, Kahn, and Williams, 2014; Ceci, Ginther, Kahn, and Williams, 2015; Kahn and Ginther, 2018). A recent survey conducted by the American Economic Association about the professional climate in the field found that 48 percent of women reported being discriminated against based on their sex (American Economic Association, 2019). One particular aspect where women experience discrimination is course evaluations, with 47 percent of women reporting experiences of discrimination or unfair treatment, compared to 8 percent of men. Nowadays, most universities use student evaluations of teachers (SET) to evaluate their professors (Becker and Watts, 1999; Hamermesh and Parker, 2005). Some studies suggest that SET are not good at measuring teaching quality (Braga, Paccagnella, and Pellizzari, 2014; Uttl, White, and Gonzalez, 2017), while others have found that characteristics not related to quality, such as gender and professors' perceived beauty, also have an effect on SET (Boring, 2017; Mengel, Sauermann, and Zölitz, 2019; Hamermesh and Parker, 2005). SET are usually taken into account for promotion or hiring decisions, and can therefore be crucial for the representation of women in academia.

In this paper we analyze whether dominance has an effect on student evaluations of teaching with a dataset of all professors of a school of business and economics in Chile.[1] In line with previous studies that have reported a backlash effect, we expect female professors who are perceived as dominant to have lower SET scores than professors that are perceived

[1] The Merriam-Webster dictionary defines dominance as "the fact or state of being dominant: such as controlling, prevailing, or powerful position especially in a social hierarchy". For more discussion on what is understood as dominant behavior, see Geniole, Denson, Dixson, Carré, and McCormick (2015) and Williams and Tiedens (2016).

to conform to more traditional gender stereotypes. The opposite should be the case for men, where dominance might be perceived as a desirable attribute.

Dominance can be measured by asking subjects to rate their dominance using a dominance scale such as the one included in the Revised Interpersonal Adjective Scale (IAS-R, Wiggins, Trapnell, and Phillips, 1988). Another alternative is to ask independent raters to grade individuals based on their pictures (Valentine, Li, Penke, and Perrett, 2014). To refrain from asking professors to complete a dominance scale, to measure dominance we use the facial width-to-height ratio (fWHR). Several studies have found a positive relationship of this measure to dominance (Geniole, Denson, Dixson, Carré, and McCormick, 2015; Alrajih and Ward, 2014; Mileva, Cowan, Cobey, Knowles, and Little, 2014; Dixson, 2017). Valentine, Li, Penke, and Perrett (2014) found that the fWHR significantly predicts perceived dominance, evaluated by independent raters. Moreover, the fWHR has been found to predict a superior financial performance for CEOs (Wong, Ormiston, and Haselhuhn, 2011), higher endorsement income among celebrities (Huh, Yi, and Zhu, 2014), and better academic performance among students (Kausel, Ventura, Datawheel, Díaz, and Vicencio, 2018). Although some papers argue that it is a better proxy of dominant behaviour for men than for women, a meta analysis of all peer-reviewed studies done in Geniole, Denson, Dixson, Carré, and McCormick (2015) shows that the fWHR predicts dominance behaviour in both sexes.

Our results show that dominance, proxied by the fWHR, has a statistically significant and economically meaningful effect on SET scores. In our preferred specification, an increase of 1 s.d. in the fWHR raises SET scores by 0.16 s.d. for the principal component of all questions included in the SET, 0.12 s.d. for overall grade given by students to the professor, and 0.17 s.d. for the average of questions 1, 3 and 4 included in the SET, which are the questions that the school uses to assess teaching quality. However, this average effect masks significant heterogeneity across genders: while a higher fWHR increases SET scores for men, the opposite is true for women. Since male professors comprise 88 percent of the faculty in our sample, the effect for this group is similar to the average effect. For the case of female professors, however, an increase of 1 s.d. in the fWHR decreases SET scores by 0.55 s.d. for the principal component.

Even though previous literature has found that the fWHR is a good proxy for dominance, some studies have found that fWHR also correlates with attractiveness (Sadalla, Kenrick,

and Vershure, 1987; Valentine, Li, Penke, and Perrett, 2014; Geniole, Denson, Dixson, Carré, and McCormick, 2015). We test for the robustness of our results by including controls for perceived beauty.[2] Although beauty does have an impact on SET scores, it does not drive our results.

It may also be the case that dominance has an impact on SET scores by increasing professors' productivity. To address this issue, we control for grades in coordinated courses as a proxy of teaching quality. When controlling for teaching productivity, we find that the effect of dominance is smaller in magnitude than in our main specification (an increase of 1 s.d. in the fWHR increases SET scores by 0.10-0.12 s.d. for men and decreases SET scores by 0.29-0.45 s.d. for women). However, the point estimates are not statistically different from each other, suggesting that productivity alone cannot explain our results.

The results presented in this paper are relevant beyond the educational sector. They imply that facial features such as the fWHR, a proxy for dominance, have a differential effect by gender on SET scores independent of professors' productivity. Any context in which individuals such as managers, co-workers, or employees evaluate another individual are subject to this bias, particularly in those contexts where there is already evidence of an effect of facial attributes on career success (see Ling, Luo, and She, 2019 for public servants or Lawson, Lenz, Baker, and Myers, 2010 for elected politicians).

The rest of the paper is organized as follows. In section 2 we describe the setting and present the data. In section 3 we discuss the empirical strategy and present the results. Section 4 discusses alternative mechanisms such as differences in beauty or productivity. Finally, section 5 states the conclusion.

2 Data

This study employs data from the universe of students who were enrolled in the School of Business and Economics of a selective university in Chile between 2012 and 2018. This school offers three 5-year undergraduate programs, commercial engineering (IC), engineering in information systems and management control (IICG) and accounting and auditing. All programs include courses in economics, business, accounting, information systems, quantitative methods, general interest, English and sports. During the time frame analyzed, 4,533 courses were offered in the School.

[2] A beauty score is calculated from geometric proportions of facial features.

The school in our study uses SET to evaluate instructors. At the end of the course, all students must complete a evaluation of teaching. SETs in this school consist of an electronic 12 question survey about the instructor, 23 question survey about the course and a section of comments about the teaching assistants. Students need to complete the survey in order to have access to the course webpage. Therefore, we have data from student evaluations for all classes taught in the school of economics and business between 2012 and 2018. Questions included in the SET are presented in Table 1.

As described above, the student evaluation consists of 12 questions, that capture different aspects of teachings. In our analysis we construct three dependent to measure SET. First, we extract the first principal component from these questions. Second, we include an overall grade given by students at the end of the questionnaire. Finally, we take the average of questions 1, 3 and 4. This average is used by the school to rank professors and give monetary awards. All three measures are standardized to have a mean of 0 and standard deviation of 1 in the sample of professors. Table 2 shows professor-level averages for all questions, as well as for our three outcomes. It also shows these averages by gender of the professor. Female and male professors receive on average the same scores in all measures, and in all individual questions.

As a proxy of dominance, we use the instructors' facial width to height ratio (fWHR). This ratio is the facial height divided by the facial width, were the height is the distance between the upper lip and the highest point in the middle of the eyebrows and the width is the maximum horizontal distance between the left and right facial boundary. We used FACE++ to compute both measures (see Kausel, Ventura, Datawheel, Díaz, and Vicencio, 2018 for more details on the calculation procedure).

The fWHR was calculated from the instructors' pictures in the university webpage. Therefore, we can only compute the measure for instructors that belong to one of the three departments, since the picture of other instructors is not available.[3] In what follows, we will refer to instructors (full time or part time) that belong to one of the three departments as professors.[4] In total, we have fWHR measures for all the professors in our sample, 83

[3]Department of Economics, Department of Management and Department of Management Control and Information Systems

[4]Professors correspond to 13% of all instructors and taught 31.7% of the total number of classes during this period, which is equivalent to 1,438 courses. Professors generally teach core courses, compared to lecturers. For example, 100% of English and sport courses were offered by lecturers. During this time period, professors taught 501 courses in Business, 383 courses in Economics, 247 courses in Accounting, 141 courses in quantitative methods, 102 courses in information systems, and only 65 courses in general

men and 12 women. The fWHR ranges in our sample from 1.44 to 2.15, with a mean of 1.76 and a standard deviation of 0.14.

Finally, the school provided class level information such as number of students in the class, semester and year the class was taught, percentage of female students, percentage of failing students, the average grade in the class and the average math admission score in the class.

3 Results

3.1 Identification

Our main regression is the following

$$y_{ic} = \alpha_s + \beta fWHR + X_i\gamma + Z_{ic}\psi + \varepsilon_{ic} \tag{1}$$

where i denotes the professor, and c denotes the class. X include professor characteristics such as gender and a dummy for whether the professor has a PhD degree.[5] Z includes classroom controls: we include controls for the number of students, the share of female students, and the average admission math score. There are several papers that have found that larger classes have a negative effect on student learning (for example, Angrist and Lavy, 1999 and Krueger, 1999). Therefore, we expect that students in larger classes give their professors lower evaluations. We include the share of female students because it may be the case that male and female students give their professors different scores (both Mengel, Sauermann, and Zölitz, 2019, and Boring, 2017 find evidence on gender bias in teaching evaluations that is driven by male students' evaluations). We also include course fixed effects, to capture differences in the course content that may explain differences in evaluations. Finally, we include semester and year fixed effects. Standard errors are clustered at the professor level.

One identification concern is that students are not randomly assigned to professors Paredes (2014). For example, if higher achieving students select into classrooms of professors

interest. Professors and lecturers also differ in some observable characteristics. Table A-2 compares the characteristics of the 95 professors versus the 652 lecturers and Table A-3 compares the characteristics of the 1,439 classes taught by professors versus the characteristics of the 3,094 classes taught by lecturers. The percentage of female professors is lower than the percentage of female lecturers. Professors teach almost 10 courses more than lecturers. Compared to lecturers, professors have a statistically significantly higher score in factor 3, but are not different than lecturers regarding factors 1 and 2 (Table A-2).

[5]This information was obtained from each professor's résumé.

with higher fWHR, our results would be biased. To address this concern, we control for the average admission math score. This score predicts students' performance, but is not affected by professors' characteristics.[6]

3.2 Main results

Panel A in Table 3 shows the results of estimating equation 1 for all professors and the three measures.[7] Columns (1), (4) and (7) estimates equation 1 without the controls (only including course fixed effects). Columns (2), (5) and (8) include semester and year fixed effects in addition to course fixed effects. Finally, columns (3), (6) and (9) include the previous fixed effects and other professor and classroom controls. The coefficient on the Female dummy is negative, indicating that female professors receive worse evaluations. However, in none of the specifications it is statistically significant. As expected, larger classrooms have a negative effect on SET scores.

Regarding the fWHR, in all columns the effect is positive and significant. In our preferred specification (including all controls), an increase of 1 s.d. in the fWHR (0.145) raises SET scores by 0.16 s.d. for the principal component, 0.12 s.d. for the professor grade, and 0.17 s.d. for the average of questions 1, 3 and 4. Using the fWHR as a proxy for dominance, professors who exhibit a more agentic behavior are better evaluated. The magnitude of the effect is comparable with results found in Kausel, Ventura, Datawheel, Díaz, and Vicencio (2018), who find that an increase of 1 s.d. in fWHR of students raises their grades in non-quantitative courses by 0.2 s.d.

Our main hypothesis is that agentic behavior is not equally valued in men than in women. Therefore, we reestimate equation 1 separating female and male professors. Panel B in Table 3 shows the results of estimating equation 1 for male professors, while panel C does the same for female professors. For male professors, the results are similar to those for all professors, although they are larger in magnitude. For instance, an increase of 1 s.d. in the fWHR raises the principal component by 0.17 s.d. This is not surprising given that male professors account for 90% of classes in the sample. However, when we estimate equation 1 for female professors, we find that a higher fWHR decreases the student evaluation score

[6] See more details of the university admission system in Chile in Bordon and Fu (2015) and Solis (2017).

[7] In Table A-4 we repeat the analysis for individual questions. The results show that for male professors the effect of fWHR is positive for all questions, while the opposite is true for female professors. Thus, the results found in table 3 do not seem to be driven by a single question.

(Panel C). The results are robust to the inclusion of controls, and are larger in absolute magnitude with respect to those for men. For example, an increase of 1 s.d. in the fWHR reduces principal component by 0.55 s.d.

Regarding the effect of control variables by gender, larger classes have a negative effect for both male and female professors, although the effect is larger for women. Women holding a PhD degree have better SET scores, while there is no significant effect for their male counterparts.

Figures A-1-A-3 in the Appendix show graphically the relationship between the fWHR and the principal component for both men and women. That is, we plotted the residuals from the regressions in Columns (3), (6) and (9) in Panel B and C of Table 3 against the fWHR. As previously discussed, the relationship is positive for men and negative for women. We can also see that women with low fWHR have higher evaluations than men with similar fWHR, and the opposite is true for women and men with high fWHR. This explains why, on average, we find no significant difference in evaluations of men and women.

3.3 Robustness checks

One possible concern regarding the results found in the previous section is that our regressions include only 12 female professors. Therefore, the negative effect of the fWHR on SET scores for women could be driven by an outlier. To further test the robustness of the results, we reestimate equation 1 excluding one professor at the time. Figures A-4-A-9 graph the coefficients and 95% confidence interval for the three different dependent variables, for male and female professors. Although the graphs show some variation in the coefficients, by and large the results hold: The fWHR coefficient is always positive (and statistically significant at a 5% significance level 98% of times) for male professors. For female professors, the effect is always negative and statistically significant for both the principal component and the overall grade, while for the average of questions 1, 3 and 4 we lose significance in one case (although the coefficient is significant at 10%). In addition, table A-5 in the appendix shows estimates of equation 1 using quantile regression with course fixed effects (Machado and Santos Silva, forthcoming). The results are similar to our baseline specification in both sign and magnitude. Overall, the results are not driven by one professor.

Another potential issue is that the fWHR might have some measurement error when professors do not directly face the camera in their pictures. To solve this, the software

computes the angle of rotation, and thus we compute an angle-corrected fWHR. Table 4 shows the results of estimating equation 1 using the corrected fWHR. In addition, in Columns 2, 4 and 6 we exclude professors whose angle of rotation is over 15 degrees. We find that the results are robust to the correction in fWHR, although we lose precision for male professors, particularly in the principal component measure.

4 Discussion

Our main hypothesis is that the fWHR, a proxy for dominant behavior, has different effects for the evaluation of men and women. Alternatively, fWHR may be capturing other professor characteristics. For example, the fWHR may correlate with beauty, and therefore, what we find is not driven by agentic behavior, but by perceived beauty. The fWHR may also correlate with differences in productivity, and therefore, students may be not responding to the dominant behavior but to productivity.

4.1 Differences in beauty

Several papers have found that beauty can have an impact on different outcomes, such as wages, labor force participation (Hamermesh and Biddle, 1994) and teacher evaluations (Hamermesh and Parker, 2005). If the fWHR correlates with beauty, especially if the correlation is positive for men and negative for women, our results could be driven by differences in beauty and not by differences in agentic behavior. Indeed, some studies find that the fWHR also correlates with attractiveness for men. For example, Sadalla, Kenrick, and Vershure (1987) argue that dominant men are sexually attractive to women, and Valentine, Li, Penke, and Perrett (2014) show that the fWHR and perceived dominance correlates with attractiveness in the context of speed dating.

To test if our results are driven by differences in attractiveness, we control for a beauty score calculated by FACE++. The program gives two beauty scores: one that would be given by a male, and one that would be given by a woman. Both scores range form 0 to 100, with a higher score indicating a more beautiful face. In general, the male score is lower and has higher variance than the female score. Because we have the percentage of female students in each class, we use both scores to build the beauty score that would be given by the class.

Table 5 shows the results of estimating equation 1 controlling for the beauty score. A higher beauty score does not have a significant effect on student evaluations scores for men, while it decreases SET scores for women.[8] More importantly, our previous results hold. The fWHR has a positive effect for male professors and a negative effect for female professors.

4.2 Differences in productivity

An important determinant of student evaluations that is omitted from all the previous regressions is teaching quality. A higher quality professor, holding everything else constant, should receive higher evaluations than lower quality professors. And if teaching quality correlates with the fWHR, omitting it from our regressions would bias our results.

A proxy for teaching quality could be grades. Students assigned to a higher quality professor should obtain higher grades. Unfortunately, professors might have different grading standards. Moreover, professors may inflate grades to get higher SET scores (McPherson, 2006; Ewing, 2012). Therefore, the average student grade will capture both the grading standard and the teaching quality.

Luckily, in the school where our study was conducted, courses taught in the same semester are usually coordinated. That is, although the courses are taught by different professors, the exam is the same for all sections. Because the exam is the same, grades within coordinated courses should reflect teaching quality. Therefore, we can control for students grades within coordinated courses using course×semester×year fixed effects to control for teaching quality.

Table 6 shows the baseline results (Columns 1, 4 and 7), the results after controlling for students' grades and the share of failing students (Columns 2, 5, and 8) and the results controlling for students' grades, the share of failing students and course×semester×year fixed effects (Columns 3, 6 and 9). Overall, our results hold. The effect of fWHR is positive for male professors, although we lose significance in specifications including course×semester×year fixed effects. For the case of female professors, results are negative and significant in all specifications. The average student grade is positive and significant for male professors, however, there is no effect of students' grades on female professors evaluations. In terms of the size of the coefficient, when we control for teaching quality, the fWHR for men decreases

[8]This result is not robust to estimating equation 1 using quantile regression. Results are available upon request.

in 33, 24 and 41 percent for the principal component, the overall grade and the average of questions 1, 3 and 4, respectively. Therefore, it appears that dominance measured by the fWHR positively correlates with teaching quality for men. That is, 33 percent of the effect on the principal component is explained because dominance seems to improve teaching quality, while the other 67 percent is explained by students giving higher ratings to professors with higher levels of perceived dominance, independently of their quality. We find a similar result for female professors: When we control for teaching quality, the fWHR coefficient decreases in 18, 19 and 20% respectively. Finally, when comparing coefficients with and without controlling for productivity, we find that they are not statistically different from each other, except for the average of questions 1, 3 and 4 for female professors.[9]

5 Conclusion

In this paper we analyze whether dominance has an effect on student evaluations of teaching (SET). Our results show that dominance, proxied by the fWHR, has a statistically significant and economically meaningful effect on SET. In our preferred specification, an increase of 1 s.d. in the fWHR raises SET scores by 0.12-0.17 s.d. More important, we find significant differences between male and female professors. While a higher fWHR increases SET scores for men, the opposite is true for women.

Throughout the paper, we have assumed that the fWHR is a good proxy of dominance. This is supported by the meta analysis done in Geniole, Denson, Dixson, Carré, and McCormick (2015). Even though the fWHR has been found to correlate with other characteristics such as attractiveness, in our analysis we control for factors such as beauty and productivity, so that our results could be attributed to dominance. However, even if our results were driven by other teacher characteristic that correlates with fWHR, we can still conclude that characteristics not related to teacher quality have an impact on SET.

Our results shed light on the use of SET to evaluate instructors, and more generally to its use as a tool for promotion or hiring. Even though in our context we do not observe an average gender gap in student evaluations, the results suggest that dominant men and women are evaluated differently by students. For instance, a male professor in the 90th

[9]We also estimate the impact of the fWHR on students' grades. That is, we replace the dependent variable on equation 1 for students' grade. We find no significant effect, providing additional evidence that the effect of the fWHR on students' evaluations score is not driven by differences in productivity (Table A-6).

percentile of the fWHR distribution has an SET score of 0.15 s.d. higher than the average professor, while a female professor with the same fWHR would see her SET score reduced by 0.79 s.d. To the extend that fWHR is a good proxy of dominance, it appears that conformity to traditional gender norms pays off for both men and women. The cost of challenging these norms, however, is much larger for women than men.

References

Alrajih, S., and J. Ward (2014): "Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders," *British Journal of Psychology*, 105(2), 153–161.

American Economic Association (2019): "AEA Professional Climate Survey: Main Findings," Discussion paper, American Economic Association.

Angrist, J. D., and V. Lavy (1999): "Using Maimonides' rule to estimate the effect of class size on scholastic achievement," *The Quarterly journal of economics*, 114(2), 533–575.

Becker, W. E., and M. Watts (1999): "How departments of economics evaluate teaching," *American Economic Review*, 89(2), 344–349.

Bertrand, M. (2011): "New Perspectives on Gender," in *Handbook of Labor Economics*, ed. by D. Card, and O. Ashenfelter, vol. 4B, pp. 1543–1590. North Holland.

Bordon, P., and C. Fu (2015): "College-major choice to college-then-major choice," *The Review of economic studies*, 82(4), 1247–1288.

Boring, A. (2017): "Gender biases in student evaluations of teaching," *Journal of Public Economics*, 145, 27–41.

Braga, M., M. Paccagnella, and M. Pellizzari (2014): "Evaluating students' evaluations of professors," *Economics of Education Review*, 41, 71–88.

Butler, D., and F. L. Geis (1990): "Nonverbal affect responses to male and female leaders: Implications for leadership evaluations.," *Journal of Personality and social Psychology*, 58(1), 48–59.

Ceci, S., D. Ginther, S. Kahn, and W. Williams (2015): "Women in science: The path to progress," *Scientific American Mind*, 26(1), 62–69.

Ceci, S. J., D. K. Ginther, S. Kahn, and W. M. Williams (2014): "Women in academic science: A changing landscape," *Psychological Science in the Public Interest*, 15(3), 75–141.

Costrich, N., J. Feinstein, L. Kidder, J. Marecek, and L. Pascale (1975): "When stereotypes hurt: Three studies of penalties for sex-role reversals," *Journal of Experimental Social Psychology*, 11(6), 520–530.

Dixson, B. J. W. (2017): "Facial Width to Height Ratio and Dominance," in *Encyclopedia of Evolutionary Psychological Science*, ed. by T. K. Shackelford, and V. A. Weekes-Shackelford, pp. 1–4. Springer International Publishing, Cham.

Eagly, A. H., M. G. Makhijani, and B. G. Klonsky (1992): "Gender and the evaluation of leaders: A meta-analysis.," *Psychological bulletin*, 111(1), 3–22.

Ewing, A. M. (2012): "Estimating the impact of relative expected grade on student evaluations of teachers," *Economics of Education Review*, 31(1), 141–154.

Geniole, S. N., T. F. Denson, B. J. Dixson, J. M. Carré, and C. M. McCormick (2015): "Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat," *PloS one*, 10(7), e0132726.

Ginther, D. K., and S. Kahn (2009): "Does Science Promote Women? Evidence from Academia 1973-2001," in *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, ed. by R. B. Freeman, and D. L. Goroff, pp. 163–194. University of Chicago Press.

Hamermesh, D. S., and J. E. Biddle (1994): "Beauty and the labor market," *The American Economic Review*, 84(5), 1174–1194.

Hamermesh, D. S., and A. Parker (2005): "Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity," *Economics of Education Review*, 24(4), 369–376.

Huh, H., D. Yi, and H. Zhu (2014): "Facial width-to-height ratio and celebrity endorsements," *Personality and Individual Differences*, 68, 43–47.

Kahn, S., and D. K. Ginther (2018): "Women and Science, Technology, Engineering, and Mathematics (STEM): Are Differences in Education and Careers Due to Stereotypes, Interests, or Family?," in *The Oxford Handbook of Women and the Economy*, ed. by S. L. Averett, L. M. Argys, and S. D. Hoffman, pp. 767–798. Oxford University Press.

Kausel, E. E., S. Ventura, L. Datawheel, D. Díaz, and F. Vicencio (2018): "Does facial structure predict academic performance?," *Personality and Individual Differences*, 129, 1–5.

Krueger, A. B. (1999): "Experimental estimates of education production functions," *The quarterly journal of economics*, 114(2), 497–532.

Lawson, C., G. S. Lenz, A. Baker, and M. Myers (2010): "Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies," *World Politics*, 62(4), 561–593.

Ling, L., D. Luo, and G. She (2019): "Judging a book by its Cover: The influence of physical attractiveness on the promotion of regional leaders," *Journal of Economic Behavior and Organization*, 158, 1–14.

Machado, J. A. F., and J. M. C. Santos Silva (forthcoming): "Quantiles via moments," *Journal of Econometrics*.

McPherson, M. A. (2006): "Determinants of how students evaluate teachers," *The Journal of Economic Education*, 37(1), 3–20.

Mengel, F., J. Sauermann, and U. Zölitz (2019): "Gender bias in teaching evaluations," *Journal of the European Economic Association*, 17(2), 535–566.

Mileva, V. R., M. L. Cowan, K. D. Cobey, K. Knowles, and A. Little (2014): "In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men," *Personality and Individual Differences*, 69, 115–118.

Paredes, V. (2014): "A teacher like me or a student like me? Role model versus teacher bias effect," *Economics of Educacion Review*, 39, 38–49.

Powers, T. A., and D. C. Zuroff (1988): "Interpersonal consequences of overt self-criticism: A comparison with neutral and self-enhancing presentations of self.," *Journal of Personality and Social Psychology*, 54(6), 1054.

Rudman, L. A. (1998): "Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management.," *Journal of personality and social psychology*, 74(3), 629–645.

Rudman, L. A., and P. Glick (2001): "Prescriptive gender stereotypes and backlash toward agentic women," *Journal of social issues*, 57(4), 743–762.

Sadalla, E. K., D. T. Kenrick, and B. Vershure (1987): "Dominance and heterosexual attraction.," *Journal of Personality and Social Psychology*, 52(4), 730.

Solis, A. (2017): "Credit access and college enrollment," *Journal of Political Economy*, 125(2), 562–622.

Uttl, B., C. A. White, and D. W. Gonzalez (2017): "Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related," *Studies in Educational Evaluation*, 54, 22–42.

Valentine, K. A., N. P. Li, L. Penke, and D. I. Perrett (2014): "Judging a Man by the Width of His Face," *Psychological science*, 25(3), 806–811.

Wiggins, J. S., P. Trapnell, and N. Phillips (1988): "Psychometric and Geometric Characteristics of the Revised Interpersonal Adjective Scales (IAS-R)," *Multivariate Behavioral Research*, 23(4), 517–530.

Williams, M. J., and L. Z. Tiedens (2016): "The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior.," *Psychological Bulletin*, 142(2), 165–197.

Wong, E. M., M. E. Ormiston, and M. P. Haselhuhn (2011): "A face only an investor could love: CEOs' facial structure predicts their firms' financial performance," *Psychological Science*, 22(12), 1478–1483.

Tables

Table 1: Teacher Evaluation Survey Questions

		Mean	Standard Dev	Min	Max
Question 1	Shows security and mastery over the subject	6.05	0.53	1.64	7.00
Question 2	Prepares the class	5.90	0.56	1.64	7.00
Question 3	Exhibits clearly	5.68	0.68	1.64	7.00
Question 4	Solves doubts and problems	5.83	0.58	1.68	7.00
Question 5	Encourages discussion and participation	5.67	0.66	1.64	7.00
Question 6	Allows asking questions and expressing ideas	5.89	0.57	1.68	7.00
Question 7	Stimulates interest	5.58	0.69	1.64	7.00
Question 8	Fair and reasonable evaluations	5.48	0.71	1.68	7.00
Question 9	Attends classes promptly	5.81	0.71	1.68	7.00
Question 10	Meets deadlines and established standards	5.77	0.64	1.68	7.00
Question 11	Treats students with respect	5.95	0.58	1.68	7.00
Question 12	Is available for students	5.78	0.60	1.64	7.00

Notes: Questions from the Student Evaluation of Teachers (SET) survey. Table reports averages at the professor-course-semester level.

Table 2: Professors Observed Characteristics

Variables	All Professors	Male Professors	Female Professors	Difference
fWHR	1.749	1.746	1.769	-0.023
	(0.132)	(0.132)	(0.132)	(0.041)
Female	0.126	0.000	1.000	-1.000
	(0.334)	(0.000)	(0.000)	(0.000)
PhD	0.716	0.723	0.667	0.056
	(0.453)	(0.450)	(0.492)	(0.141)
Number of courses	15.147	15.470	12.917	2.553
	(10.949)	(11.159)	(9.482)	(3.389)
Question 1	6.033	6.052	5.897	0.155
	(0.375)	(0.372)	(0.380)	(0.115)
Question 2	5.924	5.922	5.932	-0.010
	(0.385)	(0.394)	(0.333)	(0.120)
Question 3	5.687	5.690	5.666	0.024
	(0.509)	(0.517)	(0.475)	(0.158)
Question 4	5.865	5.871	5.823	0.049
	(0.400)	(0.395)	(0.454)	(0.124)
Question 5	5.692	5.702	5.624	0.078
	(0.497)	(0.498)	(0.508)	(0.154)
Question 6	5.939	5.945	5.897	0.048
	(0.378)	(0.370)	(0.447)	(0.117)
Question 7	5.598	5.610	5.515	0.094
	(0.516)	(0.512)	(0.559)	(0.160)
Question 8	5.553	5.561	5.493	0.068
	(0.456)	(0.465)	(0.406)	(0.141)
Question 9	5.861	5.851	5.935	-0.085
	(0.489)	(0.507)	(0.354)	(0.152)
Question 10	5.832	5.838	5.793	0.045
	(0.401)	(0.411)	(0.337)	(0.125)
Question 11	6.002	6.003	5.991	0.013
	(0.362)	(0.342)	(0.501)	(0.112)
Question 12	5.836	5.829	5.889	-0.060
	(0.390)	(0.387)	(0.427)	(0.121)
Principal Component	0.062	0.070	0.005	0.065
	(0.697)	(0.698)	(0.721)	(0.216)
Grade	0.058	0.038	0.194	-0.156
	(0.632)	(0.541)	(1.100)	(0.196)
Average Question 1, 3 and 4	0.018	0.034	-0.098	0.133
	(0.726)	(0.729)	(0.732)	(0.225)
Observations	95	83	12	

Notes: Table shows averages for each variable and standard errors in parenthesis. The unit of observation is a professor. fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Number of courses is the total number of courses taught in years 2012-2018. Questions and outcomes are computed as averages of professor-course-semester observations. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table 3: Effect of the fWHR on Students' Evaluations

Panel A: All professors									
	Principal Component			Grade			Average Q1-Q3-Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	0.996**	0.955**	1.076**	0.759**	0.743**	0.849**	1.145**	1.098**	1.204**
	(0.438)	(0.458)	(0.490)	(0.302)	(0.320)	(0.342)	(0.444)	(0.457)	(0.482)
Number of Students			-0.004**			-0.002			-0.003*
			(0.002)			(0.001)			(0.002)
Share of Female Students			0.087			-0.088			0.035
			(0.352)			(0.250)			(0.355)
PhD			0.102			0.106			0.035
			(0.137)			(0.094)			(0.136)
Average PSU scores			-0.000			0.000			-0.000
			(0.003)			(0.003)			(0.004)
Female			-0.127			-0.135			-0.263
			(0.266)			(0.207)			(0.303)
Observations	1439	1439	1439	1439	1439	1439	1439	1439	1439
Panel B: Male professors									
	Principal Component			Grade			Average Q1-Q3-Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	1.158***	1.124**	1.198**	0.850***	0.844***	0.922***	1.302***	1.265***	1.281***
	(0.434)	(0.453)	(0.501)	(0.298)	(0.317)	(0.346)	(0.442)	(0.453)	(0.483)
Number of Students			-0.004**			-0.002			-0.003*
			(0.002)			(0.001)			(0.002)
Share of Female Students			-0.069			-0.248			-0.124
			(0.373)			(0.264)			(0.368)
PhD			0.042			0.080			-0.037
			(0.163)			(0.111)			(0.159)
Average PSU scores			-0.001			0.000			-0.001
			(0.004)			(0.003)			(0.004)
Observations	1284	1284	1284	1284	1284	1284	1284	1284	1284
Panel C: Female professors									
	Principal Component			Grade			Average Q1-Q3-Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	-3.893***	-3.922***	-4.235***	-2.604***	-2.616***	-2.750***	-3.551***	-3.542***	-3.765***
	(0.530)	(0.488)	(0.665)	(0.363)	(0.313)	(0.439)	(0.722)	(0.654)	(0.682)
Number of Students			-0.020**			-0.014**			-0.019**
			(0.009)			(0.005)			(0.008)
Share of Female Students			0.230			0.386			0.356
			(0.976)			(0.558)			(1.008)
PhD			0.429***			0.215**			0.365**
			(0.097)			(0.071)			(0.149)
Average PSU scores			0.002			0.001			0.004
			(0.006)			(0.004)			(0.007)
Observations	155	155	155	155	155	155	155	155	155
Course FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year and semester FE	no	yes	yes	no	yes	yes	no	yes	yes

Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for all professors (Panel A), male professors (Panel B) and female professors (Panel C). fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table 4: Effect of the fWHR on Students' Evaluations, controlling for head rotation

Panel A: Male professors	Principal Component		Grade		Average Q1-Q3-Q4	
	(1)	(2)	(3)	(4)	(5)	(6)
Corrected fWHR	1.039	1.041	0.833*	0.820*	1.191*	1.179*
	(0.682)	(0.688)	(0.484)	(0.487)	(0.669)	(0.674)
Number of Students	-0.004**	-0.003*	-0.002	-0.001	-0.003	-0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Share of Female Students	-0.038	0.006	-0.226	-0.195	-0.095	-0.042
	(0.370)	(0.374)	(0.263)	(0.267)	(0.367)	(0.371)
PhD	0.014	0.013	0.061	0.059	-0.061	-0.064
	(0.172)	(0.171)	(0.118)	(0.118)	(0.165)	(0.164)
Average PSU scores	-0.001	-0.001	0.000	0.000	-0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Observations	1284	1275	1284	1275	1284	1275
Panel B: Female professors	Principal Component		Grade		Average Q1-Q3-Q4	
	(1)	(2)	(3)	(4)	(5)	(6)
Corrected fWHR	-4.339***	-4.339***	-2.802***	-2.802***	-4.033***	-4.033***
	(0.668)	(0.668)	(0.419)	(0.419)	(0.673)	(0.673)
Number of Students	-0.020*	-0.020*	-0.013**	-0.013**	-0.018*	-0.018*
	(0.009)	(0.009)	(0.005)	(0.005)	(0.008)	(0.008)
Share of Female Students	0.370	0.370	0.479	0.479	0.455	0.455
	(0.957)	(0.957)	(0.540)	(0.540)	(0.979)	(0.979)
PhD	0.334	0.334	0.154	0.154	0.278	0.278
	(0.209)	(0.209)	(0.146)	(0.146)	(0.208)	(0.208)
Average PSU scores	0.003	0.003	0.002	0.002	0.005	0.005
	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)
Observations	155	155	155	155	155	155
Course FE	yes	yes	yes	yes	yes	yes
Year and semester FE	yes	yes	yes	yes	yes	yes
$\lvert\text{Angle}\rvert \leq 15°$	no	yes	no	yes	no	yes

Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for male professors (Panel A) and female professors (Panel B). Corrected fWHR is the facial Width to Height Ratio, corrected for the head's angle of rotation in the picture. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. Columns 2, 4 and 6 restrict the sample to professors with an angle of rotation between $[-15°, 15°]$. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table 5: Controlling for Beauty score

	All professors			Male professors			Female professors		
	Principal Component	Grade	Average Q1-Q3-Q4	Principal Component	Grade	Average Q1-Q3-Q4	Principal Component	Grade	Average Q1-Q3-Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	1.189**	0.912***	1.254***	1.356***	0.999***	1.373***	-4.617***	-3.034***	-4.259***
	(0.477)	(0.330)	(0.473)	(0.476)	(0.323)	(0.465)	(0.620)	(0.378)	(0.627)
Number of Students	-0.004**	-0.002	-0.003*	-0.004**	-0.002	-0.003*	-0.021**	-0.014**	-0.020**
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.009)	(0.005)	(0.008)
Share of Female Students	0.071	-0.097	0.028	-0.106	-0.266	-0.145	0.130	0.312	0.227
	(0.359)	(0.255)	(0.363)	(0.385)	(0.273)	(0.384)	(1.004)	(0.583)	(1.030)
PhD	0.061	0.084	0.016	-0.039	0.041	-0.084	0.281***	0.105*	0.174
	(0.151)	(0.106)	(0.152)	(0.199)	(0.140)	(0.200)	(0.060)	(0.053)	(0.097)
Average PSU scores	-0.000	0.000	-0.000	-0.001	0.000	-0.001	0.001	0.000	0.003
	(0.003)	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.007)	(0.004)	(0.007)
Female	-0.101	-0.121	-0.252						
	(0.257)	(0.201)	(0.297)						
Beauty score	0.006	0.003	0.003	0.008	0.004	0.005	-0.021***	-0.016***	-0.028**
	(0.008)	(0.006)	(0.008)	(0.010)	(0.007)	(0.011)	(0.007)	(0.005)	(0.009)
Observations	1439	1439	1439	1284	1284	1284	155	155	155
Course fixed effect	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year and semester FE	yes	yes	yes	yes	yes	yes	yes	yes	yes

Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for all professors (columns 1-3), male professors (columns 4-6) and female professors (columns 7-9). fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. The Beauty Score is computed as the average of the beauty scores that would have been given by a male and female rater, weighted by the share of female students in the course. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table 6: Controlling for Students Grades as a proxy of Professor Quality

Panel A: Male professors									
	Principal Component			Grade			Average Q1-Q3-Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	1.198**	1.072**	0.805	0.922***	0.836**	0.698	1.281***	1.164**	0.761
	(0.501)	(0.474)	(0.774)	(0.346)	(0.329)	(0.570)	(0.483)	(0.481)	(0.788)
Number of Students	-0.004**	-0.005***	0.000	-0.002	-0.003**	-0.001	-0.003*	-0.004**	0.001
	(0.002)	(0.002)	(0.006)	(0.001)	(0.001)	(0.004)	(0.002)	(0.002)	(0.007)
Share of Female Students	-0.069	0.030	0.777	-0.248	-0.184	0.300	-0.124	-0.027	0.667
	(0.373)	(0.328)	(0.671)	(0.264)	(0.241)	(0.452)	(0.368)	(0.324)	(0.706)
PhD	0.042	-0.072	0.053	0.080	0.002	0.065	-0.037	-0.141	-0.023
	(0.163)	(0.170)	(0.294)	(0.111)	(0.116)	(0.199)	(0.159)	(0.168)	(0.287)
Average PSU scores	-0.001	-0.001	-0.001	0.000	-0.000	-0.000	-0.001	-0.001	-0.001
	(0.004)	(0.003)	(0.007)	(0.003)	(0.002)	(0.005)	(0.004)	(0.003)	(0.008)
Average Student Grade		0.645***	0.665*		0.430***	0.442		0.613***	0.712*
		(0.150)	(0.389)		(0.100)	(0.266)		(0.149)	(0.412)
Share of Failing Students		-0.333	-0.398		-0.351	-0.154		-0.119	0.109
		(0.394)	(1.056)		(0.278)	(0.746)		(0.456)	(1.163)
Observations	1284	1284	1284	1284	1284	1284	1284	1284	1284
Panel B: Female professors									
	Principal Component			Grade			Average Q1-Q3-Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	-4.235***	-4.255***	-3.489***	-2.750***	-2.811***	-2.216***	-3.765***	-3.788***	-3.021***
	(0.665)	(0.591)	(0.684)	(0.439)	(0.404)	(0.516)	(0.682)	(0.642)	(0.774)
Number of Students	-0.020**	-0.021**	-0.028	-0.014**	-0.015***	-0.020	-0.019**	-0.020**	-0.023
	(0.009)	(0.008)	(0.017)	(0.005)	(0.004)	(0.012)	(0.008)	(0.007)	(0.018)
Share of Female Students	0.230	0.026	1.779	0.386	0.280	1.406	0.356	0.222	1.828
	(0.976)	(0.988)	(2.153)	(0.558)	(0.561)	(1.174)	(1.008)	(1.019)	(2.209)
PhD	0.429***	0.428***	0.369	0.215**	0.225**	0.185	0.365**	0.366**	0.360
	(0.097)	(0.101)	(0.244)	(0.071)	(0.076)	(0.175)	(0.149)	(0.155)	(0.327)
Average PSU scores	0.002	0.001	0.024	0.001	0.000	0.014	0.004	0.003	0.022
	(0.006)	(0.007)	(0.018)	(0.004)	(0.005)	(0.012)	(0.007)	(0.007)	(0.018)
Average Student Grade		0.341	0.052		0.115	0.018		0.213	-0.181
		(0.221)	(0.808)		(0.144)	(0.506)		(0.220)	(0.805)
Share of Failing Students		0.318	0.907		-0.357	0.149		0.116	0.537
		(0.648)	(3.340)		(0.570)	(2.446)		(0.667)	(3.220)
Observations	155	155	155	155	155	155	155	155	155
Course FE	yes	yes	no	yes	yes	no	yes	yes	no
Year and semester FE	yes	yes	no	yes	yes	no	yes	yes	no
Course×Year×semester FE	no	no	yes	no	no	yes	no	no	yes

Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for male professors (Panel A) and female professors (Panel B). fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. Average Student Grade ranges from 1 to 7, with 4 as the passing grade. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Appendix

Table A-1: Professor-course averages

Variable	Panel	Mean	Sd	Min	Max	Observations
Principal Component	Overall	0	1	-7.309	2.107	N = 1439
	Between		.697	-1.878	1.643	n = 95
	Within		.769	-6.514	2.847	T = 15.147
Grade	Overall	0	1	-5.080	13.175	N = 1439
	Between		.632	-1.454	3.317	n = 95
	Within		.867	-4.454	12.522	T = 15.147
Average Questions 1, 3 and 4	Overall	0	1	-7.323	2.003	N = 1439
	Between		.726	-2.108	1.577	n = 95
	Within		.765	-6.443	2.722	T = 15.147
Female teacher	Overall	.108	.310	0	1	N = 1439
	Between		.334	0	1	n = 95
	Within		0	.108	.108	T = 15.147
fWHR	Overall	1.759	.145	1.444	2.154	N = 1439
	Between		.132	1.444	2.154	n = 95
	Within		0	1.759	1.759	T = 15.147
Average Student Grade	Overall	4.830	.544	1	6.75	N = 1439
	Between		.326	3.811	5.597	n = 95
	Within		.452	.904	6.596	T = 15.147
Number of Students	Overall	43.89	19.33	1	165	N = 1439
	Between		10.93	15	73	n = 95
	Within		17.33	-17.66	155.78	T = 15.147
Share of Failing Students	Overall	.135	.114	.007	1	N = 1439
	Between		.062	.036	.385	n = 95
	Within		.095	-.092	1.045	T = 15.147
Share of Female Students	Overall	.381	.130	.0294	1	N = 1439
	Between		.082	.220	.721	n = 95
	Within		.112	.045	.892	T = 15.147
PhD	Overall	.723	.448	0	1	N = 1439
	Between		.453	0	1	n = 95
	Within		0	.723	.723	T = 15.147
Average PSU scores	Overall	559.88	299.06	0	787.8	N = 1439
	Between		160.71	0	751.11	n = 95
	Within		271.62	-122.73	1116.36	T = 15.147
PSU score missing	Overall	.221	.415	0	1	N = 1439
	Between		.226	0	1	n = 95
	Within		.374	-.579	1.181	T = 15.147
Beauty score	Overall	51.752	11.441	24.016	79.280	N = 1439
	Between		11.574	24.318	79.201	n = 95
	Within		.610	48.569	55.247	T = 15.147

Notes: Table shows averages, standard deviations, minimum and maximum for each variable. The unit of observation is a professor-course. fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Number of courses is the total number of courses taught in years 2012-2018.

Table A-2: Professors vs. Lecturers: Differences in characteristics

	Not in sample	Sample	Difference
Female teacher	0.290	0.126	0.163***
	(0.454)	(0.334)	(0.048)
	[639]	[95]	
Number of classes	4.745	15.147	-10.402***
	(5.828)	(10.949)	(0.735)
	[652]	[95]	
Principal Component	0.037	0.062	-0.025
	(0.978)	(0.697)	(0.104)
	[652]	[95]	
Grade	-0.003	0.058	-0.061
	(0.727)	(0.632)	(0.079)
	[652]	[95]	
Average Q1-Q3-Q4	-0.064	0.018	-0.082
	(1.042)	(0.726)	(0.111)
	[652]	[95]	

Notes: Table shows averages for each variable and standard deviation in parenthesis. Square brackets indicate the number of observations. Each observation is an instructor. PhD is a dummy for professors with a PhD degree. Number of courses is the total number of courses taught in years 2012-2018. Outcomes are computed as averages of professor-course-semester observations. Professors in our sample are faculties who belong to one of the School's Departments. Professors not in sample are lecturers. We were unable to identify the gender of 13 lecturers. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table A-3: Professors vs. Lecturers: Differences in classroom characteristics

	Not in sample	Sample	Difference
Average classroom grade	4.998	4.830	0.168***
	(0.654)	(0.544)	(0.020)
	[2,890]	[1,439]	
Classroom size	36.232	43.888	-7.656***
	(20.652)	(19.327)	(0.646)
	[3,094]	[1,439]	
% failing students	0.150	0.135	0.015***
	(0.129)	(0.114)	(0.004)
	[3,094]	[1,439]	
% female students	0.400	0.381	0.020***
	(0.136)	(0.130)	(0.004)
	[3,094]	[1,439]	
Average PSU scores	643.713	559.883	83.830***
	(214.308)	(299.063)	(7.799)
	[3,094]	[1,439]	
Year 2012	0.146	0.108	0.038***
	(0.353)	(0.310)	(0.011)
	[3,094]	[1,439]	
Year 2013	0.153	0.138	0.015
	(0.360)	(0.345)	(0.011)
	[3,094]	[1,439]	
Year 2014	0.157	0.140	0.017
	(0.364)	(0.347)	(0.011)
	[3,094]	[1,439]	
Year 2015	0.154	0.158	-0.004
	(0.361)	(0.365)	(0.012)
	[3,094]	[1,439]	
Year 2016	0.157	0.154	0.003
	(0.364)	(0.361)	(0.012)
	[3,094]	[1,439]	
Year 2017	0.149	0.195	-0.046***
	(0.356)	(0.396)	(0.012)
	[3,094]	[1,439]	
Year 2018	0.085	0.108	-0.024**
	(0.278)	(0.311)	(0.009)
	[3,094]	[1,439]	
Fall semester	0.499	0.505	-0.006
	(0.500)	(0.500)	(0.016)
	[3,094]	[1,439]	
Spring semester	0.389	0.352	0.037**
	(0.488)	(0.478)	(0.015)
	[3,094]	[1,439]	
Summer semester	0.045	0.031	0.015**
	(0.208)	(0.172)	(0.006)
	[3,094]	[1,439]	

Notes: Table shows averages for each variable and standard deviation in parenthesis. Square brackets indicate the number of observations. Each observation is an instructor–course-semester. Average Student Grade ranges from 1 to 7, with 4 as the passing grade. Average PSU scores is the average admission math score in the class. Professors in our sample are faculties who belong to one of the School's Departments. Professors not in sample are lecturers. There were 204 courses for which we do not observe student grades. These are tutorial classes taught during summer term. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table A-4: Effect of the fWHR on Students' Evaluation: Individual Questions

Panel A: Male professors	(Q1)	(Q2)	(Q3)	(Q4)	(Q5)	(Q6)	(Q7)	(Q8)	(Q9)	(Q10)	(Q11)	(Q12)
fWHR	0.661**	0.686**	0.902***	0.640**	0.589	0.496*	0.664*	1.119***	0.550	0.727**	0.487*	0.709**
	(0.260)	(0.265)	(0.321)	(0.293)	(0.357)	(0.290)	(0.362)	(0.321)	(0.421)	(0.306)	(0.290)	(0.322)
Number of Students	-0.000	-0.001	-0.002	-0.003**	-0.003**	-0.002**	-0.003**	-0.004***	-0.000	-0.003**	-0.003***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Share of Female Students	-0.005	0.062	-0.112	-0.096	-0.176	-0.078	-0.244	-0.165	0.248	0.126	-0.052	0.044
	(0.211)	(0.212)	(0.230)	(0.214)	(0.236)	(0.200)	(0.242)	(0.240)	(0.233)	(0.231)	(0.241)	(0.221)
PhD	-0.038	-0.058	-0.104	0.078	-0.009	0.111	0.035	0.067	-0.065	-0.015	0.149	0.098
	(0.073)	(0.082)	(0.112)	(0.101)	(0.118)	(0.100)	(0.121)	(0.093)	(0.118)	(0.088)	(0.107)	(0.111)
Average PSU scores	-0.000	-0.000	-0.000	-0.001	-0.000	-0.001	0.000	-0.000	0.001	-0.001	-0.002	-0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1284	1284	1284	1284	1284	1284	1284	1284	1284	1284	1284	1284
Mean of Dep Variable	6.051	5.888	5.666	5.824	5.666	5.889	5.581	5.474	5.797	5.766	5.937	5.763
Panel B: Female professors	(Q1)	(Q2)	(Q3)	(Q4)	(Q5)	(Q6)	(Q7)	(Q8)	(Q9)	(Q10)	(Q11)	(Q12)
fWHR	-1.411***	-1.660***	-2.388***	-2.676***	-3.069***	-2.621***	-3.393***	-2.673***	-1.376***	-2.143***	-2.679***	-2.692***
	(0.314)	(0.195)	(0.454)	(0.476)	(0.506)	(0.369)	(0.452)	(0.770)	(0.391)	(0.564)	(0.395)	(0.515)
Number of Students	-0.010**	-0.010*	-0.012*	-0.010*	-0.012*	-0.010*	-0.015**	-0.015*	-0.009	-0.014	-0.011**	-0.011**
	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.008)	(0.004)	(0.004)
Share of Female Students	0.403	0.282	0.254	-0.045	0.120	0.037	0.377	-0.001	-0.170	0.127	0.161	-0.055
	(0.484)	(0.471)	(0.677)	(0.608)	(0.725)	(0.560)	(0.641)	(0.580)	(0.602)	(0.603)	(0.571)	(0.623)
PhD	0.194*	0.251***	0.223*	0.211**	0.265**	0.173***	0.178*	0.116	0.681***	0.316***	0.097*	0.308***
	(0.092)	(0.060)	(0.111)	(0.071)	(0.093)	(0.053)	(0.092)	(0.122)	(0.083)	(0.058)	(0.053)	(0.074)
Average PSU scores	0.003	0.002	0.004	-0.000	0.003	0.001	-0.001	-0.001	-0.001	-0.000	0.000	0.000
	(0.004)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.006)	(0.005)	(0.004)	(0.003)
Observations	155	155	155	155	155	155	155	155	155	155	155	155
Mean of Dep Variable	6.016	5.993	5.794	5.860	5.716	5.933	5.611	5.500	5.927	5.848	6.025	5.905

Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for all professors (Panel A), male professors (Panel B) and female professors (Panel C). fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table A-5: Quantile Regressions

	All professors			Male professors			Female professors		
	Principal Component	Grade	Average Q1-Q3-Q4	Principal Component	Grade	Average Q1-Q3-Q4	Principal Component	Grade	Average Q1-Q3-Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
fWHR	1.024*	0.799***	1.112	1.188***	0.910***	1.274***	-4.017**	-2.624***	-3.610***
	(0.530)	(0.168)	(2.053)	(0.292)	(0.168)	(0.254)	(1.587)	(0.713)	(1.176)
Number of Students	-0.004	-0.002	-0.003	-0.004*	-0.002	-0.003	-0.020**	-0.013***	-0.019***
	(0.004)	(0.001)	(0.016)	(0.002)	(0.001)	(0.002)	(0.008)	(0.003)	(0.005)
Share of Female Students	0.126	-0.061	0.103	-0.028	-0.220	-0.064	0.269	0.396	0.377
	(0.602)	(0.211)	(2.269)	(0.354)	(0.229)	(0.295)	(1.127)	(0.456)	(0.761)
PhD	0.104	0.106**	0.034	0.053	0.089	-0.021	0.413	0.218**	0.356**
	(0.161)	(0.050)	(0.609)	(0.096)	(0.055)	(0.080)	(0.260)	(0.110)	(0.170)
Average PSU scores	-0.000	0.000	-0.000	-0.001	-0.000	-0.001	0.002	0.001	0.004
	(0.004)	(0.001)	(0.017)	(0.002)	(0.001)	(0.002)	(0.009)	(0.004)	(0.006)
Observations	1439	1439	1439	1284	1284	1284	155	155	155
Course fixed effect	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year and semester FE	yes	yes	yes	yes	yes	yes	yes	yes	yes

Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for all professors (columns 1-3), male professors (columns 4-6) and female professors (columns 7-9). fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. The Beauty Score is computed as the average of the beauty scores that would have been given by a male and female rater, weighted by the share of female students in the course. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Table A-6: Effect of fWHR on Students' Grades

	All professors	Male professors	Female professors
	(1)	(2)	(3)
fWHR	0.167	0.187	0.170
	(0.172)	(0.182)	(0.326)
Number of Students	0.002**	0.001	0.004
	(0.001)	(0.001)	(0.003)
Share of Female Students	-0.011	-0.162	0.674*
	(0.128)	(0.130)	(0.337)
PhD	0.157***	0.166***	-0.018
	(0.050)	(0.058)	(0.113)
Average PSU scores	0.001	0.000	0.003
	(0.001)	(0.001)	(0.004)
Female	0.031		
	(0.072)		
Observations	1439	1284	155
Mean of Dep Variable	4.83	4.82	4.92
Course fixed effect	yes	yes	yes
Year and semester FE	yes	yes	yes

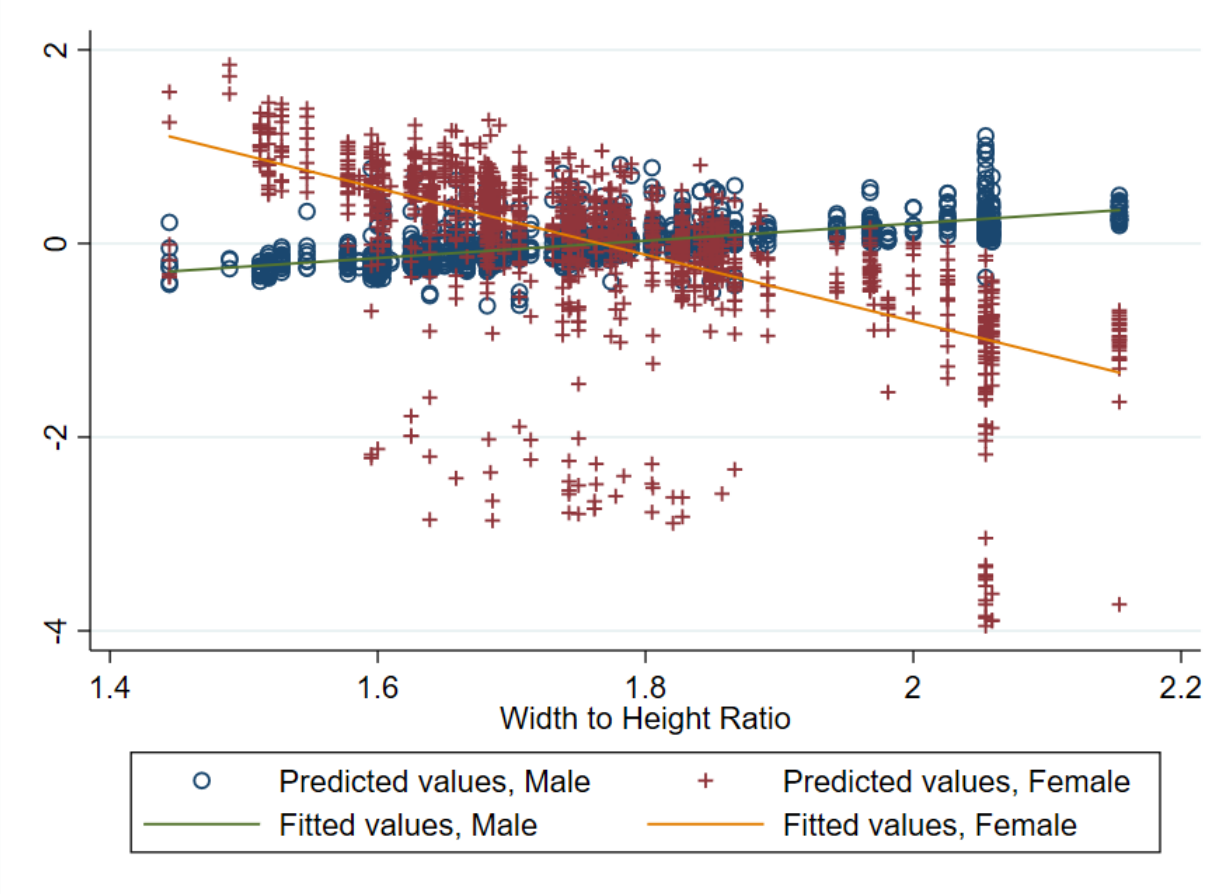
Notes: Robust standard errors in parentheses, adjusted for clustering on professors. Table shows estimates for equation 1 for all professors (column 1), male professors (column 2) and female professors (column 3), replacing the dependent variable with the average student grade. fWHR is the facial Width to Height Ratio. PhD is a dummy for professors with a PhD degree. Average PSU scores is the average admission math score in the class. The unit of observation is at the professor-course-semester level. ***, ** and * indicate statistical significance at the 99%, 95% and 90%, respectively.

Additional Figures

Figure A-1: Regressions for male and female teachers (principal component)

Notes: The figure shows predicted values of equation 1 for male and female professors, along with fitted values.

Figure A-2: Regressions for male and female teachers (grade)

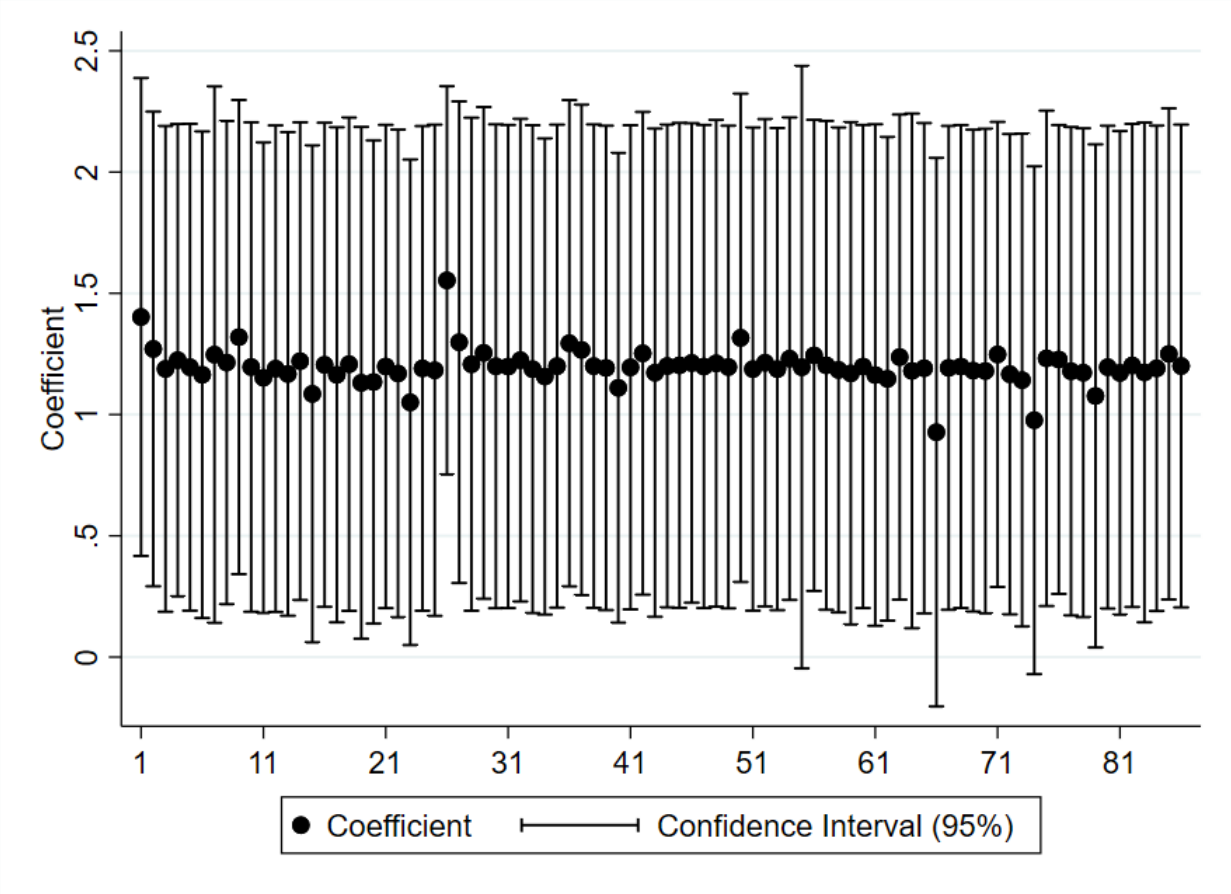


Notes: The figure shows predicted values of equation 1 for male and female professors, along with fitted values.

Figure A-3: Regressions for male and female teachers (average Q1-Q3-Q4)

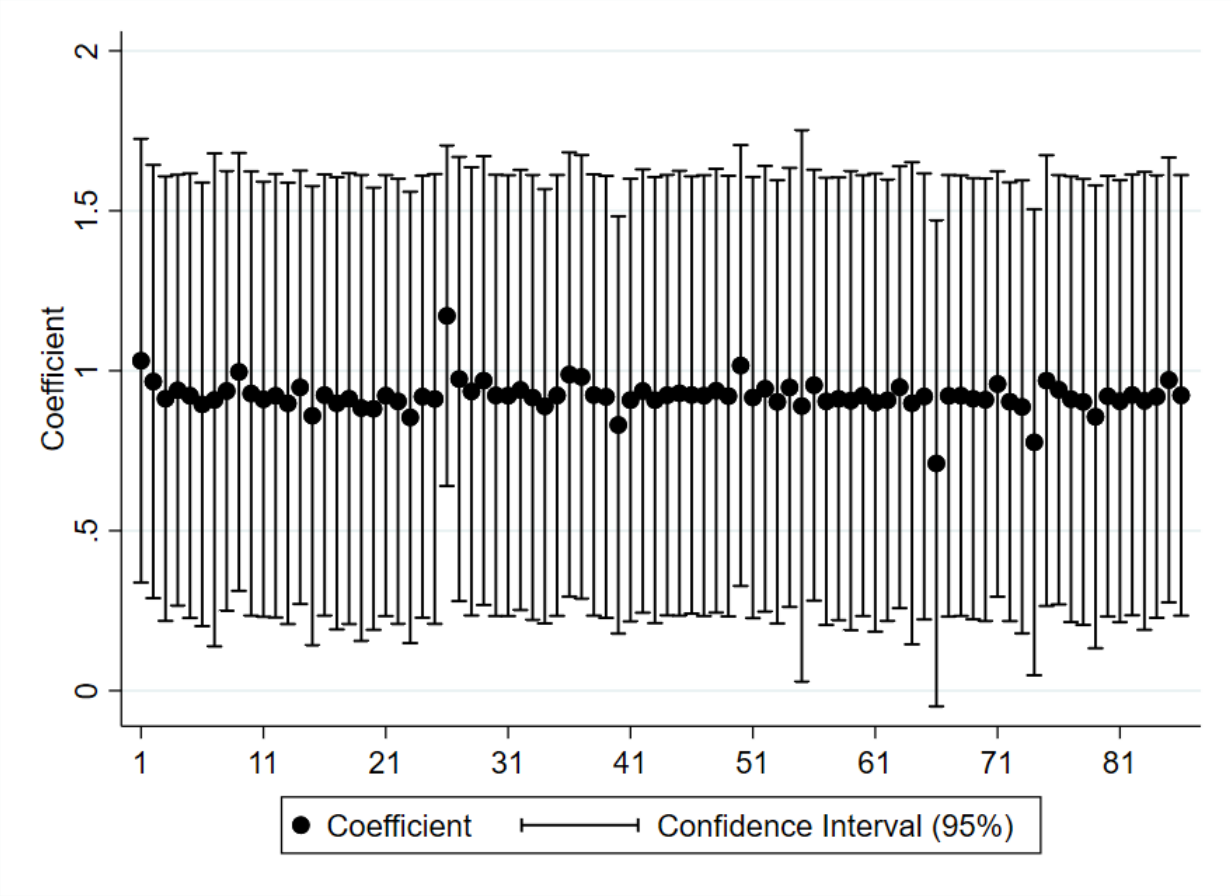
Notes: The figure shows predicted values of equation 1 for male and female professors, along with fitted values.

Figure A-4: Coefficients excluding one teacher at the time: Male teachers (principal component)



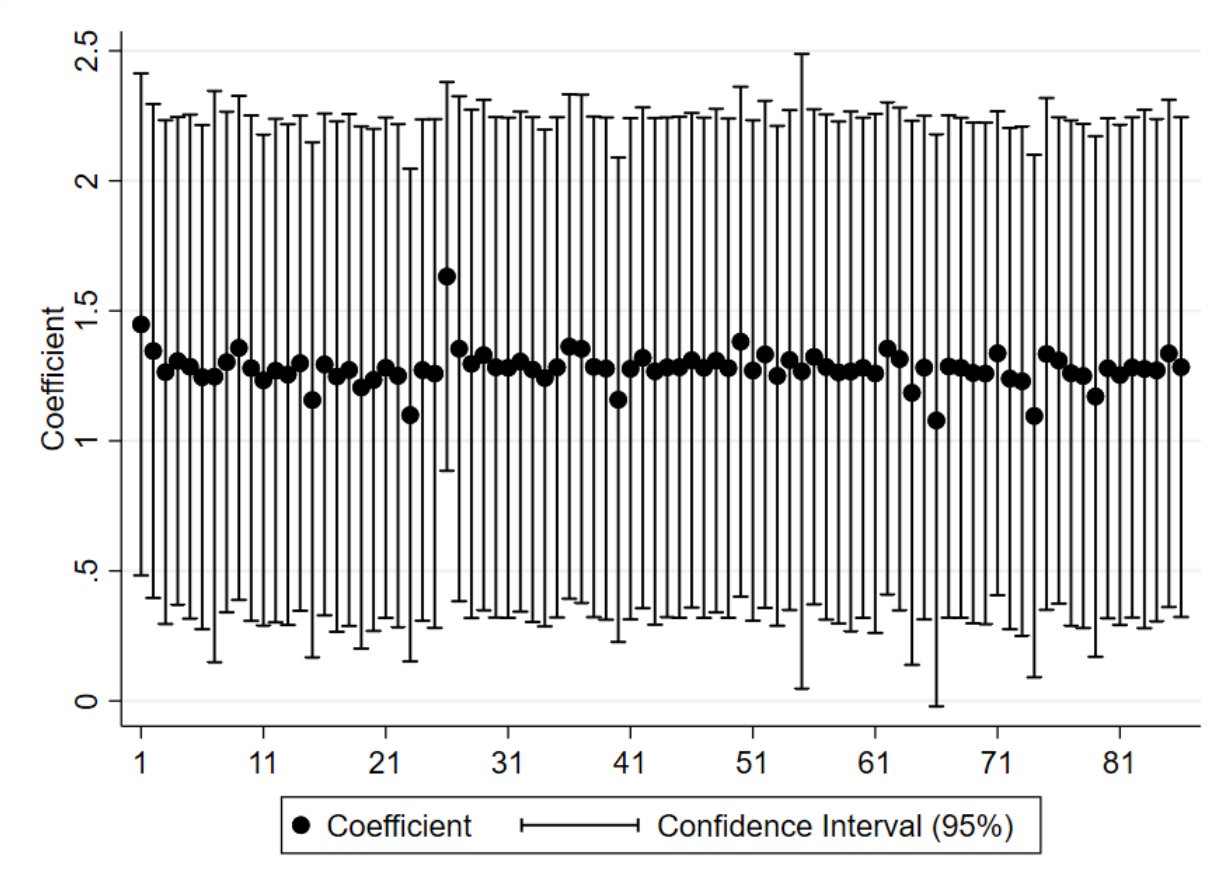
Notes: The figure shows the coefficient of fWHR in equation 1, along with its 95% confidence interval, when one professor is excluded from the sample. The horizontal axis shows the id number of the professor excluded.

Figure A-5: Coefficients excluding one teacher at the time: Male teachers (grade)



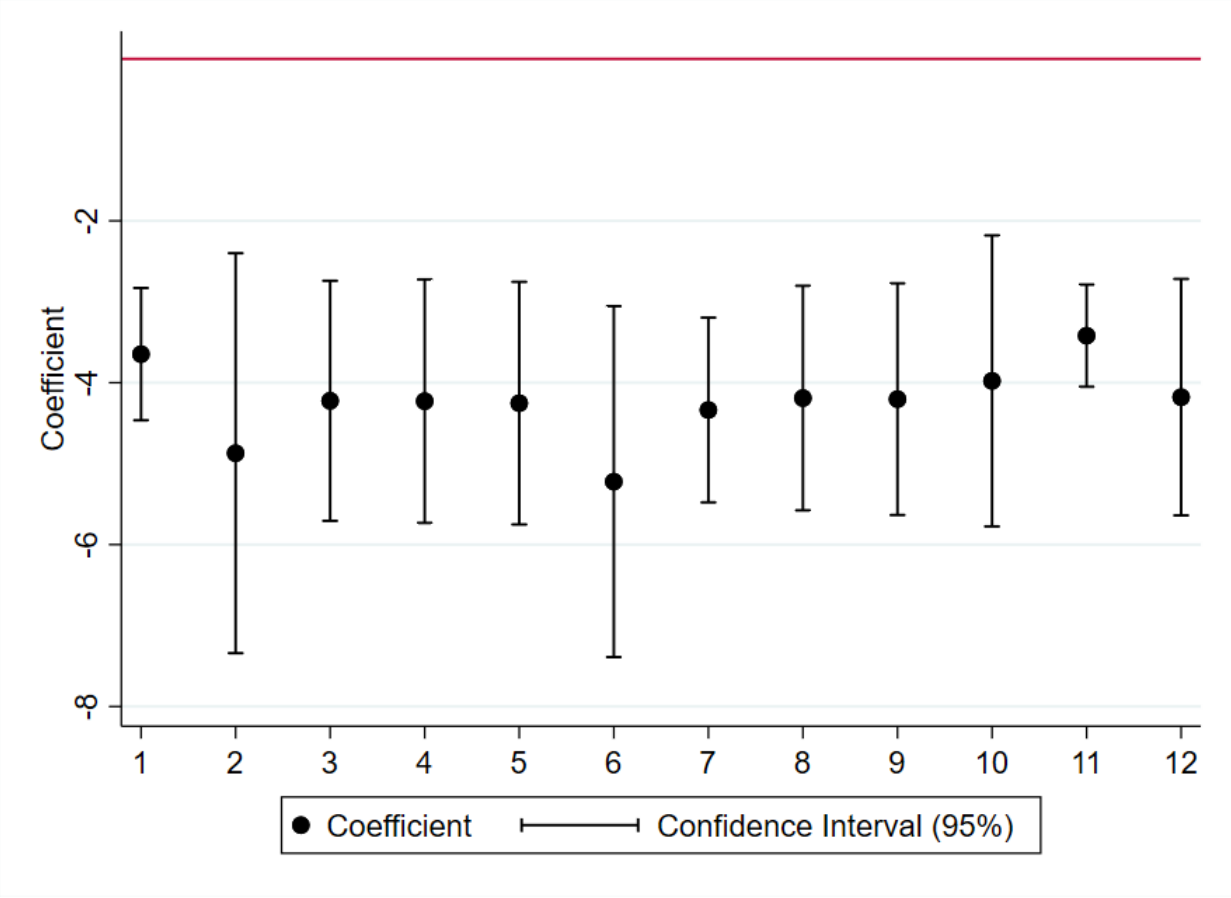
Notes: The figure shows the coefficient of fWHR in equation 1, along with its 95% confidence interval, when one professor is excluded from the sample. The horizontal axis shows the id number of the professor excluded.

Figure A-6: Coefficients excluding one teacher at the time: Male teachers (average Q1-Q3-Q4)



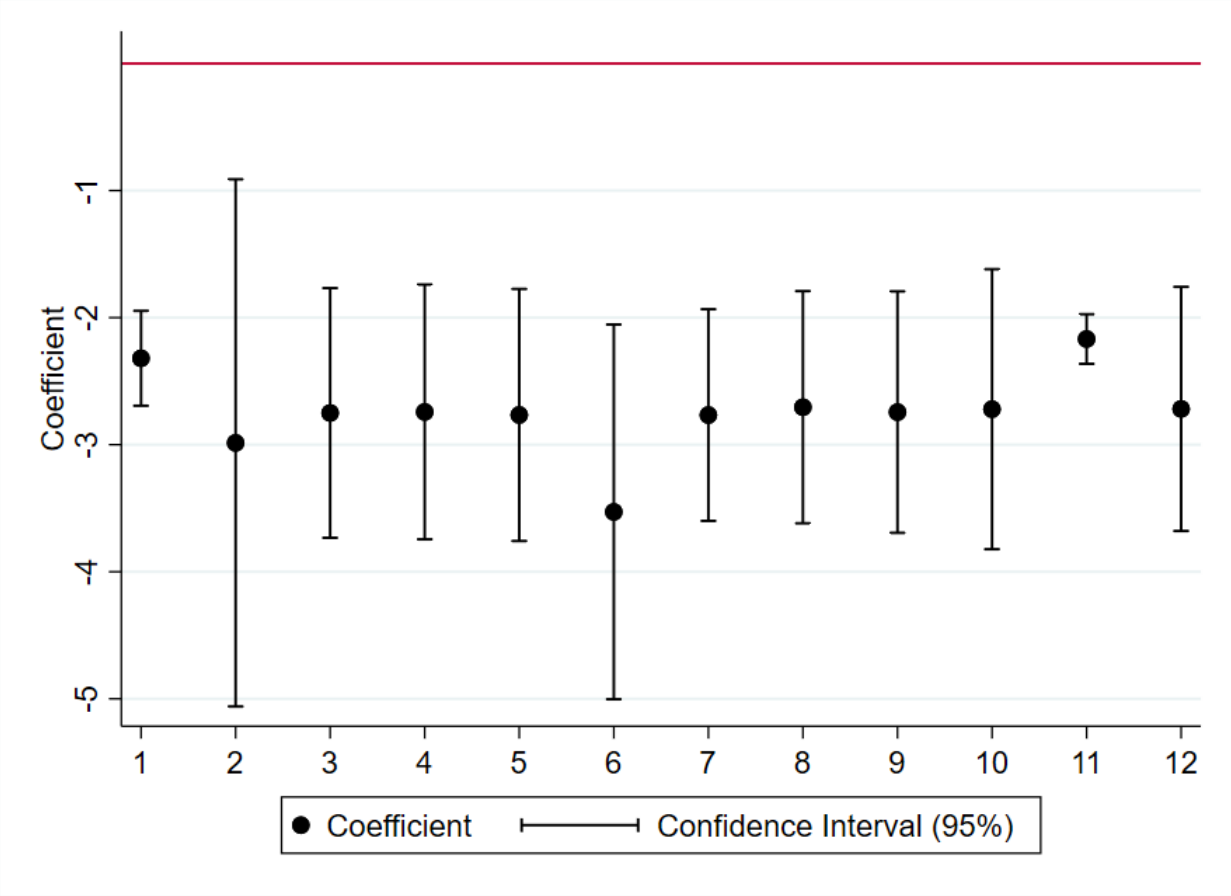
Notes: The figure shows the coefficient of fWHR in equation 1, along with its 95% confidence interval, when one professor is excluded from the sample. The horizontal axis shows the id number of the professor excluded.

Figure A-7: Coefficients excluding one teacher at the time: Female teachers (principal component)



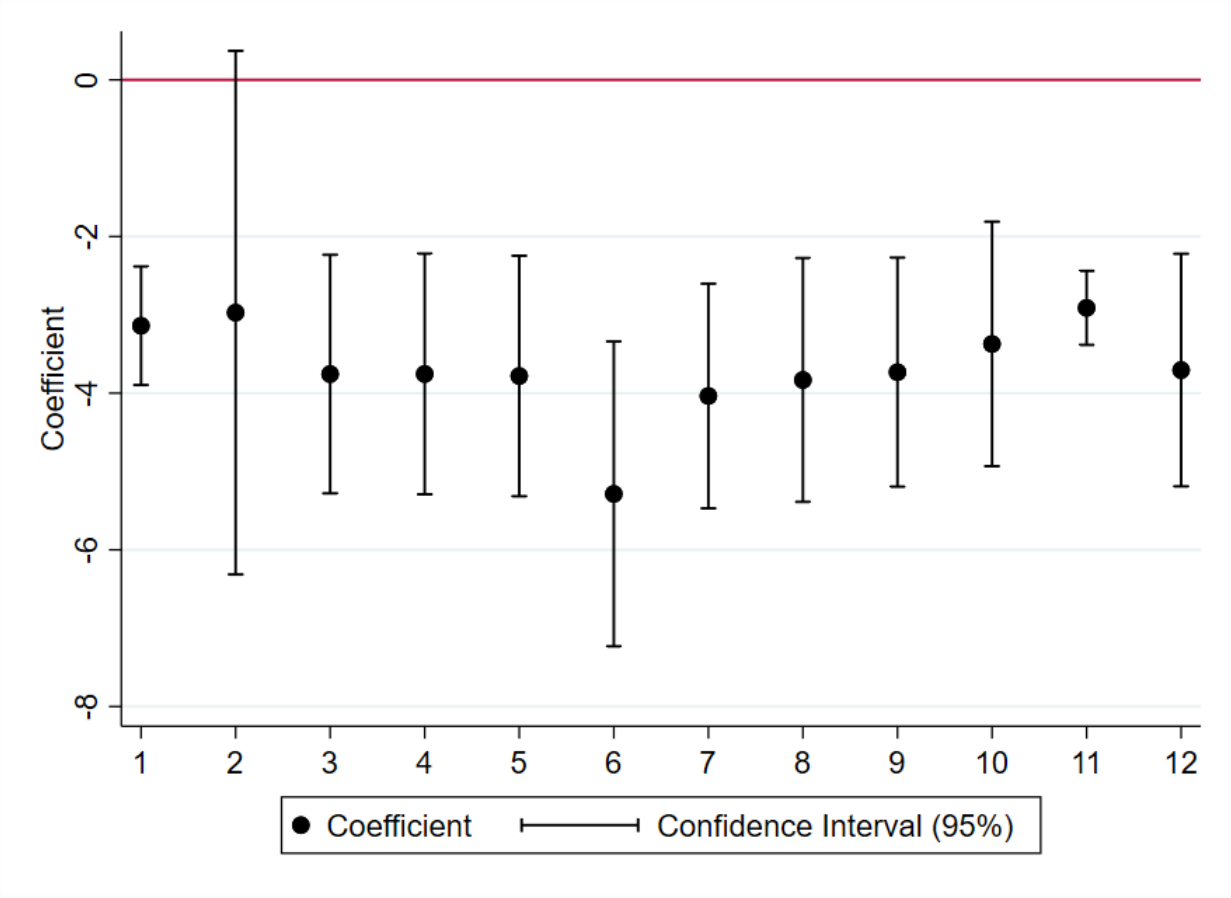
Notes: The figure shows the coefficient of fWHR in equation 1, along with its 95% confidence interval, when one professor is excluded from the sample. The horizontal axis shows the id number of the professor excluded.

Figure A-8: Coefficients excluding one teacher at the time: Female teachers (grade)



Notes: The figure shows the coefficient of fWHR in equation 1, along with its 95% confidence interval, when one professor is excluded from the sample. The horizontal axis shows the id number of the professor excluded.

Figure A-9: Coefficients excluding one teacher at the time: Female teachers (average Q1-Q3-Q4)



Notes: The figure shows the coefficient of fWHR in equation 1, along with its 95% confidence interval, when one professor is excluded from the sample. The horizontal axis shows the id number of the professor excluded.