Maurer, Stephan; Schwerdt, Guido; Wiederhold, Simon

Working Paper

Do Role Models Matter in Large Classes? New Evidence on Gender Match Effects in Higher Education

IZA Discussion Papers, No. 15860

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Maurer, Stephan; Schwerdt, Guido; Wiederhold, Simon (2023) : Do Role Models Matter in Large Classes? New Evidence on Gender Match Effects in Higher Education, IZA Discussion Papers, No. 15860, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/272487

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15860

Do Role Models Matter in Large Classes? New Evidence on Gender Match Effects in Higher Education

Stephan Maurer
Guido Schwerdt
Simon Wiederhold

JANUARY 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15860

Do Role Models Matter in Large Classes? New Evidence on Gender Match Effects in Higher Education

Stephan Maurer
University of Konstanz and CEP

Guido Schwerdt
University of Konstanz, CESifo, IZA and ROA

Simon Wiederhold
KU Eichstätt-Ingolstadt, ifo Institute, CESifo and ROA

JANUARY 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Do Role Models Matter in Large Classes? New Evidence on Gender Match Effects in Higher Education*

We study whether female students benefit from being taught by female professors, and whether such gender match effects differ by class size. We use administrative records of a German public university, covering all programs and courses between 2006 and 2018. We find that gender match effects on student performance are sizable in smaller classes, but do not exist in larger classes. This difference suggests that direct and frequent interactions between students and professors are important for the emergence of gender match effects. Instead, the mere fact that one's professor is female is not sufficient to increase performance of female students.

JEL Classification: I21, I23, I24, J16

Keywords: gender gap, role models, tertiary education, professors

Corresponding author:
Guido Schwerdt
Department of Economics
University of Konstanz
78457 Konstanz
Germany
E-mail: guido.schwerdt@uni.kn

* We thank Jörn-Steffen Pischke, Ulf Zölitz, seminar participants at Edinburgh and Hohenheim, and participants of the conference of the VfS Education Committee for helpful comments and discussions. We acknowledge that this research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2035/1 - 390681379. Christian König provided excellent research assistance.

1. Introduction

It is widely believed that female students benefit from being taught by female professors (see, for example, Bingham (2012), Warrell (2020)), but most of the causal evidence on gender match effects in higher education is limited to settings with small classes. If such effects primarily arise because female professors serve as role models for female students, they can also be expected to occur in large classes. But if such effects additionally require more direct and frequent interactions between students and professors, results on gender match effects in smaller classes may not carry over to larger class settings, which are typical for public universities worldwide.

In this paper, we study female gender match effects on student performance in a public university in Germany. Our analysis is based on administrative records for the universe of programs and courses in the period 2006 to 2018, providing considerable variation in class sizes. These data allow us to estimate female gender match effects by class size conditional on a rich set of student characteristics and program, course, semester, and lecturer fixed effects. Since grading in our setting is mostly anonymous, we do not have to be concerned about gender bias in grading. Moreover, we can account for potential ability-based sorting of students to professors by using information on students' high school GPA, which is a powerful measure of students' academic ability in Germany.[1] Finally, because our data include a large number of *compulsory* courses with different sizes, we can further address concerns of student sorting to courses.

Our results show that female students benefit more than their male peers from being taught by female professors. Overall, we find a female gender match effect of 7% of a standard deviation in grades. This estimate lies in between previously documented female gender match effects in tertiary education (Hoffmann and Oreopoulos (2009), Carrell et

[1] In about half of the study programs at German universities, the high school GPA also determines whether applicants can get admitted to oversubscribed programs.

al. (2010)). Importantly, however, this average gender match effect masks a pronounced heterogeneity by class size. In small classes, gender match effects are substantial, implying performance gains for female students of 13% of a standard deviation and a reduction in the probability of failing an exam by 1.4 percentage points. In large classes, gender match effects do not exist.[2]

We conduct a series of further analyses to show mechanisms and probe the robustness of our results. In particular, we find that the gender match effects on exam grades are largely driven by performance gains at the top of the grade distribution, pushing female students to obtain excellent rather than just good grades. However, being paired with a female lecturer also reduces the probability to fail a course for female students. In terms of robustness checks, we show that results are very similar for compulsory courses, ruling out selective course choices by students as a main driver of our results. Results are also robust to controlling for student fixed effects.

Our paper contributes to a growing literature that investigates female gender match effects in education.[3] Several papers have shown that gender match effects matter for the educational production in schools (Dee (2005, 2007), Cho (2012), Parades (2014), Antecol et al. (2015), Muralidharan and Sheth (2016), Lim and Meer (2017, 2020)). Moreover, exposure to role models in the form of advisors, mentors, or successful practitioners affects study choices and educational success in higher education, as well as occupational selection (Blau et al. (2010), Lyle and Smith (2014), Breda et al. (2018), Kofoed and McGovney (2019), Porter and Serra (2020), Canaan and Mouganie (2021, forthcoming)). Similarly, gender match effects at the student-professor level have been shown to influence major and course choices at university (Dynan and Rouse (1997), Rask and Bailey (2002), Bettinger

[2]We use the class size median to determine "small" (at or below median) and "large" (above median) classes. We also show results by class size decile and verify the robustness to changes in the large-class cutoff.

[3]In a recent study, de Gendre et al. (2022) perform a meta analysis of 538 estimates of role model effects in schools and universities. They complement this with an own investigation of role model effects in schools across 90 countries, using large-scale, standardized assessment data on 3 million students.

and Long (2005)) as well as student performance (Hoffmann and Oreopoulos (2009), Carrell et al. (2010)). Importantly, however, the existing evidence on gender match effects is heavily skewed towards settings with smaller class sizes, because most well-identified studies exploit random assignment of students to teachers rarely happening in larger-class settings. Our findings add to the literature by providing the first assessment of gender match effects in small versus large classes within the same university. Our results suggest that the findings of previous studies, which were mostly conducted in the context of small class sizes, cannot be generalized to settings with larger classes, which are common in public universities around the world.

The terms "gender match effects" and "role model effects" are sometimes used interchangeably and are often not precisely defined. Narrowly defined, role models effects may arise simply because just seeing that a female professor teaches a specific course may inspire female students in ways that lead to an increase in performance. However, our finding of a zero female-lecturer female-student interaction in large classes casts doubt on role model effects operating in this narrow sense. If just seeing that the professor is female would be sufficient to trigger sizable role model effects, we should observe them in both small and large classes. Instead, our main result that gender match effects are only present in small classes points towards the importance of classroom interactions between students and professors in generating economically meaningful gender match effects.[4]

Our findings also have important implications for policies aiming at reducing gender gaps in higher education or, more specifically, to increase the share of females who successfully complete STEM programs. Given our results, a policy to attract more female professors in STEM programs may be effective in achieving these goals if applied in settings with smaller classes that facilitate student-professor interactions. However, in larger education programs at public universities or in massive open online courses with little

[4]In fact, according to data from student evaluations at the university we study, we observe that there are more frequent and intense classroom interactions in small classes compared to large classes.

interaction between students and professors, an increase in the share of female professors in STEM may not have similar effects.

The remainder of the paper is structured as follows. Section 2 discusses the institutional background, describes our data, and lays out our empirical strategy. We present our results and robustness checks in Section 3. There, we also discuss student-lecturer interactions as a mechanism explaining gender match effects. Section 4 concludes.

2. Empirical Setup

2.1. Data and Institutional Background

We draw on the universe of bachelor-level exams taken at a medium-sized public university in Germany between 2006 and 2018. The university has 13 academic departments that offer different degree programs, which we henceforth call majors or programs.[5] For administrative purposes, the departments are further organized into three "sections": STEM, Humanities (which includes several social sciences), and a third section consisting of Political Science, Law, and Economics. Undergraduate majors are designed to be completed in three years, but it is quite common for students to take longer. Majors generally require a combination of compulsory courses, core elective courses, free elective courses, and a final thesis for a total of 180 ECTS. However, the proportion of each of these components may vary among different majors.[6] Students choose their major prior to enrollment. It is possible to change majors later on, but doing so may extend the duration of one's studies, as not all previously completed courses necessarily count towards the new major.

In our setting, an observation is the exam result in a given class taken by a given student in a given program and semester. We exclude law-related majors, as they have a

[5]We exclude programs that are only taken as minors.

[6]ECTS stands for European Credit Transfer and Accumulation System. One ECTS point corresponds to 25 to 30 hours of studying.

very different grading scheme from other programs. This leaves us with 27 majors that cover STEM fields, Humanities, social sciences, Political Science, and Economics.

Exams are graded on a scale from 1 to 5, with a total of 11 different possible grades.[7] Grades between the top grade of 1.0 and 4.0 are passing grades, the grade of 5.0 indicates a fail. To facilitate comparison, we standardize exam grades at the exam-semester level with mean 0 and standard deviation 1. We also reversed the usual German ordering so that higher values indicate better outcomes.

For every exam, there are usually at least two sittings, one immediately after the course and one several weeks later. Students who fail the first sitting can register for the second one, but students can also choose to take only the second sitting. In most courses, students can take at most two sittings. Failing a compulsory course twice typically means students have to leave their program and cannot enroll into the same program at any other public university in Germany. We exclude retries, second attempts, and any later attempts within the same course.[8] Courses can have up to two lecturers. We consider a course as female-taught if at least one of the lecturers is female, but we show below that our results are robust to alternative codings.

In addition to exam results, our data also contain rich student-level background characteristics that include gender, age, citizenship, and a student's experience in their major (which we proxy by the academic year in which the first exam is taken). Based on the location where students finished high school, we can also construct a dummy for "local" students, which takes a value of 1 if students completed their high school education in the county where the university is located. Importantly, our data provide information on the GPA of the high school leaving exam, which we use as control for student ability.[9] Previ-

[7]Grades starting with 1, 2, and 3 can take three values each (e.g., 1.0, 1.3, and 1.7).

[8]In some cases, our data have several entries for a given exam-major-student-semester combination that are all coded as first attempt. If one of the grades is a fail, we consider the course as a fail. When there are several non-fails, we average the grades over all the attempts.

[9]Most of the students at the university we study come from federal states with centralized final high school exams, which facilitates the comparability of grades. Moreover, we also have information on the

ous research has shown that high school grades in Germany are informative about student ability, as they correlate strongly with earnings (Schwerdt and Woessmann (2017)) and standardized test scores (Neumann et al. (2011)). In our data, we also observe a clear positive link between high school and university grades: For a one standard deviation increase in the high school grade, university grades on average improve by one third of a standard deviation (also see Appendix Figure A1).[10] We thus consider high school GPA to be a powerful measure of academic ability and a strong predictor of university exam performance.

We exclude observations where information on any of the student characteristics is missing. These restrictions remove 24,331 exam results and leave us with a final sample of 313,843 exam results from 18,598 distinct students.

Summary statistics are shown in Appendix Table A1. While there are slightly more female than male students in our sample, female lecturers only account for roughly a quarter of the courses taken. Female students take more courses taught by female lecturers than their male counterparts. Female students usually have better exam grades, and come to university with better high school GPAs and at a slightly younger age. The vast majority of bachelor students are German citizens, and about 13% of them attended high school in the county of their university.

2.2. Empirical Strategy

We are interested in female gender match effects, i.e., whether female students perform better when taught by female lecturers. Since students typically choose their program of study and many of their courses, there are several potential confounders. However, our

type of high school students attended and on the year in which they took the high school leaving exam. Our results are robust to allowing the association between high school grades and university exam grades to vary by the place of the high school, graduation year, and type of high school leaving exam (see Appendix Table A4).

[10]Luis Silva et al. (2022) even find that high school grades in Portugal are on average better at predicting study success at university than university admission tests.

data allow us to follow the same lecturers and courses over time, exploiting changes in who teaches which courses. Specifically, for student i enrolled in program p (e.g., Economics) taking course c (e.g., Microeconomics I) in semester t (e.g., winter semester 2006/07), we set up the following model:

$$\begin{aligned} grade_{ipct} &= \beta FemaleLecturer_{ct} \times FemaleStudent_i \\ &+ \gamma' StudentChars_{it} + \lambda' LecturerSet_{ct} \\ &+ \omega_p + \xi_c + \tau_t + \epsilon_{ipct}, \end{aligned} \tag{1}$$

The outcome of interest, *grade*, is exam grades, standardized to mean 0 and standard deviation 1 at the exam-semester level. $FemaleLecturer$ is a dummy for whether the lecturer of course c in semester t is female (if there are two lecturers: if at least one of the lecturers is female). $FemaleStudent$ is a dummy for whether student i is female, and the product of the two dummies is our key variable of interest with associated coefficient β. $StudentChars$ is a vector of student characteristics: gender, final high school grade (standardized to mean 0 and standard deviation 1 in the overall sample), age, dummies for having a German citizenship and for having completed high school in the county where the attended university is located, respectively, and the starting year in the major (coded as the academic year in which we observe the first exam). With the exception of age, student characteristics are time-invariant. $LecturerSet$ are fixed effects for the combination of first and second lecturer. ω denotes fixed effects for the program as part of which student i takes the course, ξ are course fixed effects, and τ are semester fixed effects. Standard errors are twoway-clustered at the student and course level.

Including this demanding set of fixed effects allows us to address multiple possible confounders in the estimation of gender match effects. For instance, we can account for different grading standards and gender shares across programs, courses, and over time,

systematic selection of students into courses that are perceived as easy or hard, and lecturers' teaching abilities. We identify effects from over-time changes in the lecturer(s) who teach a specific course, which could be due to, for example, sabbaticals, recruitment of new professors, or within-department reshuffling of teaching duties.

One remaining concern is that students systematically respond to changes in lecturer gender based on their own ability and gender. Below, we therefore also show results for compulsory courses and courses offered early in the study program, where students have little or no choice. Moreover, in Appendix Table A2, we assess whether the female-male student difference in various student characteristics differs between courses taught by female professors and courses taught by male professors. To do so, we use five pre-determined student characteristics as outcome variables in the main estimation model outlined above (Pei et al. (2019)). We show the results of this balancing test across all classes as well as by class size and type of course (all vs. compulsory). We find little evidence for systematic differences: From the 30 coefficients of interest, only 3 are statistically significant at the 5% level, and all coefficients are economically small.[11] Most importantly, we do not observe any sorting of students based on ability as measured by high school GPA. In addition, Appendix Table A3 shows that female students do not systematically sort into courses taught by female professors.

3. Results

3.1. Main Results

Figure 1 provides a graphical illustration of our main result. It shows female gender match effects along deciles of class size. For class sizes in the lowest 5 deciles (correspond-

[11] In particular, the female-male difference in age of students taught by a female lecturer is somewhat smaller than the female-male age difference of students taught by a male lecturer. However, the magnitude of the difference is small and we always control for student age in our regressions.

ing to 73 or fewer students), we find positive, sizable, and statistically significant effects.[12] Pairing a female student with a female lecturer improves the student performance by 10–18% of a standard deviation in smaller classes. In terms of magnitude, the estimated gender match effects amount to 3–4 times the gender gap in exam performance.[13] However, for class sizes above the median, estimated female gender match effects decrease substantially in size. For the 6th, 7th and 8th decile in the class size distribution, we still find positive and marginally significant coefficients of around 5–8% of a standard deviation, whereas for the two highest deciles, coefficients are close to 0 and statistically insignificant. The heterogeneity by class size is also illustrated by the solid black lines, which depict a separate estimate of female gender match effects for class sizes below and above the median, respectively.

Table 1 shows our main result in regression table format. In Columns 1 and 2, we estimate female gender match effects in the whole sample, without or with controlling for a student's high school GPA. In both cases, we find statistically significant average effects of around 7% of a standard deviation. This effect size falls in between previous estimates of female gender match effects in tertiary education: Hoffmann and Oreopoulos (2009) find gains of up to 5% of a standard deviation for the University of Toronto, while Carrell et al. (2010) report effects of 10% of a standard deviation for the US Air Force Academy. Columns 3 and 4 correspond to the solid lines in Figure 1: They show that average female gender match effects are entirely driven by courses below the class size median, where we find an effect of 12.8% of a standard deviation. Above the median, the estimate is close to zero and statistically insignificant.

[12]Note that we proxy class size by the number of students taking the final exam. The actual number of students regularly attending the lectures is likely smaller than the number of exam-takers, as attendance is typically not compulsory at German universities.

[13]Conditional on other student characteristics and our set of fixed effects, female students perform 4% of a standard deviation worse than male students.

In Table 2, we examine from which part of the grade distribution the estimated female gender match effects come from. To do so, we replace the continuous grade outcome by a series of dummies that indicate whether students got an A, B, C, D, or failed. As can be seen in Panel A, female students that are paired with a female lecturer in a small class are 4.4 percentage points more likely to get an A, are 1.8 percentage points less likely to get a C, and are 1.4 percentage points less likely to fail a course. Female gender match effects thus seem to be present along the entire grade distribution: at the top, female students benefit from having a female lecturer by being more likely to receive excellent rather than just good grades; at the bottom, gender match effects materialize through a reduced risk of failing a course. For large courses, we do not find gender match effects for any grade category.

3.2. Robustness

One main worry is that our results simply reflect selection patterns, for instance, because high-ability female students systematically choose programs or courses with female lecturers. However, such systematic sorting is unlikely to explain our results. First, we control for students' academic ability measured by high school GPA. Second, due to the inclusion of program and course fixed effects, we can rule out that our effects are driven by selection into programs or courses. One remaining concern is that high-ability female students take more courses with female lecturers. If they are particularly likely to do this in small courses — but not in large courses —, this could potentially explain our results.

We provide two additional analyses to address this concern. First, in a specification analogous to our main empirical model, we can show that female students are not more likely to take female-taught courses. This holds both among higher-ability and lower-ability students as well as in small and large courses (see Appendix Table A3). Moreover, we repeat our main analysis for compulsory courses.[14] Table 3 shows female gender

[14]In some programs, students can choose *when* to take a compulsory course.

match effects of the same magnitude in small compulsory courses (Column 1) as in small elective courses (Column 2). In large courses, compulsory or elective, we cannot detect any gender match effects (Columns 3 and 4). However, some programs have very few compulsory courses, especially in Humanities. We thus also look at courses taken in the first academic year of the study program, i.e., in the first two semesters. These are often basic courses, serving as the foundation of the more advanced courses in the second and third years of the program. Thus, even though not all of these early courses are actually compulsory, there may be the implicit (or even explicit) recommendation to take these courses early on. Table 3 reveals the same pattern for courses in the first two semesters as for compulsory courses: We find a sizable female gender match effect in small courses (Column 5), and a zero effect in large courses (Column 6). Given that our results also hold in courses that students are required or recommended to take, we conclude that systematic selection of high-ability female students to female lecturers is no major concern for our analysis.

Another worry is that our results are driven by gender-biased grading, i.e., female lecturers giving better grades to female students. For instance, Jansson and Tyrefors (2022) find evidence for same-sex bias in grading when exams are not anonymous. However, the institutional setting in our study render gender-biased grading unlikely. Written exams are usually graded blind, with graders only knowing the student ID of the examinees, not their name or gender.[15] In addition, large exams are often graded by teaching assistants and not by the lecturers themselves. Lecturers are therefore unlikely to know the gender of a student who wrote a given exam. The one major exception to this are so-called "seminars," where students usually write and present a term paper. In these courses, a student's identity is known to the grader. However, the class size of seminars is usually very small. In the Economics Department, for example, seminars are capped at 12 stu-

[15]Oral examinations are a possibility, but occur rarely. For instance, in the Economics Department there is no class where the grade is exclusively determined by an oral examination.

dents. Given that gender match effects are also present in courses with 20, 30, and even 70 students (see Figure 1), gender-biased grading is unlikely to explain our findings.

A number of additional exercises, discussed in detail in the appendix, confirm the robustness of the results. These robustness checks include adding student fixed effects or program-by-semester fixed effects, applying alternative definitions of "female-taught" or "large" courses, allowing the effect of high school GPA to vary by high school type, location, and graduating year, and excluding students who drop out early. We also find that female gender match effects in small courses do not differ much along the three broad academic fields of the studied university (Economics/Political Science, STEM, Humanities) or along students' ability distribution.

3.3. The Role of Student-Lecturer Interactions

Our evidence suggests that female gender match effects in higher education exist, but are strongly dependent on class size as they are not present in large courses. But why do these gender match effects exist, and why do they depend on class size? On the former question, we believe we have ruled out preferential grading and non-random assignment of students to lecturers. However, this still leaves at least two potential explanations: One explanation is gender-specific teaching skills, i.e., women might be better at teaching women (vice versa for men). Another explanation is role model effects in a narrow sense, i.e., female lecturers motivating female students to do better. The difference between the two explanations is subtle, and we cannot distinguish between them empirically.

What can explain the class size gradient in female gender match effects? We believe that the intensity of student-teacher interactions is important. These interactions are likely more frequent and of higher quality in smaller classes than in anonymous mass lectures. Appendix Table A7 corroborates this claim based on data from course evaluations in the Economics Department. We observe that the larger the course, the less students

feel that they can make comments, get useful feedback, or have the opportunity to ask questions.

The psychological literature also suggests the importance of student-teacher interactions. For instance, Buck et al. (2008) find that feeling a strong personal connection is necessary for being seen as a role model. Naturally, it seems easier to develop a personal connection with a lecturer in a small class compared to a large class. Additionally, Stout et al. (2010) show that female students are more likely to participate in class and seek help if their professor is female. It is likely that this effect is stronger in small classes, where there is more opportunity to ask questions and interact with the professors.

4. Conclusion

We study whether female gender match effects in higher education depend on class size. To do so, we exploit rich administrative records from a German university, which cover all programs and courses in the period 2006 to 2018. We find that female gender match effects are substantial in small classes, implying performance gains of 13% of a standard deviation and a reduction in the probability of failing an exam by 1.4 percentage points if female students are taught by a female professor. In contrast, there are no female gender match effects in large classes.

We are the first to show this quantitatively important interaction between female gender match effects and class size. Our results complement the growing empirical literature that investigates gender match effects in education, which, however, is heavily skewed towards settings with smaller classes. In particular, our findings call into question the generalizability of findings on female gender match effects from studies that exploit random assignments of students to several classes of smaller size.

Our findings also offer insights into the nature of female gender match effects. The mere knowledge that one's professor is female, which also students in large classes have, is apparently in itself not enough to increase the performance of female students. This

suggests that the idea that gender match effects occur simply because female students are inspired by seeing another woman excel in a subject to the point of becoming a professor is too simplistic. Rather, our results are in line with a more complex mechanism driving gender match effects that require direct and frequent interactions between students and professors, which is more typical in smaller classes.

Finally, our results also have important policy implications. Enrollment in tertiary education has increased in many countries in recent years, and the COVID-19 pandemic has led to an increase in online education options in tertiary education, including massive open online courses. These developments may result in more settings with larger class sizes and less direct and frequent interactions between students and professors. Our results suggest that this trend towards more online education may weaken the impact of policies designed to increase female graduation rates in traditionally male-dominated fields (such as STEM) by increasing gender diversity among professors.

References

[1] Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik, "The Effect of Teacher Gender on Student Achievement in Primary School", *Journal of Labor Economics* 33 (2015): 63-89.

[2] Bettinger, Eric P., and Bridget Terry Long, "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students", *AEA Papers and Proceedings* 95 (2005): 152-157.

[3] Bingham, Liz, "Role models are essential to help women reach the top", *The Guardian* 2018, Sep 27 `https://www.theguardian.com/careers/role-models-gender-barrier`.

[4] Blau, Francine D., Janet M. Currie, Rachel T.A. Croson, and Donna K. Ginter, "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial", *AEA Papers and Proceedings* 100 (2010): 348-352.

[5] Breda, Thomas, Julie Grenet, Marion Monnet, Clementine Van Effenterre, "Can Female Role Models Reduce the Gender Gap in Science? Evidence from Classroom Interventions in French High Schools", PSE Working Papers Nr. 2018-06 (2018).

[6] Buck, Gayle A., Vicki L. Plano Clark, Diandra Leslie-Pelecky, Yun Lu, and Patricia Cerda-Lizarraga, "Examining the Cognitive Processes Used by Adolescent Girls and Women Scientists in Identifying Science Role Models: A Feminist Approach"*Science Education* 92 (2008): 688-707.

[7] Canaan, Serena, and Pierre Mouganie, "Does Advisor Gender Affect Women's Persistence in Economics?"*AEA Papers and Proceedings* 111 (2021): 112-116.

[8] Canaan, Serena, and Pierre Mouganie, "The Impact of Advisor Gender on Female Students' STEM Enrollment and Persistence"*Journal of Human Resources* forthcoming.

[9] Cho, Insook, "The Effect of Teacher-Student Gender Matching: Evidence from OECD Countries", *Economics of Education Review* 31 (2012): 54-67.

[10] Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap", *Quarterly Journal of Economics* 125 (2010): 1101-1144.

[11] Thomas, S., "A Teacher Like Me: Does Race, Ethnicity or Gender Matter?", *AEA Papers and Proceedings* 95 (2005): 158-165.

[12] Dee, Thomas S., "Teachers and the Gender Gaps in Student Achievement", *Journal of Human Resources* 42 (2007): 529-554.

[13] de Gendre, Alexandra, Jan Feld, Nicolas Salamanca, and Ulf Zölitz, "Do Same-Sex Teachers Affect Test Scores and Job Preferences? A Super-Study and a Meta-Analysis on Role Model Effects in Education", Working Paper, 2022

[14] Dynan, Karen E., and Cecilia Elena Rouse, "The Underrepresentation of Women in Economics: A Study of Undergraduate Economics Students", *Journal of Economic Education* 28 (1997): 350-368.

[15] Hoffmann, Florian, and Philip Oreopoulos, "A Professor Like Me. The Influence of Instructor Gender on College Achievement", *Journal of Human Resources* 44 (2009): 480-494.

[16] Jansson, Joakim, and Björn Tyrefors, "Grading Bias and the Leaky Pipeline in Economics: Evidence from Stockholm University", *Labour Economics* 78 (2022).

[17] Kofoed, Michael S., and Elizabeth McGovney, "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice", *Journal of Human Resources* 54 (2019): 430-467.

[18] Lim, Jaegeum, and Jonathan Meer, "The Impact of Teacher-Student Gender Matches. Random Assignment Evidence from South Korea", *Journal of Human Resources* 52 (2017): 979-997.

[19] Lim, Jaegeum, and Jonathan Meer, "Persistent Effects of Teacher-Student Gender Matches", *Journal of Human Resources* 55 (2020): 809-835.

[20] Lyle, David S., and John Z. Smith, "The Effect of High-Performing Mentors on Junior Officer Promotion in the US Army", *Journal of Labor Economics* 32 (2014): 229-258.

[21] Muralidharan, Karthik, and Ketki Sheth, "Bridging Education Gender Gaps in Developing Countries. The Role of Female Teachers", *Journal of Human Resources* 51 (2016): 269-297.

[22] Neumann, Marko, Ulrich Trautwein, and Gabriel Nagy, "Do Central Examinations Lead to Greater Grading Comparability? A Study of Frame-of-Reference Effects on the University Entrance Qualification in Germany", *Studies in Educational Evaluation* 37 (2011): 206-217.

[23] Paredes, Valentina, "A Teacher Like Me or a Student Like Me? Role Model Versus Teacher Bias Effect", *Economics of Education Review* 34 (2014): 38-49.

[24] Pei, Zhuan, Jörn-Steffen Pischke, and Hannes Schwandt. "Poorly Measured Confounders Are More Useful on the Left Than on the Right", *Journal of Business and Economic Statistics* 37 (2019): 205-216.

[25] Porter, Catherine, and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Female Role Models", *American Economic Journal: Applied Economics* 12 (2020): 226-254.

[26] Rask, Kevin N., and Elizabeth M. Bailey, "Are Faculty Role Models? Evidence from Major Choice in an Undergraduate Institution", *Journal of Economic Education* 33 (2002): 99-124.

[27] Schwerdt, Guido, and Ludger Wößmann, "The Information Value of Central School Exams", *Economics of Education Review* 56 (2017): 65-79.

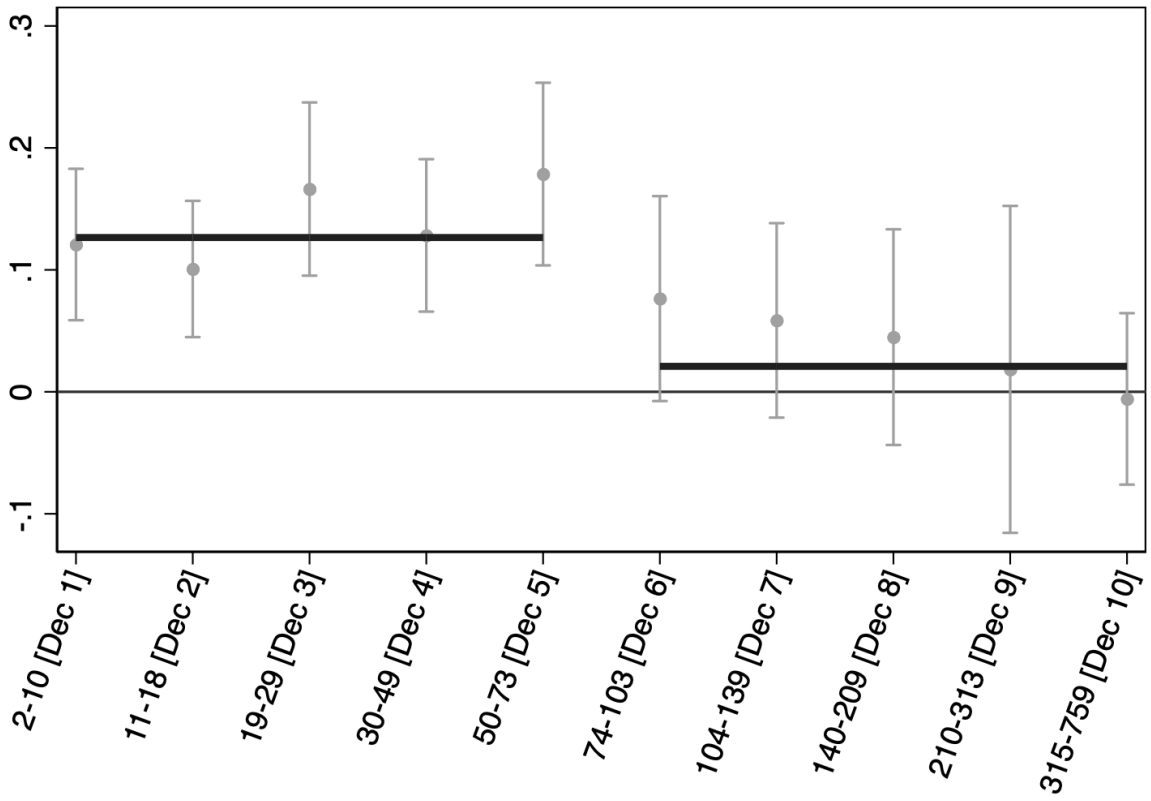
[28] Silva, Pedro Luis, Carla Sá, and Rciardo Biscaia, "High School and Exam Scores: Does Their Predictive Validity for Academic Performance Vary with Programme Selectivity?", IZA DP No. 15350 (2022).

[29] Stout, Jane G., Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus, "STEMing the Ide: Using Ingroup Experts to Inoculate Women's Self-Concept in Science, Technology, Engineering, and Mathematics (STEM)", *Journal of Personality and Social Psychology* 100 (2010): 255-270.

[30] Warrell, Margie, ""Seeing Is Believing: Female Role Models Inspire Girls To Think Bigger", *Forbes* 2020, Oct 9, `https://www.forbes.com/sites/margiewarrell/2020/10/09/seeing-is-believing-female-role-models-inspire-girls-to-rise/`.

Figures and Tables

Figure 1: Class size heterogeneity of female gender match effects



Notes: Figure shows estimated female gender match effects and their 95% confidence intervals by class size decile. Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level. Estimations control for student characteristics (gender, high school GPA, age, German citizenship, being a local student, and first year in major) and for class size decile, program, course, semester, and lecturer set fixed effects. Black lines depict average female gender match effects for class sizes below and above the median, respectively (see Columns 3 and 4 of Table 1 for details). *Data source:* Administrative student records.

Table 1: Female gender match effects by class size

	(1)	(2)	(3)	(4)
Female lecturer	0.069***	0.072***	0.128***	0.019
× female student	(0.019)	(0.018)	(0.018)	(0.028)
Student characteristics				
High school GPA		0.434***	0.378***	0.493***
		(0.007)	(0.007)	(0.009)
Female student	0.035**	-0.060***	-0.068***	-0.060***
	(0.014)	(0.013)	(0.014)	(0.017)
Student age	-0.035***	-0.007***	-0.004	-0.012***
	(0.003)	(0.002)	(0.002)	(0.003)
Native student	0.375***	0.226***	0.201***	0.237***
	(0.031)	(0.027)	(0.032)	(0.036)
Local student	-0.138***	-0.127***	-0.111***	-0.143***
	(0.017)	(0.013)	(0.015)	(0.018)
First year in major	0.017***	0.014**	0.006	0.027***
	(0.007)	(0.006)	(0.006)	(0.010)
Class size	All	All	Small	Large
Observations	313,843	313,843	157,100	156,726

Notes: Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level. High school GPA standardized to mean 0 and std. dev. 1 in the overall sample. All regressions control for program, course, semester, and lecturer set fixed effects. Small courses have 73 or fewer students, large courses have 74 or more students. Students' migration background is based on citizenship. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

Table 2: Female gender match effects for different grade categories in small classes

Dep Var:	(1) A	(2) B	(3) C	(4) D	(5) Fail
Panel A: Small Classes					
Female lecturer	0.044***	-0.007	-0.018***	-0.005	-0.014***
× female student	(0.007)	(0.007)	(0.005)	(0.003)	(0.004)
Mean dependent variable	0.291	0.381	0.177	0.065	0.086
Observations	157,100	157,100	157,100	157,100	157,100
Panel B: Large Classes					
Female lecturer	0.002	0.006	-0.007	-0.001	0.000
× female student	(0.006)	(0.010)	(0.006)	(0.006)	(0.007)
Mean dependent variable	0.135	0.293	0.266	0.136	0.169
Observations	156,726	156,726	156,726	156,726	156,726

Notes: Dependent variable: Binary variables indicating the four major grade categories (Columns 1–4) and binary variable taking a value of 1 if the student failed the exam, zero otherwise (Column 5). All regressions control for student characteristics (gender, high school GPA, age, German citizenship, being a local student, and first year in major) and for program, course, semester, and lecturer set fixed effects. In Panel A, sample is restricted to classes with 73 or fewer students; in Panel B, sample is restricted to classes with 74 or more students. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

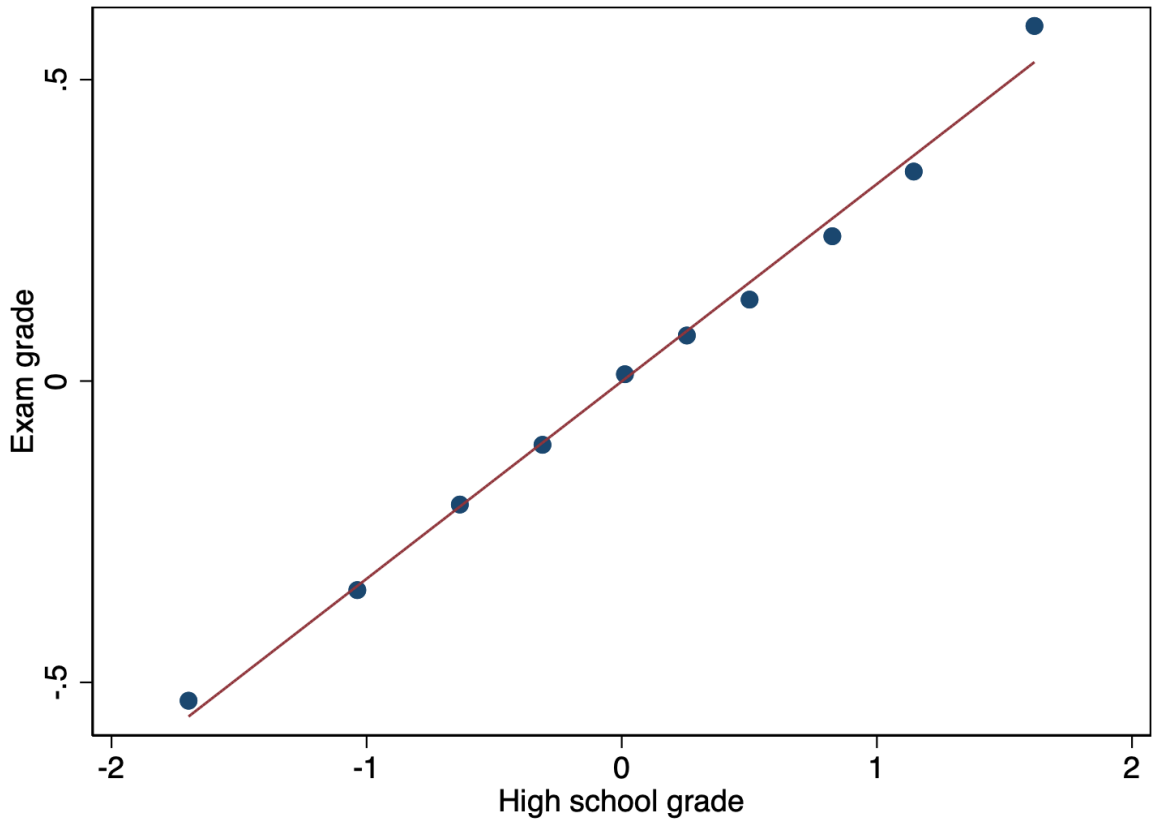
Table 3: Female gender match effects by class size: Compulsory vs. elective courses

	(1)	(2)	(3)	(4)	(5)	(6)
Female lecturer	0.113***	0.129***	0.030	0.004	0.132***	0.012
× female student	(0.042)	(0.018)	(0.028)	(0.047)	(0.038)	(0.033)
Course type	Comp.	Elect.	Comp.	Elect.	First 2 semesters	
Class size	Small	Small	Large	Large	Small	Large
Observations	21,924	135,143	101,744	54,959	35,730	91,711

Notes: Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level. All regressions control for student characteristics (gender, high school GPA, age, German citizenship, being a local student, and first year in major) and for program, course, semester, and lecturer set fixed effects. Small courses have 73 or fewer students, large courses have 74 or more students. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

Appendix

Figure A1: Relationship between exam grades and high school GPA



Notes: Binned scatterplot of the bivariate relationship between exam grades and final high school GPA. Exam grades are standardized to mean 0 and std. dev. 1 at the exam-semester level. High school GPA is standardized to mean 0 and std. dev. 1 in the overall sample. For both grade variables, the usual German ordering is reversed so that higher values indicate better outcomes. *Data source:* Administrative student records.

Table A1: Summary statistics

	Overall			Females			Males		
	Mean	Med	SD	Mean	Med	SD	Mean	Med	SD
Panel A: Exam-level variables									
Female lecturer	0.273	0	0.445	0.322	0	0.467	0.213	0	0.409
Exam grade	0	0.149	0.976	0.032	0.185	0.946	-0.038	0.100	1.010
Failed exam	0.128	0	0.334	0.101	0	0.302	0.159	0	0.366
Class size	120.1	73	130.8	110.0	67	121.8	132.2	79	139.7
Observations		313,843			170,924			142,919	
Panel B: Student-level variables									
Female student	0.532	1	0.499						
HS GPA	-0.145	-0.070	1.017	-0.045	-0.070	1.003	-0.258	-0.233	1.022
Age student	21.181	21	2.983	21.064	20	3.053	21.313	21	2.896
Native student	0.967	1	0.180	0.964	1	0.185	0.969	1	0.173
Local student	0.127	0	0.333	0.111	0	0.314	0.147	0	0.354
First year in major	2011.7	2012	3.9	2011.7	2012	4.0	2011.8	2012	3.9
Observations		18,598			9,893			8,705	

Notes: Table presents summary statistics for exam-level variables (Panel A) and student-level variables (Panel B). HS GPA refers to the final high school GPA; standardized to mean 0 and std. dev. 1 in the overall sample. Students' migration background is based on citizenship. Local students completed their high school in the county where the university is located. First year refers to the first academic year in which a student appears in our data in a given major. *Data source:* Administrative student records.

Table A2: Balancing tests

	Female gender match effect coefficient in					
	All classes	Small classes	Large classes	Comp.	Small comp.	Large comp.
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome variable:						
HS GPA	-0.006	-0.010	0.005	0.000	-0.027	0.006
	(0.012)	(0.019)	(0.011)	(0.015)	(0.047)	(0.015)
Age student	-0.105**	-0.171**	-0.004	-0.047	-0.302*	-0.006
	(0.051)	(0.070)	(0.044)	(0.061)	(0.155)	(0.057)
Native student	-0.001	-0.001	-0.002	-0.001	0.016**	-0.003
	(0.002)	(0.004)	(0.003)	(0.003)	(0.008)	(0.004)
Local student	0.004	0.007	0.001	0.002	0.006	0.000
	(0.005)	(0.007)	(0.004)	(0.006)	(0.016)	(0.005)
First year in major	0.002	0.019	-0.017	0.007	0.019	0.005
	(0.013)	(0.016)	(0.017)	(0.010)	(0.035)	(0.010)

Notes: Table shows results from regressing a number of predetermined student characteristics on the interaction of female student and female lecturer. All regressions control for the student being female and for program, course, semester, and lecturer set fixed effects, as well as for the four student characteristics that are not used as outcome variable in the respective regression. Columns 1–3 report results for compulsory and elective courses, Columns 4–6 report results for compulsory courses only. Small courses have 73 or fewer students, large courses have 74 or more students. HS GPA refers to the final high school GPA; standardized to mean 0 and std. dev. 1 in the overall sample. Students' migration background is based on citizenship. Local students completed their high school in the county where the university is located. First year refers to the first academic year in which a student appears in our data in a given major. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

Table A3: Female course choice effects

	(1)	(2)	(3)	(4)	(5)
Female student	0.000	0.001	0.001	0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Class size	All	All	All	Small	Large
Student high school GPA	All	Above median	Below median	All	All
Observations	313,843	139,234	172,228	313,843	313,843

Notes: Dependent variable: Dummy for whether the course is taught by at least one female lecturer (Columns 1–3). In Column 4 (Column 5), the dummy is set to 1 if the course is female-taught and the course size is 73 students and below (above 73 students). In Column 2 (Column 3), we consider only student with an above-median (below-median) final high school GPA. All regressions control for student characteristics (high school GPA, age, German citizenship, being a local student, and first year in major) and for program, course, and semester fixed effects. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

Table A4 shows the robustness of our results to different specifications and parameterizations. As discussed in Section 2.2, classes can have up to two lecturers. In our main analysis, we keep all classes and define a course as female-taught if at least one lecturer is female. We check the robustness of our results to this definition in Panels A and B. In Panel A, we restrict the sample to classes with only one lecturer, where the definition of a female-taught course is unambiguous. In Panel B, we keep all classes, but define a course as female-taught if both lecturers are female. Our results are robust to both alternative definitions of female-taught courses.

In Panel C of Table A4, we account for variation in school quality in Germany over time and across space by interacting the high school GPA with indicators of the location of the high school, graduation year, and broad types of high school leaving exam.[16] Again, our results remain essentially unchanged.

[16]Location is measured by the county of high school graduation for students who graduated from high school in Germany. For students who completed high school abroad, we use the country of graduation. The most common type of high school leaving exam is the regular "Abitur" taken at standard upper secondary high schools. Other common types include Abitur at more specialized high schools, diplomas that allow university attendance only in some specific programs ("fachgebundene Hochschulreife") or various types of vocational or second-chance education programs that award a university entrance qualification.

In a similar vein, Panel D of Table A4 makes use of the fact that we observe several exams per student, allowing us to account for student fixed effects. Coefficients decrease by about half in this specification, however without altering our basic pattern: A sizable female gender match effect in small classes and no effect in large classes. Another potential confounder could be that specific departments hired more female lecturers over time and also changed exam standards, entry requirements, or other aspects of teaching. In Panel E, we therefore include major-by-semester fixed effects. Results are virtually identical to our baseline findings. The same holds for Panel F, where we exclude students who studied less than three semesters in their major. The latter check shows that our results are not driven by students who drop out early.

In our main analysis, we have defined large and small classes based on the median of the overall class size distribution. Based on the idea that the intensity of student-teacher interaction depends on class size, we consider this to be the most sensible approach. This is also in line with the pattern observed in Figure 1, showing that female gender match effects strongly decrease above the median of the class size distribution. However, one disadvantage of this approach is that some majors are very small and might thus not have many large classes, whereas for other majors, most classes might be large. In Panel G of Table A4, we therefore use major-specific medians to define large and small classes. This change in the definition of the class size cutoff leaves our results for small classes unchanged, as we continue to find a large female gender match effect. However, we now also observe a statistically significant, albeit much smaller, effect in large classes. This is likely due to the fact that in some majors, "large" classes by our definition are in fact small. In the programs "Slavistic"and "Cultural Studies of Antiquity", for example, the median number of exam takers is 7. In "French Studies"and "Italian Studies", the median is 8.

Finally, we check whether our results depend on the standardization of exam grades at the exam-semester level. In Panel H of Table A4, we use raw exam grades that follow

the German system from 1 (very good) to 5 (fail). In line with this new ordering, we now obtain a negative point estimates on the female lecturer female student interaction, but otherwise the same qualitative result: In small classes, female students paired with female lecturers receive significantly better (i.e., lower) grades, which is not the case in large classes.

Table A4: Robustness

Class size	All	Small	Large
	(1)	(2)	(3)
Panel A: Classes with only one lecturer			
Female lecturer	0.081***	0.130***	0.030
× female student	(0.019)	(0.018)	(0.033)
Observations	281,483	142,458	139,008
Panel B: Alternative treatment definition: Both lectures female			
Female lecturer	0.077***	0.128***	0.025
× female student	(0.019)	(0.018)	(0.032)
Observations	313,843	157,100	156,726
Panel C: Additional high school controls			
Female lecturer	0.064***	0.114***	0.015
× female student	(0.017)	(0.017)	(0.027)
Observations	313,753	156,864	156,503
Panel D: Controlling for student fixed effects			
Female lecturer	0.037**	0.071***	0.003
× female student	(0.016)	(0.016)	(0.025)
Observations	313,381	155,905	155,470
Panel E: Controlling for major × semester fixed effects			
Female lecturer	0.072***	0.128***	0.020
× female student	(0.018)	(0.018)	(0.028)
Observations	313,839	157,096	156,688
Panel F: Dropping students who study less than 3 semesters in major			
Female lecturer	0.077***	0.126***	0.026
× female student	(0.018)	(0.018)	(0.030)
Observations	289,995	152,032	137,944
Panel G: Using major-specific medians to define large courses			
Female lecturer	0.072***	0.141***	0.061***
× female student	(0.018)	(0.027)	(0.021)
Observations	313,843	54,182	259,234
Panel H: Raw exam grades			
Female lecturer	-0.059***	-0.111***	-0.014
× female student	(0.018)	(0.017)	(0.030)
Observations	313,843	157,100	156,726

Notes: Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level (raw exam grades in Panel H). All regressions control for student characteristics (gender, high school GPA, age, German citizenship, being a local student, and first year in major — the only exception being Panel D, where these get captured by the student fixed effects) and for program, course, semester, and lecturer set fixed effects. With the exception of Panel G, small courses have 73 or fewer students, large courses have 74 or more students. In Panel C, we allow the effect of the high school GPA to vary by the place of the high school, graduation year, and type of high school leaving exam. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<.10$, ** $p<.05$, *** $p<.01$. *Data source:* Administrative student records.

In Table A5, we explore the heterogeneity of our results by broad academic field. The university we study is divided into three sections: (i) Political Science & Economics, (ii) STEM (including Psychology), and (iii) Humanities.[17] As can be seen in the bottom of the table, the three sections differ substantially in female lecturer share and class size. While most of the classes in the Economics & Political Science section are above the median in size, the opposite is true for the Humanities section. However, in spite of these differences, our key results hold in all three sections: Positive female gender match effects in small classes, and no effects in large classes. Intriguingly, we find female gender match effects to be strongest for STEM disciplines, which is the section with the lowest share of female lecturers (Columns 2 and 5). In STEM fields, there is even a positive and sizable female gender match effect in large classes (Column 5), just shy of statistical significance at conventional levels (p=0.15).

[17]Part of the section of Political Science and Economics is also the Law Department. However, as explained in Section 2.1, we exclude law programs from our analysis.

Table A5: Female gender match effects by class size and broad field

	(1)	(2)	(3)	(4)	(5)	(6)
Female lecturer	0.074**	0.173***	0.118***	-0.004	0.047	-0.023
× female student	(0.030)	(0.036)	(0.026)	(0.041)	(0.033)	(0.052)
Class size	Small	Small	Small	Large	Large	Large
Broad field	Econ & PolSci	STEM	Humanities	Econ & PolSci	STEM	Humanities
Fem. lecturer share	0.333	0.218	0.404	0.232	0.174	0.285
Observations	26,759	55,456	76,284	77,745	65,824	17,554

Notes: Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level. All regressions control for student characteristics (gender, high school GPA, age, German citizenship, being a local student, and first year in major) and for program, course, semester, and lecturer set fixed effects. Small courses have 73 or fewer students, large courses have 74 or more students. The allocation of programs to broad fields follows the administrative division of the university. Econ & PolSci includes the programs Economics and Political & Administration Sciences. STEM includes the programs Biological Sciences, Chemistry, Computer Science, Information Engineering, Life Science, Financial Mathematics, Mathematics, Molecular Materials Science, Nanoscience, Physics, Psychology. Humanities includes the programs British and American Studies, German Literature, French Studies, History, Italian Studies, Cultural Studies of Antiquity, Literature-Art-Media, Philosophy, Slavistik/Literature, Sociology, Spanish Studies, Linguistics, Sports science. Financial Mathematics is offered jointly by the Department of Mathematics and the Department of Economics and is allocated to both Econ & PolSci and STEM. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

Do the benefits of being matched with a female lecturer accrue rather to high-ability or to low-ability female students? We investigate this question in Table A6, using students' high school GPA as a measure of academic ability.[18] Columns 1 and 2 of Table A6 report results for students with a high school GPA above the median, Columns 3 and 4 restrict the sample to students with a below-median high school GPA. In both groups, we find that female students benefit from being paired with a female lecturer in a small class, while high-ability students benefit even somewhat more (14.3% of a standard deviation, compared to 12.3% of a standard deviation for low-ability students) (Columns 1 and 3). High-ability female students even benefit from having a female lecturer in large classes, albeit to a smaller extent than in small classes (Column 2).

Table A6: Female gender match effects by class size and high school GPA

	(1)	(2)	(3)	(4)
Female lecturer	0.143***	0.060*	0.123***	-0.005
× female student	(0.026)	(0.031)	(0.022)	(0.032)
Class size	Small	Large	Small	Large
Student high school GPA	Above median		Below median	
Observations	65,094	73,941	89,324	82,764

Notes: Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level. All regressions control for student characteristics (gender, high school GPA, age, German citizenship, being a local student, and first year in major) and for program, course, semester, and lecturer set fixed effects. Small courses have 73 or fewer students, large courses have 74 or more students. Standard errors, twoway-clustered at the student and course level, in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Administrative student records.

[18] As shown above in Figure A1, high school GPA is strongly correlated with subsequent university performance.

Table A7: Differences in lecturer-student interactions by class size from student evaluations

Dep Var:	Can make questions and comments		Get useful feedback		Opportunities to ask questions	
	(1)	(2)	(3)	(4)	(5)	(6)
Class size	-0.001***		-0.002***		-0.002**	
	(0.000)		(0.001)		(0.001)	
Large class		-0.232***		-0.404***		-0.327**
		(0.080)		(0.131)		(0.122)
Mean Dep Var	4.574		4.258		4.469	
Observations	71		71		46	

Notes: Dependent variable: Course average of student replies to the questions indicated in the column header. The full questions read: "I feel I can ask questions and make comments at any time" (Columns 1 and 2); "I get useful feedback and advice from the lecturer when I ask" (Columns 3 and 4); "I have enough opportunities to ask questions" (Columns 5 and 6). Responses were given on a 5-point Likert scale, ranging from "strongly agree" (=5) to "strongly disagree" (=1). *Large class* is a binary variable, taking a value of 1 if the course had more than 73 filled-out evaluations, zero otherwise. Robust standard errors in parentheses. Significance levels: * $p<$ .10, ** $p<$ .05, *** $p<$ .01. *Data source:* Student evaluations from all Economics classes for the winter semester 2018/19.