IWH
Halle Institute for Economic Research
Member of the Leibniz Association

Discussion Papers

No. 14
May 2023

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

Stephan Maurer, Guido Schwerdt, Simon Wiederhold

Authors

Stephan Maurer
University of Konstanz and CEP
E-mail: stephan.maurer@uni-konstanz.de

Guido Schwerdt
University of Konstanz, ifo Institute, CESifo, IZA, and ROA
E-mail: guido.schwerdt@uni-konstanz.de

Simon Wiederhold
Halle Institute for Economic Research (IWH) – Member of the Leibniz Association, Department of Structural Change and Productivity, Martin Luther University Halle-Wittenberg, ifo Institute Munich, and CESifo
E-mail: simon.wiederhold@iwh-halle.de
Tel +49 345 7753 840

Editor

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

Address: Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

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

Abstract

It is well established that female students perform better when taught by female professors. However, little is known about the mechanisms explaining these gender match effects. Using administrative records from a German public university, which cover all programs and courses between 2006 and 2018, we show that gender match effects are sizable in smaller classes, but are absent in larger classes. These results suggest that direct and frequent interactions between students and professors are crucial for gender match effects to emerge. In contrast, the mere fact that one's professor is female is not sufficient to increase performance of female students.

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

JEL classification: I21, I23, I24, J16

* We thank Sabrina Eisenbarth, Alexander Giessing, Jörn-Steffen Pischke, Ludger Woessmann, Ulf Zölitz, seminar participants at Edinburgh, Hohenheim, Munich, and St. Gallen, 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.[1] 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 set fixed effects. Since grading in our setting is mostly anonymous, we do not have to be concerned about gender bias in grading. Moreover, because our data include a large number of *compulsory* courses with different sizes, we can account for potential ability-based sorting of students to professors. Finally, we can further address concerns of student sorting to courses by using information on students' high school GPA, which is a powerful measure of students' academic ability in Germany.

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] For instance, Carrell et al.'s (2010) seminal study demonstrated the effects of gender matching in higher education in classes with an average size of 19 students.

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 1.5 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 present along the whole grade distribution, pushing female students to obtain excellent rather than just good grades and reducing their probability of failure. 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, and to restricting the sample to lecturers that teach both small and large courses in order to account for potential sorting of female instructors to small versus large classes.

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), Agurto et al (2022)). Similarly, gender match effects at the student-professor level have been

[2]We use the class size median to determine "small" (at or below median) and "large" (above median) classes, but verify the robustness to changes in the large-class cutoff. We also show results by class size decile, finding that gender match effects also exist in medium-sized classes.

[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, using large-scale, standardized assessment data across 90 countries.

shown to influence major and course choices at university (Dynan and Rouse (1997), Rask and Bailey (2002), Bettinger 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. Such random assignment, however, rarely happens 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 larger 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 smaller and larger classes. Instead, our main result that gender match effects are only present in smaller classes points towards the importance of classroom interactions between students and professors in generating economically meaningful gender match effects. This is also corroborated by data from student evaluations at the university we study, according to which smaller classes have more frequent and intense classroom interactions.

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 set-

tings with smaller classes that facilitate student-professor interactions. However, in larger education programs at public universities or in massive open online courses with little interaction between students and professors, an increase in the share of female professors in STEM may not have equally positive 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.[4] 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.[5] However, the proportion of each of these components may vary among different majors. Students choose their major prior to enrollment. It is possible to change majors later on; however, this may prolong the duration of one's studies, as not all previously completed courses are necessarily credited in the new major.

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

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

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 very different grading scheme from other programs. This leaves us with 27 majors that cover STEM, 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.[6] 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.[7] 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

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

[7]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.

GPA of the high school leaving exam, which we use as control for student ability.[8] Previous 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).[9] 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. After applying these restrictions, 23,552 exam results were dropped, resulting in a final sample of 310,554 exam results from 18,592 distinct students.

Summary statistics are shown in Appendix Table A1. While our sample has 14% more female than male students, 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.

[8]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 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).

[9]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.

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 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 standardized exam grades. $FemaleLecturer$ 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. They are similar to lecturer fixed effects, but differentiate between situations where a lecturer teaches alone or with

different other lecturers.[10] 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.[11]

Including this demanding set of fixed effects allows us to address multiple possible confounders in the estimation of gender match effects. In particular, 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 gender of 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 4 are statistically significant at the 5% level, and all coefficients are economically small.[12] Most importantly, we do not observe any sorting of students based on ability as measured by

[10]This way, all courses taught only by lecturer A get a different lecturer set dummy from courses taught by lecturer A together with lecturer B.

[11]To estimate these models with many sets of fixed effects, we use the Stata command *reghdfe* (Correia 2014).

[12]Three of the four significant coefficients pertain to student age. In particular, we observe that 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.

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 (corresponding to 74 or fewer students), we find positive, sizable, and statistically significant effects.[13] 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.[14] 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 sometimes marginally significant coefficients of around 4–8% 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)

[13]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.

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

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.[15] Columns 3 and 4 correspond to the solid lines in Figure 1: They show that average female gender match effects are mostly driven by courses below the class size median, where we find an effect of 13.1% of a standard deviation. Above the median, the estimate is close to zero and statistically insignificant. We also report the p-value of a test for equality of the gender match effects for small and large courses, and can safely reject this hypothesis.

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 4.5 percentage points more likely to get an A, are 1.9 percentage points less likely to get a C, and are 1.5 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.[16] 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

[15] Consistent with our findings, the sample in Hoffmann and Oreopoulos (2009) has larger average class sizes compared to that in Carrell et al. (2010).

[16] Additionally, we also run unconditional quantile regressions for grade quantiles 10, 25, 50, 75 and 90 (see Appendix Table A5).

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.[17] Table 3 shows female gender 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 compulsory *de jure*, 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 potential issue in any study of gender match effects using end-of-course grades as a measure of performance (e.g., Hoffmann and Oreopoulos (2009), Carrell et al. (2010))

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

is gender-biased grading, i.e., female graders 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 renders 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.[18] In addition, exams are typically 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 students. 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. Moreover, while we have addressed the potential sorting of (high-ability) female students to courses taught by female instructors, another worry might be that (high-ability) female lecturers sort to small courses. However, we show that our results are robust to restricting the sample to lecturers who teach both large and small courses. We can also rule out that our results are driven by the fact that

[18]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.

female lecturers are typically younger than male lecturers, potentially affecting their style of teaching, by allowing lecturer age to have a differential effect by student gender.

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 than men at teaching women. 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. Appendix Table A8 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 than in an anonymous mass lecture. 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 such behavior is more pronounced in small classes, where there is more opportunity to ask questions and interact with the professors.

It is worth considering potential alternative explanations for why female gender match effects mainly exist in smaller classes, although we cannot provide direct evidence for them. One possibility is that room size and climate may play a role, with larger classes potentially being louder and more difficult to concentrate in. Additionally, attendance may be lower in larger classes, reducing opportunities for students to interact with their professors. Another potential explanation is that the importance of teaching assistants (vis-à-vis professors) may be higher in larger classes, so the gender (and teaching style) of the professor matters less. Lastly, it is possible that students simply learn less in larger classes, which could also account for the absence of gender match effects.

These potential mechanisms reinforce the main message of the paper, namely that the female gender match effects observed in the literature are unlikely due to a simple role model effect of "seeing is believing". If such role model effects were the only explanation, we would also expect to observe them in larger classrooms. However, the fact that gender match effects depend on classroom size suggests that the mechanisms underlying these effects may be more complex and context-dependent than previously thought.

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 smaller classes, implying performance gains of 13% of a standard deviation and a reduction in the probability of failing an exam by 1.5 percentage points if female students are taught by a female professor rather than a male 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 suggest that gender match effects 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] Agurto, Marcos, Muchin Bazan, Siddarth Hari, and Sudipta Sarangi, "To Inspire and to Inform: The Role of Role Models", Working Paper, 2022.

[3] 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.

[4] 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`.

[5] 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.

[6] 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).

[7] 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.

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

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

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

[11] 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.

[12] Correia, Sergio. "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects." *Statistical Software Components* S457874, Boston College Department of Economics (2014, revised 2019).

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

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

[15] 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

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] Rios-Avila, Fernando. "Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition", *Stata Journal* 20 (2020): 51-94.

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

[31] 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).

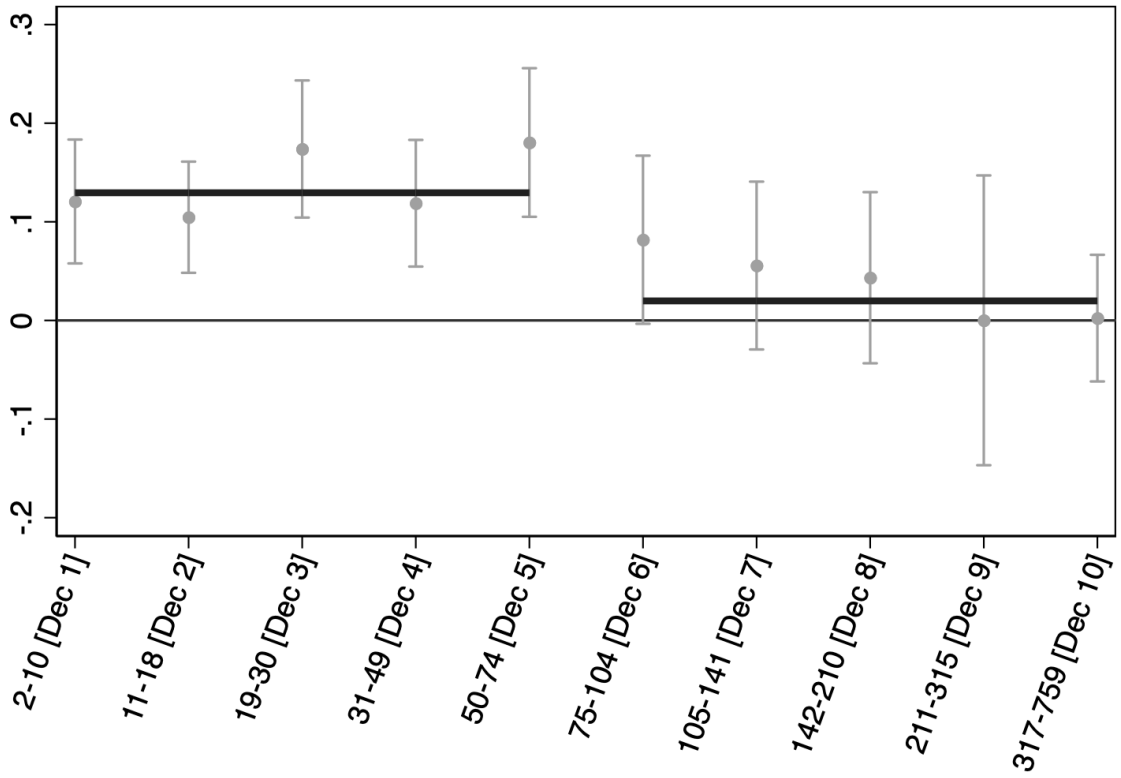
[32] Stout, Jane G., Nilanjana Dasgupta, Matthew Hunsinger, and Melissa A. McManus, "STEMing the Ide: Using Ingroup Experts to Inoculate Women's Self-Concept in Science, Technology, Engineering, and Mathematics (STEM)", *Journal of Personality and Social Psychology* 100 (2010): 255-270.

[33] 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	0.071***	0.073***	0.131***	0.018
× female student	(0.019)	(0.018)	(0.018)	(0.029)
Student characteristics				
High school GPA		0.435***	0.379***	0.494***
		(0.007)	(0.007)	(0.009)
Female student	0.035**	-0.060***	-0.069***	-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	0.376***	0.227***	0.205***	0.234***
	(0.031)	(0.027)	(0.032)	(0.036)
Local student	-0.138***	-0.128***	-0.112***	-0.143***
	(0.017)	(0.013)	(0.015)	(0.018)
First year in major	0.016**	0.014**	0.005	0.028***
	(0.007)	(0.006)	(0.006)	(0.010)
Class size	All	All	Small	Large
p-value large = small			<.001	
Observations	310,554	310,554	155,591	154,960

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 74 or fewer students, large courses have 75 or more students. Students' migration background is based on citizenship. Standard errors, twoway-clustered at the student and course level, in parentheses. In the bottom of the table, we report the p-value of a test for the equality of the gender match effects in small and large classes, based on a model that includes interactions between all variables and a dummy for small courses. 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	0.045***	-0.006	-0.019***	-0.005	-0.015***
× female student	(0.007)	(0.007)	(0.005)	(0.003)	(0.004)
Mean dependent variable	0.286	0.380	0.179	0.066	0.089
Observations	155,591	155,591	155,591	155,591	155,591
Panel B: Large Classes					
Female lecturer	0.002	0.005	-0.006	-0.002	0.001
× female student	(0.006)	(0.010)	(0.006)	(0.006)	(0.007)
Mean dependent variable	0.134	0.293	0.267	0.136	0.169
Observations	154,960	154,960	154,960	154,960	154,960

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 74 or fewer students; in Panel B, sample is restricted to classes with 75 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.

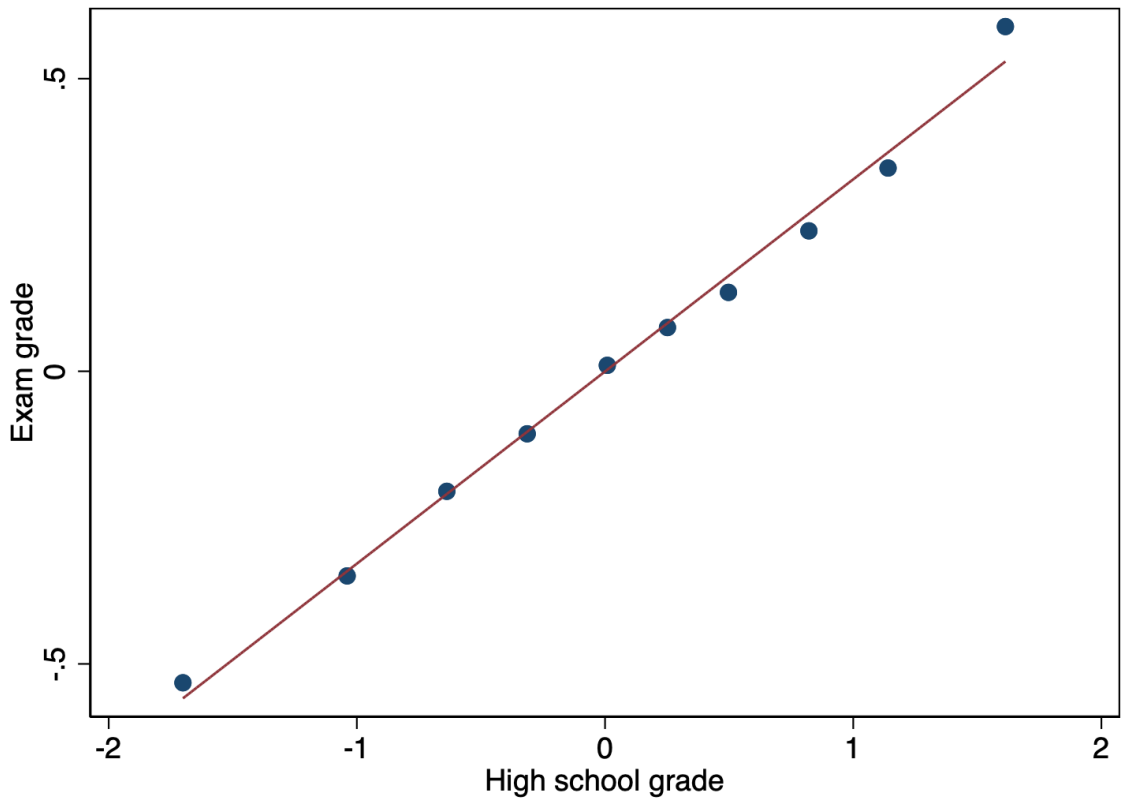
Table 3: Female gender match effects by class size: Compulsory vs. elective courses

	(1)	(2)	(3)	(4)	(5)	(6)
Female lecturer	0.118***	0.132***	0.029	0.002	0.139***	0.010
× female student	(0.041)	(0.019)	(0.028)	(0.048)	(0.038)	(0.033)
Course type	Comp.	Elect.	Comp.	Elect.	First 2 semesters	
Class size	Small	Small	Large	Large	Small	Large
Observations	22,444	133,114	101,224	53,713	36,474	90,964

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 74 or fewer students, large courses have 75 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	0.272	0	0.445	0.321	0	0.467	0.213	0	0.410
Exam grade	0	0.149	0.976	0.032	0.185	0.946	-0.038	0.099	1.010
Failed exam	0.129	0	0.335	0.103	0	0.303	0.160	0	0.367
Class size	121.1	74	131.1	111.0	70	122.3	133.1	80	140.0
Observations		310,554			168,788			141,766	
Panel B: Student-level variables									
Female student	0.532	1	0.499						
HS GPA	-0.149	-0.073	1.017	-0.049	-0.073	1.002	-0.262	-0.236	1.022
Age student	21.377	20.933	2.879	21.264	20.780	2.975	21.505	21	2.759
Native student	0.967	1	0.180	0.964	1	0.185	0.969	1	0.173
Local student	0.127	0	0.334	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	3.8
Observations		18,592			9,890			8,702	

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	0.006	0.000	-0.033	0.007
	(0.012)	(0.019)	(0.011)	(0.015)	(0.047)	(0.015)
Age student	-0.104**	-0.176**	0.007	-0.047	-0.355**	0.004
	(0.050)	(0.072)	(0.043)	(0.061)	(0.168)	(0.055)
Native student	-0.001	-0.001	-0.002	-0.001	0.015**	-0.003
	(0.002)	(0.004)	(0.003)	(0.003)	(0.008)	(0.004)
Local student	0.004	0.008	0.000	0.002	0.006	0.000
	(0.005)	(0.007)	(0.004)	(0.006)	(0.016)	(0.005)
First year in major	0.002	0.021	-0.017	0.007	0.021	0.005
	(0.013)	(0.016)	(0.017)	(0.010)	(0.035)	(0.010)
Observations	310,554	155,591	154,960	123,673	22,444	101,224

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 74 or fewer students, large courses have 75 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	0.000	-0.000	0.000	-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	310,554	138,233	169,940	310,554	310,554

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 74 students and below (above 74 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.

One worry regarding the marked difference in the gender match effects for small and large courses is that the two types of courses may be taught by systematically different lecturers. To alleviate this concern, we restrict the sample to courses where at least one of the lecturers taught at least one large and one small course over our period of observation. The size of the interaction term, shown in panel C of Table A4, is somewhat smaller than in the baseline, but the results show the same pattern: Positive and significant female gender match effects for small courses, none for larger ones. Panel D further explores the role of teacher characteristics by considering teacher age. Given that female lecturers are on average younger than male ones, our estimated gender match effects could also be

due to female students reacting differentially to younger lecturers. We therefore interact teacher age with student gender and include this as an additional control.[19] Our results are not affected by adding this control, indicating that our gender match effects are not simply age effects.

In Panel E of Table A4, we account for variation in school quality in Germany over time and across space by interacting the student high school GPA with indicators of the location of the high school, graduation year, and broad types of high school leaving exam.[20] Again, our results remain essentially unchanged.

In a similar vein, Panel F 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 G, we therefore include major-by-semester fixed effects. Results are virtually identical to our baseline findings.

One possibility for students to differentially select into courses taught by men or women might be by postponing courses. In Panel H of Table A4, we thus define the usual study semester in which a given course is taken by calculating the modal study semester in which students in a given major take this exam. Based on this, we create a dummy variable indicating whether students take the course in the usual study semester (or earlier). Including this indicator as an additional control variable does not affect our

[19]When there are two lecturers, we take the average of their ages. We lose observations since our dataset does not provide birth year information for every lecturer.

[20]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.

results. The same holds for Panel I, 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 J 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", the median is 8.

Finally, in Panel K, we check whether our results depend on the standardization of exam grades at the exam-semester level. 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***	0.132***	0.029
× female student	(0.019)	(0.018)	(0.034)
Observations	279,930	142,550	137,377
Panel B: Alternative treatment definition: Both lectures female			
Female lecturer	0.077***	0.129***	0.024
× female student	(0.019)	(0.018)	(0.033)
Observations	310,554	155,591	154,960
Panel C: Only lecturers that teach both large and small courses			
Female lecturer	0.054**	0.112***	0.033
× female student	(0.026)	(0.029)	(0.031)
Observations	207,603	64,339	143,261
Panel D: Controlling for lecturer age × student gender			
Female lecturer	0.087***	0.135***	0.035
× female student	(0.019)	(0.019)	(0.028)
Observations	266,039	129,483	136,556
Panel E: Additional high school controls			
Female lecturer	0.065***	0.117***	0.014
× female student	(0.017)	(0.017)	(0.027)
Observations	310,465	155,359	154,737
Panel F: Controlling for student fixed effects			
Female lecturer	0.038**	0.071***	0.002
× female student	(0.016)	(0.016)	(0.025)
Observations	310,095	154,393	153,700
Panel G: Controlling for major × semester fixed effects			
Female lecturer	0.073***	0.132***	0.018
× female student	(0.018)	(0.018)	(0.028)
Observations	310,550	155,587	154,922
Panel H: Controlling for taking a course at the usual time			
Female lecturer	0.074***	0.133***	0.019
× female student	(0.019)	(0.018)	(0.029)
Observations	299,616	145,638	153,975
Panel I: Dropping students who study less than 3 semesters in major			
Female lecturer	0.079***	0.129***	0.025
× female student	(0.019)	(0.018)	(0.030)
Observations	286,683	150,429	136,249
Panel J: Using major-specific medians to define large courses			
Female lecturer	0.073***	0.141***	0.063***
× female student	(0.018)	(0.027)	(0.021)
Observations	310,554	53,511	256,746
Panel K: Raw exam grades			
Female lecturer	-0.059***	-0.114***	-0.012
× female student	(0.019)	(0.018)	(0.030)
Observations	310,554	155,591	154,960

Notes: Dependent variable: Exam grades, standardized to mean 0 and std. dev. 1 at the exam-semester level (raw exam grades in Panel K). 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 F, where these get captured by the student fixed effects) and for program, course, semester, and lecturer set fixed effects. With the exception of Panel J, small courses have 74 or fewer students, large courses have 75 or more students. In Panel E, 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 2, we explored the size of gender match effect for different grade categories, finding effects along the whole distribution. In Table A5, we complement this analysis by running unconditional quantile regressions for the sample of small courses.[21] We find positive and significant gender match effects at the 10th, 25th, 50th, 75th, and 90th percentile of the grade distribution. The effects are most pronounced at the top and median of the grade distribution.

Table A5: Quantile regression results for small classes

	(1)	(2)	(3)	(4)	(5)
Female lecturer	0.128***	0.172***	0.168***	0.084***	0.047***
× female student	(0.026)	(0.021)	(0.019)	(0.011)	(0.015)
Percentile	10th	25th	50th	75th	90th

Notes: 155,591 observations. 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 74 or fewer students, large courses have 75 or more students. Standard errors based on a bootstrap with 100 repetitions. Significance levels: * $p<.10$, ** $p<.05$, *** $p<.01$. *Data source:* Administrative student records.

In Table A6, 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.[22] 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

[21] To implement this, we use the *rifhdreg* command for Stata (Rios-Avila 2020). Quantiles are over the within-course standardized grade.

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

female gender match effect in large classes (Column 5), just shy of statistical significance at conventional levels (p=0.15).

Table A6: Female gender match effects by class size and broad field

	(1)	(2)	(3)	(4)	(5)	(6)
Female lecturer	0.078**	0.181***	0.118***	-0.005	0.046	-0.017
× female student	(0.030)	(0.036)	(0.026)	(0.041)	(0.032)	(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	0.220	0.404	0.232	0.174	0.280
Observations	27,050	55,765	74,243	77,446	64,589	17,248

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 74 or fewer students, large courses have 75 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 A7, using students' high school GPA as a measure of academic ability. Columns 1 and 2 of Table A7 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 (14.8% of a standard deviation, compared to 12.4% of a standard deviation for lower-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 A7: Female gender match effects by class size and high school GPA

	(1)	(2)	(3)	(4)
Female lecturer	0.148***	0.058*	0.124***	-0.006
× female student	(0.027)	(0.031)	(0.023)	(0.032)
Class size	Small	Large	Small	Large
Student high school GPA	Above median		Below median	
Observations	65,198	72,974	87,921	81,977

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 74 or fewer students, large courses have 75 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 A8: 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 74 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.

Halle Institute for Economic Research –
Member of the Leibniz Association

Kleine Maerkerstrasse 8

D-06108 Halle (Saale), Germany

Postal Adress: P.O. Box 11 03 61

D-06017 Halle (Saale), Germany

Tel +49 345 7753 60

Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188