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

DISCUSSION PAPER SERIES

IZA DP No. 12491

Students are Almost as Effective as Professors in University Teaching

Jan Feld
Nicolás Salamanca
Ulf Zölitz

JULY 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12491

Students are Almost as Effective as Professors in University Teaching

Jan Feld
Victoria University of Wellington and IZA

Nicolás Salamanca
Melbourne Institute and IZA

Ulf Zölitz
University of Zurich and IZA

JULY 2019

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

Students are Almost as Effective as Professors in University Teaching*

In a previous paper, we have shown that academic rank is largely unrelated to tutorial teaching effectiveness. In this paper, we further explore the effectiveness of the lowest-ranked instructors: students. We confirm that students are almost as effective as senior instructors, and we produce results informative on the effects of expanding the use of student instructors. We conclude that hiring moderately more student instructors would not harm students, but exclusively using them will likely negatively affect student outcomes. Given how inexpensive student instructors are, however, such a policy might still be worth it.

JEL Classification: I21, I24, J24
Keywords: student instructors, university, teacher performance

Corresponding author:
Ulf Zölitz
Department of Economics
University of Zurich
Schönberggasse 1
8001 Zurich
Switzerland
E-mail: ulf.zoelitz@uzh.ch

* We would like to thank Luke Chu, Alexandra de Gendre, Ingo Isphording, Julie Moschion, Cain Polidano, Chris Ryan, Derek Stemple, Barbara Wolfe and Jeffrey Yusof for helpful comments. We thank Sophia Wagner for providing outstanding research assistance. This research was partly supported by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project number CE140100027). The Centre is administered by the Institute for Social Science Research at The University of Queensland, with nodes at The University of Western Australia, The University of Melbourne, and The University of Sydney.

> Comedian John Mulaney's reaction to finding out that his babysitter was only three years older than him: *"So why was she in charge? All she could do was dial the telephone a little better than I could . . . [T]hat would be like if you're going out of town for the week and you paid a horse to watch your dog."*

1 Introduction

Many universities offer courses that include tutorials (also called TA sessions or exercise sessions), which are small group sessions that complement course lectures. These sessions are staffed quite differently in different universities: while US institutions typically rely on PhD students, institutions in other OECD countries often use professors or a mixture of professors and students for tutorial teaching (Feld, Salamanca, & Zölitz, in press). The advantage of using student instructors, that is, bachelor's and master's students who teach tutorials, is that they are cheaper than more-senior instructors.[1] However, student instructors are also less experienced and much less qualified than professors. It is therefore important to ask whether student instructors can provide the same quality of education, or whether—in the words of the above quoted John Mulaney—they are like horses watching over dogs.

In this paper, we investigate how student instructors affect their students' academic performance and labor market outcomes. We use data from a Dutch business school with two key features. First, each course has several tutorial sections, some of which are taught by students and others of which are taught by more senior instructors. This gives us the necessary variation in instructor type to answer our research question. Second, systematic assignment of students to instructors is not a concern in this business school; students assigned to student instructors and

[1] The median wage of all types of postsecondary teachers in the US is $76,000, while median wage of graduate teaching assistants is only $ 37,720 (Bureau of Labor Statistics, 2017b, 2017a). In our setting, the hourly wage rate of student instructors is less than half of that of assistant professors and less than a third of that of full professors.

students assigned to senior instructors are very similar in a host of characteristics, including academic ability. This allows us to estimate differences in effectiveness between both types of instructors without worrying about selection bias.

This paper builds on our previous paper (Feld, et al., in press), which uses data from the same environment to compare the teaching effectiveness of tutorial instructors with various academic ranks, including student instructors but also PhD students, postdocs, lecturers, and assistant, associate, and full professors. In that paper, we find no meaningful differences between instructors of different academic ranks in how they affect students' grades, affect grades in subsequent related courses, or affect students' earnings or job satisfaction after graduation. We only find that higher-ranked instructors receive better course evaluations, but even this difference is fairly small. These findings led us to conclude that universities should rely more on inexpensive student instructors for tutorial teaching.

In this paper, we focus on the role of student instructors to make three additional contributions. First, we estimate the effect of instructor type using a different methodology. In our previous paper, we use Chetty, Friedman, and Rockoff's (2014) value-added measures, which force us to restrict our sample to instructors who teach the same course at least twice. This restriction could introduce selection bias in our estimates if worse instructors end up teaching only once. In this paper, we estimate the effect of being assigned a student instructor by regressing individual student outcomes, as opposed to instructor value-added measures, on instructor type. This approach allows us to include all instructors, even those who taught only once, which shields our estimates from—and allows us to quantify—this bias. Second, we look at a broader set of student outcomes and explore more thoroughly whether student instructors affect students in

particular courses or particular kinds of students. Finally, we explore how hiring more and hiring *only* student instructors would affect student outcomes.

As in Feld et al. (in press), we find that students taught by student instructors receive only marginally worse course grades and no worse grades in subsequent related courses. The effect on grades is driven by larger, yet small, negative effects of student instructors in non-mathematical courses, in non-first-year courses, and on low-ability students. We find slightly larger negative effects of student instructors on course evaluations than in our previous paper. This difference is partly driven by worse-rated student instructors being more likely to teach only once, illustrating that selection bias is a valid concern when using value-added measures. We again find no significant effect on earnings and job satisfaction, nor do we find any differences in length of job search after graduation or retrospective study satisfaction. Our results further dispel several other concerns about student instructors: they do not affect students' chances of failing a course, nor do they make students study longer hours outside the classroom. Overall, we confirm that student instructors are almost as effective as more-senior instructors in tutorial teaching.

We then explore the potential consequences of hiring more student instructors. One concern is that hiring more students for tutorial teaching would mean hiring lower-quality ones, which would naturally occur if the business school recruits the best available instructors from a limited pool of applicants. We explore this concern by testing whether student instructors perform worse in the teaching terms in which many of them are hired. This is not the case. Another concern is that increasing the number of student instructors would expose students to too many student instructors throughout their studies. Although having a single student instructor may not be harmful, the effect of having multiple instructors throughout one's studies may add up. We do find some suggestive evidence that this is the case, though only for students exposed to seven or more

student instructors (3 percent of our estimation sample). Together, our findings suggest that moderately increasing student instructors would not harm students in any meaningful way. We finish by discussing the potential effects of an extreme policy in which the business school would *only* use student instructors for tutorial teaching. Such a policy would likely have more negative effects on students than our estimates suggest. However, given the large savings potential, both in time and in money, of relying on cheaper instructors, such a policy might still be worth it.

Our paper relates to a few studies on how the origin and ethnicity of graduate teaching assistants (TAs) affect student performance. Lusher, Campbell, and Carrell (2018) find that students' grades increase when they are assigned to same-ethnicity graduate TAs. Borjas (2000) and Fleisher, Hishimoto and Weinberg (2002) study the effect of foreign-born graduate TAs, and reach opposing conclusions; Borjas (2000) finds that foreign-born TAs negatively affect student grades, whereas Fleisher et al. (2002) find that they have negligible, or sometimes positive, effects on student grades. Bettinger, Long and Taylor (2016) find that students are more likely to major in a subject if their first courses in that subject are taught by a PhD student. None of these studies, however, compares the effectiveness of student and non-student instructors.

Another related strand of literature looks at the effect of instructor characteristics on student outcomes at the university level. Bettinger and Long (2010) and Figlio, Shapiro and Soter (2015) find a positive effect of adjunct instructors, compared to tenure track and tenured instructors, on student performance. Hoffmann and Oreopoulus (2009) find that objective instructor characteristics, such as academic rank and salary, do not predict student performance, yet students' evaluations of their teachers are positively correlated with student performance. De Vlieger, Jacob and Stange (2018) find that instructor performance in a college algebra course at a large for-profit university grows modestly with course-specific teaching experience, but is unrelated to pay.

Fairlie, Hoffmann and Oreopoulos (2014) find that minority students benefit from minority instructors. None of these studies explicitly consider student instructors despite their importance in higher education. Our paper fills this gap.

2 Institutional Background and Data

2.1 Institutional Environment

To estimate the effect of student instructors on student outcomes, we use data from a Dutch business school from the academic years 2009–10 through 2014–15.[2] The bulk of teaching at this institution is done in four regular teaching terms of eight weeks each, in which students typically take two courses simultaneously. We use 'course' throughout to refer to a subject-year-term combination. For example, we consider Microeconomics in term 1 of 2011 and Microeconomics in term 1 of 2012 as two separate courses. Over the entire eight-week teaching term, students typically receive three to seven lectures for each course. The bulk of the teaching, however, happens over twelve two-hour tutorials. These tutorials are at the center of our analysis.

Tutorials are organized in groups of up to 16 students who are assigned to one instructor—either a student instructor or a more-senior one. In these tutorials, the instructor, who is always present, guides students' discussions of course material and exercise solutions. While instruction styles differ between universities, discussing course material and exercise solutions in tutorials is done in many institutions at the undergraduate and graduate level. This business school takes tutorials very seriously: attendance is compulsory, recorded, and often graded by the instructor, and nonattendance can easily result in failing the course. Business school guidelines explicitly prohibit switching between assigned tutorial groups. Within a course, tutorials are also quite

[2] For more detailed information on the institutional environment see Feld and Zölitz (2017) and Feld et al. (in press).

homogeneous in that they use identical course material, have the same assigned readings and exercise questions, and follow the same course plan.

In many courses, a mixture of student instructors and more- senior instructors teach tutorial groups. This within-course variation in instructor type identifies our estimates of the effectiveness of student instructors compared to senior instructors, which include postdocs, lecturers, assistant professors, associate professors, and full professors. We treat PhD instructors as a separate sub-group, but since they are not the focus of this paper, we do not explicitly report their estimates.

Student instructors are typically recruited by an education manager and approved by the course coordinators. The most important characteristics in the recruitment process are the students' grades, previous experience with the course, and a sufficient command of English, which is the language of instruction for all courses. Student instructors stand out due to their low academic rank and because they lack teaching experience. In the six-year period covered by our data, student instructors taught an average of 2.4 courses, compared to the average of 4.5 and 7.9 courses taught by PhD students and senior instructors.

It is by far cheaper for the business school to employ student instructors than any other staff type. Student instructors' hourly wages are only €14, compared to €31 for assistant professors and €47 for full professors. These differences in salary ignore overhead costs, which are also lower for student instructors. The search and hiring costs of student instructors are also close to zero. They can easily be recruited from the constantly renewing pool of students taking each course, and they are offered standard short-term contracts, usually for a single teaching term. Thus, student instructors are elastic, convenient, and low-cost labor for the university.

Table 1. Characteristics of all courses and courses that use at least one student and one non-student instructor (sample courses)

	Courses:		
	All (N=1,323)	Sample (N=206)	Difference
	(1)	(2)	(3)
Student instructor	0.10	0.42	0.32
PhD student instructor	0.21	0.20	-0.01
Senior instructor	0.69	0.37	-0.32
Bachelor course	0.53	0.83	0.30
Mathematical course	0.34	0.29	-0.05
First-year course	0.13	0.35	0.22
Student GPA	6.42	6.17	-0.25
Student grade	7.01	6.70	-0.31
No. students	78.01	218.74	140.73
No. instructors	2.42	5.82	3.40
No. tutorials	6.13	16.23	10.10
No. students per tutorial	11.64	13.21	1.57

This table shows summary statistics at the course level based on the complete data set of 1,323 courses (column 1) compared to the sample to courses that staff some (but not all) of their tutorials with student instructors (column 2).

2.2 Data and Summary Statistics

Our core data set has information on 103,664 student-course observations. However, we limit our estimation sample to courses with one student instructor and one non-student instructor teaching tutorials. This is our main sample restriction. We also exclude some data from our analyses because of nonstandard assignment of students to tutorial groups and impute missing values for some of our control variables. Appendices 1 and 2 describe our sample restrictions and data imputations in detail.

The sample restrictions above make sample courses—for which our estimates describe a local average treatment effect—differ from the average course at the business school. We show these differences in Table 1. The average course in our estimation sample uses 42 percent of student instructors compared to only 10 percent for the average course offered at the business school. The business school disproportionately uses student instructors in large bachelor-level

courses, which likely reflects the larger need for teaching staff in these courses. Student instructors are also used more in courses with lower-achieving students and in which students get lower final grades, yet these differences are not very large. These differences should be considered when interpreting our results throughout the paper.

In Table 2, we show summary statistics aggregated at the instructor level (Panel A) and the student level (Panel B) for our estimation sample. In total, we have 485 instructors in our estimation sample, of which 49 percent are students, 26 percent are PhD students and 25 percent are senior instructors. Among senior instructors, lecturers are the largest group (12 percent of all instructors), followed by assistant professors (6 percent), associate professors (2 percent), postdocs (2 percent), and full professors (1 percent). Instructors' nationalities are quite diverse, with the largest nationalities being German (43 percent) and Dutch (29 percent). Thirty-eight percent of instructors are women. Each instructor teaches, on average, five courses, 13 tutorial groups and 170 different students in our sample period. Our estimation sample contains 9,323 students who are on average 21 years old. Their most common nationalities are German (41 percent) and Dutch (26 percent), and 39 percent of them are women. Seventy-seven percent of our sample are bachelor's students and 13 percent are exchange students. We observe the average student in nine distinct courses.

The student outcome data in our main estimation is based on 45,060 course enrollments. Out of these course enrollments, 7.6 percent of students dropped out of the course, so we ultimately observe 41,619 final course grades, which is our main academic outcome. Final course grades usually consist of multiple graded components, with the highest weight typically placed on the final exam. Some of the components of the final grade, such as group work or tutorial participation and sometimes even parts of the final exam, are directly graded by the students' own instructor. In

our data, we only observe final grades. Differences in grading standards between instructor types could therefore partially drive our estimated grade differences between instructors, a concern we discuss in Section 4.2. The Dutch grading scale ranges from 1 to 10, with 5.5 as the lowest passing grade. In Figure 1 we show that the distribution of the course final grades covers the entire range of possible grades, giving us plenty of variation in our main measure of academic performance. Throughout our analyses, we account for differences in student ability using students' GPA, constructed as the average of all grades prior to the current course, weighted by course credit points.

Table 2. Summary statistics for instructors and students

Panel A	Instructors (N = 485)			
	Mean	SD	Min	Max
Student instructor	0.49	0.49	0	1
PhD student instructor	0.26	0.42	0	1
Senior instructor	0.25	0.42	0	1
Dutch	0.29	0.45	0	1
German	0.43	0.5	0	1
Belgian	0.02	0.15	0	1
Other nationality	0.24	0.43	0	1
Female	0.38	0.49	0	1
No. of courses taught	4.88	5.93	1	46
No. of tutorial groups taught	13.12	16.55	1	120
No. of students taught	170.34	215.28	9	1,586
Panel B	**Students (N = 9,323)**			
	Mean	SD	Min	Max
Age	20.72	2.25	16.34	41
Dutch	0.26	0.44	0	1
German	0.41	0.49	0	1
Belgian	0.05	0.21	0	1
Other nationality	0.14	0.35	0	1
Female	0.39	0.49	0	1
Bachelor	0.77	0.37	0	1
Exchange	0.13	0.33	0	1
No. of courses taken	9.21	6.57	1	33

This table shows summary statistics at the instructor (Panel A) and student (Panel B) level based on our main estimation sample. Nationalities of students and instructors do not add up to one due to missing nationality information.

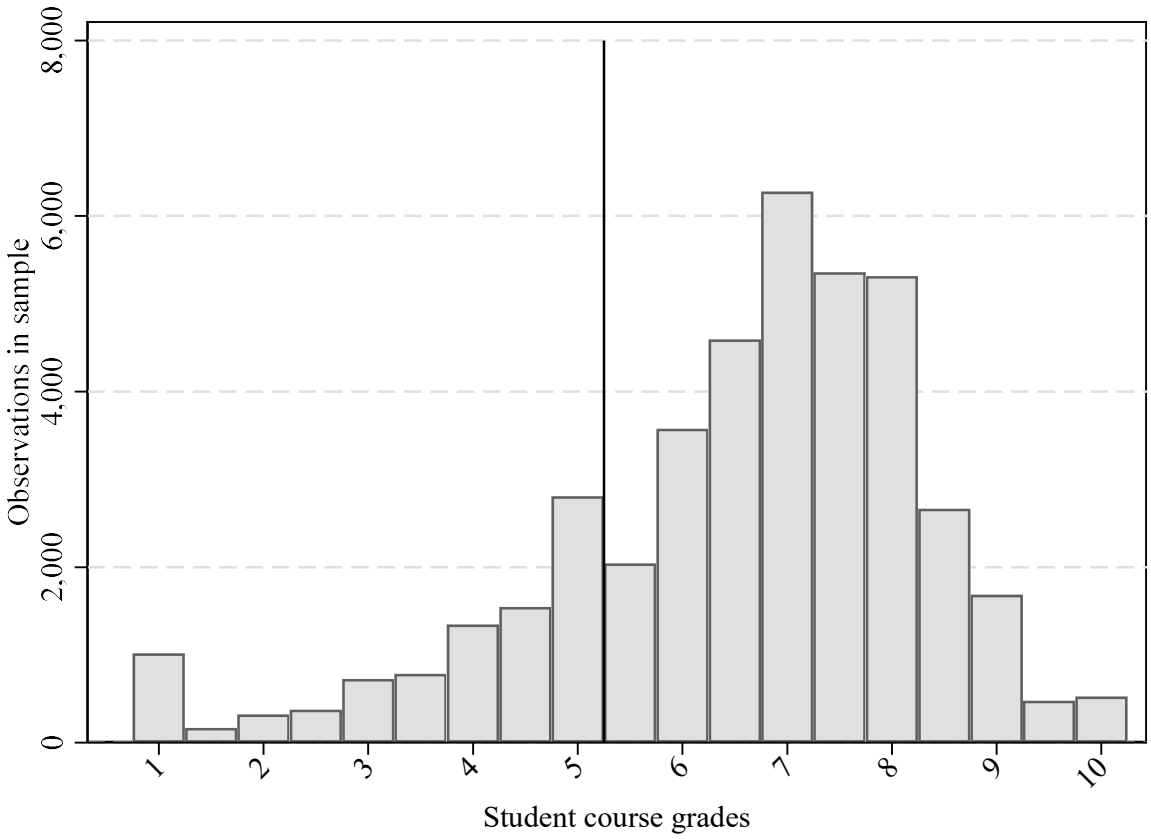

Fig. 1. Distribution of course grades

This figure is based on our estimation sample. The vertical line at 5.5 shows the lowest possible passing grade. This grade distribution has a mean of 6.51 and a standard deviation of 1.82.

2.3 Assignment of Instructors and Students to Tutorial Groups

Schedulers randomly assign students who enroll in a course to tutorial groups, considering students' scheduling conflicts. The main scheduling conflict comes from students' enrollment in different elective courses. Lectures and tutorials in core courses are usually scheduled in a way that avoids conflicts.[3] For example, first-year compulsory courses that students take in parallel are

[3] A small number of students have other scheduling conflicts because they take language courses, work as student instructors, have regular medical appointments, or are top athletes and need to accommodate inflexible training schedules. One exception from the random assignment process is that before the fall of 2015, students could opt out of participating in tutorials that started at 6:30 p.m. Students in those evening tutorials represent only 3.2 percent of our observations.

scheduled on different days. If scheduling conflicts occur, the schedulers manually reallocate students to different tutorial groups. From 2010–11 onward, the schedulers balanced tutorial groups by nationality for bachelor's students (making sure that the proportion of German, Dutch, and other nationality students were the same across tutorial groups in each course), but otherwise, the assignment remained random. Previous work with data from this business school has shown that tutorial group composition looks as we would expect under conditional random assignment of students to tutorials (see Feld & Zölitz, 2017, and Zölitz & Feld, 2018).

After assigning students, schedulers allocate instructors to tutorial groups. Instructors are typically allocated into consecutive time slots, and in each term about 10 percent of instructors indicate some slots in which they are not available for teaching. However, this happens prior to any scheduling of students or other instructors and requires approval from the department chair.

Random assignment of students to tutorial groups and unsystematic assignment of instructors to tutorials imply that instructor characteristics are, on average, unrelated to observable and unobservable student characteristics. We test this implication by estimating if, in our sample, instructor type is related to six 'preassignment' student characteristics: previous GPA, gender, age, the rank of the student ID (a proxy for tenure at the business school), tutorial group size, and exchange student status. We first regress each of these preassignment characteristics on student instructor and PhD student instructor dummies (keeping senior instructors as the base group), as well as fixed effects for time-of-the-day and day-of-the-week and fixed effects for all course-by-parallel-course combinations as controls. The course combination fixed effects account for any systematic student and instructor sorting into courses and any nonrandom assignment of students to tutorial groups driven by their course schedule.

In Figure 2, we show the results of these balancing tests (see also Table A1 in the appendix). Most important for our study, we see that student instructors are not assigned to students with systematically higher or lower ability. Our estimates show that students taught by a student instructor have only an insignificant 0.2 percent of a standard deviation lower GPA than students taught by a senior instructor. We also see no economically or statistically significant differences in the age, student ID number, tutorial size, or share of exchange students. We do, however, see that student instructors teach fewer female students. Although the difference is tiny (1.5 percentage points or 0.2 of a woman in a typical tutorial group of 15 students), it is statistically significant.

To check whether this gender imbalance is due to chance, we follow Pei, Pischke and Schwandt (2018) and jointly test whether being assigned to a student instructor compared to a PhD student or senior instructor is significantly related to *all six* preassignment characteristics. We implement their "right-hand side" version of their balancing test with multiple covariates (p. 20) by regressing a dummy for student instructor on all six preassignment characteristics and testing the joint significance of their coefficient estimates. Rejecting the null hypothesis of this F-test would be evidence that student instructors systematically differ from other instructors in at least one of these characteristics. The p-value of the F-test is 0.0962, providing only marginally significant evidence against balancedness.[4] The weakness of this evidence, along with the small size of the characteristic unbalance reassures us that systematic assignment of students to instructor is not a concern in our setting. As further reassurance, in Section 4.1 we confirm that the tiny

[4] This implementation of the Pei et al. (2018) test implicitly pools PhD student instructors and senior instructors in the base group. To allow for this difference between base groups while also accounting for their interdependence in the balancing test, we implement the right-hand side balancing test twice more using PhD instructor and senior instructor dummies as dependent variables. We then use Fisher's method to combine the p-values of all three tests and account for their positive dependence by using the Benjamini–Hochberg adjustment to produce a comparison significance level (Benjamini & Hochberg, 1995; Fisher, 1992). This more involved test cannot reject the null of balancedness at conventional significance levels.

gender imbalance we observe is unlikely to introduce any bias in our estimates. And, as an extra precaution, we include an extensive list of student- and tutorial-level controls in our analyses.

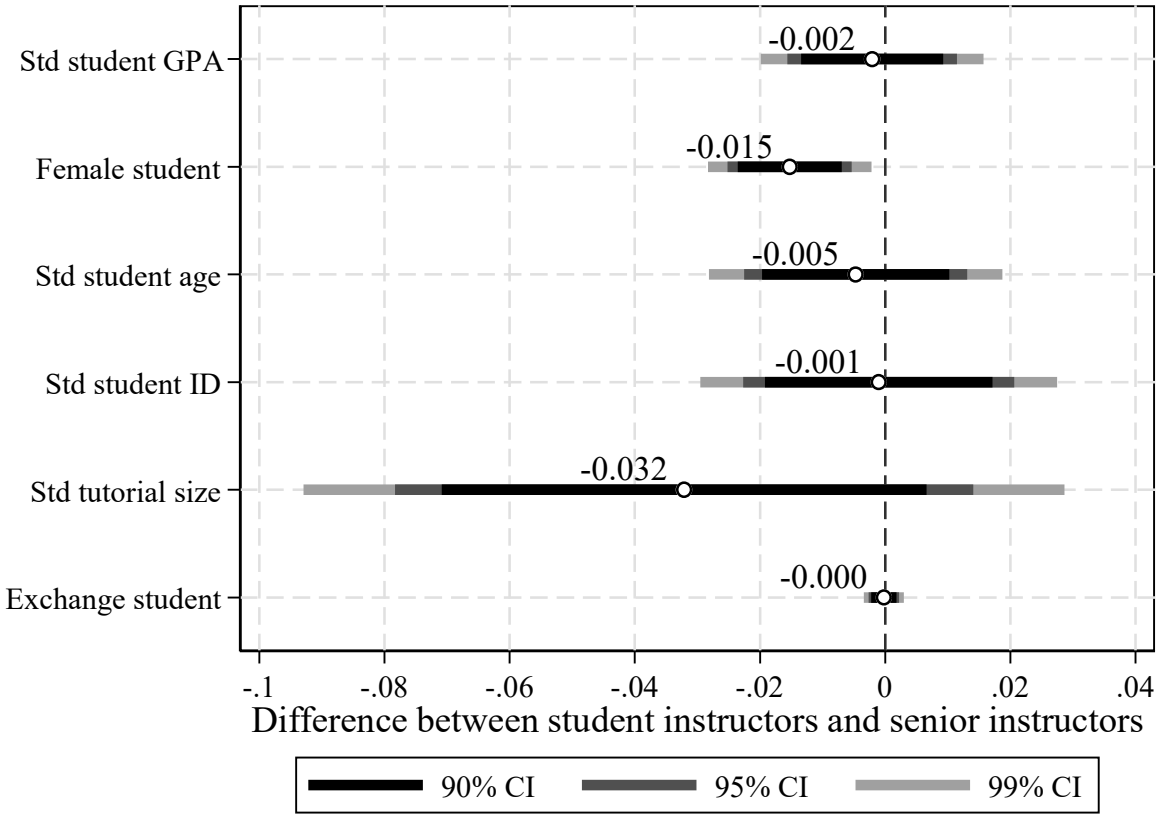

Fig. 2. Balancing test

This figure is based on regression estimates shown in Table A1. Confidence intervals are based on standard errors clustered at the instructor level.

3 Empirical Strategy

We estimate the effects of student instructors on student outcomes via variations of the following model

$$y_{ic} = \beta\ \text{student instructor}_{ic} + \delta \text{PhD}_{ic} + \boldsymbol{\gamma}'\mathbf{X_{ic}} + \varepsilon_{ic}, \quad (1)$$

where y_{ic} is the outcome of student *i* in course *c*, and the main regressor of interest is $\text{student instructor}_{ic}$, an indicator of whether student *i* in course *c* is taught by a student instructor.

We control for PhD_{ic}, an indicator of whether the instructor is a PhD student instructor, which leaves academically more-senior instructors (postdocs, lecturers, and assistant, associate, and full professors) as the base group. As in our balancing tests, in all specifications the vector $\mathbf{X_{ic}}$ includes fixed effects for all unique course-by-parallel-course combinations to account for any systematic student and instructor sorting into courses and any nonrandom assignment of students to tutorial groups driven by their course schedule. $\mathbf{X_{ic}}$ also includes scheduling controls, student-level controls and tutorial-level controls. *Scheduling controls* are exhaustive day-of-week and time-of-day dummies for each tutorial meeting. S*tudent-level controls* are dummies for student gender, student nationality, whether a student registered late for the course, whether the student is repeating the course, whether the student is part of the honors program, whether the student is part of a special business school-wide research program, whether the student is on exchange from another university, the student's study program (e.g. economics or business), and cubic polynomials in students' age and students' GPA before taking the course. *Tutorial-level controls* are the tutorial means of all student-level controls. Finally, ε_{ic} is an idiosyncratic error term, which we assume to be uncorrelated with all the regressors. We cluster the standard errors at the instructor level.

When estimating Equation (1) we use student-level outcomes as dependent variables in contrasts to the instructor-level value-added measures that we used in Feld et al. (in press). These value-added measures have the advantages of being free of contemporaneous model error and being able to incorporate estimation noise to produce the best linear predictions of the effect of individual instructors on future student outcomes. In Feld et al. (in press), we use these value-added measures to show that there are individual differences between instructors in teaching effectiveness. Value-added measures, however, have one key drawback for identifying the effect of different types of instructors: they can only be constructed for instructors who we observe

teaching the same course at least twice. This implicit sample restriction can create selection bias if, for example, low-quality student instructors only teach once. This source of selection bias is not a concern in this paper because we regress grades (instead of grade value-added measures) on student instructor dummies, which allows us to include information from once-only instructors.

We also estimate fully interacted versions of Equation (1) to measure heterogenous effects of student instructors by course type, student type, or teaching term. For example, when estimating whether the effect of student instructors differs between mathematical and non-mathematical courses, we interact all explanatory variables except for the course-combination fixed effects with a mathematical course dummy.

When estimating the effect of instructors on outcomes that we measure through the course evaluation survey or the graduate survey, we only observe some of the students in our estimation sample: around 36 percent of our estimation sample completed the course evaluation survey, and about 32 percent of those who were eligible to receive the graduate survey (those who obtained their bachelor's degree at the business school between September 2010 and September 2015) responded. There are some differences in the kinds of students answering each question. We see, for example, that high-GPA students are more likely to answer both surveys, and we also see some heterogeneity by student nationality. Importantly, we find no significant relationship between instructor type and response rates for any of the questionnaire items. To minimize the effects of selective response on our estimates, we weight all our observations by the predicted probability of responding to the respective survey item following Wooldridge (2007). More specifically, we first estimate students' probability of responding to each survey item with linear probability models for the relevant subpopulations (see Tables A2-A4 in the appendix). We then windsorize these predicted probabilities at the 1st and 99th percentile of all positive predicted values. Finally, we

estimate the effect of student instructors on survey-reported student outcomes weighting each observation by the inverse of the windsorized predicted response probabilities.

To ease the interpretation of our results, we standardize all dependent variables to have means of zero and standard deviations of one across our estimation sample. The exceptions are having found a job at the time of graduation, which we keep as a dummy variable, time needed to find a job, which we keep as months, and current and first-job yearly earnings which we include in log-points.

4 The Effect of Student Instructors on Student Outcomes

4.1 Effect of Student Instructors on Course Grades

In Figure 3, we show estimates of having a student instructor instead of a senior instructor on students' grades from four different specifications with an increasing number of control variables (see also Table A5 in the appendix). The point estimates are small, precise, and barely move across the different specifications. In our preferred specification at the bottom of the figure, which includes course combination fixed effects, scheduling controls, tutorial-level controls, and student-level controls, we see that having a student instructor instead of a senior instructor reduces students' grades by a mere 1.7 percent of a standard deviation. This effect is marginally significant, and we can rule out that being assigned to a student instructor reduces students' grades by more than 3.7 percent of a standard deviation based on the 95 percent confidence interval.

Our reference category of senior instructors contains lecturers, postdocs and professors of all levels. In an unreported regression we have estimated the difference between each senior instructor type and student instructors. While we lack the statistical power to provide conclusive evidence in this regression, these results suggest that lecturers are most similar to student

instructors: they increase grades by a statistically insignificant 0.6 percent of a standard deviation over student instructors.

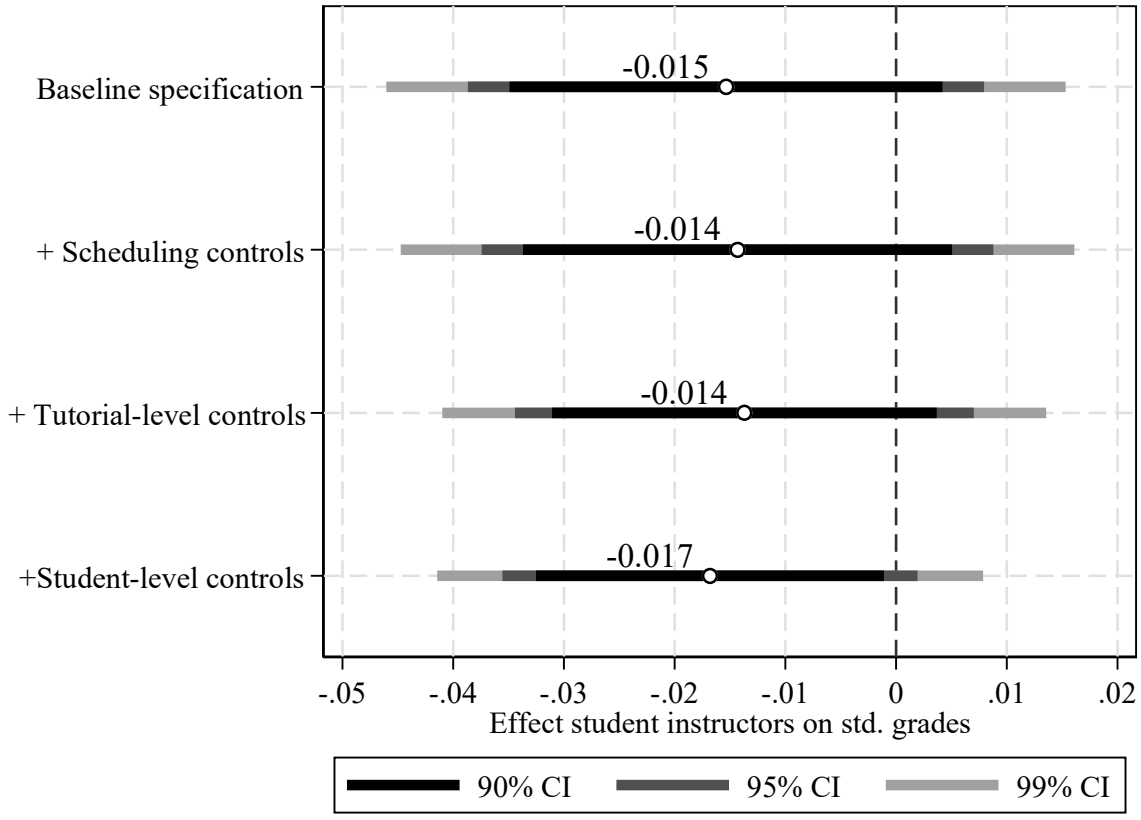

Fig. 3. Effects of student instructors on grades

This figure is based on regression estimates shown in Table A5. Confidence intervals are based on standard errors clustered at the instructor level.

The estimated effect of student instructors on student grades is small and slightly more negative than estimates in Feld et al. (in press), where differences in effectiveness between student instructors and more-senior instructors range from 0.2 and 0.8 percent of a grade standard deviation, depending on which senior instructor type we use as comparison. In terms of the Dutch 1to10 grade scale, the effect size is equivalent to a reduction of 0.03 points, which is less than the gap between the median and the 52nd percentile of student GPA. The effect is similar in magnitude

to the effect in the same environment of having tutorial peers with one standard deviation higher GPA which increase students grades by 1.3 percent of a standard deviation (Feld & Zölitz, 2017). This effect is small compared to the impact of other determinants of student grades in the same environment, such as the 17 percent of a standard deviation grade premium received by students with the same nationality as their graders (Feld, Salamanca, & Hamermesh, 2016). The effect is also small compared to estimates from Lusher et al. (2018), who find that Asian students' grades increase by 7.7 percent of a standard deviation when exposed to TAs of their own ethnicity.

Student instructors have no significant effects on course dropout or course failure, and we can rule out effect sizes of more than 1 percentage point in dropout rates and 1.6 percentage points in fail rates based on the 95 percent confidence intervals (see Table A6 in the appendix).

We also explore whether selection bias—specifically, the possibility that lower-quality instructors do not teach more than once—could have biased the estimated effects of student instructors in our previous work (Feld et al., in press). To do this, we estimate the effect of student instructors on dropout rates, fail rates, and student grades separately for student instructors who we observe teaching just once and those who we observe teaching at least twice. The latter are the only instructors identifying the value-added estimates in Feld et al. (in press). In Figure 4 we show that once-only student instructors achieve significantly worse outcomes for their students (right-hand-side figure): they increase dropout rates by 1 percentage point, increase fail rates by 2.3 percentage points, and lower student grades by 10 percent of a standard deviation (see also Table A7 in the appendix). These effects cannot be explained by potential positive effects of teacher experience; we see in unreported regressions that there are no differences between the student instructor effects when they teach for the first time and subsequent times on for any of these outcomes. If anything, student instructors become less effective after they have taught for the first

time – though this difference is not statistically significant. The negative effects of once-only student instructors are substantial, especially when compared to the smaller effects of students who teach at least twice (left-hand-side figure). Yet this selection bias barely affects our previous findings for one simple reason: less than 10 percent of our estimation sample is taught by single-term student instructors. Our concern for selection bias was therefore warranted, but in the end, was not empirically important in our setting. Our results, however, should serve as a cautionary tale of how selection could introduce bias in other applications in which value-added measures are used to estimate the relationship between teacher quality and other teacher characteristics.

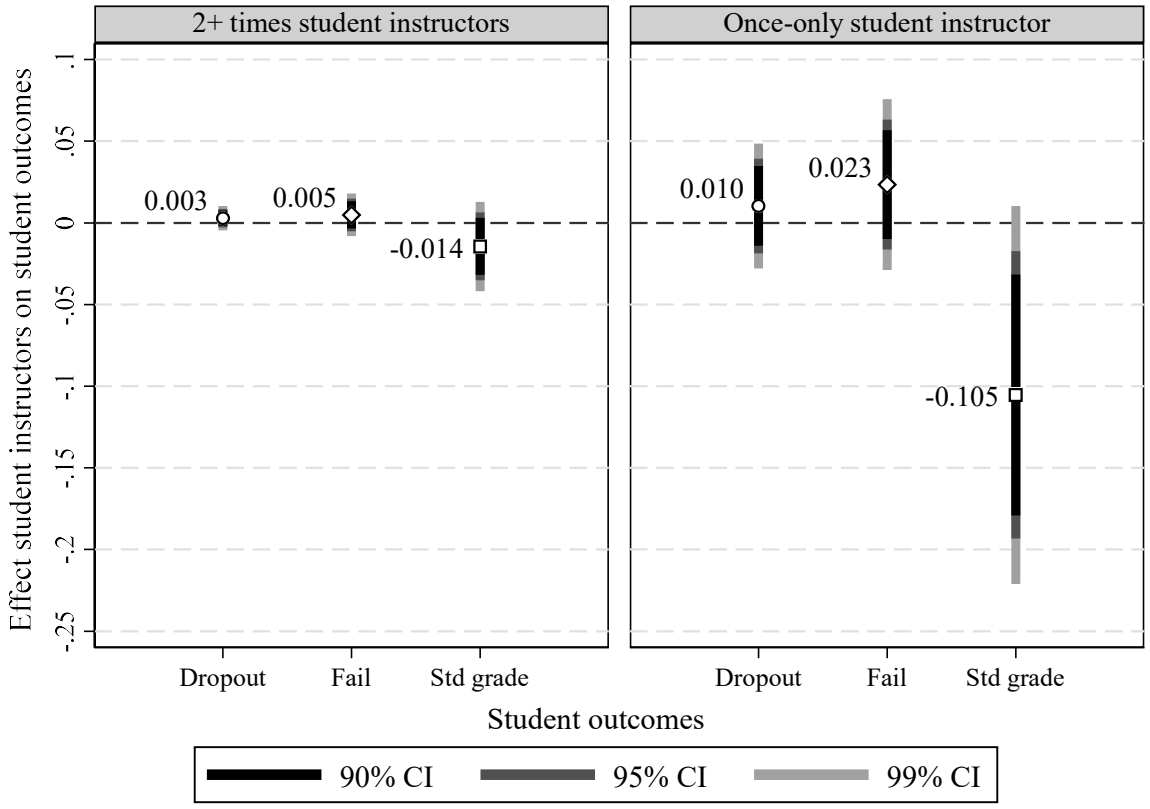

Fig. 4. Selection bias on the effects of student instructors on student academic outcomes

This figure is based on regression estimates shown in Table A7. Confidence intervals are based on standard errors clustered at the instructor level.

4.2 Heterogenous Effects: First-year Courses, Math Courses, and Student Ability

In which situations could student instructors matter more? The answer to this question is important because subgroup analyses can often uncover potential drivers of average treatment effects. Moreover, there could be subgroups of students meaningfully affected by student instructors that are missed by our average treatment effects. Identifying these subgroups is important to characterize possible distributional effects of increasing the use of student instructors.

Our estimated difference between student instructors and senior instructors could be driven downward if student instructors were worse, but the difference would also have compensating grading biases, canceling out any detectable difference. Student instructors could, for example, give students higher participation grades. In first-year courses, however, instructors have very little influence on grading as the final course grade consists entirely of the final exam grade. These final exams contain many machine-graded multiple-choice questions. And, while some instructors help with the grading of the nonmachine graded part of exams, they usually mark the same question for all students in the course so that grading biases would affect all students equally across tutorials. If the student instructor effects are small because of compensating grading biases, we would expect a more negative impact in first-year courses.

We show in Figure 5 that the effect of student instructors in first-year courses is even closer to zero than the effect for the whole sample, whereas in non-first-year courses, having a student instructor reduces students' grades by 2.8 percent of a standard deviation (see also Table A8 in the appendix). This difference could indicate that student instructors grade *less* generously, or it could be the result of other differences in the role of instructors between first-year and non-first-year courses. More importantly, these results dispel the concern that our main effects are small because student instructors grade more generously.

Student instructors could also be worse for students in mathematical courses, as these tend to be more difficult and require the instructor to be more active in explaining the course material. Contrary to this hypothesis, however, in Figure 5 we see virtually no difference between the effectiveness of student and senior instructors in mathematical courses. In non-mathematical courses, however, students score 3 percent of a standard deviation lower when taught by a student instructor. One reason for this difference could be that student instructors lack the experience or broader knowledge to effectively teach less technical courses.

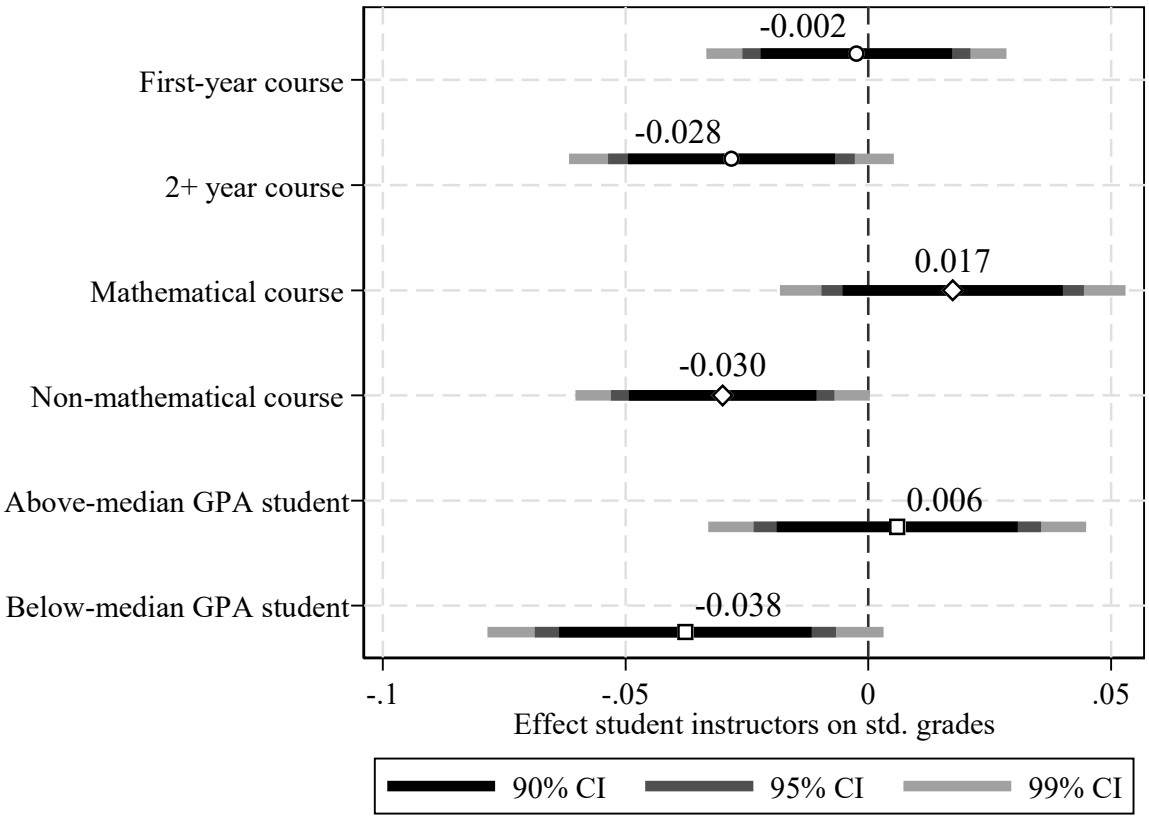

Fig. 5. Heterogenous effects of student instructors on grades

This figure is based on regression estimates shown in Table A8. Confidence intervals are based on standard errors clustered at the instructor level.

Finally, it may be that student instructors are more harmful to lower-ability students, who are less capable of independently understanding the course material. To test this hypothesis, we categorize students as *lower ability* or *higher ability* based on whether their preassignment GPA is in the bottom half or the top half of the course-specific GPA distribution. Consistent with the hypothesis, in Figure 5 we see that student instructors are harmful for lower ability students, lowering their grades by 3.8 percent of a standard deviation, yet they do not affect higher ability students.

Putting the results of this section together, we gather that student instructors could be more harmful to student grades in non-first-year courses, non-mathematical courses, and for lower ability students. Yet the effect on student grades remains comparatively small—never larger than 4 percent of a standard deviation. Nevertheless, these small differences should be considered for assessing the potential distributional effects of increasing the number of student instructors.

4.3 Effects on Follow-On Grades

It could be that current grades cannot tell the whole story; student instructors could affect learning in ways that only reveal themselves later. For example, student instructors may teach more to the test, whereas senior instructors help the students get a deeper understanding of the course material. If this is the case, we would expect that student instructors negatively affect students' *follow-on grades*—which we define as their grades in the next course students take that is offered by the same business school department. Effects on follow-on grades would identify persistent effects of student instructors on student performance, a measure closely related to the effects of teachers on 'deep learning' (Carrell & West, 2010, p. 412).

In Figure 6 we show the effects of having a student instructor on follow-on grades overall, by course type, and by student ability (see also Table A9 in the appendix). As in Feld et al. (in press), we see no effect of student instructors on follow-on grades on average, nor do we see significant heterogeneity by course type. Regarding student ability, our estimates show that students with a GPA above median perform 3.8 percent of a standard deviation *better* in their next related course if they have been taught by a student instructor. Overall, however, we see little evidence that student instructors matter for students' follow-on grades and that any effects on current grades do not seep through to follow-on grades.

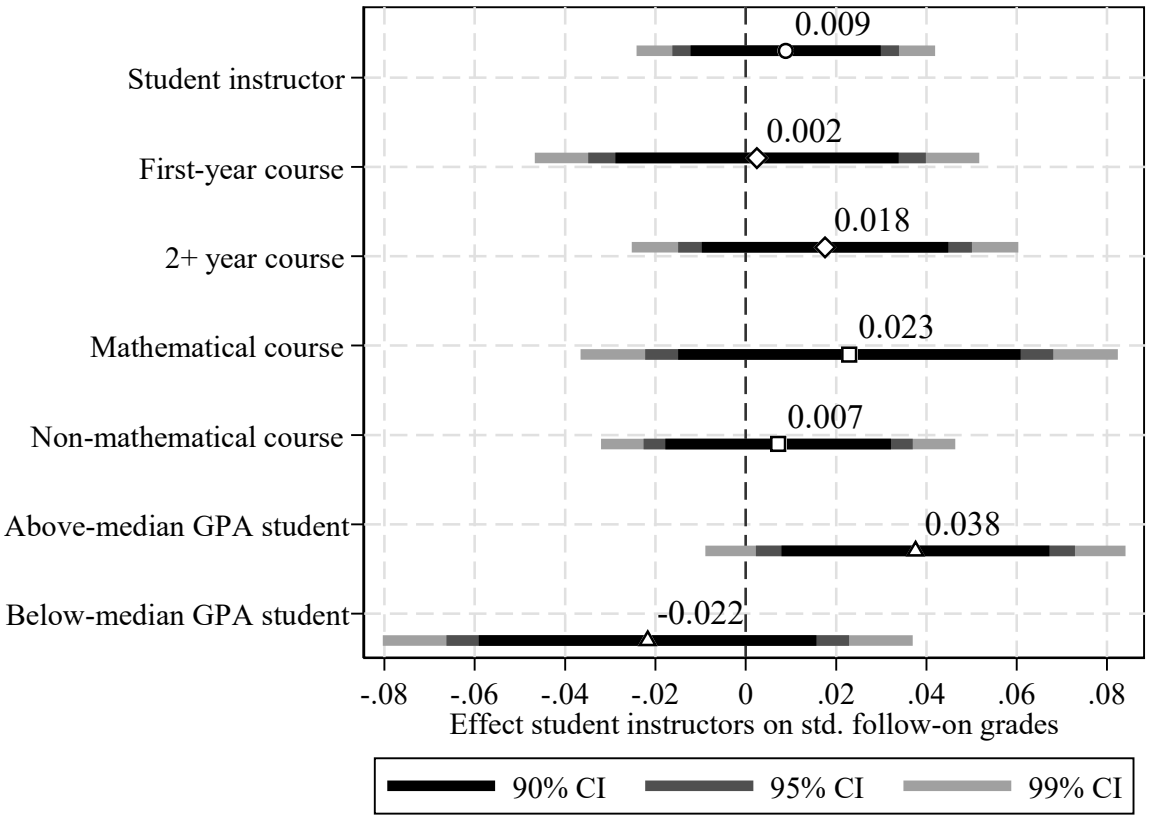

Fig. 6 Effects of student instructors on follow-on course grades

This figure is based on regression estimates shown in Table A9. Confidence intervals are based on standard errors clustered at the instructor level.

4.4 Effects on Students' Course Evaluations

Even though student instructors only have a small effect on grades, they may well affect other aspects of students' experiences at the university. The small effect on grades may, for example, be a result of students' compensating for bad student instructors by studying more at home. More generally, student instructors could decrease their students' nonpecuniary benefits of education. If this is the case, increasing the number of student instructors would impose a cost on students that we do not see by only looking at grades and follow-on grades. To explore these issues, we use data from the course evaluation survey. This survey asks students several questions about different facets of their experience with the course at the end of each teaching term (but before they take their final exams).[5] These data allow us to peek inside the "black box" of instructor effectiveness.

We first estimate the effects of student instructors on four instructor-related evaluation items: the overall instructor evaluation, how well the instructor encouraged group discussions (as is often required at the business school), how well the instructor stimulated knowledge transfer to other contexts, and how well the instructor mastered the course content. We see in Figure 7 that student instructors are perceived as worse at transferring knowledge to other contexts, a result that is not surprising given their lack of experience. The estimated effects on overall evaluation, encouragement of group work, and mastery of course content are also negative, but not statistically significant (see also Table A10 in the appendix).

We then estimate the effect of student instructors on four other evaluation items which, while not directly related to instructor performance, could be affected by it: the overall course evaluation, the evaluation of the tutorial group functioning, the evaluation of the course material, and students' self-reported study hours outside the classroom. In Figure 8 we see that being

[5] See Appendix 2 for the wording of all course evaluation questions and Feld and Zölitz (2017) for more detailed description of the course evaluation procedure.

assigned to student instructors leads students to evaluate the overall course 9.5 percent of a standard deviation worse than students evaluate a course taught by senior instructors (see also Table A11 in the appendix). This estimate is somewhat larger than estimates in Feld et al. (in press), in which we find that student instructors receive between 0.7 percent and 4.6 percent of a standard deviation lower course evaluations compared to senior instructors. Selection bias plays an important part in this difference: the few students taught by once-only student instructors in our sample rate the course 37 percent of a standard deviation lower than similar students assigned to senior instructors. School administrators might trace these dramatically lower course evaluations to particular instructors, thereby affecting their chances of teaching again. This situation introduces selection bias into our previous analyses with course evaluation value-added measures (which do not use once-only instructor information). The estimated effect of student instructors excluding once-only instructors is a still larger but more comparable 8 percent of a standard deviation reduction in course evaluations.

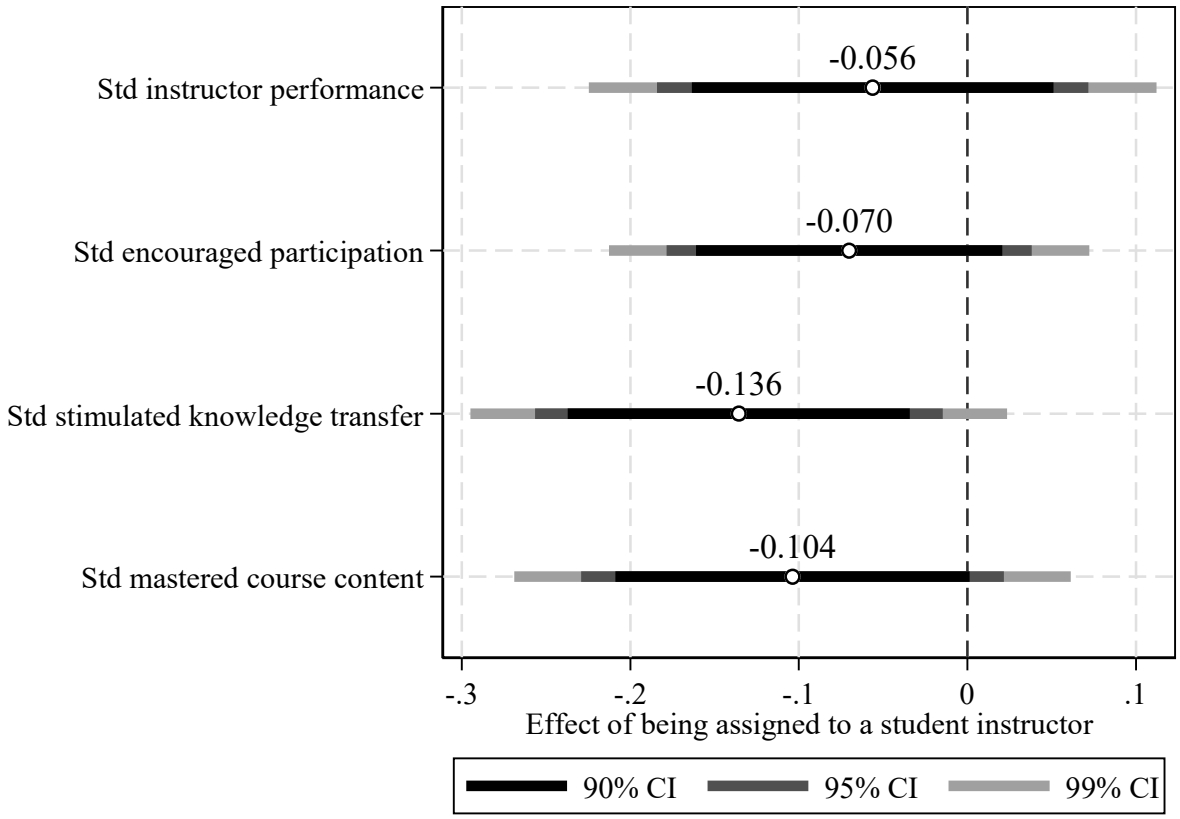

Fig. 7. Effect of student instructor on instructor-related evaluation outcomes

This figure is based on regression estimates shown in Table A10. Confidence intervals are based on standard errors clustered at the instructor level.

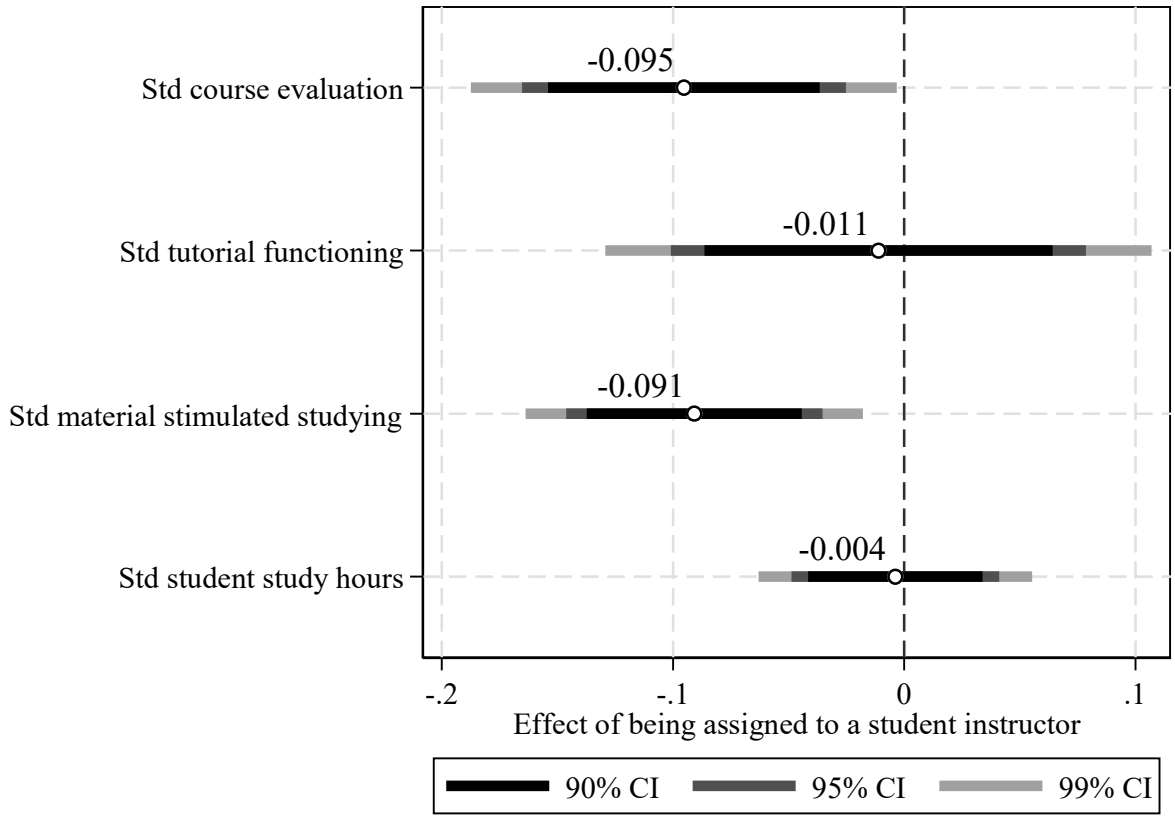

Fig. 8. Effect of student instructor on other evaluation outcomes

This figure is based on regression estimates shown in Table A11. Confidence intervals are based on standard errors clustered at the instructor level.

We also find that students of student instructors find the course material less stimulating. This suggests that an important difference between student instructors and senior instructors is their ability to encourage students to engage with the course material. Importantly, student instructors do not affect their students' study hours outside the classroom. This is evidence against the idea that students compensate for low instructor quality by studying more, which was our primary concern. Overall, the results in this section show that students perceive student instructors as being of lower quality in some regards.

4.5 Effects on Postgraduation Outcomes

Could student instructors affect their students' long-run outcomes? Despite having little impact on student academic outcomes, student instructors could affect their students' outcomes after graduation. Student instructors could, for example, be less able to provide their students with the skills, knowledge, or referrals necessary for beginning a successful career. Moreover, because students of student instructors perceive them to be somewhat worse than senior instructors, this perception could cascade to students' satisfaction with the overall quality of their education after graduation.

To estimate the effect on labor market outcomes and retrospective study satisfaction, we use data from a graduate survey of students who obtained their bachelor's degree between September 2010 and September 2015.[6] This graduate survey included questions about job search length after graduation, earnings in the first job after graduation, current earnings, and job satisfaction as well as retrospective study satisfaction (see Appendix 2 for more details on these variables).

In Figure 9 we show that having a student instructor has no effect on any of these postgraduation outcomes (see also Table A12 in the appendix). This result echoes and expands our previous findings in Feld et al. (in press), in which we report no differences across instructor academic ranks on earnings and job satisfaction. However, some of these estimates are too imprecise to draw any strong conclusions.

[6] We conducted the survey in cooperation with the business school's alumni office, which provided us with contact details for 4,215 bachelor's degree students. We first contacted the graduates via email and provided them with a link to the online survey. We then hired a team of current students who called the graduates who did not respond to the online survey to conduct the survey over the phone. Out of the contacted graduates, 1,693 responded to either the email or phone survey. See also Zölitz and Feld (2018) for a more detailed description of the alumni survey.

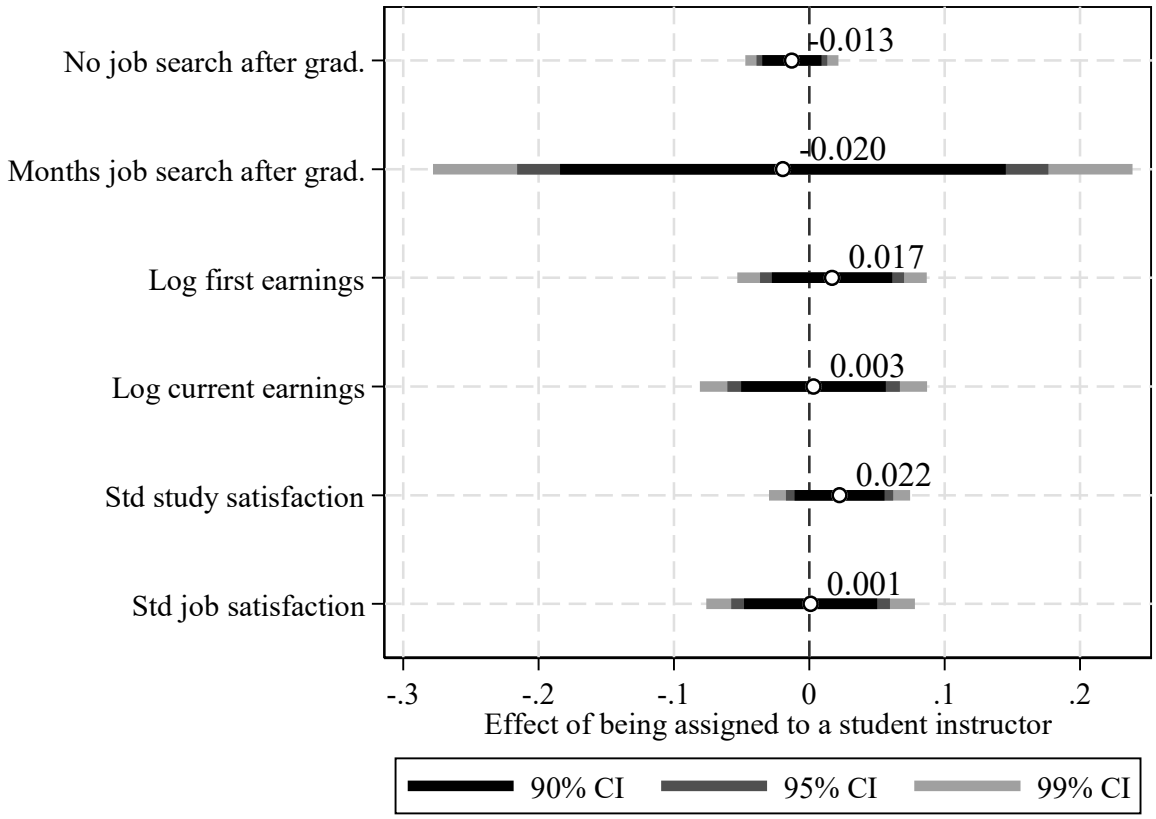

Fig. 9. Effect of student instructor on postgraduation outcomes

This figure is based on regression estimates shown in Table A12. Confidence intervals are based on standard errors clustered at the instructor level.

5 Effects of Hiring More Student Instructors

5.1 Effects in Teaching Terms with High Demand for Student Instructors

Given that student instructors can deliver tutorial instruction similar to senior instructors for a fraction of the cost, one may wonder how hiring more student instructors would affect students. If the teaching quality of each additional student instructor hired decreases—as would naturally happen if the business school recruits the best-suited students as instructors from a limited pool—it could be that such policy still lowers student performance. Jepsen and Rivkin (2009) show this

to be a valid concern by analyzing a class size reduction policy in California. They find that the positive effect of a reduction in class size is partly offset by a decrease in teacher quality.

To explore whether hiring more student instructors would mean hiring students who are worse teachers, we compare the effectiveness of student instructors in teaching terms with *high demand* (above-median demand) or *low demand* (below-median demand) for them. Our measure of demand for student instructors is the number of student instructors employed each period divided by the total number of students eligible to become student instructors (that is, master's students and third year or above bachelor's students) enrolled in the academic year. In high-demand teaching terms, there are, on average, 51 additional tutorials taught, and the average number of student instructors is 32, compared with 14 in low-demand terms. In contrast, the number of potentially employable students remains stable, at around 1,400, for both low- and high-demand terms. Shifts in our student instructor demand measure thus come from staffing demands rather than a short supply of employable students.

We do see some evidence that student instructors hired in high-demand terms have lower academic ability compared to those hired in low-demand terms: within academic years, the average first-year GPA of student instructors is 0.5 grade points lower in high-demand terms than in low-demand terms (p-value = 0.030, although we only observe GPA for 92 of our student instructors). This aligns with schedulers' claims that grades are the main hiring criteria for student instructors. However, it is not obvious that student instructors with lower academic ability are also worse instructors. The absence of the relationship between academic rank and teaching effectiveness that we find above shows that this need not be the case.

In Figure 10 we show the estimated effect of student instructors in terms with low and high demand for them (see also Table A13 in the appendix). We cannot rule out that student instructors

are equally effective in low- and high-demand terms for any of the outcomes we consider. The estimates for grades and follow-on grades look nearly indistinguishable across those types of terms. And, if anything, the point estimates for all outcomes suggest that student instructors perform *better* in high-demand terms than in low-demand terms. These results suggest that student instructors' academic ability is unrelated to their teaching ability. While these estimates could be confounded by the possibility that courses taught in high-demand terms are more suitable for student instructors to teach, we view these findings as suggestive evidence that hiring more student instructors does not mean hiring instructors of worse quality.

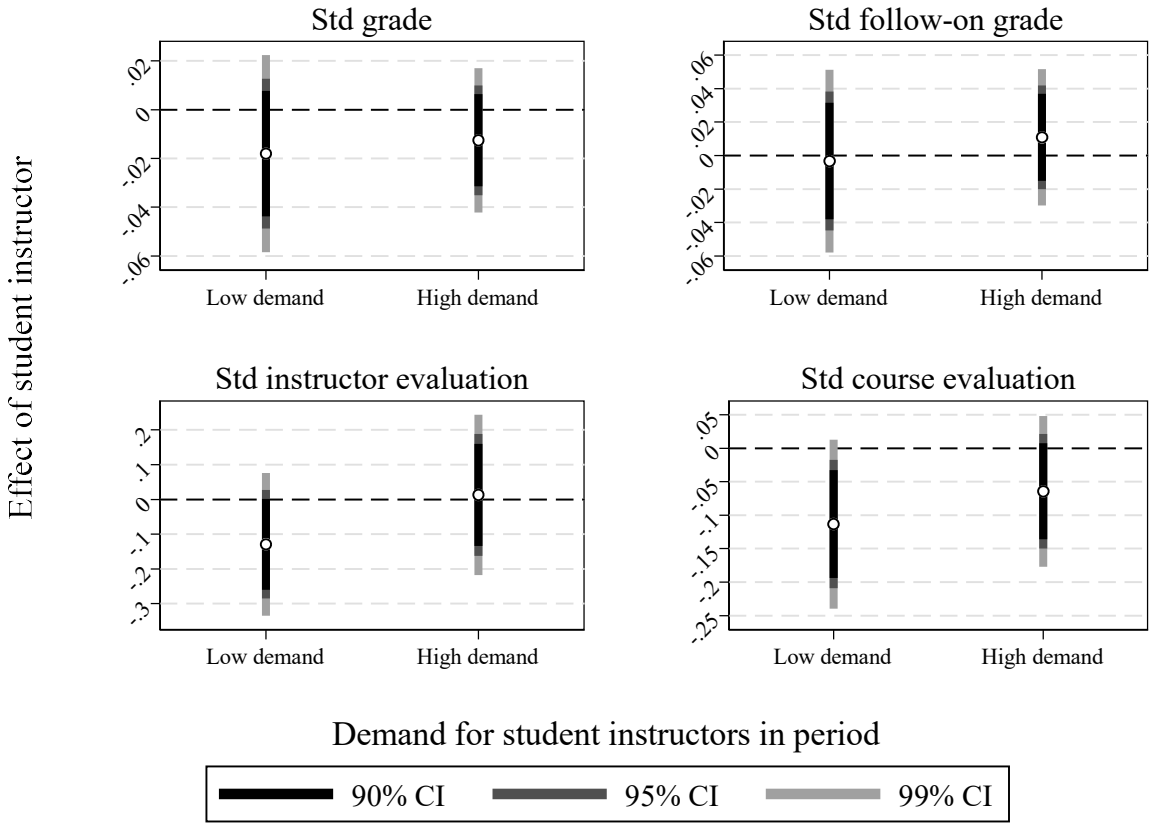

Fig. 10. Effect of student instructor in high- and low-demand teaching terms

This figure is based on regression estimates shown in Table A13. Confidence intervals are based on standard errors clustered at the instructor level.

5.2 Cumulative Effects of Student Instructors

Hiring more student instructors would also mean that students would be exposed to more student instructors throughout their studies. While having *one* student instructor in any given course has a negligible effect on students, the effect of having several student instructors could add up. For example, student instructors could be leaving their students with some knowledge gaps. If these knowledge gaps are small enough and if they were being filled-in eventually by more-capable instructors down the road, we would not be able to detect them by simply looking at grades or follow-on grades. However, the accrued effect of these gaps would be evident in students who were exposed to several student instructors throughout their studies.

To test for these potential cumulative effects of student instructors on student grades, we allow the effect of student instructors to vary with the number of previous student instructors each student has been exposed to. Students in our data differ widely in the number of student instructors they have had, with 45 percent of our data coming from students exposed to no student instructors, and over 10 percent of our data coming from students exposed to at least four student instructors.

In Figure 11 we show that for students who have been previously taught by up to five student instructors, an additional student instructor does not affect their grades (see also Table A14 in the Appendix). However, for students who have been taught by six or more student instructors, an additional student instructor does seem to lower their grades, with the estimated penalty being as high as 10 percent of a standard deviation for students exposed to seven or more student instructors (3 percent of our estimation sample). There are only a few students exposed to that many student instructors in our sample, but this number of students would increase if the business school were to increase its use of student instructors in tutorials. While an F-test does not reject the null hypothesis that the effect of student instructors is the same across the number of previous

student instructors, we interpret these estimates as suggestive evidence that exposure to many student instructors can harm students' grades.

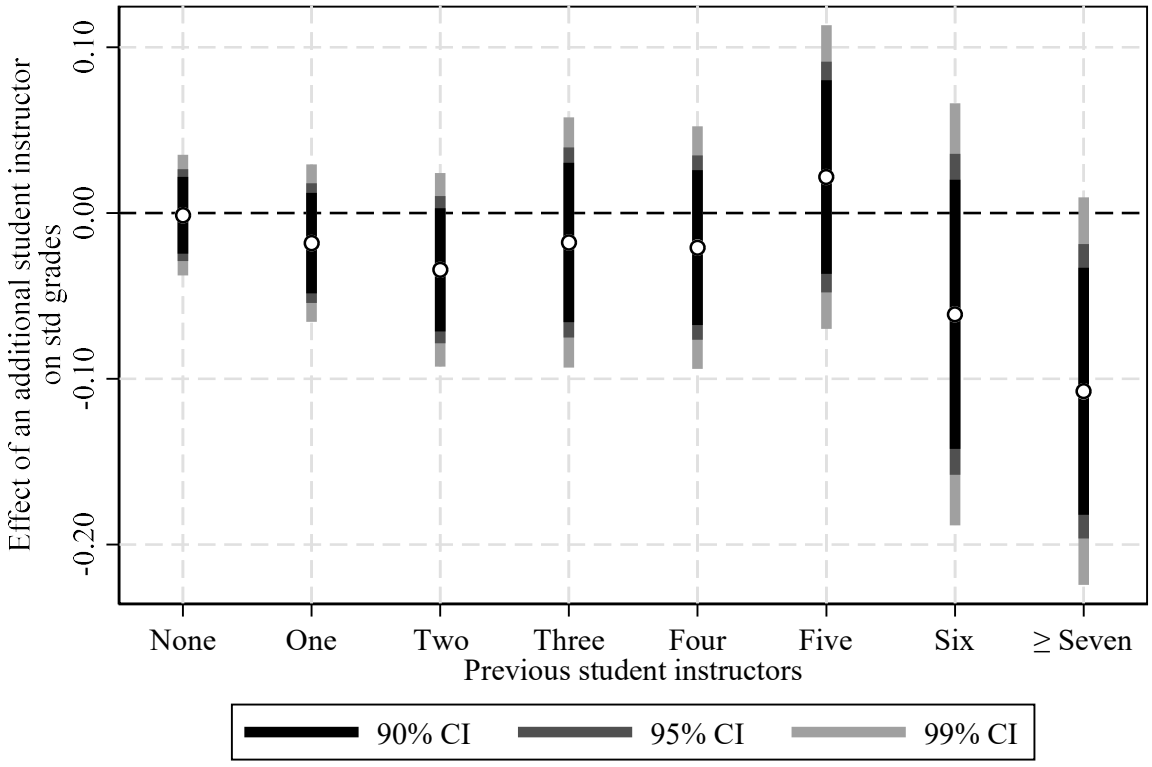

Fig. 11. Cumulative effect of student instructor on grades

Note: This figure is based on regression estimates shown in Table A14. Confidence intervals are based on standard errors clustered at the instructor level.

5.3 Discussion: The Effects of Using Only Student Instructors

The results in the previous subsection give us confidence that relatively large increases in the use of student instructors would not have any meaningful negative effects on students. But what would happen if the business school implemented an extreme policy of *only* using student instructors in tutorials? Such a policy would lead to substantially decreased wage costs or increased research time for senior instructors. Such a policy is, in fact, quite common: about 25 percent of OECD

universities only use students or PhD student instructors for tutorials (Feld et al., in press). Yet, we have two reasons to believe that exclusively relying on student instructors would lead to worse student outcomes than our results in this paper suggest.

First, such a drastic increase in the number of student instructors would still likely mean hiring lower teaching quality ones if education managers are at least somewhat able to discern prospective student instructors' teaching quality. Moreover, we have already seen that increasing the demand for student instructors from 14 to 32 per term leads to hiring student instructors with 0.5 points lower GPAs. Even if students' academic ability is completely unrelated to their teaching ability among the currently used student instructors, there is surely a minimum level of academic ability below which students would not be able to teach effectively. At the current teaching load per student instructor, the business school would have to hire about 74 student instructors per term. Given that the school currently uses about 21 student instructors per term, we would be guessing on the effect of almost quadrupling this amount. Extrapolating from our estimates to such a drastic change in student instructor usage would be going beyond the available support in the data. Consequently, we still have concerns that drastically increasing the demand for student instructors would force the school to hire students who should not be teaching.

Second, only using student instructors would also mean allocating them to courses in which they are less effective. This concern goes beyond the heterogeneous effects discussed in Section 4.2. All our estimates are based on courses that have both student and non-student instructors, and these courses are likely the ones which student instructors are best able to teach. For example, it could be that only courses with material simple enough to be taught by student instructors would use them to begin with. Or it could be that the course content of such courses has been gradually adapted to better fit the teaching style of student instructors. Either way, using student instructors

in all courses can result in students teaching courses in which they are less effective instructors or forcing course coordinators to adjust the course material so that students are able to teach it. Both outcomes would impose costs that are invisible through our current estimates.

6 Conclusion

In previous work, we have shown there are negligible differences in the effectiveness of tutorial teaching across academic ranks and concluded that universities should hire more students to teach tutorials. In this paper, we have tested whether this conclusion survives additional scrutiny. Using an alternative methodology and looking at a host of other outcomes, we have shown that student instructors fare slightly worse than our previous paper suggests when compared to senior instructors, yet they remain surprisingly good tutorial instructors overall. They are only a tiny bit worse at improving students' grades and receive just moderately lower course evaluations from students. In our previous paper, we underestimate these differences because we used value-added measures, which forced us to exclude lower quality, once-only student instructors. Yet, overall, the differences we find are tiny when considering the differences in experience and qualifications between student and senior instructors.

There are a few reasons student instructors are performing as well as more academically senior instructors. They could, for example, compensate for their lack of knowledge and experience by being better able to relate to their fellow students. Having the course material fresh in their heads—because they just learned it—may also make them better at explaining it. They could also be working harder than senior instructors, who might often feel like they are wasting their time teaching tutorials. Student instructors may be doing so well because course coordinators have worked hard to accommodate their lack of knowledge and experience by, for example,

providing well-crafted solution manuals. The smaller effect of student instructors in first-year courses, in which instructors typically receive more guidance including more extensive tutor manuals, is consistent with this interpretation. Convincingly isolating any one of these mechanisms goes beyond the scope of this paper, but our results do suggest that isolating the mechanism could be the secret ingredient needed to turn unqualified and inexpensive students into effective teachers.

We have also shown that hiring more student instructors does not necessarily mean hiring worse quality ones. However, this does not mean that drastically increasing the use of student instructors would have no impact on student outcomes. We have seen some evidence that student instructors are more harmful to students in non-mathematical and non-first-year courses, that lower-ability students fare worse when assigned to student instructors, and that students exposed to many student instructors throughout their studies also get lower grades. The effects of student instructors are not so negative as to be concerning in any of these analyses, yet they suggest that some students would be harmed if all tutorials were taught by student instructors.

Our findings raise the question of why senior instructors, like professors, teach tutorials at all. We have argued above that replacing all senior instructors with students could mean hiring some students who should not be teaching and using student instructors in some courses they should not teach. The difference between student and senior instructors could be larger or smaller in other institutions where tutorials are taught differently or where students are different. It thus makes sense to evaluate tutorial staffing policies for each institution and course separately. But even with all these caveats, it is worth asking whether (and why) students can do that part of professors' job just as well as trice as cheaply.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Ournal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bettinger, E. P., & Long, B. T. (2010). Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes. *Review of Economics and Statistics*, *92*(3), 598–613. https://doi.org/10.1162/REST_a_00014

Bettinger, E. P., Long, B. T., & Taylor, E. S. (2016). When Inputs are Outputs: The Case of Graduate Student Instructors. *Economics of Education Review*, *52*, 63–76. https://doi.org/10.1016/j.econedurev.2016.01.005

Borjas, G. J. (2000). Foreing-Born Teaching Assistants and the Academic Performance of Undergraduates. *American Economic Review*, *90*(2), 355–359. Retrieved from http://www.jstor.org/stable/117250

Bureau of Labor Statistics. (2017a). Occupational Employment Statistics - Graduate Teaching Assistants. Retrieved December 13, 2018, from https://www.bls.gov/oes/2017/may/oes251191.htm

Bureau of Labor Statistics. (2017b). Postsecondary Teachers : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics. Retrieved December 13, 2018, from https://www.bls.gov/ooh/education-training-and-library/postsecondary-teachers.htm

Carrell, S. E., & West, J. E. (2010). Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy*, *118*(3), 409–432. https://doi.org/10.1086/653808

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates. *American Economic Review*, *104*(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

De Vlieger, P., Jacob, B., & Stange, K. (2018). Measuring Effectiveness in Higher Education. In C. M. Hoxby & K. Stange (Eds.), *Productivity in Higher Education*. University of Chicago Press. Retrieved from http://www.nber.org/chapters/c13880.pdf

Fairlie, R., Hoffmann, F., & Oreopoulos, P. (2014). A Community College Instructor Like Me: Race and Ethnicity Interaction in the Classroom. *American Economic Review*, *104*(8), 2567–2591. https://doi.org/10.1257/aer.104.8.2567

Feld, J., Salamanca, N., & Hamermesh, D. S. (2016). Endophilia or Exophobia: Beyond Discrimination. *The Economic Journal*, *126*(594), 1503–1527. https://doi.org/10.1111/ecoj.12289

Feld, J., Salamanca, N., & Zölitz, U. Are Professors Worth It? The Value-added and Costs of Tutorial Instructors. *Journal of Human Resources*. https://doi.org/10.3368/jhr.55.3.0417-8752R2

Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics*, *35*(2), 387–428. https://doi.org/10.1086/689472

Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2015). Are Tenure Track Professors Better Teachers? *Review of Economics and Statistics*, *97*(4), 715–724. https://doi.org/10.1162/REST_a_00529

Fisher, R. A. (1992). Statistical Methods for Research Workers. In Kotz S. & J. N.L. (Eds.), *Breakthroughs in Statistics* (pp. 66–70). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4380-9_6

Fleisher, B., Hashimoto, M., & Weinberg, B. A. (2002). Foreign GTAs can be Effective Teachers of Economics. *Journal of Economic Education*, *33*(4), 299–325. https://doi.org/10.1080/00220480209595329

Hoffmann, F., & Oreopoulos, P. (2009). Professor Qualities and Student Achievement. *Review of Economics and Statistics*, *91*(1), 83–92. https://doi.org/10.1162/rest.91.1.83

Jepsen, C., & Rivkin, S. (2009). Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher. *Journal of Human Resources*, *44*(1), 223–250. Retrieved from https://www.jstor.org/stable/20648893

Lusher, L., Campbell, D., & Carrell, S. (2018). TAs like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates. *Journal of Public Economics*, *159*(March 2016), 203–224. https://doi.org/10.1016/j.jpubeco.2018.02.005

Pei, Z., Pischke, J.-S., & Schwandt, H. (2018). Poorly Measured Confounders are More Useful on the Left than on the Right. *Journal of Business & Economic Statistics*, 1–12. https://doi.org/10.1080/07350015.2018.1462710

Wooldridge, J. M. (2007). Inverse Probability Weighted Estimation for General Missing Data Problems. *Journal of Econometrics*, *141*(2), 1281–1301. https://doi.org/10.1016/J.JECONOM.2007.02.002

Zölitz, U., & Feld, J. (2018). *The Effect of Peer Gender on Major Choice*. *University of Zurich, Department of Economics Working Paper Series*. https://doi.org/10.2139/ssrn.3071681

APPENDIX

A1 Data Restrictions

In this appendix, we list the observations we exclude from our estimation sample because they either represent exceptions from the standard tutorial group assignment procedure or because they are out of the scope of the estimates in this paper.

- We exclude eight courses in which the course coordinator or other education staff actively influenced the tutorial group composition. One course coordinator, for example, requested to balance student gender across tutorial groups. The scheduling department informed us about these courses.
- We exclude 21 tutorial groups that consisted mainly of students who registered late for the course. Before April 2014, the business school reserved one or two slots per tutorial group for students who registered late. In exceptional cases in which the number students who registered late substantially exceeded the number of empty spots, new tutorial groups were created that mainly consisted of those students who registered late. This late registration policy was abolished in April 2014.
- We exclude 46 repeater tutorial groups. One course coordinator explicitly requested to assign repeater students who failed his courses in the previous year to special repeater tutorial groups.
- We exclude 17 tutorial groups that consist mainly of students from a special research-based program. For some courses, students in this program were all assigned to separate tutorial groups with a more-experienced teacher.
- We exclude 95 part-time MBA students because these students are typically scheduled for special evening classes with only part-time students.

- We exclude observations from 107 instructors who we could not categorize as either student instructor, PhD student instructor, or senior instructor. We suspect these are student instructors but cannot be sure.
- Missing data on covariates is imputed and flagged as missing (see Appendix A2 for details), and we lose only five observations for which there is some remaining data missing.
- Finally, we limit our estimation sample to courses that had at least one student instructor and one non-student instructor. Table 1, however, includes information on those excluded courses for comparison.

After our sample restrictions and imputations, our estimation sample includes complete information on 45,060 course enrollments from 9,323 students who took 206 courses, taught by 485 instructors over 24 teaching periods between the academic years 2009–10 and 2014–15. In the analyses we lose observations due to missing values of the dependent variables, which we never impute.

A2 Variable Definitions and Covariate Imputations

In this appendix we describe the nontrivial variable coding decisions and the imputation methods we use throughout the paper.

- Our main time variable in the data combines the academic year (between 2009–10 and 2014–15) and the (regular) teaching terms (1, 2, 4, and 5) into a discrete variable with 24 unique values. We exclude nonregular teaching terms 3 and 6 from our data. In these nonregular teaching terms, courses are only two weeks long and not always graded.
- We code 107 distinct nationality codes for students and 60 distinct nationality codes for staff into main categories: Dutch, German, Belgian, Other nationality, and Not recorded.

- We have information on student GPA for 114 staff members whom we also observe as teachers. We use this information for student instructors in Section 5.1, where we explore whether student instructors in high-demand terms have lower GPAs.
- We code dropout students as students who are registered in a course but who receive no grade at the end of the course.
- We impute missing values for students' age using iterative conditional means, restricting conditional means to be calculated with at least 5 observations. We first impute students' missing age as the mean age of other students enrolled in their tutorial. When age is still missing, we impute it as the mean age of students in the same course, nationality, and gender group. If that is not possible, we impute it as the mean age of students in the same course alone. If that is still not possible, we impute it based on the mean age of students in the same subject matter (as defined by the course code), academic year, and study program. The few remaining missing age observations are imputed as the mean age of students in the same academic year and study program.
- Student GPA is constructed based on our data. For the first year and first period of our data window, we have no grades to construct it, so we impute it as zero. The remaining missing values for students' GPA are imputed using iterative conditional means, restricting conditional means to be calculated with at least 5 observations. We first impute students' missing GPA as the mean GPA of other students enrolled in the same course who have the same gender and nationality. When GPA is still missing, we impute it as the mean GPA of students who took the same subject (as defined by the course code) and enrolled in the same academic year and study program. If that is still not possible, we impute it as the mean GPA of students enrolled in the same academic year, study program, gender,

nationality and age group. The remaining missing GPA observations are imputed as the mean GPA of students in the same study program and enrolled in the same academic year. We drop five student-term observations with missing GPAs.

- The formulation of instructor and course evaluation questions and their answer scale is:
  - *Overall performance*. Question: "Evaluate the overall functioning of this tutor with a grade on a ten-point scale (1 = very bad, 6 = sufficient, 10 = very good)." Scale: Very bad (1) – very good (10)
  - *Encouraged participation*. Question: "The tutor encouraged all students to participate in the (tutorial) group discussions." Scale: Totally disagree (1) – Totally agree (5)
  - *Knowledge transfer*. Question: "The tutor stimulated the transfer of what I learned in this course to other contexts." Scale: Totally disagree (1) – Totally agree (5).
  - *Mastered content*. Question: "The tutor sufficiently mastered the course content." Scale: Totally disagree (1) – Totally agree (5)
  - *Course evaluation*. Question: "Please give an overall grade for the quality of this course (1 = very bad, 6 = sufficient, 10 = very good)?". Scale: Very bad (1) – very good (10)
  - *Tutorial functioning*. Question: "My tutorial group has functioned well." Scale: Totally disagree (1) – Totally agree (5)
  - *Material stimulated studying*. Question: "The learning materials stimulated me to start and keep on studying." Scale: Totally disagree (1) – Totally agree (5)
  - *Student study hours*. Question: "How many hours per week on the average **(excluding contact hours)** did you spend on self-study (presentations, cases,

assignments, studying literature, etc)?" Scale: Open numerical answers ranging from 0 to 90

- The formulation of the graduate survey questions and their answer scale is:
  - *Unemployment*. Question: "After graduation, how long did it take you to find your first job?" Scale: "I already had a job lined up," "0–1 months," "3–4 months," "4–-6 months," "6–12 months," "More than 12 months." From this question we construct two measures of job search length after graduation. The first one is an indicator of whether the student had a job waiting for her after graduation. The second one is a continuous measure of the job search length after graduation in months using the midpoint of each of the answer categories as values, not including zero job search lengths, and top-coding job search length at 12 months.
  - *First earnings*. Question: "Looking back at your first job after university, what was your entry salary? What was your yearly income before taxes? (including bonuses and holiday allowances)." Scale: €0– €1,000,000. We use the natural logarithm of this measure in our analyses.
  - *Current earnings*. Question: "What is your yearly income before taxes from your main job? (including bonuses and holiday allowances)" Scale: €0– €1,000,000. We use the natural logarithm of this measure in our analyses.
  - *Job satisfaction*. Question: "How satisfied are you, overall, with your current work?" Scale: Totally unsatisfied (1) – Totally satisfied (10)
  - *Study satisfaction*. Question: "Looking back at your time as a bachelor's student, how satisfied are you with your overall study experience at [name of business school]?" Scale: Totally unsatisfied (1) – Totally satisfied (10)

Table A1. Balancing test

Dep. variable:	Std student GPA	Female student	Std student age	Std student ID	Std tutorial size	Exchange student
	(1)	(2)	(3)	(4)	(5)	(6)
Student instructor	-0.002	-0.015***	-0.005	-0.001	-0.032	-0.000
	(0.007)	(0.005)	(0.009)	(0.011)	(0.024)	(0.001)
PhD student instructor	0.001	-0.009	-0.024**	-0.009	-0.001	0.000
	(0.009)	(0.007)	(0.011)	(0.013)	(0.030)	(0.001)
Joint significance [p-value]	[.925]	[.011]	[.076]	[.781]	[.306]	[.910]
Mean of dep. var.	0.00	0.38	0.00	0.00	0.00	0.04
SD of dep. var.	1.00	0.49	1.00	1.00	1.00	0.20
R-Squared	0.65	0.06	0.41	0.04	0.66	0.64
Course combination FE	Yes	Yes	Yes	Yes	Yes	Yes
Scheduling controls	Yes	Yes	Yes	Yes	Yes	Yes
Instructors	485	485	485	485	485	485
Observations	45,060	45,060	45,060	45,060	45,060	45,060

This table shows coefficients from OLS course-by-parallel-course fixed effects regressions. Dependent variables are pretutorial-assignment student characteristics standardized over the estimation sample when indicated. Scheduling controls include time-of-day and day-of-week dummies. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2. Predictors of survey response: instructor evaluation items

Dep. variable: 1 if student answered the instructor evaluation question on:	Overall performance	Encouraged participation	Knowledge transfer	Mastered content
	(1)	(2)	(3)	(4)
Student instructor	0.012*	0.011	0.012	0.011
	(0.007)	(0.007)	(0.007)	(0.007)
PhD student instructor	0.006	0.004	0.005	0.005
	(0.009)	(0.009)	(0.009)	(0.009)
German student	0.052***	0.054***	0.055***	0.055***
	(0.006)	(0.006)	(0.006)	(0.006)
Belgian student	0.086***	0.089***	0.090***	0.090***
	(0.012)	(0.012)	(0.012)	(0.012)
Female student	0.071***	0.069***	0.070***	0.070***
	(0.005)	(0.005)	(0.005)	(0.005)
Age	0.004*	0.004*	0.004*	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
GPA	0.033***	0.033***	0.033***	0.033***
	(0.003)	(0.003)	(0.003)	(0.003)
Outcome mean	0.36	0.36	0.36	0.36
R-Squared	0.15	0.15	0.16	0.16
Controls	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes
Instructors	485	485	485	485
Observations	45,060	45,060	45,060	45,060

This table shows average marginal effects from OLS course-by-parallel-course fixed effects regressions. Dependent variables are dummies marking whether a student answered the respective instructor evaluation question. The comparison group is male, Dutch students taught by senior instructors. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A3. Predictors of survey response: course evaluation items

Dep. variable: 1 if student answered the course evaluation question on:	Course evaluation	Tutorial functioning	Material stimulated studying	Student study hours
	(1)	(2)	(3)	(4)
Student instructor	0.013*	0.012*	0.010	0.009
	(0.007)	(0.007)	(0.007)	(0.007)
PhD student instructor	0.007	0.006	0.003	0.005
	(0.009)	(0.009)	(0.009)	(0.009)
German student	0.055***	0.055***	0.053***	0.043***
	(0.006)	(0.006)	(0.006)	(0.006)
Belgian student	0.095***	0.089***	0.085***	0.067***
	(0.012)	(0.012)	(0.012)	(0.012)
Female student	0.074***	0.069***	0.071***	0.074***
	(0.005)	(0.005)	(0.005)	(0.005)
Age	0.005**	0.004*	0.004*	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
GPA	0.036***	0.032***	0.032***	0.034***
	(0.003)	(0.003)	(0.003)	(0.003)
Outcome mean	0.38	0.36	0.36	0.34
R-Squared	0.13	0.15	0.15	0.15
Controls	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes
Instructors	485	485	485	485
Observations	45,060	45,060	45,060	45,060

This table shows average marginal effects from OLS course-by-parallel-course fixed effects regressions. Dependent variables are dummies marking whether students answered the course evaluation question. The comparison group is male, Dutch students taught by senior instructors. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A4. Predictors of graduate survey response

Dep. variable: 1 if student answered the graduate survey question on:	Unemployment after graduation:	Log earnings after graduation:		Satisfaction after graduation:	
		First	Current	Std study	Std job
	(1)	(2)	(3)	(4)	(5)
Student instructor	0.001	-0.003	-0.002	-0.002	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
PhD student instructor	0.006	0.003	0.009	0.003	0.003
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
German Student	-0.054***	-0.048***	-0.021***	-0.018***	-0.045***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)
Belgian student	-0.007	-0.017	0.003	0.031***	-0.027***
	(0.010)	(0.011)	(0.011)	(0.011)	(0.010)
Female student	-0.010**	-0.014***	-0.017***	-0.004	-0.006
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Age	-0.003	-0.003	-0.005***	-0.011***	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
GPA	0.034***	0.029***	0.043***	0.063***	0.037***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Outcome mean	0.29	0.26	0.31	0.38	0.27
R-Squared	0.15	0.13	0.14	0.16	0.14
Controls	Yes	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes	Yes
Instructors	480	480	480	480	480
Observations	42,307	42,307	42,307	42,307	42,307

This table shows average marginal effects from OLS course-by-parallel-course fixed effects regressions. Dependent variables are dummies marking whether a student answered the respective graduate survey question. The comparison group is male Dutch students taught by senior instructors. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A5. Effects of student instructors on grades

Dep. variable: Std grade	Baseline specification	+ Scheduling controls	+ Tutorial-level controls	+ Student-level controls
	(1)	(2)	(3)	(4)
Student instructor	-0.015	-0.014	-0.014	-0.017*
	(0.012)	(0.012)	(0.011)	(0.010)
R-Squared	0.19	0.19	0.21	0.50
Course combination FE	Yes	Yes	Yes	Yes
Scheduling controls		Yes	Yes	Yes
Tutorial-level controls			Yes	Yes
Student-level controls				Yes
Instructors	485	485	485	485
Observations	41,619	41,619	41,619	41,619

This table shows coefficients from OLS course-by-parallel-course fixed effects regressions. The dependent variable is student course grade standardized over the estimation sample. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Section 3 explains scheduling controls, tutorial-level controls, and student-level controls in detail. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A6. Effects of Student Instructors on Course Dropout and Fail

Dep. variable:	Student dropout	Student fail course
	(1)	(2)
Student instructor	0.004	0.006
	(0.003)	(0.005)
Mean of dep. var.	0.08	0.22
SD of dep. var.	0.27	0.41
R-Squared	0.21	0.36
Course combination FE	Yes	Yes
Controls	Yes	Yes
Instructors	485	485
Observations	45,060	41,619

This table shows coefficients from OLS course-by-parallel-course fixed effects regressions. Dependent variables are student course dropout and student course failure dummies. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7. Selection bias on the effects of student instructors on student academic outcomes

Dep. variable:	Dropout	Fail	Std. grade
	(1)	(2)	(3)
Effect of a student instructor if they are:			
Once-only student instructor	0.010	0.023	-0.105**
	(0.015)	(0.020)	(0.045)
At least twice student instructor	0.003	0.005	-0.014
	(0.003)	(0.005)	(0.011)
F-Test equal effects [p-value]	[.616]	[.371]	[.047]
R-Squared	0.21	0.37	0.50
Controls	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes
Instructors	483	483	483
Observations	44,538	41,145	41,145

This table shows average marginal effects from fully interacted OLS course-by-parallel-course fixed effects regressions. The dependent variables are a student dropout dummy, a student fail dummy, and student course grades standardized over the estimation sample. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A8. Heterogenous effects of student instructors on grades

Dep. variable: Std grade	Effect of student instructors for:		
	First-year course	Mathematical course	Student GPA above median
	(1)	(2)	(3)
Yes	-0.002	0.017	0.006
	(0.012)	(0.014)	(0.015)
No	-0.028**	-0.030**	-0.038**
	(0.013)	(0.012)	(0.016)
F-Test equal effects [p-value]	[.837]	[.207]	[.693]
R-Squared	0.51	0.51	0.51
Course combination FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Instructors	485	485	485
Observations	41,619	41,619	41,619

This table shows average marginal effects from fully interacted OLS course-by-parallel-course fixed effects regressions. The dependent variable is student course grade standardized over the estimation sample. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). 'Yes' and 'No' refer to whether the sample specified in the columns title applies or not. For Column (1), for example, 'Yes' means that the estimates come the subsample of first-year courses and 'No' means the estimates come for the subsample of non-first-year courses. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9. Effects of student instructors on follow-on grades

Dep. variable: Std follow-on course grade		Effect of student instructors for:		
	Preferred specification	First-year course	Mathematical course	Student GPA above median
	(1)	(2)	(3)	(4)
Student instructor	0.009			
	(0.013)			
No		0.018	0.007	-0.022
		(0.017)	(0.015)	(0.023)
Yes		0.002	0.023	0.038**
		(0.019)	(0.023)	(0.018)
F-Test equal effects [p-value]	-	[.289]	[.636]	[.339]
R-Squared	0.36	0.36	0.36	0.36
Course combination FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Instructors	446	446	446	446
Observations	21,322	21,322	21,322	21,322

This table shows an OLS coefficient (column 1) and average marginal effects (columns 2–4) from separate fully interacted OLS course-by-parallel-course fixed effects regressions. The dependent variable is student course grade standardized over the estimation sample. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). 'Yes' and 'No' refer to whether the sample specified in the columns title applies or not. For Column (2), for example, 'Yes' means that the estimates come the subsample of first-year courses and 'No' means the estimates come for the subsample of non-first-year courses. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10. Effect of student instructor on instructor evaluations

Dep. variable: Standardized evaluation of instructors':	Overall performance	Encouraged participation	Knowledge transfer	Mastered content
	(1)	(2)	(3)	(4)
Student instructor	-0.056	-0.070	-0.136**	-0.104
	(0.065)	(0.055)	(0.062)	(0.064)
R-Squared	0.19	0.17	0.19	0.17
Controls	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes
Instructors	476	472	472	472
Observations	16,200	16,300	16,344	16,376

This table shows coefficients from WLS course-by-parallel-course fixed effects regressions using question-specific predicted answering probabilities as inverse probability weights (see Section 3 for details). Dependent variables are standardized instructor evaluation ratings. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Effect of student instructor on course evaluations

Dep. variable: Standardized evaluation of:	Course evaluation	Tutorial functioning	Material stimulated studying	Student study hours
	(1)	(2)	(3)	(4)
Student instructor	-0.095***	-0.011	-0.091***	-0.004
	(0.036)	(0.046)	(0.028)	(0.023)
R-Squared	0.23	0.17	0.19	0.21
Controls	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes
Instructors	481	476	471	476
Observations	16,974	16,337	16,191	15,126

This table shows coefficients from WLS course-by-parallel-course fixed effects regressions using question-specific predicted answering probabilities as inverse probability weights (see Section 3 for details) in each column. Dependent variables are course evaluation items, standardized over the estimation sample where indicated. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A12. Effect of student instructor on student postgraduation outcomes

Dep. variable: student answers of graduate survey questions on:	Job search length after graduation:		Log of earnings after graduation:		Satisfaction after graduation:	
	None	Months	First	Current	Std study	Std job
	(1)	(2)	(3)	(4)	(5)	(6)
Student instructor	-0.013	-0.020	0.017	0.003	0.022	0.001
	(0.013)	(0.100)	(0.027)	(0.032)	(0.020)	(0.030)
R-Squared	0.13	0.11	0.14	0.17	0.12	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes	Yes	Yes
Instructors	437	427	437	438	440	437
Observations	11,777	6,019	10,528	12,788	15,678	11,205

This table shows coefficients from WLS course-by-parallel-course fixed effects regressions using question-specific predicted answering probabilities as inverse probability weights (see Section 3 for details) in each column. Dependent variables are graduate survey answers, standardized over the estimation sample where indicated. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A13. Effect of Student Instructor in High- and Low-Demand Teaching Terms

Dep. variable:	Std grade	Std follow-on course grade	Std instructor evaluation	Std course evaluation
	(1)	(2)	(3)	(4)
	Effect of a student instructor if student instructors are in:			
Low demand	-0.018	-0.003	-0.129	-0.113**
	(0.016)	(0.021)	(0.079)	(0.049)
High demand	-0.013	0.011	0.013	-0.064
	(0.011)	(0.016)	(0.089)	(0.044)
F-Test equal effects [p-value]	[.775]	[.583]	[.170]	[.408]
R-Squared	0.5	0.36	0.19	0.24
Controls	Yes	Yes	Yes	Yes
Course combination FE	Yes	Yes	Yes	Yes
Instructors	485	446	476	481
Observations	41,619	21,322	16,200	16,974

This table shows average marginal effects from separate fully interacted course-by-parallel-course fixed effects regressions in each column. Columns 3 and 4 use question-specific predicted answering probabilities as inverse probability weights (see Section 3 for details). Dependent variables are student outcomes standardized over the estimation sample. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A14. Cumulative effect of student instructor on grades

Dep. variable: Std grade	Effect of an additional student instructor
	(1)
Number of previous student instructors:	
None	-0.001
	(0.014)
One	-0.018
	(0.018)
Two	-0.034
	(0.023)
Three	-0.018
	(0.029)
Four	-0.021
	(0.028)
Five	0.022
	(0.035)
Six	-0.061
	(0.049)
≥Seven	-0.108**
	(0.045)
F-Test equal effects [p-value]	[.229]
R-Squared	0.51
Controls	Yes
Course combination FE	Yes
Instructors	485
Observations	41,619

This table shows average marginal effects from a fully interacted OLS course-by-parallel-course fixed effects regression. The dependent variable is student course grade standardized over the estimation sample. All specifications include a PhD student instructor dummy, making senior instructors the comparison group. Controls include scheduling controls, tutorial-level controls, and student-level controls (see Section 3 for details). Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1