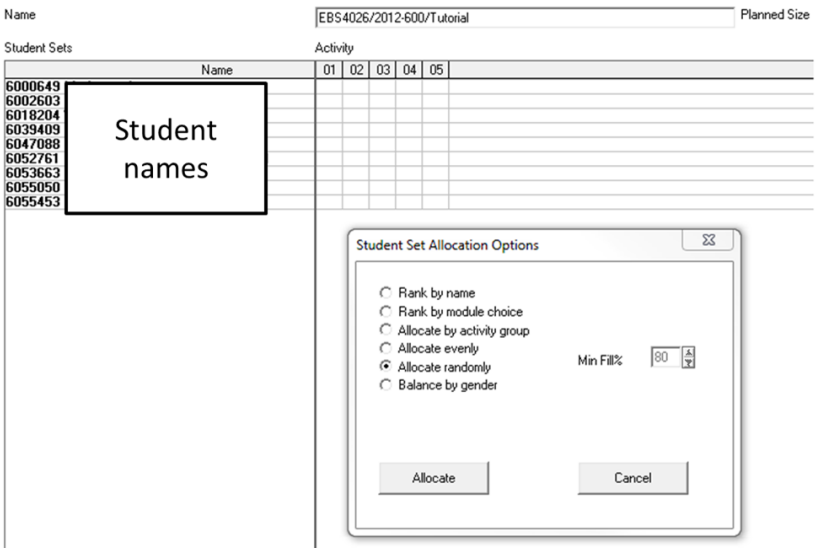
Online Appendix

Gender Bias in Teaching Evaluations (Friederike Mengel, Jan Sauermann, and Ulf Zölitz)

(September 1, 2017)

Appendix A: Figures

Figure A1: Screenshot of the scheduling software used by the SBE Scheduling Department



Note: This screenshot shows the program Syllabus Plus Enterprise Timetable.

Appendix B: Tables

Table B1: Gender bias in instructor evaluation – courses without course papers as part of assessment

	(1)
Female instructor (β_1)	-0.2443***
	(0.0399)
Female student (β_2)	-0.1209***
	(0.0261)
Female instructor * Female student (β_3)	0.1661***
	(0.0439)
Constant	0.5718**
	(0.2458)
Observations	11,014
R-squared	0.2023
$\beta_1 + \beta_3$	-0.0783*
	(0.0467)

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Instructor evaluation. All regressions include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses.

Table B2: Split sample regressions by student gender

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Instructor evaluation	Group-related	Material-related	Course-related	Hours spent	Final grade
Female students only						
Female instructor	-0.0611	0.0182	-0.0180	0.0048	-0.1787	0.0153
	(0.0394)	(0.0332)	(0.0284)	(0.0272)	(0.2297)	(0.0332)
Constant	0.2355	-0.2477	-0.5256	-1.3169**	10.3959	0.3178
	(0.4711)	(0.5204)	(0.3645)	(0.5684)	(6.6159)	(0.7396)
Observations	8,673	8,673	8,673	8,673	8,673	8,673
R-squared	0.2547	0.2232	0.3025	0.3066	0.2888	0.5642
Male students only						
Female instructor	-0.2099***	-0.0624**	-0.0634**	-0.0753***	0.0676	0.0300
	(0.0324)	(0.0275)	(0.0250)	(0.0247)	(0.1822)	(0.0327)
Constant	-0.4334	0.1020	0.8695*	0.0600	9.5223	2.2006***
	(0.7079)	(0.3236)	(0.4608)	(0.5945)	(7.2705)	(0.8279)
Observations	11,279	11,279	11,279	11,279	11,279	11,279
R-squared	0.2326	0.2022	0.2598	0.2814	0.3102	0.5071

Note: *** p<0.01, ** p<0.05, * p<0.1 All regressions include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses.

Table B3: Evaluations of graduate student instructors – by separate items

	(1)	(2)	(3)	(4)	(5)
Evaluation item	T1	T2	T3	T4	T5
Female instructor (β_1)	-0.2180***	-0.2445***	-0.1420**	-0.1913***	-0.1768***
	(0.0668)	(0.0598)	(0.0555)	(0.0627)	(0.0521)
Female student (β_2)	-0.0576	-0.0039	-0.0449	-0.0406	-0.0585
	(0.0408)	(0.0396)	(0.0381)	(0.0382)	(0.0373)
Female instructor * Female student (β_3)	0.0332	-0.0598	-0.0384	-0.0740	-0.0109
	(0.0655)	(0.0622)	(0.0579)	(0.0618)	(0.0573)
Observations	5,340	5,337	5,323	5,346	5,270
R-squared	0.2537	0.2559	0.2302	0.2475	0.2809
$\beta_1 + \beta_3$	-0.185***	-0.304***	-0.180***	-0.265***	-0.188***
	(0.0711)	(0.0663)	(0.0611)	(0.0701)	(0.0603)

Note: *** p<0.01, ** p<0.05, * p<0.1. All estimates are based on regressions which include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, nationality, field of study, age). The sample used in this regression includes graduate student instructors only. Robust standard errors clustered at the section level are in parentheses.

Table B4: Gender bias in students' evaluations – by variation in response items

	(1) Low Dispersion (SD ≤ median)	(2) High Dispersion (SD > median)
Female instructor (β_1)	-0.1718***	-0.2283***
	(0.0301)	(0.0478)
Female student (β_2)	-0.0544***	-0.1690***
	(0.0209)	(0.0310)
Female instructor * Female student (β_3)	0.0722*	0.1756***
	(0.0375)	(0.0542)
Constant	-0.5122	0.2878
	(0.4368)	(0.4536)
Observations	9,992	9,960
R-squared	0.2429	0.2583
$\beta_1 + \beta_3$	-0.0996***	-0.0527
	(0.0351)	(0.0526)

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Instructor evaluation. For defining individuals as "low dispersion" and "high dispersion," we calculated the standard deviation of a student's answers across all evaluation items within his or her evaluation sheet. Low dispersion (high dispersion) is defined as evaluations with below-median (above-median) standard deviation. All regressions include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses.

Table B5: Main results – excluding course coordinators

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Instructor-related	Group-related	Material-related	Course-related	Hours spent	Final grade
Female instructor (β_1)	-0.2223***	-0.0495*	-0.0538**	-0.0636***	0.0437	0.0069
	(0.0338)	(0.0278)	(0.0244)	(0.0242)	(0.1814)	(0.0316)
Female student (β_2)	-0.1218***	-0.0015	-0.0322	-0.0399**	1.4260***	-0.0215
	(0.0206)	(0.0211)	(0.0196)	(0.0192)	(0.1609)	(0.0242)
Female instructor * Female student(β_3)	0.1192***	0.0192	0.0167	0.0469	-0.1023	0.0402
	(0.0350)	(0.0337)	(0.0319)	(0.0313)	(0.2562)	(0.0428)
Observations	16,807	16,807	16,807	16,807	16,807	16,807
R-squared	0.1945	0.1527	0.2179	0.2290	0.2553	0.5082
$\beta_1+\beta_3$	-0.103***	-0.0303	-0.0372	-0.0167	-0.0586	0.0471
	(0.0380)	(0.0314)	(0.0267)	(0.0259)	(0.209)	(0.0328)

Note: *** p<0.01, ** p<0.05, * p<0.1.All regressions include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses. Control variables refer to students' characteristics.

Table B6: Determinants of survey response

	(1)	(2)	(3)	(4)	(5)
Female instructor (β_1)		-0.0003	-0.0067	-0.0067	-0.0083
		(0.0044)	(0.0052)	(0.0053)	(0.0060)
Female student (β_2)	0.0864***	0.0864***	0.0804***	0.0739***	0.0579***
	(0.0037)	(0.0037)	(0.0046)	(0.0048)	(0.0054)
Female instructor * Female student (β_3)			0.0170**	0.0174**	0.0181**
			(0.0076)	(0.0078)	(0.0090)
Grade (first sit)					0.0167***
					(0.0015)
GPA					0.0437***
					(0.0023)
German				0.0636***	0.0171***
				(0.0045)	(0.0052)
Other nationality				0.0710***	0.0627***
				(0.0057)	(0.0067)
Economics				-0.0140	-0.0063
				(0.0124)	(0.0135)
Other study field				0.0782***	0.0809***
				(0.0196)	(0.0248)
Age				-0.0004	0.0080***
				(0.0011)	(0.0014)
Section size				0.0004	0.0009
				(0.0016)	(0.0018)
Constant	0.3305***	0.3306***	0.3328***	0.6316***	0.0610
	(0.0021)	(0.0026)	(0.0028)	(0.2161)	(0.1294)
Observations	75,330	75,330	75,330	72,376	55,856
R-squared	0.0580	0.0580	0.0580	0.0790	0.0878
$\beta_1+\beta_3$			0.0103	0.0107	0.00985
			(0.00659)	(0.00675)	(0.00758)

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Dummy variable for survey response. All regressions include course fixed effects and parallel course fixed effects for the courses taken at the same time. Robust standard errors clustered at the section level are in parentheses.

Table B7: Selection of students into response (Heckman selection model)

	(1)	(2)	(3)	(4)	(5)
	Model 1		Model 2		Baseline
	Instructor evaluation	Response	Instructor evaluation	Response	Instructor evaluation
Female instructor (β_1)	-0.2190***	-0.0234	-0.2194***	-0.0243	-0.2185***
	(0.0299)	(0.0172)	(0.0300)	(0.0192)	(0.0305)
Female student (β_2)	-0.1160***	0.1666***	-0.1260***	0.0740***	-0.1191***
	(0.0175)	(0.0146)	(0.0178)	(0.0162)	(0.0179)
Female instructor * Female student (β_3)	0.1380***	0.0511**	0.1374***	0.0519*	0.1371***
	(0.0312)	(0.0246)	(0.0316)	(0.0271)	(0.0318)
Mean past response				1.7841***	
				(0.0184)	
Constant	0.1400	-1.9086***	0.2830	-2.1331***	0.1985
	(0.1999)	(0.1044)	(0.2067)	(0.1188)	(0.2030)
ρ	0.0295**		-0.0497***		
	(0.0141)		(0.0187)		
ln σ	-0.0626***		-0.0608***		
	(0.0081)		(0.0082)		
Observations	55,856		54,530		19,952
Pseudo R-squared	0.0573		0.2331		
R-squared					0.1682
$\beta_1+\beta_3$	-0.0809**		-0.0820**		-0.0814**
	(0.0335)		(0.0337)		(0.0341)

Note: *** p<0.01, ** p<0.05, * p<0.1. All regressions include course fixed effects; the regression shown in Column (5) also includes parallel course fixed effects for the courses taken at the same time. Column (5) also includes individual FE. Robust standard errors clustered at the section level are in parentheses. All regressions include course fixed effects, section size and students' characteristics (GPA, grade, nationality, field of study, age). Due to the large number of dummy variables, the regressions presented in this table do not contain parallel course fixed effects for the courses taken at the same time. Control variables refer to students' characteristics.

Table B8: Instructor gender and instructor characteristics

	(1) Female instructor
PhD Student	0.0265
	(0.1013)
Lecturer	0.1034
	(0.1098)
Professor	0.0101
	(0.1116)
Age	-0.0113***
	(0.0032)
Non-Dutch	0.0695
	(0.0538)
Full-time	-0.1269**
	(0.0644)
Research fellow	-0.0331
	(0.0741)
Constant	0.7348***
	(0.1332)
Observations	377
R-squared	0.0921

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Female instructor. Omitted category: student instructors. Standard errors are in parentheses.

Table B9: Effect of instructors gender on students' study hours for male students (β_1; Panel 1) and female students ($\beta_1 + \beta_3$; Panel 2) depending on instructor and student seniority

	→ Increasing Instructor Seniority →				
	Student	PhD student	Lecturer	Professor	Overall
	Panel 1: Male Students (β_1)				
1st year Bachelor	-.4427	-.9951	.7791	-.7783	-.1223
2nd year Bachelor and higher	.6486	-1.638**	.2562	.3307	.0561
Master	.9005	.8763	.2837	.2739	.2381
Overall	.0422	-.5641	.5847*	.3553	.0443
	Panel 2: Female Students ($\beta_1 + \beta_3$)				
1st year Bachelor	-.5078	.8947	1.0327	-3.6357	.0068
2nd year Bachelor and higher	.0287	.6519	-1.2892**	-.6845	-.1887
Master	2.2919	-.5425	-.101	1.9685	.2086
Overall	-.1798	.1756	-.0659	.7007	-.0393
	Panel 3: Number of observations				
1st year Bachelor	2,183	1,218	1,634	307	5,342
2nd year Bachelor and higher	2,515	1,876	2,659	1,505	8,555
Master	654	1,707	1,407	2,287	6,055
Overall	5,352	4,801	5,700	4,099	19,952

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Students' study hours. All estimates are based on regressions which include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses.

Table B10: Effect of instructors gender on grades for male students (β_1; Panel 1) and female students ($\beta_1 + \beta_3$; Panel 2) depending on instructor and student seniority

	→ Increasing Instructor Seniority →				
	Student	PhD student	Lecturer	Professor	Overall
	Panel 1: Male Students (β_1)				
1st year Bachelor	-.0218	-.0201	.0067	.0849	-.0119
2nd year Bachelor and higher	.0791	.0359	-.0057	.0337	.0681
Master	.245	.0469	-.5009***	-.0168	-.0788
Overall	.0419	.0241	-.092	.0751	.0109
	Panel 2: Female Students ($\beta_1 + \beta_3$)				
1st year Bachelor	.0788	-.0383	-.1035	-.2202	-.0091
2nd year Bachelor and higher	.1210	-.1954	.0582	.0515	.0546
Master	.0900	-.0157	-.1449	.1882	.0188
Overall	.1000*	-.0795	.0123	.1163	.0397
	Panel 3: Number of observations				
1st year Bachelor	2183	1,218	1,634	307	5,342
2nd year Bachelor and higher	2,515	1,876	2,659	1,505	8,555
Master	654	1,707	1,407	2,287	6,055
Overall	5,352	4,801	5,700	4,099	19,952

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Course grades. All estimates are based on regressions which include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses.

Table B11: Value added, instructor gender, and students' evaluations

	(1)	(2)	(3)	(4)
Female instructor	-0.0380	-0.0113		
	(0.0511)	(0.0515)		
Students' evaluations			0.0142	0.0051
			(0.0386)	(0.0385)
Constant	0.0856***	0.0260	0.0729***	0.0187
	(0.0307)	(0.0417)	(0.0249)	(0.0367)
Instructor seniority Controls	NO	YES	NO	YES
Observations	690	688	688	687
R-squared	0.0008	0.0185	0.0002	0.0189

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Teacher value added. Standard errors are in parentheses. Unit of observation: instructor level.

Table B12: Estimates of gender bias in students' evaluations of male students (β_1; Panel 1) and female students ($\beta_1 + \beta_3$; Panel 2) depending on instructor and student seniority

	→ Increasing Instructor Seniority →				
	Student	PhD student	Lecturer	Professor	Overall
	Panel 1: Male Students (β_1)				
1st year Bachelor	-.1317	-.3521**	-.1072	.1001	-.1275**
2nd year Bachelor and higher	-.3478***	.1518	-.0322	.1404	-.2404***
Master	-.4691**	-.6316***	.204	-.0478	-.2507***
All students	-.2379***	-.2798***	-.0392	.085	-.2069***
	Panel 2: Female Students ($\beta_1 + \beta_3$)				
1st year Bachelor	-.1537	-.2629	-.0403	.4645	-.0607
2nd year Bachelor and higher	-.4016***	.2286*	.1934*	.3941	-.0701
Master	-.5383**	-.4601***	.3482	.0787*	-.1179*
All students	-.274***	-.1359	.1232*	.2583**	-.076**
	Panel 3: Number of observations				
1st year Bachelor	2,183	1,218	1,634	307	5,342
2nd year Bachelor and higher	2,515	1,876	2,659	1,505	8,555
Master	654	1,707	1,407	2,287	6,055
All students	5,352	4,801	5,700	4,099	19,952

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Instructor evaluation. All estimates are based on regressions which include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses. The full table with student seniority can be found in the Online Appendix (Table **??**).

Table B13: Gender bias in instructor evaluation – by student's course grade

	(1)	(2)	(3)	(4)
Student grades	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Female instructor (β_1)	-0.1788***	-0.2061***	-0.2102***	-0.1969***
	(0.0471)	(0.0539)	(0.0621)	(0.0719)
Female student (β_2)	-0.0914***	-0.0805**	-0.2042***	-0.1272**
	(0.0337)	(0.0382)	(0.0456)	(0.0584)
Female instructor * Female student (β_3)	0.0527	0.1307*	0.1884**	0.1152
	(0.0602)	(0.0672)	(0.0773)	(0.0986)
Constant	0.3489	0.9507**	0.0746	-0.8966
	(0.6040)	(0.4142)	(0.6777)	(0.7197)
Observations	7,004	5,238	4,548	3,162
R-squared	0.2776	0.2933	0.3068	0.3374
$\beta_1 + \beta_3$	-0.126**	-0.0753	-0.0219	-0.0817
	(0.0565)	(0.0596)	(0.0647)	(0.0855)

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Instructor evaluation. Quartiles are based on the student's grade in the course and are calculated at the course level. All regressions include course fixed effects, parallel course fixed effects for the courses taken at the same time, section size and other control variables for students' characteristics (GPA, grade, nationality, field of study, age). Robust standard errors clustered at the section level are in parentheses.