

# Supplemental Material

*CBE—Life Sciences Education*

Bailey *et al.*

**Table S1. Selection of random effects to predict gender gaps in verbal participation.**

Rank	Model <sup>a</sup>	Converged?	AICc	$\Delta i$	Best Model <sup>b</sup>
1	(1 Classroom)	Yes	167.98	0	*
2	(1 Classroom) + (1 Semester)	No	170.04	2.06	
3	No random effects	Yes	170.86	2.88	
4	(1 Semester)	Yes	172.87	4.89	

<sup>a</sup> Although not shown, models include all fixed effects of interest: Class.Size + Upper.Level + Fem.Inst + Perc.Fem

<sup>b</sup> As described in Methods, if models were within 2 AICc, the model with the fewest number of parameters was chosen as the best model.

**Table S2. Selection of fixed effects to predict gender gaps in verbal participation.**

Rank	Model <sup>a</sup>	AICc	$\Delta i$	$\omega_i$	Best Model <sup>b</sup>
1	Perc.Fem	154.64	0	0.28	*
2	Perc.Fem + Upper.Level	156.39	1.75	0.12	
3	Perc.Fem + Class.Size	156.69	2.06	0.10	
4	Perc.Fem + Fem.Inst	156.86	2.22	0.09	
5	No fixed effects	157.08	2.44	0.08	
6	Perc.Fem + Upper.Level + Class.Size	157.75	3.11	0.06	
7	Upper.Level	158.37	3.73	0.04	
8	Perc.Fem + Upper.Level + Fem.Inst	158.65	4.01	0.04	
9	Upper.Level + Class.Size	158.73	4.09	0.04	
10	Class.Size	158.85	4.21	0.03	
11	Perc.Fem + Class.Size + Fem.Inst	158.95	4.32	0.03	
12	Fem.Inst	159.19	4.55	0.03	
13	Perc.Fem + Upper.Level + Class.Size + Fem.Inst	160.05	5.41	0.02	
14	Upper.Level + Fem.Inst	160.48	5.84	0.02	
15	Upper.Level + Class.Size + Fem.Inst	160.99	6.35	0.01	
16	Class. Size + Fem.Inst	161.05	6.41	0.01	

<sup>a</sup> Class.Size = # students attending class, Upper.Level = 300- or 400- level class, Fem.Inst = female instructor, Perc.Fem = % females in attendance. Although not shown, all models also include a random effect to allow for a random intercept for each class: (1|Classroom)

<sup>b</sup> As described in Methods, if models were within 2 AICc, the model with the fewest number of parameters was chosen as the best model.

**Table S3. Selection of random effects to predict gender gaps in verbal participation (controllable variables).**

Rank	Model <sup>a</sup>	Converged?	AICc	$\Delta i$	Best Model <sup>b</sup>
1	(1 Classroom)	Yes	166.15	0	*
2	(1 Classroom) + (1 Semester)	No	168.09	1.94	
3	no random effects	Yes	169.65	3.50	
4	(1 Semester)	No	171.72	5.57	

<sup>a</sup> Although not shown, models include all fixed effects of interest: Part.Req + Instructor.Ques + Group.Work + Ov.Call.Rate

<sup>b</sup> As described in Methods, if models were within 2 AICc, the model with the fewest number of parameters was chosen as the best model.

**Table S4. Selection of fixed effects to predict gender gaps in verbal participation (controllable variables).**

Rank	Model <sup>a</sup>	AICc	$\Delta i$	$\omega_i$	Best Model <sup>b</sup>
1	Ov.Call.Rate	152.44	0.00	0.24	*
2	Ov.Call.Rate + Group.Work	152.80	0.37	0.20	
3	Ov.Call.Rate + Group.Work + Instructor.Ques	153.62	1.18	0.13	
4	Ov.Call.Rate + Instructor.Ques	153.97	1.53	0.11	
5	Ov.Call.Rate + Part.Req	154.35	1.91	0.09	
6	Ov.Call.Rate + Group.Work + Part.Req	155.04	2.60	0.07	
7	Ov.Call.Rate + Instructor.Ques + Part.Req	155.76	3.33	0.05	
8	Ov.Call.Rate + Group.Work + Instructor.Ques + Part.Req	155.85	3.41	0.04	
9	No fixed effects	157.08	4.64	0.02	
10	Instructor.Ques	158.30	5.86	0.01	
11	Group.Work	158.81	6.37	0.01	
12	Part.Req	159.04	6.60	0.01	
13	Group.Work + Instructor.Ques	159.66	7.22	0.01	
14	Instructor.Ques + Part.Req	160.15	7.71	0.01	
15	Group.Work + Part.Req	160.95	8.51	0.00	
16	Group.Work + Instructor.Ques + Part.Req	161.78	9.34	0.00	

<sup>a</sup> Ov.Call.Rate = overall call rate (average # hands called on/# hands raised per student), Instructor.Ques = # questions posed by instructor to the class, Part.Req = degree to which classroom participation is required in students' grades (see Methods), Group.Work = # of times students worked in pairs or groups. Although not shown, all models also include a random effect to allow for a random intercept for each class: (1|Classroom)

<sup>b</sup> As described in Methods, if models were within 2 AICc, the model with the fewest number of parameters was chosen as the best model.

**Table S5. Results of stepwise multiple linear regression with performance gap as target (including Instructor F8).**

Model	R <sup>2</sup>	Adjusted R <sup>2</sup>	Significance (change in R <sup>2</sup> )	Variable	B (coefficient)	SE <sub>B</sub>	$\beta$ (standardized coefficient)	p value
1	0.228	0.202	0.007	(Intercept)	-0.435	0.129		0.002
				Percent female in attendance	0.821	0.280	0.478	0.007
2	0.379	0.334	0.015	(Intercept)	-0.473	0.119		<0.0001
				Female Instructor	0.200	0.077	0.394	0.015
				Percent female in attendance	0.700	0.260	0.408	0.012

**Table S6. Results of multiple linear regression to predict male and female performance separately**

Target	R <sup>2</sup>	Adjusted R <sup>2</sup>	Significance (change in R <sup>2</sup> )	Variable	B (coefficient)	SE <sub>B</sub>	β (standardized coefficient)	p value
Avg. Female Grade	0.295	0.242	0.009	(Intercept)	2.754	0.167		<0.0001
				Female Instructor	0.231	0.110	0.345	0.045
				Percent female in attendance	0.818	0.366	0.366	0.034
Avg. Male Grade	0.003	-0.070	0.955	(Intercept)	3.248	0.140		<0.0001
				Female Instructor	-0.022	0.092	-0.047	0.811
				Percent female in attendance	0.067	0.306	0.043	0.828

**Table S7. Selection of random effects to predict student performance (course grade z-scores).**

Rank	Model <sup>a</sup>	Converged?	AICc	Δi	Best Model <sup>b</sup>
1	(1 Classroom)	Yes	5329.60	0	*
2	(1 Classroom) + (1 Semester)	No	5331.60	2.01	
3	(1 Semester)	Yes	5375.92	46.32	
4	No Random Effects	Yes	5376.07	46.47	

<sup>a</sup> Although not shown, models include all fixed effects of interest: ACT + Fem.Inst + Fem.Stud + Fem.Inst\*Fem.Stud

**Table S8. Selection of fixed effects to predict student performance (course grade z-scores).**

Rank	Model <sup>a</sup>	AICc	Δi	ω <sub>i</sub>	Best Model <sup>b</sup>
1	ACT + Fem.Inst*Fem.Stud	5316.7	0	0.237	*
2	ACT + Fem.Stud + Fem.Inst*Fem.Stud	5316.7	0	0.237	
3	ACT + Fem.Inst + Fem.Inst*Fem.Stud	5316.7	0	0.237	
4	ACT + Fem.Inst + Fem.Stud + Fem.Inst*Fem.Stud	5316.7	0	0.237	
5	ACT + Fem.Stud	5321.6	4.9	0.020	
6	ACT + Fem.Inst + Fem.Stud	5322.0	5.3	0.017	
7	ACT	5323.5	6.9	0.008	
8	ACT + Fem.Inst	5324.0	7.3	0.006	

<sup>a</sup> ACT = ACT score, Fem.Inst = Female Instructor, Fem.Stud = Female Student, Fem.Inst\*Fem.Stud = Interaction between instructor gender and student gender. Only models with Δi < 10 are shown (in fact, the next best model has a Δi = 164).

Although not shown, all models also include a random effect to allow for a random intercept for each class: (1|Classroom).

<sup>b</sup> As described in Methods, if models were within 2 AICc, the model with the fewest number of parameters was chosen as the best model.