

## **SUPPLEMENTARY MATERIAL**

**Title: Do small classes reduce performance gaps in STEM?**

### **This file includes:**

**Tables S1-S3:** Summary results tables

**Tables S4-S6:** University of Minnesota (UMN) results

**Table S7:** Interactions by class size

**Figure S1:** Interactions by class size and university

**Supplemental Tables S1-S3.** Statistical results for performance metrics and accounting for potential demographic predictors across seventeen introductory biology courses in fall 2016. We used AIC model-selection statistics to determine variables to include in the models. All independent and dependent continuous variables are normalized so the coefficients show effects of independent variables' effects in terms of units of standard deviations.

**Table S1. Mixed-effect regression analysis of predictors of *exam* performance in introductory biology courses.**

Parameter Estimates					
	Estimate	Std. Error	df	t	Pr>t
Intercept	-0.223	0.081	13.332	-2.757	0.016
Gender	-0.146	0.048	1723.087	-3.046	0.002
Class size	0.071	0.039	739.098	1.815	0.070
URM	0.417	0.069	1377.147	6.012	<0.001
Gender x Class size	-0.145	0.050	1598.692	-2.894	0.004

**Table S2. Mixed-effect regression analysis of predictors of performance on *non-exam* points in introductory biology courses.**

Parameter Estimates					
	Estimate	Std. Error	df	t	Pr>t
Intercept	-0.279	0.084	16.082	-3.331	0.004
Gender	0.217	0.047	1731.752	4.605	<0.001
URM	0.262	0.069	1533.219	3.831	<0.001

**Table S3. Mixed-effect regression analysis of predictors of *total course* performance in introductory biology courses.**

Parameter Estimates					
	Estimate	Std. Error	df	t	Pr>t
Intercept	-0.282	0.085	14.431	-3.295	0.005
Gender	0.012	0.048	1727.616	0.243	0.808
Class size	0.060	0.039	868.006	1.526	0.127
URM	0.407	0.069	1521.797	5.869	<0.001
Gender x Class size	-0.108	0.050	1648.997	-2.160	0.031

### *University of Minnesota (UMN) data results*

Using AIC inference techniques, we re-analyzed the UMN data to find the best fit model that predicts exam performance, *other* performance, and whole class performance based on gender and underrepresented minority (URM) status. We chose the models that best fit the data in accordance to AIC model-selection statistics. All models also included the course ID as a random effect. For the UMN data we possess measures of incoming preparation (ACT), but did not include them to keep our analyses consistent with the larger dataset. However, when we do include them the trends are the same, with class size significantly and negatively affecting female performance.

We found the same trends as those listed in the main document. That is, as class size increased, women performed worse on exams ( $SGender \cdot class\ size\ B = -0.122$ ,  $t(1437) = -2.31$ ,  $P = 0.021$ ,  $SE = 0.053$ ) and in the total course score ( $B = -0.117$ ,  $t(1436) = -2.21$ ,  $P = 0.027$ ,  $SE = 0.053$ ). We also found that women overall performed better on non-exam assessments ( $B = -0.259$ ,  $t(1439) = 4.914$ ,  $P < 0.001$ ,  $SE = 0.053$ ), regardless of class size.

**Table S4-S6.** Statistical results for performance metrics from University of Minnesota, a subset of the whole dataset.

**Table S1.** Mixed-effect regression analysis of predictors of *exam* performance in introductory biology courses.

Parameter Estimates					
	Estimate	Std. Error	df	t	Pr>t
Intercept	-0.222	0.084	1437	-2.648	0.008
Gender	-0.153	0.053	1437	-2.906	0.004
Class size	0.073	0.039	1437	1.853	0.064
URM	0.348	0.083	1437	4.190	<0.001
Gender x Class size	-0.122	0.053	1437	-2.313	0.021

**Table S2.** Mixed-effect regression analysis of predictors of performance on *non-exam* points in introductory biology courses.

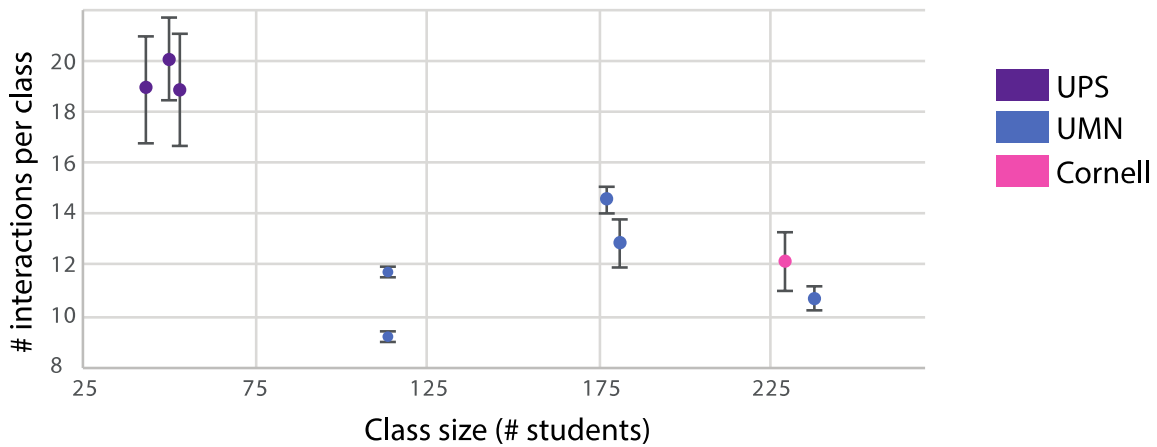
Parameter Estimates					
	Estimate	Std. Error	df	t	Pr>t
Intercept	-0.320	0.084	1439	-3.814	<0.001
Gender	0.259	0.053	1439	4.914	<0.001
Class size	0.190	0.083	1439	2.283	0.023

**Table S3.** Mixed-effect regression analysis of predictors of *total course* performance in introductory biology courses.

Parameter Estimates					
	Estimate	Std. Error	df	t	Pr>t
Intercept	-0.300	0.084	1436	-3.566	<0.001
Gender	0.027	0.053	1436	0.506	0.613
Class size	0.046	0.041	1436	1.122	0.262
URM	0.320	0.083	1436	3.846	<0.001
Gender x Class size	-0.117	0.053	1436	-2.212	0.027
age	-0.053	0.028	1436	-1.916	0.056

**Table S7.** Average (SE) interactions per 50-minute class period in nine of the seventeen courses in the current study.

Class code	Class size	University	Instructor gender	Classes observed	Interactions adjusted for 50 minute class periods	SE
UMN1	239	UMN	Multiple instructors	8	10.58	0.523
Cornell1	230	Cornell	Multiple instructors	17	12.00	1.330
UMN2	229	UMN	Multiple instructors	-	-	-
UMN3	182	UMN	Male	6	12.78	1.078
UMN4	178	UMN	Multiple instructors	5	14.40	0.559
UMN5	153	UMN	Multiple instructors	-	-	-
UMN6	115	UMN	Female	14	9.10	0.190
UMN7	115	UMN	Female	16	11.67	0.217
UMN8	95	UMN	Male	-	-	-
UMN9	90	UMN	Male	-	-	-
UMN10	87	UMN	Female	-	-	-
UMN11	87	UMN	Multiple instructors	-	-	-
CSU1	87	CSU	Female	-	-	-
UMN12	73	UMN	Female	-	-	-
UPS1	52	UPS	Male	12	18.73	2.530
UPS2	51	UPS	Female	14	20.00	1.870
UPS3	44	UPS	Male	12	18.72	2.410



**Figure S1.** Mean number of interactions per 50-minute class period by class size across University of Puget Sound (UPS; purple), University of Minnesota (UMN; blue), and Cornell University (pink).