Supplemental Material

CBE—Life Sciences Education

Ballen *et al.*

SUPPLEMENTARY MATERIAL

Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning

Authors: Cissy J. Ballen[†‡*], Carl Wieman[§‖], Shima Salehi[§], Jeremy B. Searle[†], Kelly R. Zamudio[†]
[†]Department of Ecology & Evolutionary Biology, Cornell University, Ithaca, NY 14853
[‡]Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, MN 55455
[§]Graduate School of Education, Stanford University, Stanford, CA 94305
[‖]Department of Physics, Stanford University, Stanford, CA 94305

* balle027@umn.edu

This PDF file includes:

Assessments S1 to S3
Figure S1 and S2
Full statistical models and result tables S1 to S6
Mediation analyses

Other supplementary material for this manuscript includes the following:

Raw data as an excel table

Assessment S1. 30-question content knowledge assessment covering semester content

PRE-/POST-COURSE ASSESSMENT
EVOLUTIONARY BIOLOGY & BIODIVERSITY BIOEE1780

1. Who first published the concept that organisms evolve by natural selection?

A. Darwin was first
B. Wallace was first
C. Darwin and Wallace published at the same time
D. Lamarck was first

2. Below are data on four female lizards collected over their lifetime:

	Lizard **A**	Lizard **B**	Lizard **C**	Lizard **D**
Body length (cm)	20	12	10	15
Total offspring surviving to adulthood	19	28	22	26
Age at death (years)	4	5	4	6
Comments	Lizard **A** has a low parasite load and quick reaction time.	Lizard **B** has red coloration and has short legs and tail.	Lizard **C** has dark coloration and is a fast runner.	Lizard **D** has a large territory and has mated with many males.

Which lizard would be considered the most fit, in an evolutionary sense?

A. Lizard **A**
B. Lizard **B**
C. Lizard **C**
D. Lizard **D**

3. Directional selection differs from stabilizing selection in that:

A. Directional selection operates only in small populations whereas stabilizing selection is effective in both small and large populations
B. Directional selection favors intermediate over extreme phenotypes, whereas stabilizing selection favors one end of the phenotype distribution
C. Directional selection favors one end of the phenotype distribution, whereas stabilizing selection favors intermediate over extreme phenotypes
D. Directional selection requires new mutations whereas stabilizing selection operates on existing variation
E. Directional selection operates on existing variation, whereas stabilizing selection operates on new mutations

4. How long does it take for organisms to evolve?

A. All organisms - millions of years

B. Large organisms – millions of years, smaller organisms - more quickly
C. If there is a genetic change in a population over a few generations, that is considered evolution
D. Evolution can be fast, but there needs to be substantial genetic change before it can be considered evolution

5. Which of the following statements about evolution is true?

A. Evolution happens for the good of the species
B. Evolution always goes from simple to complex
C. Similar evolutionary trends can occur in different species
D. Evolution always involves at least one of the following: natural selection, artificial selection, and/or sexual selection

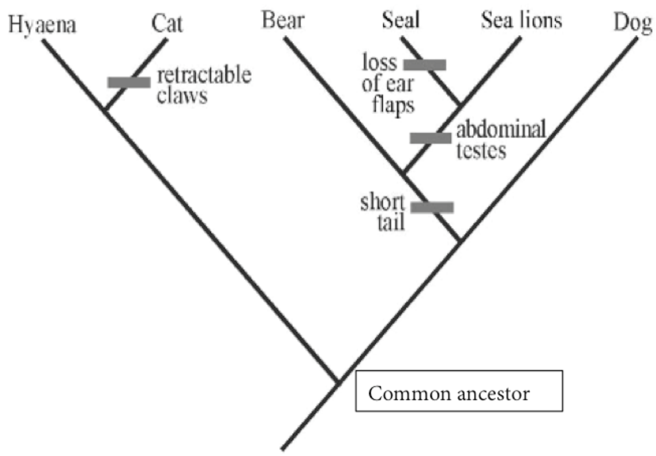
6. A species of guppy (small fish), found in streams in Venezuela, has males that exhibit a genetically determined range of coloration from brightly colored to dull gray. Brightly colored males are seen easily and consumed by predators, but plain males are not chosen by females for mating. In a stream with no predators, the proportion of brightly colored males is high. If a few predators are added to the same stream, the proportion of brightly colored males decreases within five months (3-4 generations).

What best accounts for the changes in male coloration after predators are added?

A. The traits of some guppies gradually changed as they encountered predators, so that they became less brightly colored. The offspring of these individuals inherited these changes.
B. When predators were introduced into the stream, the proportion of guppies with genes that cause dull coloration increased because those genes also control strength and speed.
C. The presence of predators led to mutations in the genes responsible for coloration, and, as a result, the proportion of brightly colored males gradually decreased.
D. The proportion of guppies with genes that cause bright coloration decreased due to strong selection by predators, even though females still preferred brightly colored mates.

7. To best maximize transmission of your genes, list the order of investment (from largest to smallest) you should make to a half-sister, a female cousin, and a niece, assuming they are the same age.

A. Cousin, niece, half-sister.
B. Equal parts to the half-sister and niece, less to the cousin.
C. Most to the half-sister, less but equal parts to the niece and cousin.
D. Half-sister, niece, cousin.



8. In the tree above, assume that the ancestor had a long tail, ear flaps, external testes, and fixed claws. Based on the tree and assuming that all evolutionary changes in these traits are shown, what traits does a sea lion have?

A. long tail, ear flaps, external testes, fixed claws
B. short tail, no ear flaps, external testes, fixed claws
C. short tail, no ear flaps, abdominal testes, fixed claws
D. short tail, ear flaps, abdominal testes, fixed claws
E. long tail, ear flaps, abdominal testes, retractable claws

9. Considering: (1) The extinction of the dinosaurs; (2) The Cenozoic era; (3) The Mesozoic era; and (4) The Paleozoic era, which of the following is placed in the correct chronological order, oldest first?

A. 1, 2, 3, 4
B. 4, 1, 2, 3
C. 3, 2, 4, 1
D. 4, 3, 1, 2

10. Which statement about HOX genes is true?

A. A shared set of HOX genes is involved in the development of all invertebrate life forms, but not in vertebrate taxa.
B. HOX gene alleles frequencies changed rapidly during the flu epidemic of 1918 and now provide a famous example of the evolution disease resistance in humans.
C. The expansion and diversification of the HOX gene cluster helped facilitate the evolution of body plan complexity in animals.
D. The high mutation rates inherent to HOX genes often result in hybrid sterility and eventual speciation.

11. In mammals and birds, how is it thought that new species normally arise?

A. A mutation leads to a new species in the same geographic area as the old species, such that you have the old species and new species in the same area.
B. The old species subdivides into two geographically separate populations each of which becomes a new species through the accumulation of mutations.
C. No genetic change is necessary. The environment in one part of the range of the old species leads to changed characteristics and a new species, with the old species also continuing.
D. In mammals and birds the variety of species that we currently see were formed a long time ago, and processes leading to the formation of new species has halted.

12. Which of the following crosses generates offspring which are all identical heterozygotes?

A. Cross between identical heterozygotes
B. Cross between identical homozygotes
C. Cross between different heterozygotes
D. Cross between different homozygotes

13. An allele is:

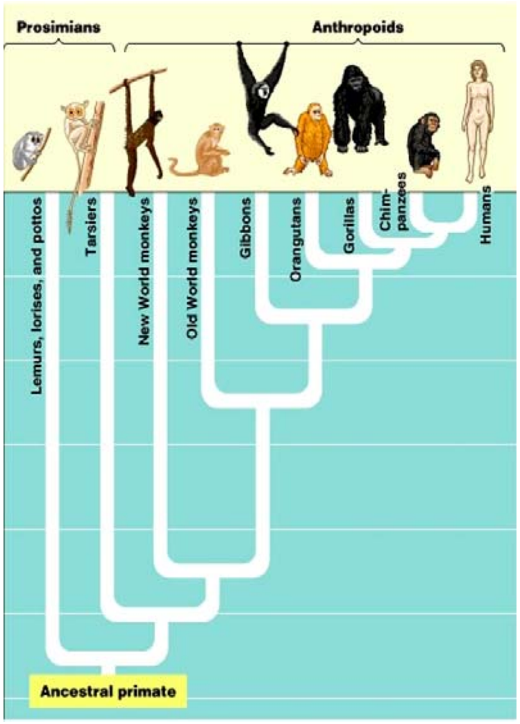
A. A gene that is favored by natural selection
B. Part of an enzyme where a substrate is cleaved
C. A version of a gene
D. A feature of an organism important for its survival

14. Which of the following is *false*?

A. Small populations are expected to lose genetic variants more rapidly than large populations
B. Small populations show higher rates of mutation than large populations
C. Small populations show chance changes in frequency of variants more than large populations
D. Small populations are more likely to show inbreeding than large populations

15. If there is a population of a species where individuals routinely reproduce at a younger age than individuals elsewhere, what would you expect of the total life span in that population?

A. They would be expected to have a longer life span than in a population where individuals reproduce late
B. They would be expected to have a shorter life span than in a population where individuals reproduce late
C. They would be expected to have the same life span as in a population where individuals reproduce late
D. It is impossible to predict the outcome



16. The diagram above indicates

A. Humans are more complex than other primates.
B. Gorillas are unrelated to humans.
C. Humans and Chimpanzees share a more recent common ancestor than Gorillas and Orangutans.
D. Gibbons and Orangutans are more closely related than Gibbons and Humans.

17. When was the last common ancestor of humans and chimpanzees?

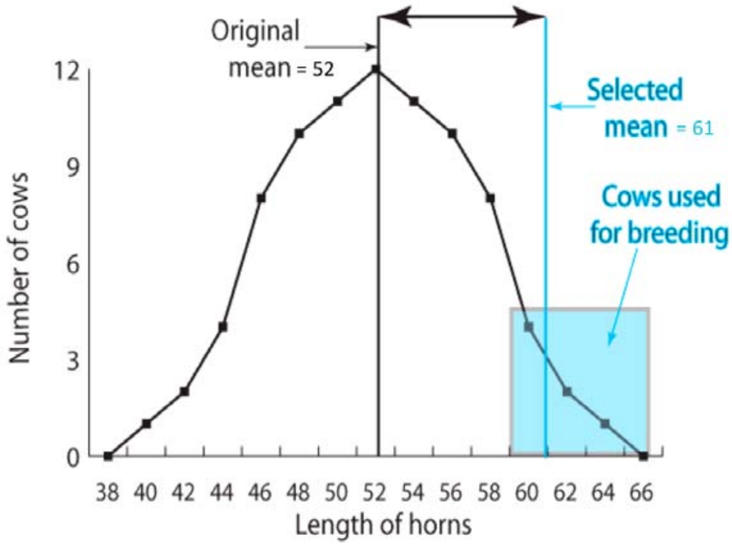
A. About 20 million years ago
B. About 7 million years ago
C. About 2 million years ago
D. About 200,000 years ago

18. Under which scenario would natural selection drive an allele to fixation most quickly?

A. The adaptive allele is recessive and rare
B. The adaptive allele is dominant and common
C. The adaptive allele is dominant and rare
D. The adaptive allele is partially dominant and intermediate frequency

19. Antibiotic resistance is considered one of the world's most pressing public health problems. Which of the following factors related to bacteria facilitates the evolution of resistance to antibiotics?

A. High mutation rates in bacteria
B. Bacterial horizontal and vertical gene transfer
C. Sub-lethal doses of antibiotics
D. All of the above



20. Texas longhorn cattle breeders are interested in artificially selecting cattle that have longer horns. The cattle (all 5 years of age) were drawn randomly from the same breeding program, so horn length is normally distributed around the population mean of 52" from tip-to-tip. Cattle breeders only bred the cows (and bulls) that had larger horns, with a mean length of 61". The heritability of horn length is 0.1.
Using the graph above and the Breeder's Equation, $R = h^2S$, what would you expect the mean horn length of the offspring to be in response to the selection we imposed?

A. 52.9"
B. 56.5"
C. 61.0"
D. 61.9"

21. Which of the following organisms always have a multicellular stage in the life cycle?

A. Animals
B. Green algae
C. Bacteria
D. Fungi

22. It is thought that eukaryotic cells possess mitochondria because

A. Mitochondria were present in the ancestor to all living life forms, including eukaryotes.
B. An archeal cell was phagocytized by an ancestral eukaryotic cell, which then harnessed the archeal cell's mitochondria.
C. Mitochondria evolved from a pre-existing type of organelle as a response to selection pressures relating to functional demands.
D. An engulfed bacterium survived within an ancestral eukaryote and formed a symbiotic relationship.

23. The scientific term for the flowering plants is:

A. Angiosperms
B. Glaucophytes
C. Gymnosperms
D. Plantae

24. Which are the most closely related to each other?
A. Animals and amoebozoans
B. Animals and bacteria
C. Animals and fungi
D. Animals and vascular plants

25. Most of the animals below show bilateral symmetry. Which of the following shows radial symmetry?
A. Butterfly
B. Human
C. Sea anemone
D. Slug

26. What is true of echinoderms?
A. They have an endoskeleton of calcareous plates
B. They have tube feet that provides motility in most species
C. They are acoelomates, or they lack a fluid-filled body cavity
D. A and B are true
E. A, B, and C are true

27. Which of the following groups has the most named species?
A. Arthropods
B. Vertebrates
C. Fungi
D. Vascular plants

28. Which of the following are most closely related to humans?
A. Crabs
B. Earthworms
C. Octopuses
D. Sea stars

29. Vertebrate jaws evolved from:
A. Gill arches
B. Skin
C. The skull
D. The tongue

30. Birds are a type of:
A. Reptile
B. Mammal

C. Pterosaur
D. Ray-finned fish

Answer key: 1. C; 2. B; 3. C; 4. C; 5. C; 6. D; 7. B; 8. D; 9. D; 10. C; 11. B; 12. D; 13. C; 14. B; 15. B; 16. C; 17. B; 18. D; 19. D; 20. A; 21. A; 22. D; 23. A; 24. C; 25. C; 26. D; 27. A; 28. D; 29. A; 30 A.

Assessment S2. Students' science self-efficacy, or confidence in their scientific ability.

Please rank each of the following according to how confident you feel in carrying out each of the activities below:
1 = not confident
2 = a little confident
3 = somewhat confident
4 = highly confident
5 = extremely confident

- Understanding scientific processes behind important scientific issues
- Making scientific arguments with friends or family
- Posing scientific questions
- Interpreting data from tables and graphs
- Determining the validity of scientific evidence
- Understanding course content

Assessment S3. Students' sense of social belonging in the classroom environment.

Please rank each of the following according to what extent you agree or disagree that:
1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

- Students in the class try to help one another understand course material (e.g. sharing lecture notes when absent)
- Students in the class consider themselves as part of a community.
- I am comfortable making a comment or asking a question during class discussions.
- Cornell demonstrates a strong institutional commitment to diversity.

Figure S1. Knowledge assessment instrument (KAI) binned by Bloom's taxonomy level, as rated by science education specialists at the Center for Teaching Excellence at Cornell University. Level 1: knowledge; 2: Comprehension; 3: Application; 4: Analysis; 5: Synthesis; 6: Evaluation. The frequency of weighted percentages matches that of assessments such as the MCAT, the GRE, and those from other undergraduate science courses (Zheng *et al.*, 2008).

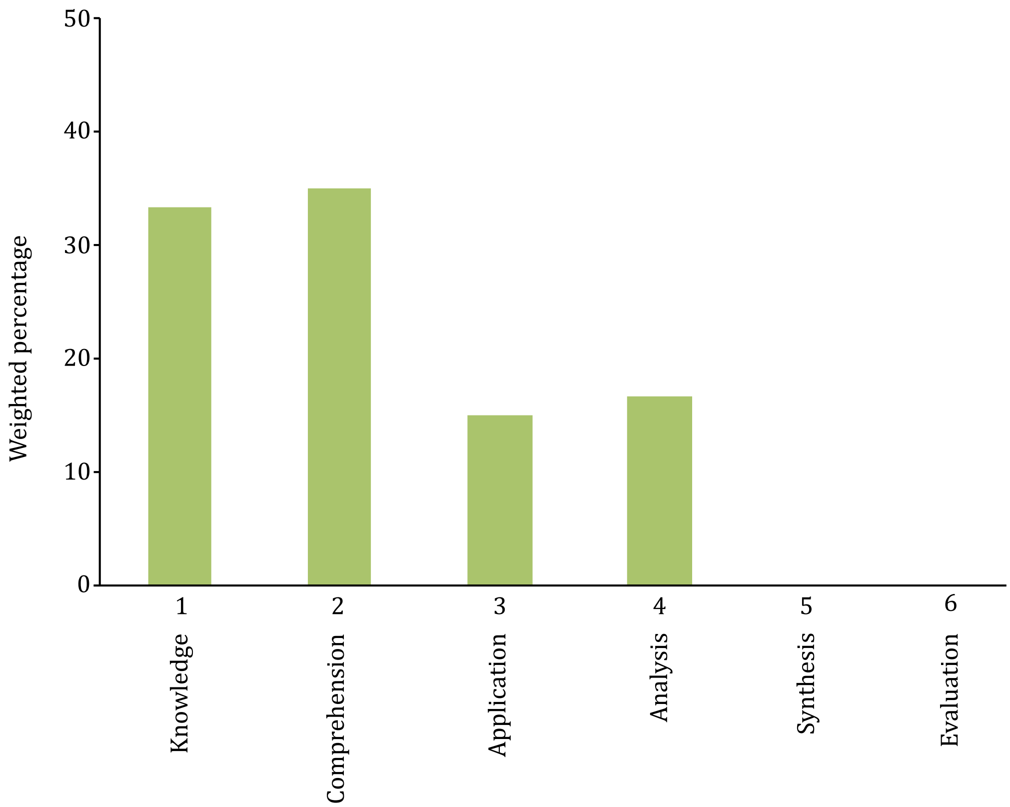


Figure S2. Contrast of full (A) and partial (B) mediation models to test mediation effects on student performance. The full mediation model tests how pedagogy and students' characteristics affect student performance indirectly via science self-efficacy of students. The partial model tests the *partial* mediation effect of scientific self-efficacy of students on students' performance. In this model, pedagogy and students' characteristics directly and indirectly affect students' performance via science self-efficacy.

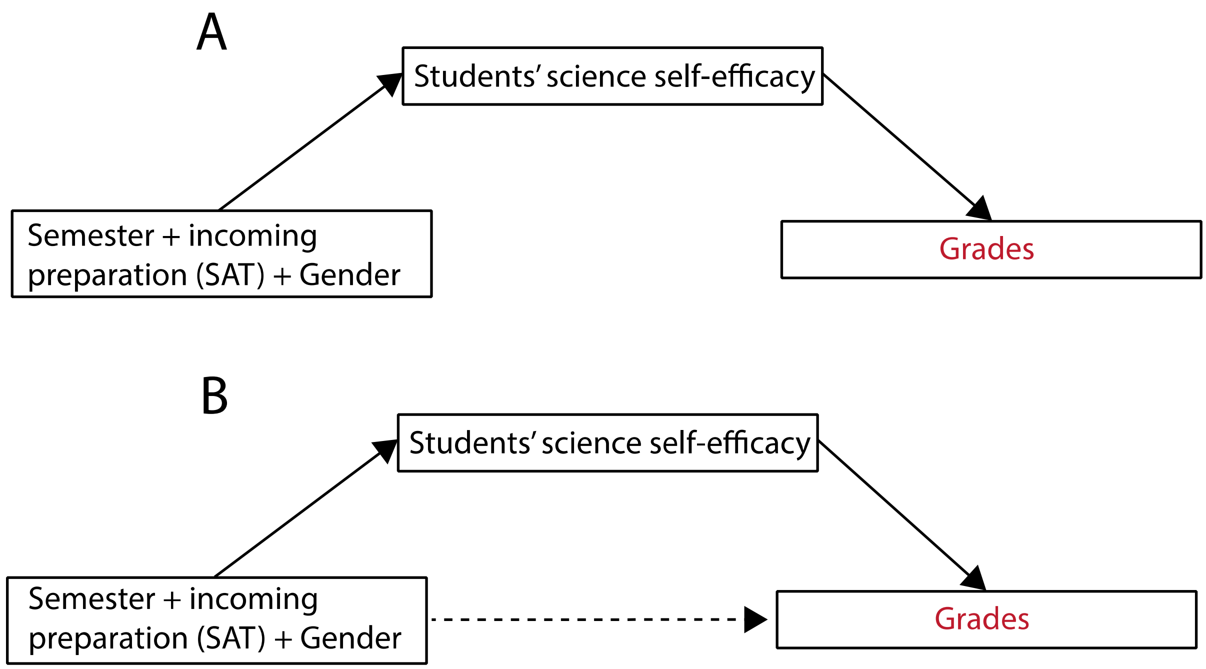


Table S1. Statistical results for changes in course grades and pre-/post-course KAI score differential across semesters (traditional vs. active) and accounting for potential demographic predictors. Both models include the same set of fixed variables, and different covariates that reflect appropriate measures of incoming preparation for the performance metric. Partial eta-squared values represent the effect sizes and relative contribution to the variance of predictor variables.

Dependent variable: Course grade					
Source	df	Mean Square	*F*	*P*	Partial Eta Squared
Semester	1	0.218	5.013	0.026	0.015
Gender	1	0.027	0.623	0.430	0.002
URM-status	1	0.280	6.425	0.012	0.020
Incoming math SAT score	**1**	**0.968**	**22.234**	**<0.001**	**0.065**
Semester * URM-status	**1**	**0.337**	**7.740**	**0.006**	**0.024**
Semester * Gender	1	0.005	0.125	0.724	<0.001

Dependent variable: KAI score differential					
Source	df	Mean Square	*F*	*P*	Partial Eta Squared
Semester	1	69.303	8.208	0.004	0.030
Gender	1	32.237	3.818	0.052	0.014
URM-status	1	56.715	6.717	0.010	0.024
Pre-course KAI score	**1**	**1195.6**	**141.6**	**<0.001**	**0.344**
Semester * URM-status	**1**	**47.620**	**5.640**	**0.018**	**0.020**
Semester * Gender	1	0.142	0.017	0.897	<0.001

Table S2. Statistical results and effect sizes for changes in science self-efficacy across semesters (traditional vs. active) and accounting for potential demographic predictors. We generated a single science self-efficacy variable after finding responses from six confidence statements to be highly correlated.

Dependent variable: Additive science self-efficacy differential					
Source	df	Mean Square	*F*	*P*	Partial Eta Squared
Semester	**1**	**1.728**	**6.523**	**0.011**	**0.025**
Gender	1	0.090	0.340	0.560	0.001
URM-status	1	0.251	0.949	0.331	0.004
Pre-semester science self-efficacy	**1**	**13.27**	**50.07**	**<0.001**	**0.166**
Semester * URM-status	1	0.001	0.005	0.945	<0.001
Semester *Gender	1	0.089	0.337	0.562	0.001

Table S3. Statistical results and effect sizes for changes in sense of social belonging responses across semesters (traditional vs. active) and accounting for potential demographic predictors. Three out of the four statements presented in this table were combined because they were specifically about the classroom microclimate and showed strong communality according to a principle component analysis. The fourth survey item gauged students' opinion about the university as a whole, which we did not expect to change based on changes in the classroom. All models include the same set of predictor variables.

Dependent variable: Additive classroom social belonging measure					
Source	df	Mean Square	*F*	*P*	Partial Eta Squared
Semester	**1**	**2.472**	**4.202**	**0.041**	**0.015**
Gender	1	0.991	1.685	0.195	0.006
URM-status	**1**	**2.311**	**3.928**	**0.048**	**0.014**
Semester * URM-status	1	0.064	0.109	0.741	<0.001
Semester * Gender	1	0.656	1.116	0.292	0.004

Dependent variable: Cornell demonstrates a strong institutional commitment to diversity.					
Source	df	Mean Square	*F*	*P*	Partial Eta Squared
Semester	1	0.071	0.102	0.750	<0.001
Gender	1	0.036	0.052	0.820	<0.001
URM-status	1	1.113	1.609	0.206	0.006
Semester * URM-status	1	0.002	0.003	0.955	<0.001
Semester * Gender	1	0.051	0.073	0.787	<0.001

Table S4. The effects of active learning on student science self-efficacy and sense of belonging in the classroom. A) Mean gains in self-reported confidence over the course of a traditional semester (fall 2014) versus an active semester (fall 2015). Due to high correlation of responses, we generated a single science self-efficacy response variable per student. Using this combined measure, we found higher gains in science self-efficacy in fall 2015, as measured by post-semester responses minus pre-semester responses. B) Mean reports of classroom and university social belonging after fall 2014 and fall 2015 semesters. In analyses we combined statements 1-3 because they address the classroom microclimate and showed strong communality according to a principle component analysis. Survey statement 4 addresses students' opinion about the university as a whole, which did not change significantly as a result of changes in the classroom.

	Fall 2014			Fall 2015		
A. Semester science self-efficacy gains: How confident are you…	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**
1. Understanding scientific processes behind important scientific issues	0.10	0.83	128	0.49	0.91	138
2. Making scientific arguments with friends or family	0.16	0.89	128	0.51	1.11	138
3. Posing scientific questions	0.23	0.95	128	0.51	1.34	138
4. Interpreting data from tables and graphs	0.28	0.84	128	0.44	1.24	138
5. Determining the validity of scientific evidence	0.49	0.84	128	0.41	1.04	138
6. Understanding the content of this course	-0.18	1.12	131	0.75	1.41	142
B. Post-semester only: To what extent do you agree…						
1. Students in the class help one another understand the course material	3.91	0.77	148	3.96	0.88	157
2. Students in the class consider themselves as part of a community	3.51	0.87	148	3.89	0.95	157
3. I am comfortable making a comment during class discussions	3.10	0.90	148	3.54	0.99	157
4. Cornell demonstrates a commitment to diversity	3.47	1.13	148	3.59	1.02	157

Table S5. Statistical results from mediation analyses split by student URM-status, with self-efficacy as the mediator. The effects of pedagogy and student characteristics (gender and incoming preparation) on student performance, and whether performance gains were mediated by changes in scientific self-efficacy.

non-URM: mediation analysis of grades				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	-0.124	0.084	-1.475	0.14
Semester	**0.347**	**0.083**	**4.173**	**0.000**
Incoming math SAT score	0.001	0.001	0.904	0.366
Grades	*b*	*SE*	*Z*	*P*
Science self-efficacy	-0.005	1.113	-0.005	0.996
Incoming math SAT score	**0.033**	**0.010**	**3.281**	**0.001**

non-URM: mediation analysis of KAI score differential				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	**-0.209**	**0.081**	**-2.590**	**0.010**
Semester	**0.445**	**0.078**	**5.669**	**0.000**
Incoming pre-course KAI score	**0.022**	**0.011**	**1.970**	**0.049**
KAI score differential	*b*	*SE*	*Z*	*P*
Science self-efficacy	0.156	0.384	0.406	0.685
Incoming pre-course KAI score	**-0.624**	**0.063**	**-9.829**	**0.000**

URM: mediation analysis of grades				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	**0.465**	**0.196**	**2.368**	**0.018**
Semester	**0.401**	**0.199**	**2.017**	**0.044**
Incoming math SAT score	0.001	0.001	0.760	0.447
Grades	*b*	*SE*	*Z*	*P*
Science self-efficacy	**3.547**	**1.845**	**1.923**	**0.054**
Incoming math SAT score	**0.037**	**0.016**	**2.257**	**0.024**

URM: mediation analysis of KAI score differential				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	0.241	0.198	1.220	0.222
Semester	**0.400**	**0.195**	**2.050**	**0.040**
Incoming pre-course KAI score	-0.002	0.022	-0.098	0.922
KAI score differential	*b*	*SE*	*Z*	*P*

Science self-efficacy	**1.655**	**0.638**	**2.593**	**0.010**
Incoming pre-course KAI score	**-0.499**	**0.093**	**-5.352**	**0.000**

Table S6. Statistical results from mediation analyses split by student URM-status, with social belonging survey measure as second mediator. Adding this second mediation path did not improve the model.

non-URM: mediation analysis of grades				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	-0.130	0.086	-1.504	0.133
Semester	**0.370**	**0.085**	**4.352**	**0.000**
Incoming math SAT score	0.001	0.001	0.897	0.370
Social belonging	*b*	*SE*	*Z*	*P*
Gender	**0.278**	**0.124**	**2.250**	**0.024**
Semester	**0.234**	**0.122**	**1.920**	**0.055**
Incoming math SAT score	**0.004**	**0.001**	**3.489**	**0.000**
Grades	*b*	*SE*	*Z*	*P*
Science self-efficacy	-0.005	1.224	-0.004	0.997
Social belonging	0.225	0.882	0.255	0.799
Incoming math SAT score	**0.036**	**0.012**	**3.042**	**0.002**

non-URM: mediation analysis of KAI score differential				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	**-0.199**	**0.080**	**-2.485**	**0.013**
Semester	**0.464**	**0.078**	**5.981**	**0.000**
Incoming pre-course KAI score	**0.024**	**0.011**	**2.120**	**0.034**
Social belonging	*b*	*SE*	*Z*	*P*
Gender	0.217	0.116	1.874	0.061
Semester	**0.232**	**0.112**	**2.073**	**0.038**
Incoming pre-course KAI score	0.025	0.016	1.511	0.131
KAI score differential	*b*	*SE*	*Z*	*P*
Science self-efficacy	0.540	0.397	1.362	0.173
Social belonging	0.052	0.298	0.175	0.861
Incoming pre-course KAI score	**-0.594**	**0.064**	**-9.305**	**0.000**

URM: mediation analysis of grades				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	**0.508**	**0.200**	**2.547**	**0.011**
Semester	**0.491**	**0.200**	**2.460**	**0.014**

Incoming math SAT score	0.002	0.001	1.166	0.244

Social belonging	*b*	*SE*	*Z*	*P*
Gender	-0.195	0.287	-0.679	0.497
Semester	0.342	0.287	1.193	0.233
Incoming math SAT score	-0.003	0.002	-1.687	0.092

Grades	*b*	*SE*	*Z*	*P*
Science self-efficacy	2.770	1.822	1.521	0.128
Social belonging	2.095	1.410	1.485	0.138
Incoming math SAT score	**0.036**	**0.017**	**2.106**	**0.035**

URM: mediation analysis of KAI score differential				
Science self-efficacy	*b*	*SE*	*Z*	*P*
Gender	0.304	0.205	1.483	0.138
Semester	**0.421**	**0.195**	**2.154**	**0.031**
Incoming pre-course KAI score	0.011	0.024	0.437	0.662

Social belonging	*b*	*SE*	*Z*	*P*
Gender	-0.251	0.266	-0.945	0.345
Semester	0.254	0.253	1.001	0.317
Incoming pre-course KAI score	-0.019	0.032	-0.588	0.556

KAI score differential	*b*	*SE*	*Z*	*P*
Science self-efficacy	1.369	0.618	2.215	0.027
Social belonging	0.904	0.500	1.807	0.071
Incoming pre-course KAI score	**-0.535**	**0.102**	**-5.267**	**0.000**

Mediation analyses for science self-efficacy and sense of social belonging

Mediation analyses for science self-efficacy split by student URM-status

Grades**:** For the full mediation model, the other fit indices were in the acceptable range: RMSEA = 0.07 (acceptable range: 0-0.08), CFI = 0.94 (acceptable range: above 0.9), SRMR = 0.03 (acceptable range: 0-0.1). According to this model, for non-URM students, the improvement in science self-efficacy was 0.35 units (0.59 SD) higher in an active learning environment ($P < 0.0001$). However, there was no significant association between changes in self-efficacy and performance for these students ($P = 0.99$). For URM students, the changes in scientific self-efficacy was also significantly higher in the active learning classroom ($b = 0.40$, $P = 0.044$), implying that in an active learning environment, the scientific self-efficacy of URM students improved 0.4 units (0.68 SD) more. Unlike non-URM students, the changes in self-efficacy of URM students was significantly associated with their grade ($b = 3.55$, $P = 0.05$); one unit of increase in efficacy led to a 3.55% (0.41 SD) increase in course grade (Figure 2D). These results suggest that for URM students, changes in self-efficacy was a mediator for the positive effect of active learning practices on students' grades; students' self-efficacy improved during the active learning semester, and the gains in self-efficacy led to improvement in academic performance. However, for non-URM students, there was no such mediation effect; while self-efficacy was higher in the active learning semester, the gains in self-efficacy did not play a role in academic performance.

Knowledge assessment instrument (KAI) learning gains: For the full mediation model, two of the fit indices were in the acceptable range: CFI = 0.93 (acceptable range: above 0.9), SRMR = 0.041. One fit index was outside the acceptable range: RMSEA = 0.15 (acceptable range: 0-0.08), which may be a result of missing data from students who did not take both the pre- and post- course KAI. For non-URM students, the improvement in self-efficacy was 0.45 units (0.76 SD) higher in an active learning environment ($P = 0.0001$), but we found no significant association between changes in self-efficacy and performance for these students ($P = 0.69$). For URM students, the changes in scientific self-efficacy was also significantly higher in the active learning classroom ($b = 0.4$, $P = 0.04$); the scientific self-efficacy of URM students improved 0.4 units (0.68 SD) more than in the traditional lecture. Importantly, the changes in self-efficacy of URM students was significantly associated with their KAI learning gains ($b = 1.66$, $P = 0.01$); one unit of increase in self-efficacy led to 1.66 points (0.46 SD; out of 30 points) increase in KAI gains. The smaller coefficient we observe for KAI gains may be due to missing assessment data (an optional in-class exercise), or because the KAI is a lower-risk evaluation. These results suggest that for URM students, changes in self-efficacy mediated the positive effect of active learning practices on students' learning gains. In other words, students' self-efficacy improved during the active learning semester, and the gains in self-efficacy led to improvement in the KAI gain. For non-URM students, there was no such mediation effect; while self-efficacy was higher in the active learning semester, the gains in self-efficacy did not play a role in KAI gain. Note that the coefficient for the effect of active learning on self-efficacy is 0.03 higher in modeling KAI gains than modeling grades. This is because in order to have more coherence between the measure of students' incoming preparation and outcome performance, we used different covariates in the analyses: we used SAT math score in modeling grades, and the pre-course KAI score in modeling KAI gains.

Pathway analyses for social belonging split by student URM-status

We also tested the possibility that the sense of social belonging served as a second mediator, in addition to increased self-efficacy, between students' preparation, characteristics, and instructional practices and their performance. Adding this second mediation path did not improve the model for non-URM students and URM students (Table S6).

Grades: For non-URM students, while social belonging was higher in the active learning course ($P = 0.055$, $b = 0.234$), social belonging was not correlated with grades ($P = 0.799$). For URM students, grades were not significantly correlated with social belonging ($P = 0.13$), nor did social belonging significantly improve after the active-learning course ($P = 0.23$).

KAI learning gains: For non-URM students, social belonging was higher in the active learning course ($b = 0.23$, $P = 0.038$). However, social belonging itself was not a significant predictor of KAI learning gains ($P = 0.86$). For URM students, social belonging did not differ across the two semesters ($P = 0.32$). However, social belonging was marginally associated with gain in the KAI ($P = 0.07$). Furthermore, the inclusion of social belonging to the mediation analysis of KAI gain reduced the fit of the model and none of the fit indices fell within the acceptable range: CFI= 0.88 (acceptable: >0.9), RMSEA=0.17 (acceptable: < 0.8), SRMR=0.057 (acceptable: <0.5).

These results indicate that social belonging is not a mediator between students' characteristics and instructional practices and students' performance for non-URM and URM students. This result is particularly notable for URM students, where science self-efficacy was a full mediating factor.

Social belonging and science self-efficacy were significantly correlated for both groups of students (r(URM) = 0.397, $P = 0.008$; r(non-URM) = 0.326, $P < 0.0001$). So we also tested for social belonging as another predictor of science self-efficacy. This addition did not significantly improve the fit of the full-mediation models of science self-efficacy for either measure of the students' performance (P(grade)= 0.0.58, P(KAI) = 0.99). While active learning practices improved the science self-efficacy of URM students, which led to improved performance, these practices did not significantly influence the sense of social belonging among URM students.

References

Zheng AY, Lawhorn JK, Lumley T, Freeman S. Application of Bloom's Taxonomy Debunks the``MCAT Myth''. Science. 2008;319(5862):414.