Supplemental Material

CBE—Life Sciences Education

Freeman *et al*.

Supplemental material

Table of Contents

3. Evolution by natural selection

Table S10 Regression output from best models: E-ACORNS analysis

I. Pre-post survey questions

The following instructions and questions were posted in the course management system.

NOTE: The items indicated by asterisks (*; these did not appear on the actual survey) relate to student attitude or intent. Because they were either taken from a variety of published instruments or created for this study, we lack rigorous validity evidence on the constructs represented so treated the data as preliminary and exploratory.

Please give your best effort to answer the following questions. Some of them ask you to gauge your feelings on different issues, so just answer these honestly. Other questions ask you about biology topics; just answer these to the best of your ability without using any references or outside sources. Remember that you are not being graded for correctness, and that your answers will never be seen by any of the course instructors.
We've assigned the survey because the results will help us improve the introductory biology series at UW. The entire assignment should take you 20-25 minutes. Thank you so much for helping us make this course better!

*The next 2 questions ask you to rate your level of interest in an activity. (Very Interested, Interested, Neither Interested nor Uninterested, Uninterested, Very Uninterested)

- How interested or uninterested are you in obtaining an undergrad research experience in the future?
- How interested or uninterested are you in pursuing a science-related research career?

*The next 2 questions ask how likely you would be to do an activity. (Very Likely, Likely, Neither Likely nor Unlikely, Unlikely, Very Unlikely)

- How likely or unlikely do you think it is that you will be able to get a position as an undergraduate researcher during your remaining time in college, including summers
- If the Biology Department offered a one-time, 60-minute session on how to get an undergrad research experience,and if it fit conveniently in your schedule, how likely or unlikely would you be to attend?

*Please reply "yes" or "no" to the next 2 questions, and explain, if asked.

- Have you done research as a UW student?
- Did you do research before coming to UW?
- If you replied "yes" to the prior question, please describe what research you did before coming to UW, and where you did this research.
  [If you replied "no" please type "no" again.]

The next 3 questions ask you what you think about 3 topics. Please just answer them honestly and thoughtfully.
(Open response)

- What does it mean to think like a scientist?
- What does it mean to do science?
- Did you perform what you would call real research in your BIOL 180 labs? Why or why not?

*Please indicate how much you agree or disagree with each of the following statements.
(Strongly disagree; Slightly disagree; Neither disagree or agree; Slightly Agree; Strongly agree)

- My BIOL *course#* lab experience taught me valuable skills.
- My BIOL *course#* lab experience helped prepare me for what I plan to do in life.
- My BIOL *course#* lab experience was <u>not</u> helpful to me.
- Experiments I did in BIOL *course#* labs will help solve a problem in the world.
- Results I obtained in BIOL *course#* labs were important to the scientific community.
- I faced challenges that I managed to overcome in my BIOL *course#* lab experiments.
- I was responsible for the outcomes of my BIOL *course#* lab experiments.
- My BIOL *course#* lab experiments addressed a question(s) that was important to me.
- The results I obtained in BIOL *course#* lab gave me a sense of personal achievement.
- My BIOL *course#* lab experiments were interesting.

*The next 12 items ask you how confident you are that you can complete a task.
(1. Not at all. 2. (blank) 3. (blank) 4. A lot)

- How confident are you in your ability to use technical science skills? (tools, instruments, and techniques)
- How confident are you in your ability to use scientific language and terminology when presenting the results of an experiment?
- How confident are you in your ability to communicate the results of an experiment to a group of your peers?
- How confident are you in your ability to communicate the results of an experiment to a group of professional scientists?
- How confident are you in your ability to effectively divide the tasks between group members when working together on an experiment?
- How confident are you in your ability to work with a team to interpret data from an experiment?
- How confident are you in your ability to propose explanations for the results of a study?
- How confident are you in your ability to design a logical next experiment, based on the results of your experiment?
- How confident are you in your ability to relate results and explanations to the work of others?
- How confident are you in your ability to contribute to science?
- How confident are you in your ability to think scientifically?
- How confident are you in your ability to do science?

*The following 5 questions ask how you think about yourself and your personal identity. Please indicate how much you agree or disagree with the statement.
(Strongly disagree; Slightly disagree; Neither disagree or agree; Slightly Agree; Strongly agree)

- I feel like I belong in the field of science.
- I have a strong sense of belonging to the community of scientists.
- Being able to do science is an important part of who I am.
- I am more like a scientist than I was before participating in BIOL *course#* labs.
- I have come to think of myself as a 'scientist'.

Please answer the next 2 questions to the best of your ability, without consulting any references.

- A species of snail (an animal) is poisonous. How would biologists explain how this species evolved from anancestral species of snail that was not poisonous? In your answer, be sure to connect what is happening at the molecular (genetic) level to the level of the whole organism.

- A species of flightless bird (flightless birds, such as penguins, cannot fly) is closely related to bird species that are able to fly. How would biologists explain how a flightless bird species originated from an ancestral bird species that couldfly? In your answer, be sure to connect what is happening at the molecular (genetic) level to the level of the whole organism.

On the post-survey, these prompts were changed to:

- One species of prosimians (animals) has long tarsi. How would biologists explain how this species with long tarsi evolved from an ancestral species of prosimian that had short tarsi? In your answer, be sure to connect what is happening at the molecular (genetic) level to the level of the whole organism.
- In one species of Suricata (animals), a pollex is absent. How would biologists explain how the Suricata species without a pollex evolved from an ancestral species of Suricata with a pollex? In your answer, be sure to connect what is happening at the molecular (genetic) level to the level of the whole organism.

Please answer the final question (below) to the best of your ability, without consulting any references.

- The claim has been made that women may be able to achieve significant improvements in memory by taking iron supplements. Prior to accepting this claim, and to determine whether or not this claim is fraudulent, you decide to perform a scientific experiment. Describe your proposed experiment and provide justifications for each aspect of your experimental design. Lastly, state whether the results of your experiment could prove the hypothesis that iron supplements enhance memory.

On the post-survey, this prompt was changed to:

- Advocates of herbal medicine claim that echinacea helps fight upper respiratory tract infections (colds and flu). Prior to accepting this claim, and to determine whether or not this claim is fraudulent, you decide to perform a scientific experiment. Describe your proposed experiment and provide justifications for each aspect of your experimental design. Lastly, state whether the results of your experiment could prove the hypothesis that echinacea is effective against colds and flu."

II. Overview of CURE activities

Table S1 The CURE sequence

Course 1 (Bio 180)	**Course 2 (Bio 200)**
Introduction to the model system, research question, and experimental design Transfer cells to selective medium	PCR candidate gene (antibiotic target) Send PCR products out for sequencing
Select antibiotic-resistant cells Begin daily transfers (experimental evolution)	Analyze sequence data from antibiotic-resistant and -sensitive strains; identify differences if present
Introduction to data analysis in the software program R Continue daily transfers	Analyze 3-D structure of inferred protein product
Introduction to assays used to assess fitness and level-of-resistance Practice with R Conclude daily transfers	Prepare poster (one poster per team of four students)
Set up fitness and resistance assays	Poster session in lobby of main biology building attended by classmates, department faculty and staff, members of students' families
Perform fitness and resistance assays	
Analyze fitness and resistance assays	

III. Scoring rubrics for open-response prompts

We used the prompts and rubrics provided in Tables S2, S3, and S4 to document changes in student understanding of three measures of learning other than exam scores.

Table S2 Prompts and rubrics used to evaluate student understanding of the culture of scientific research.

These prompts and rubrics were developed by Wachtell et al. (in review) and evaluate the culture of scientific research framework of Dewey et al. (2020).

A. *Prompt 1: What does it mean to think like a scientist?* (6 points possible)

Category	Explanation/examples
Asking questions	Curiosity, extend frontier of knowledge
Process thinking	Hypothesis testing, experimental design
Critical thinking	Skepticism, demanding evidence, quality assurance, rigor
Evidence-based conclusions	Data-based reasoning
Open-minded	Consider alternatives, multiple perspectives
Multiple approaches	Most convincing evidence is based on multiple independent sources

B. *Prompt 2: What does it mean to do science?* (15 points possible)

Category	Sub-element
Investigate	Consult prior studies
	Observe natural world
	Ask a question
Collect data	Perform an experiment or collect observational data
	Test a hypothesis
	Repeat the experiment to verify the result
Analyze data	Analyze data (include visualization)
	Interpret data
	Patterns may lead to models
Collaborate	Work in a team
	Exchange information and ideas among team members
	Jointly produce information for dissemination
Communicate	Share results with community (papers, posters, etc.)
	Undergo peer review
	Replicate other teams' findings

(Table S2, continued)

C. *Prompt 3: Did you perform what you would call real research in your BIOL 180 labs? Why or why not?* (15 points possible)

Category	**Sub-element**
Authenticity	New knowledge
	Relevance to scientific community
Processes	Collaboration
	Used publication-standard techniques
	Understand how and why the techniques work
	No right/wrong data
Iteration	Troubleshoot
	Repeat experiments
Connections to other work	Work continued over course of term
	Work will continue beyond the class
	Communicate results
Ownership	Work on own question and/or hypothesis
	Design the experiment
	Carry out the experiment or observations
	Be responsible for the integrity of the data

Table S3 Prompts and rubrics used to evaluate student understanding of experimental design

These rubrics evaluate prompts that were variations on the following example: “Advertisements for an herbal product, ginseng, claim that it promotes endurance. Prior to accepting this claim, and to determine whether or not this claim is fraudulent, you decide to perform a scientific experiment. Describe your proposed experiment and provide justifications for each aspect of your experimental design. Lastly, state whether the results of your experiment could prove the hypothesis that ginseng promotes endurance. This should take you approximately 10-15 minutes to complete.” There are 17 points possible; the complete rubric with examples is given in Appendix C.2 in Brownell et al. (2014).

	0 points answer	**1 point answer**	**2 point answer**
1. Identifies variable that will be manipulated	Other than ginseng	Ginseng OR herbal product	N/A
2. Identifies variable that will be measured.	Other than endurance	Endurance	N/A
3. Describes how dependent variable will be measured.	Not mentioned or too subjective to be verified	Reasonable outcome measure but no specifics/units.	Reasonable outcome measure with specifics/units.
4. Realization that other variables need to be held constant.	Not mentioned OR related to independent variable	Stated one reasonable variable that could be controlled	Stated two or more reasonable variables that could be controlled
5. Control for vehicle effect	Not mentioned	Recognize need for placebo but no/insufficient reasoning	Recognize need for placebo and supply correct reasoning
6. Sample size	Not mentioned	State “large sample size” but provide no/vague reasoning	State “large sample size” and provide correct reasoning
7a. Repeat experiment	Not mentioned OR “NO”, OR a possibility	Yes, recognizes need	N/A
7b. Reasoning for repeating experiment	No explanation given OR incorrect reasoning	“Increase validity of results” but vague	Provide appropriate justification
8a. Conclusions that could be drawn	Not mentioned OR stated only as part of hypothesis/prediction	States what conclusion can be drawn but does not qualify the conclusion	States what conclusion can be drawn and qualifies the conclusion (e.g. sources of error, limits to generalization)
8b. Results cannot prove your hypothesis	Not mentioned OR YES, can prove hypothesis	Recognition that you “cannot prove a hypothesis: but did not provide any reasoning/explanation	Recognition that you can only “disprove a hypothesis” or “build support for a hypothesis”

Table S4 Prompts and rubrics used to evaluate student understanding of evolution by natural selection

These rubrics evaluate prompts that had the form “A species of *taxon name* is *trait state*. How would biologists explain how this species evolved from anancestral species of *taxon name* is that was not *trait state*? In your answer, be sure to connect what is happening at the molecular (genetic) level to the level of the whole organism.” There 15 points possible in the expert-like assessment and 4 points possible in the naïve ideas assessment (Sievers et al., accepted with minor revisions).

A. Expert-like ideas

Core Concept	**Novice**	**Intermediate**	**Advanced**
1. Nature of mutation	Mutation occurs,	creates heritable variation,	and is random with respect to fitness.
2. Variation in populations	Variation in populations exists,	is based on a diversity of alleles,	and exists independently of environmental conditions.
3. Genotype to phenotype	Mutations change genotypes	and may change gene products,	and, if so, change phenotypes.
4. Phenotype to fitness (natural selection)	Traits vary in their impact on fitness,	leading to differential reproductive success	in a specific environment.
5. Evolution	Evolution occurs when trait frequencies change,	or more precisely when allele frequencies change,	due to the fitness advantage of a trait.

B. Naïve ideas

	0 points	**-1 points**
Teleological or anthropomorphic causation (purposeful/“conscious” change)	No mention	Mutations occur in response to a change in the environment, or traits change due to want or need.
Inheritance of acquired characters	No mention	Traits change due to use/disuse, “exertion,” or interaction with the environment, with an implication that these changes are inherited.
Naïve group selectionism	No mention	Changes happen “for the good of the species.”
Essentialism	No mention	All individuals in a population change at once, or adaptation is conflated with speciation.

IV. Sample sizes used in the analyses

Average totals for the total number of students in each treatment who responded to each construct or prompt are reported in Table 1 in the main text. The sample sizes reported here, in Table S7, are relevant to the power analysis and to the models that tested for disproportionate impacts on minoritized students. In each table, “Trad” refers to traditional labs; “ContGen” refers to continuing generation students—meaning not-1stGen. Numbers vary among constructs and prompts due to missing data.

Table S5 Sample sizes by construct, disaggregated by demographic groups of interest

A) Culture of scientific research prompts

	“What does it mean to do science?”			“What does it mean to think like a scientist?”			“Did you do real research in your *coursename* lab?	
	Treatment group			Treatment group			Treatment group	
	CURE	Trad		CURE	Trad		CURE	Trad
URM	14	67		14	67		16	70
NonURM	149	159		149	159		154	159
LowSES	28	119		28	119		32	122
HighSES	150	118		150	118		154	118
Female	114	158		114	158		118	160
Male	64	79		64	79		68	80
First Gen	37	79		37	79		21	81
ContGen	140	156		140	156		144	158

B) Experimental design prompt

	E-EDAT	
	Treatment group	
	CURE	Trad
URM	14	58
NonURM	148	141
LowSES	28	103
HighSES	148	106
Female	111	142
Male	65	67
First Gen	37	69
ContGen	138	138

C) Evolution by natural selection prompts

	E-ACORNS, trait gain			E-ACORNS, trait loss	
	Treatment group			Treatment group	
	CURE	Trad		CURE	Trad
URM	14	66		15	66
NonURM	149	156		149	142
LowSES	28	114		29	108
HighSES	151	118		150	110
Female	114	154		112	141
Male	65	78		67	77
First Gen	37	75		37	69
ContGen	141	155		141	147

V. Power analysis results

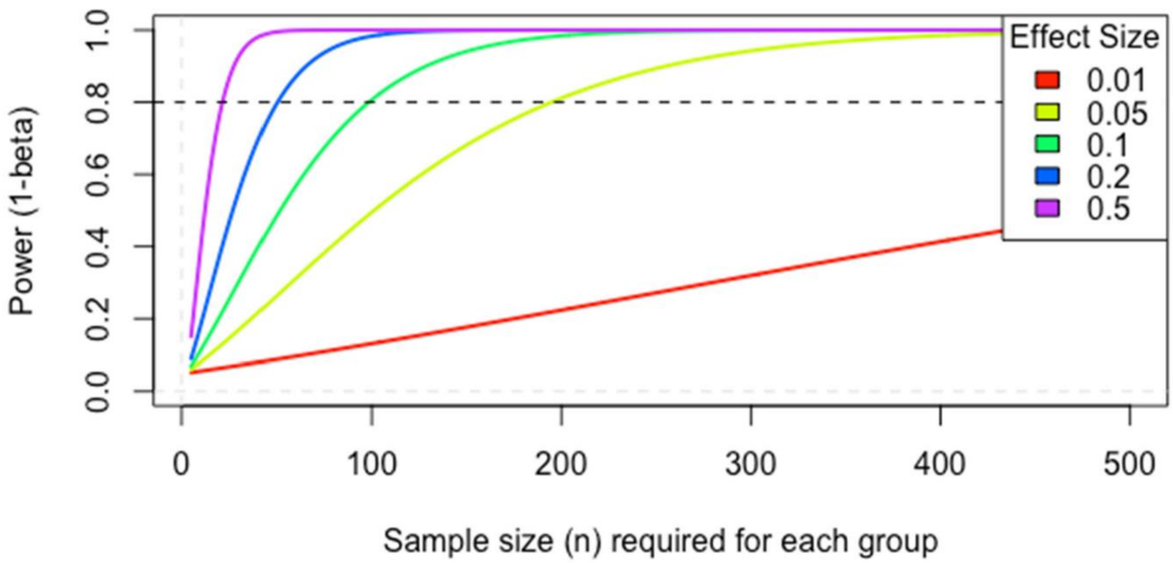

Figure 1 Power analysis indicates the minimum size of each group required to detect an array of effect sizes for this dataset

Following Kraft (2020), we considered effect sizes of 0.20 and above as large (lavendar to blue lines), 0.05 to less than 0.20 as medium (blue to chartreuse line), and less than 0.05 as small (chartreuse to red lines). Note that a lime-green line at effect size 0.1 is present to aid interpretation. Effect size is measured as Cohen's f^2 at a two-tailed $p = 0.05$.

VI. Regression output from the best models, each analysis

Data for the analysis of exam scores, as an index of learning and of course performance, are given in Table 2 of the main paper.

Other measures of learning

1. *Culture of scientific research*

Table S6 Regression output from best model: Thinking Like a Scientist analysis

The dependent variable was the sum of points scored on a 6-point rubric (see Wachtell et al., in prep). Binary sex, URM status, SES status, first-generation status, and SAT score were not retained as predictors in the best model; SAT score and Treatment were also not retained as a predictor in the best model.

	Estimate	SE	*t*-value	*p*-value
Intercept	-1.94	0.09	-22.2	<<0.001
PreScore	0.17	0.06	2.8	0.006

Table S7 Regression output from best model: What it Means to Do Science analysis

The dependent variable was the sum of points on a 15-point rubric (see Wachtell et al., in prep). Binary sex, URM status, SES status, and first-generation status were not retained as predictors in the best model.

	Estimate	SE	*z*-value	*p*-value
(Intercept)	-2.87	0.08	-36.2	<<0.001
PreScore	0.12	0.03	4.2	<0.001
SAT total score	0.002	0.0003	7.2	<<0.001
Treatment (Ref:CURE)	-0.15	0.08	-1.92	0.055

Table S8 Regression output from best models: Did You Do Real Research in Lab? analysis

a) The initial model in this analysis was a binomial regression assessing whether students were more likely to answer yes or no to this question, based on the predictors. Binary sex, URM status, SES status, first-generation status, and SAT total score were not retained as predictors in the best model.

	Estimate	SE	*z*-value	*p*-value
Intercept	0.28	0.10	2.2	0.03
Treatment	1.29	0.23	5.5	<<0.0001

b) The second model in this analysis was a binomial regression with the outcome variable being whether students were more likely to provide valid warrants on the 15-point "Real Research" rubric explaining why labs represented real research, based on the predictors. Binary sex, URM status, SES status, first-generation status, and SAT total score were not retained in the best model.

	Estimate	SE	*z*-value	*p*-value
Intercept	-0.66	0.14	-4.8	<0.001
Treatment	1.11	0.20	5.5	<<0.0001

c) The third and final model in this analysis was a binomial regression assessing whether students were more likely to provide valid warrants on the "Real Research" rubric, explaining why labs did *not* represent real research, based on the predictors. Binary sex, URM status, SES status, first-generation status, and SAT total score were not retained in the best model.

	Estimate	SE	*z*-value	*p*-value
Intercept	-0.71	0.14	-5.2	<0.0001
Treatment	-1.35	0.27	5.0	<<0.0001

2. *Experimental design*

Table S9 Regression output from best model: E-EDAT analysis

These data are from a linear model testing the impact of an array of predictors on the total E-EDAT score (15-point rubric).

	Estimate	SE	*z*-value	*p*-value
Intercept	-1.54	0.10	-17.6	<<0.0001
PreScore	0.07	0.01	5.2	<0.0001
SAT total score	0.0008	0.0002	3.7	0.0002

3. *Evolution by natural selection*

Table S10 Regression output from best models: E-ACORNS analysis

a) The initial model in this analysis was a linear regression assessing which variables best-predicted total score on the E-ACORNS rubric, in response to a question about trait gain. Binary sex, URM status, SES status, first-generation status, and SAT total score were not retained as predictors in the best model.

	Estimate	SE	*z*-value	*p*-value
Intercept	-1.86	0.08	-24.7	<<0.0001
PreScore	0.04	0.02	3.9	<0.0001
Treatment	0.40	0.07	6.2	<<0.0001

b) The second model in this analysis was a linear regression assessing which variables best-predicted the total number of misconceptions on the E-ACORNS rubric that students declared in response to a question about trait gain. Treatment, binary sex, URM status, SES status, and first-generation status were not retained as predictors in the best model. Students with higher SAT scores were more likely to declare misconceptions, independent of treatment.

	Estimate	SE	*z*-value	*p*-value
Intercept	-10.33	2.93	-3.5	<<0.0001
SAT total score	0.005	0.002	2.6	0.009

c) The third model in this analysis was a linear regression assessing which variables best-predicted total score on the E-ACORNS rubric, in response to a question about trait loss. URM status, SES status, first-generation status, and SAT total score were not retained as predictors in the best model. Female students had higher scores than male students, on average, independent of treatment.

	Estimate	SE	*z*-value	*p*-value
Intercept	-2.04	0.08	-25.5	<<0.0001
PreScore	0.05	0.02	3.0	0.003
SAT total score	0.0009	0.0002	3.4	0.0006
Treatment	0.26	0.07	3.8	0.0002
Sex	0.16	0.07	2.2	0.025

d) The final model in this analysis was a linear regression assessing which variables best-predicted the total number of misconceptions on the E-ACORNS rubric that students declared in response to a question about trait loss. The null model provided the best fit to the data.

	Estimate	SE	*z*-value	*p*-value
Intercept	-2.47	0.19	-13.2	<<0.0001