# Supplemental Material

*CBE—Life Sciences Education*

Lachance *et al.*

**Supplemental Figures**

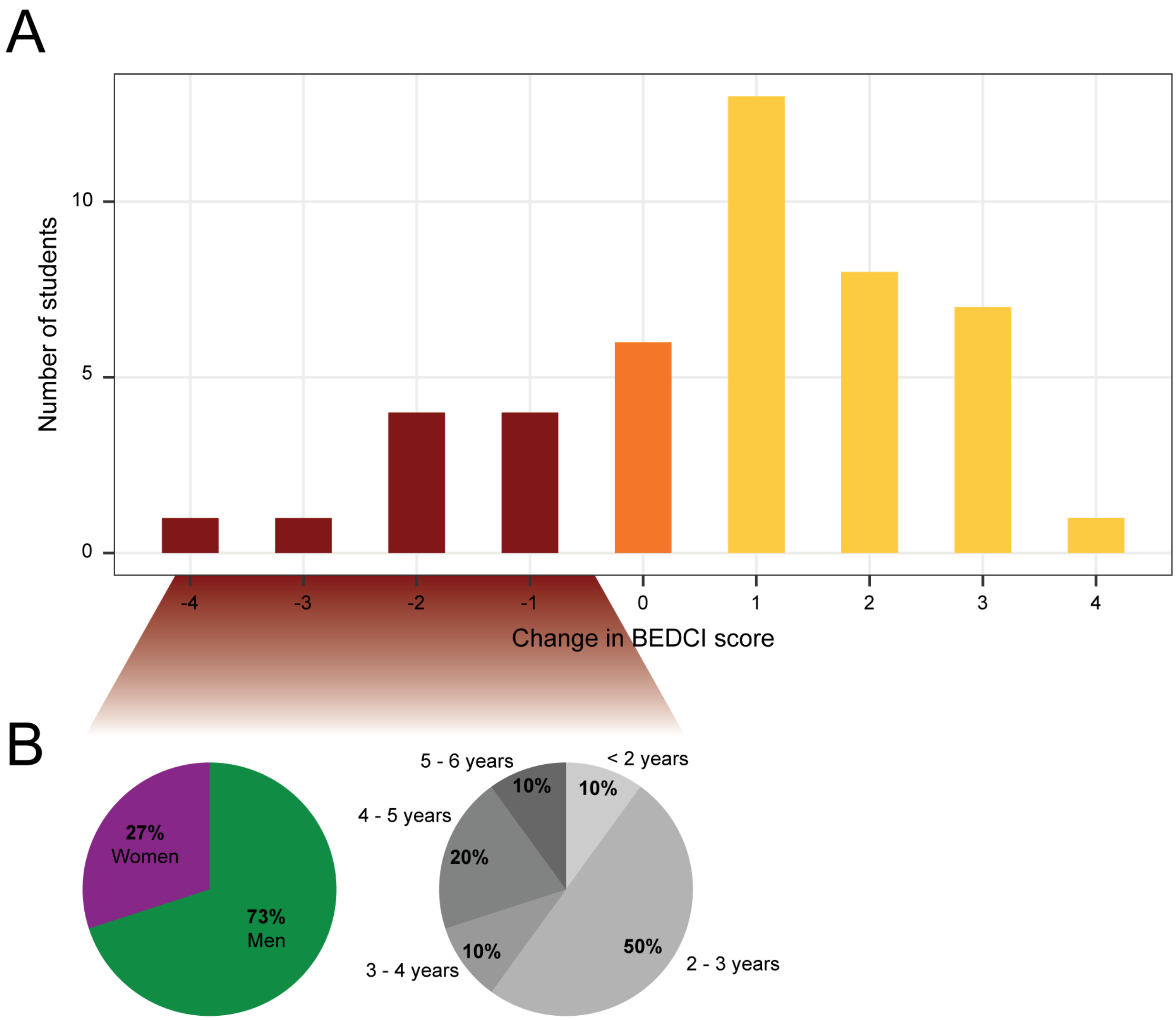
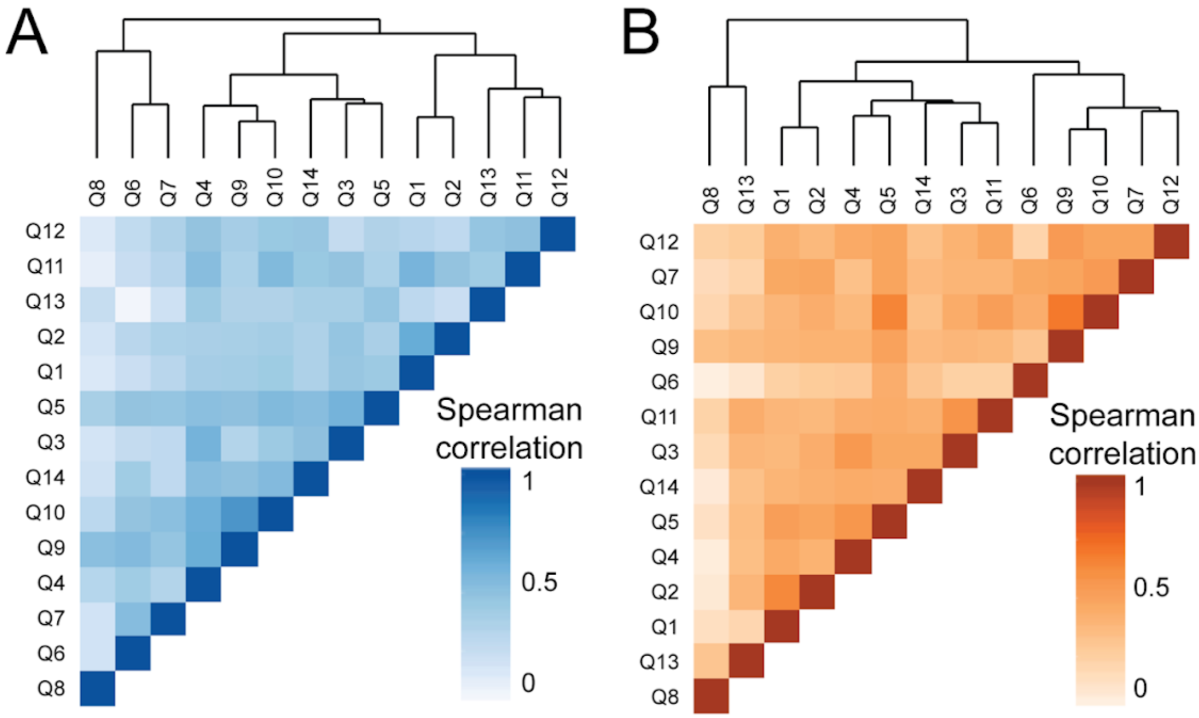



**Figure S1. 65% of students had a net increase in performance of experimental design over the semester.** A) Experimental design performance was assessed with the BEDCI and the change in BEDCI score was calculated for each student as the total number of questions correctly on the pre-test (out of 14) subtracted from the total score of each student on the post-test. Yellow bars indicate improvement over the semester, orange bars represent students that did not change over the semester, and maroon bars indicate those students that decreased over the semester. B) Distributions of the gender (left) and previous years worked in lab (right) for the students who had negative changes in BEDCI scores. For the 22% of students whose change in BEDCI score was negative ($n$ = 10), their demographic distribution is not exactly representative of that of the entire class. However, the number of students in this subset is so low that even a single student can greatly affect the reported percentages.



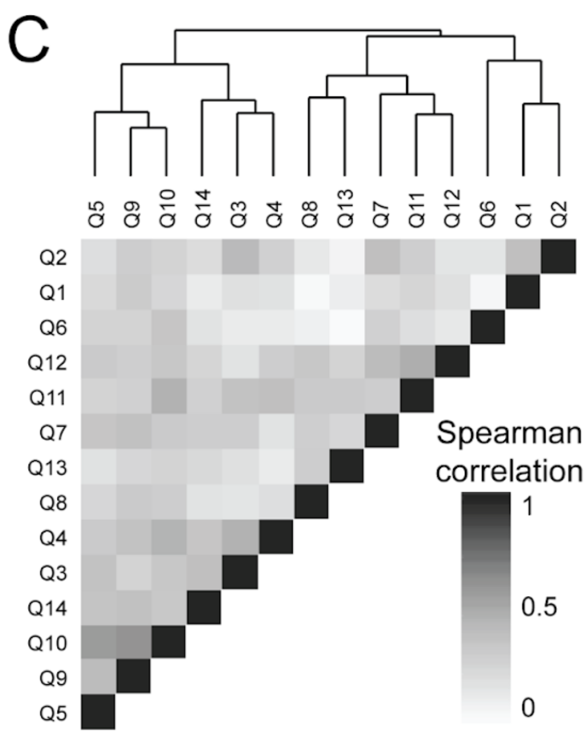




**Figure S2. Correlation generated seven topical groups of research skills self-efficacy questions.** Heat maps of the Spearman correlations between individual students' responses on the 14 self-efficacy items. Correlations were performed for the pre-test (A), post-test (B), and the change between pre- and post-tests (C). Q1 and Q2 consistently grouped together and represented the field knowledge factor. Q3, Q4, Q5, and Q14 always grouped together and were the components of the experimental design factor. Q9 and Q10 came together for the interpretation and iteration group and Q11, Q12, and Q13 formed the science communication factor. The remaining three items, Q6, Q7, and Q8 varied independently from the other items and were each given their own categories (controls, experimentation, and statistics, respectively).

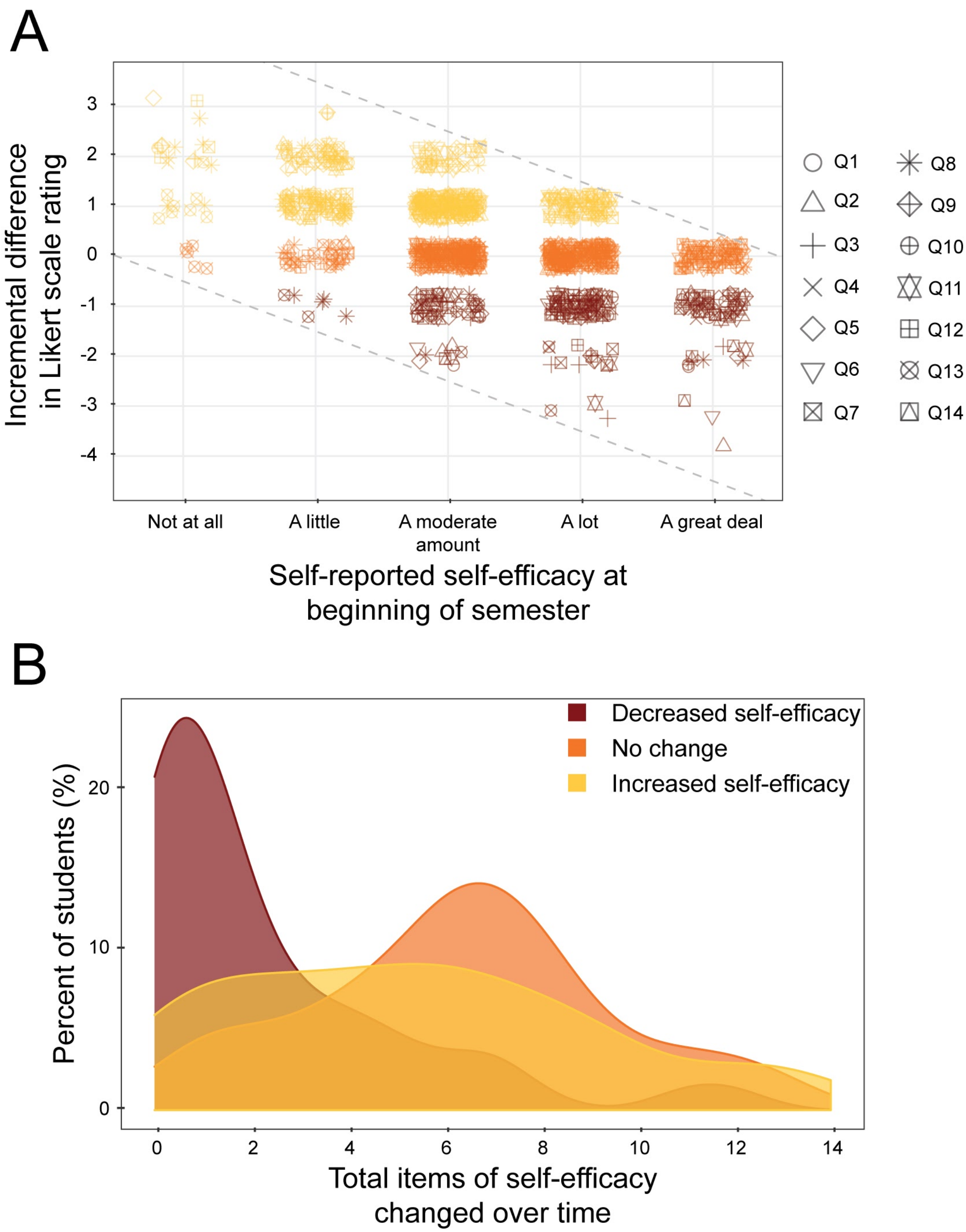



**Figure S3. Over 20% of students decrease on at least one individual self-efficacy item from the beginning to the end of the semester.** A) Dot plot showing the change in self-efficacy as a function of self-efficacy at the beginning of the semester. The majority of decreases in self-efficacy are most common on items where students initially rated themselves in one of the two highest categories. B) Plot of the number of items on which individual students had decreased (maroon), increased (yellow), or no change (orange) in self-efficacy from the beginning to the end of the semester. Many students had decreases in self-efficacy in one or a few skills over the course of the semester. However, less than 5% of the students had decreases in self efficacy on 4 or more items.

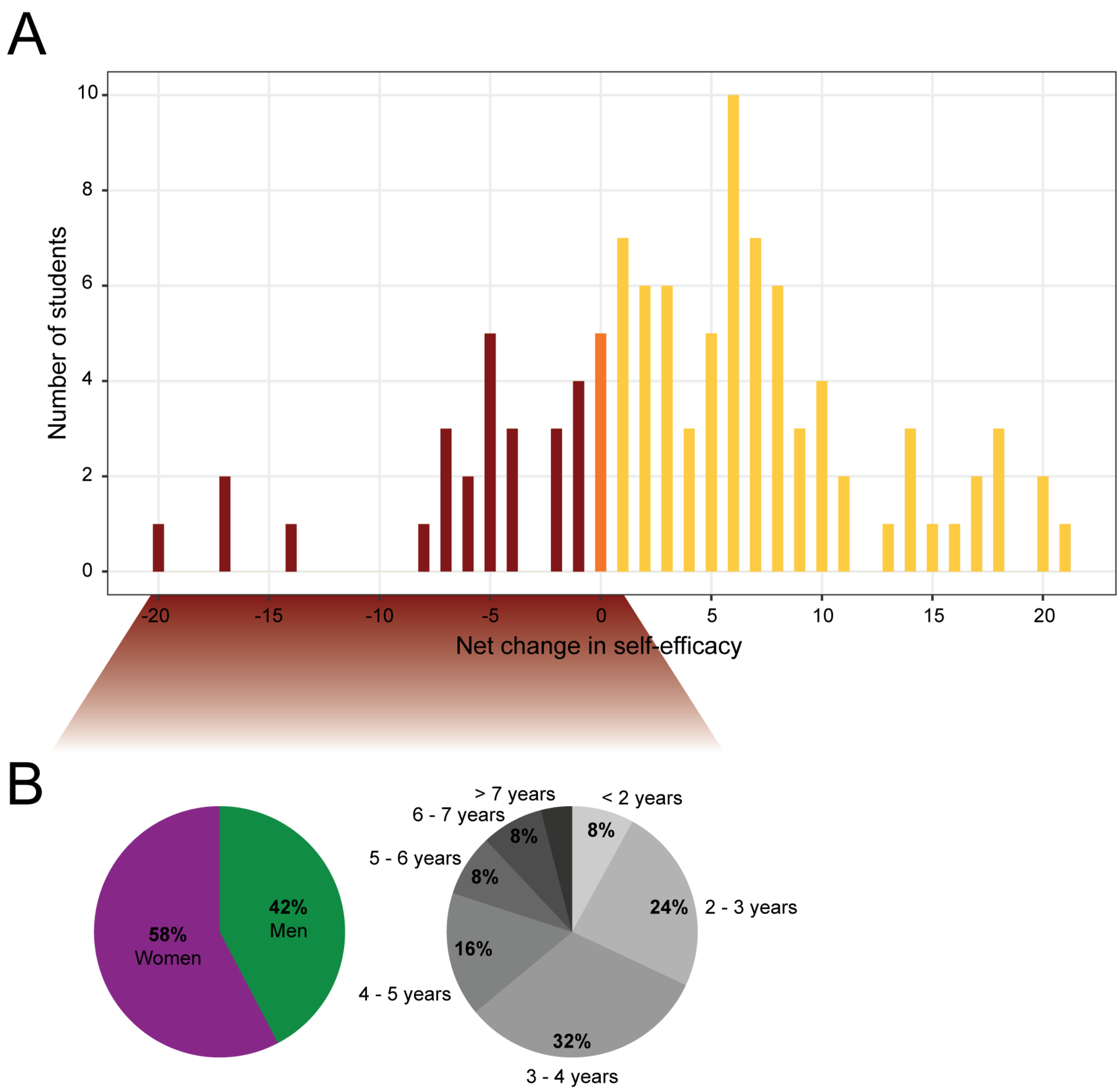



**Figure S4. Over 70% of students had a net increase in self-efficacy in research skills over the semester.** A) Net changes in self-efficacy obtained from summing changes of each student on all 14 of the self-efficacy items. Yellow bars indicate improvement over the semester, maroon bars indicate those students that decreased over the semester, and orange bars indicate a net self-efficacy change of zero. B) Distributions of the gender (left) and previous years worked in lab (right) for the students who had a negative net change in self-efficacy. Women were slightly overrepresented in the population of students who had a net negative change in self-efficacy. The was distribution of previous length of time doing research for the students who decreased in self-efficacy was similar to that of our participant population ($\chi^2$ = 3.6022, p = 0.7303), indicating that this effect was uniform across our population with respect to this metric.

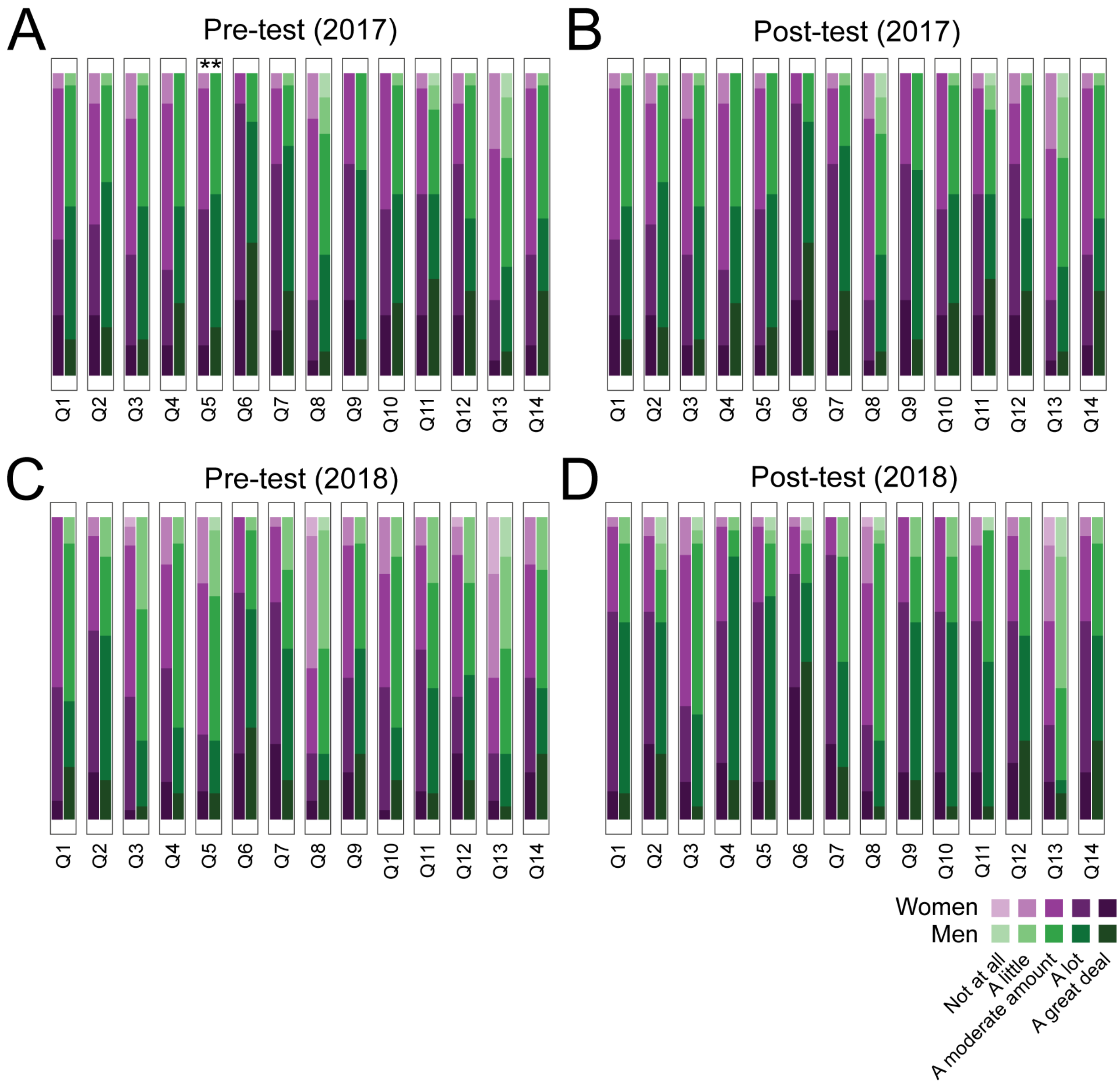



**Figure S5. The only significant difference between men and women's self-efficacy is in experimental design at the beginning of the semester in 2017.** Histograms of student responses to the 14 self-efficacy items on the 2017 pre-test (A), 2017 post-test (B), 2018 pre-test, and 2018 post-test comparing the responses of men ($n$ = 25 in 2017, $n$ = 23 in 2018) and women ($n$ = 20 in 2017, $n$ = 32 in 2018). Women have significantly lower self-efficacy than men in designing experiments (Q5) at the beginning of the 2017 semester, but otherwise do not significantly differ from men. Significance was determined with ordinal logistic regressions where pre- and post-test ratings were the response variables and gender was the predictive factor (* $p < 0.05$).

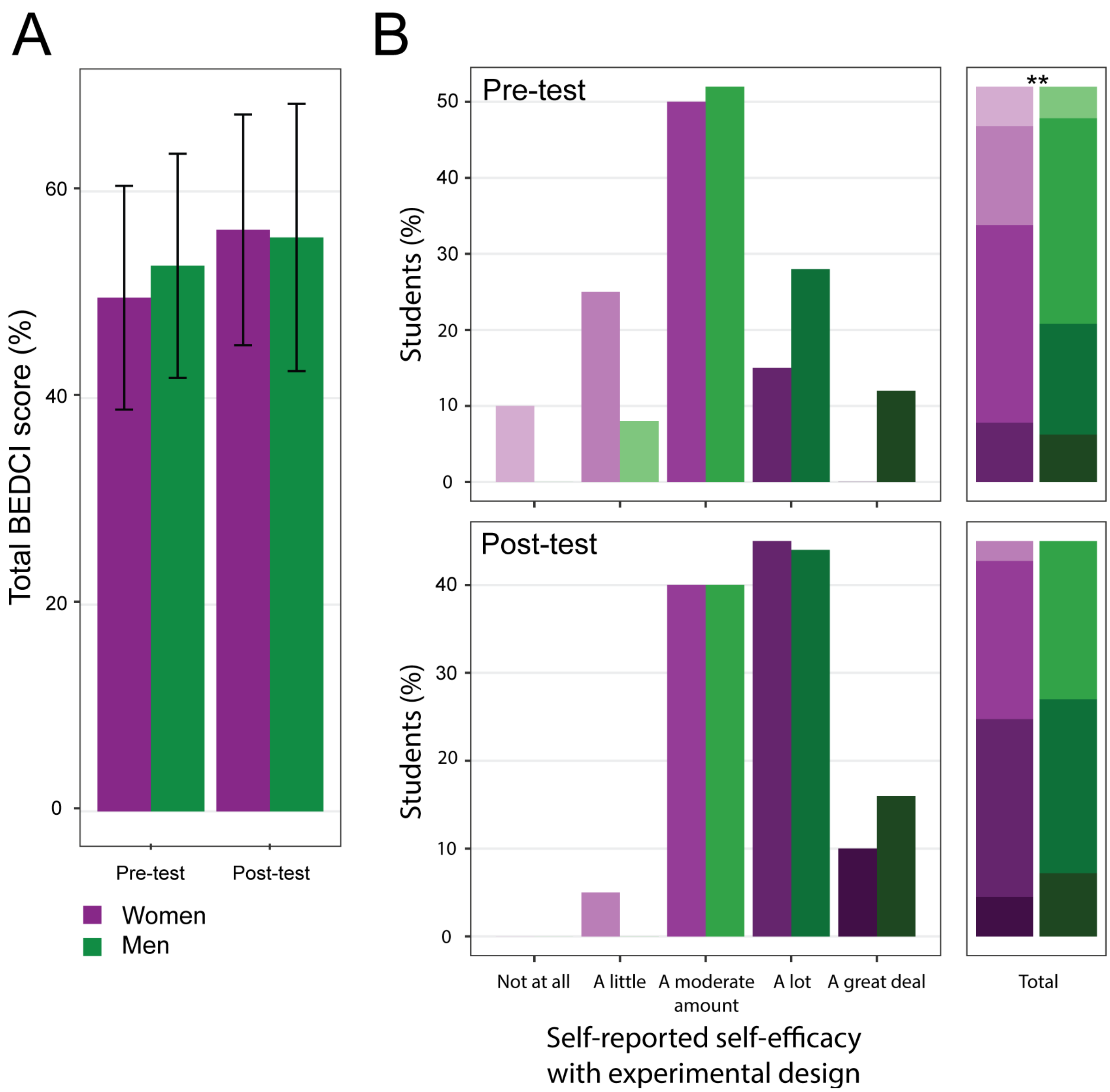



**Figure S6. Women in the 2017 student cohort were less confident in designing experiments at the beginning the semester, despite performing comparably on an assessment of experimental design aptitude. The gender gap in self-efficacy disappears by the end of the course.** A) Histogram of total BEDCI scores for men and women on the pre- and post-tests. Women do not significantly differ from men on their performance on the pre and post-tests. Significance was determined using the Student's *t*-test (* $p < 0.05$). B) Histograms separated by gender of the student responses to the self-efficacy item on designing experiments ($n = 25$ men, $n = 20$ women). Women have lower self-efficacy than men on the pre-test, but both groups increase in self-efficacy on the post-test such that there is no longer a significant difference between the groups. Significance was determined with ordinal logistic regressions where pre- and post-test ratings were the response variables and gender was the predictive factor (* $p < 0.05$).

**Supplemental Tables**

**Table S1.** Demographics of student participants in the study from 2017 and 2018.

| | 2017 | 2018 | p Value |
|---|---|---|---|
| **Gender** | | | 0.171 |
| Men | 25 (56%) | 23 (42%) | |
| Women | 20 (44%) | 32 (58%) | |
| **Race**[a] | | | 0.687 |
| URM | 6 (9%) | 7 (11%) | |
| Non-URM | 64 (91%) | 59 (89%) | |
| **Previous Years of Research Lab Experience** | | | 0.065 |
| < 2 years | 6 (13%) | 13 (24%) | |
| 2 - 3 years | 11 (24%) | 7 (13%) | |
| 3 - 4 years | 3 (7%) | 3 (5%) | |
| 4 - 5 years | 2 (4%) | 4 (7%) | |
| 5 - 6 years | 17 (38%) | 16 (29%) | |
| 6 - 7 years | 13 (29%) | 3 (5%) | |
| > 7 years | 2 (4%) | 0 (0%) | |
| **Previous Degree Subject**[b] | | | 0.433 |
| Biology | 37 (82%) | 47 (85%) | |
| Chemistry | 13 (29%) | 23 (42%) | |
| Engineering | 7 (16%) | 3 (5%) | |
| Math | 4 (9%) | 5 (9%) | |
| Physics | 3 (7%) | 2 (4%) | |
| Computer Science | 3 (7%) | 3 (5%) | |
| Other | 7 (16%) | 16 (29%) | |
| **Program**[a] | | | 0.040 * |
| Biological Sciences in Public Health | 0 (0%) | 4 (6%) | |
| Bioinformatics & Integrative Genomics | 5 (7%) | 3 (5%) | |
| Biology & Biomedical Sciences | 61 (87%) | 47 (71%) | |
| Biophysics | 0 (0%) | 2 (3%) | |
| Chemical Biology | 1 (1%) | 0 (0%) | |
| Systems Biology | 0 (0%) | 1 (2%) | |
| Virology | 3 (4%) | 9 (14%) | |
| **Experience with Experimental Design** | | | 0.905 |
| None | 2 (4%) | 2 (4%) | |
| Minimal | 11 (24%) | 10 (18%) | |
| Some | 10 (22%) | 18 (33%) | |
| Moderate | 17 (38%) | 19 (35%) | |
| Extensive | 5 (11%) | 6 (11%) | |
| **Comfort with Experimental Design** | | | 0.932 |
| Extremely uncomfortable | 3 (7%) | 2 (4%) | |
| Moderately uncomfortable | 4 (9%) | 4 (7%) | |
| Slightly uncomfortable | 9 (20%) | 12 (22%) | |
| Neither uncomfortable nor comfortable | 2 (4%) | 9 (16%) | |
| Slightly comfortable | 10 (22%) | 8 (15%) | |
| Moderately comfortable | 15 (33%) | 18 (33%) | |
| Extremely comfortable | 2 (4%) | 2 (4%) | |
| **Pre-test total self-efficacy in research skills** | | | 0.577 |
| **Post-test total self-efficacy in research skills** | | | 0.673 |
| **Net change in research skill self-efficacy** | | | 0.768 |

**Table S1. Demographics of student participants in the study from 2017 and 2018.** The student population in both years was not significantly different with the exception of graduate program affiliation (* $p < 0.05$). Differences in proportions of gender, race, previous degree subject, and program were tested with a chi-squared statistical test. Differences in distributions of self-reported experience and comfort with experimental design, as well as total and net changes in research skills self-efficacy, were assessed using the Wilcoxon signed-rank test. [a] Denotes that demographics were for all students enrolled in the course in 2017 and 2018, not just the students that opted into the study. [b] Some students reported more than one previous degree subject.

**Table S2.** Students significantly improved in many aspects of research skills self-efficacy during their first semester of graduate school.

| Research skill | Survey question | Wilcoxon's V | pAdj Value |
|---|---|---|---|
| Understand contemporary concepts in your field | 1 | 524.5 | 0.012 * |
| Make use of the primary scientific research literature in your field (e.g., journal articles) | 2 | 488.0 | 0.092 |
| Identify a specific question for investigation based on the research in your field | 3 | 677.0 | 0.023 * |
| Formulate a research hypothesis based on a specific question | 4 | 429.5 | < 0.001 *** |
| Design an experiment or theoretical test of the hypothesis | 5 | 339.5 | < 0.001 *** |
| Understand the importance of "controls" in research | 6 | 389.0 | 0.007 ** |
| Observe and collect data | 7 | 607.5 | 0.307 |
| Statistically analyze data | 8 | 498.0 | 0.001 *** |
| Interpret data by relating results to the original hypothesis | 9 | 385.5 | 0.007 ** |
| Reformulate your original research hypothesis (as appropriate) | 10 | 286.0 | < 0.001 *** |
| Relate results to the "bigger picture" in your field | 11 | 506.0 | 0.129 |
| Orally communicate the results of research projects | 12 | 413.0 | 0.007 ** |
| Write a research paper for publication | 13 | 411.0 | 0.027 * |
| Think independently | 14 | 619.0 | 0.018 * |

**Table S2. Students significantly improved in many aspects of research skills self-efficacy during their first semester of graduate school** ($n$ = 103)**.** For each of the 14 self-efficacy items, students were asked to respond "to what extent do you feel you can do each of the following?" Responses were collected on a Likert scale with prompts ranging from "Not at all" to "A great deal." Significance was determined using a Wilcoxon signed-rank test for differences in students' pre- to post-course scores, with a false discovery correction for multiple hypotheses (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two-tailed).

**Table S3.** Students significantly improved in concept inventory questions relating to controls, hypotheses, biological variation, and accuracy.

| Core concept | BEDCI question | McNemar's $\chi^2$ | pAdj Value |
|---|---|---|---|
| Controls | 1 | 4.0 | 0.159 |
| | 5 | 0.1 | 1.0 |
| Hypotheses | 2 | 5.9 | 0.159 |
| | 9 | 0.0 | 1.0 |
| Biological variation | 3 | 0.0 | 1.0 |
| | 10 | 4.0 | 0.159 |
| Accuracy | 4 | 4.9 | 0.159 |
| Extraneous factors | 6 | 0.1 | 1.0 |
| | 14 | 0.1 | 1.0 |
| Independent sampling | 7 | 0.0 | 1.0 |
| | 12 | 0.8 | 1.0 |
| Random sampling | 8 | 0.0 | 1.0 |
| | 13 | 0.0 | 1.0 |
| Purpose of experiments | 11 | 0.1 | 1.0 |

**Table S3. Students trended towards improvement in concept inventory questions relating to controls, hypotheses, biological variation, and accuracy.** Significance was determined by McNemar's chi-squared test for differences in students' pre- to post-course scores, with a false discovery correction for multiple hypotheses (1 degree of freedom).

**Table S4.** Backgrounds of women and men participants in the study.

| | Women | Men | p Value |
|---|---|---|---|
| **Number of Students** | 52 | 49 | |
| **Race**[a] | | | 0.990 |
| URM | 6 (10%) | 7 (10%) | |
| Non-URM | 57 (90%) | 66 (90%) | |
| **Previous Years of Research Lab Experience** | | | 0.040 * |
| < 2 years | 5 (10%) | 15 (31%) | |
| 2 - 3 years | 15 (29%) | 9 (18%) | |
| 3 - 4 years | 3 (6%) | 4 (8%) | |
| 4 - 5 years | 1 (2%) | 4 (8%) | |
| 5 - 6 years | 14 (27%) | 14 (29%) | |
| 6 - 7 years | 9 (17%) | 3 (6%) | |
| > 7 years | 4 (8%) | 1 (2%) | |
| **Previous Degree Subject**[b] | | | 1.0 |
| Biology | 42 (81%) | 39 (80%) | |
| Chemistry | 15 (31%) | 15 (27%) | |
| Engineering | 5 (10%) | 5 (10%) | |
| Math | 3 (6%) | 2 (4%) | |
| Physics | 2 (4%) | 2 (4%) | |
| Computer Science | 1 (2%) | 1 (2%) | |
| Other | 18 (35%) | 16 (33%) | |
| **Program**[a] | | | 0.900 |
| Biological Sciences in Public Health | 2 (3%) | 2 (3%) | |
| Bioinformatics & Integrative Genomics | 4 (6%) | 4 (5%) | |
| Biology & Biomedical Sciences | 50 (79%) | 58 (79%) | |
| Biophysics | 1 (2%) | 1 (1%) | |
| Chemical Biology | 0 (0%) | 1 (1%) | |
| Systems Biology | 1 (2%) | 0 (0%) | |
| Virology | 5 (8%) | 7 (10%) | |
| **Experience with Experimental Design** | | | 0.905 |
| None | 3 (6%) | 3 (6%) | |
| Minimal | 9 (18%) | 9 (17%) | |
| Some | 17 (33%) | 14 (29%) | |
| Moderate | 19 (39%) | 19 (37%) | |
| Extensive | 4 (8%) | 4 (8%) | |
| **Comfort with Experimental Design** | | | 0.939 |
| Extremely uncomfortable | 2 (4%) | 2 (4%) | |
| Moderately uncomfortable | 3 (6%) | 3 (6%) | |
| Slightly uncomfortable | 12 (23%) | 11 (22%) | |
| Neither uncomfortable nor comfortable | 6 (12%) | 6 (12%) | |
| Slightly comfortable | 12 (33%) | 11 (22%) | |
| Moderately comfortable | 16 (33%) | 15 (29%) | |
| Extremely comfortable | 1 (2%) | 1 (2%) | |
| **Pre-test total self-efficacy in research skills** | | | 0.809 |
| **Post-test total self-efficacy in research skills** | | | 0.990 |
| **Net change in research skill self-efficacy** | | | 0.823 |

**Table S4. Backgrounds of women and men students participating in the study.** The student populations of women and men were not significantly different with the exception of previous years of research lab experience (* $p < 0.05$). Differences in proportions of race, previous degree subject, and program were tested with a chi-squared statistical test. Differences in distributions of self-reported experience and comfort with experimental design, as well as total and net changes in research skills self-efficacy, were assessed using the Wilcoxon signed-rank test. [a] Denotes that demographics were for all students enrolled in BCMP 200 in 2017 and 2018, not just the students that opted into the study. [b] Some students reported more than one previous degree subject.

**Instruments for Data Collection**

**Questions on Pre- and Post-Course Surveys**

Your responses to this assessment will not be linked to your name and ID. Unless you specify otherwise, your responses to this assessment will be analyzed without any personal identifying information as a part of an education research study on graduate training in experimental design and science communication. Participation in this research study does not include any additional work beyond completing the requirements of the course, but may help educators better understand how to train graduate students in the life sciences. Since your responses will be collected anonymously, they will pose no risks to disclosure of your personal identity, your grades, or any other aspect of your present or future education or employment.

Allowing your responses to be included in the study is completely voluntary and you will not be penalized for choosing to withdraw from the study at any time. If you have any questions about this study or would like to withdraw from the study at a later time, please contact the study director, Madhvi Venkatesh, at madhvi_venkatesh@hms.harvard.edu.

I agree to having my responses to this assessment included in the aforementioned study?

- o Yes
- o No

Your responses for this assessment will be collected and analyzed anonymously. In order to ensure that your responses to this study can be collected anonymously, but also can be linked to your answers on the other assessments, please answer the following questions that will generate a unique identifier tied to your responses. You will be asked to complete the same questions on all of the pre- and post-course assessments.

First letter of father's first name:

First letter of mother's first name:

First letter of your middle name:

Birthdate:

Last digit of birth year (ie. 9 for 1989):

RESEARCH EXPERIENCE SELF-RATING

To what extent do you feel that you can do each of the following?

| | Not at all | A little | A moderate amount | A lot | A great deal |
|---|---|---|---|---|---|
| Understand contemporary concepts in your field | o | o | o | o | o |
| Make use of the primary scientific research literature in your field (e.g., journal articles) | o | o | o | o | o |
| Identify a specific question for investigation based on the research in your field | o | o | o | o | o |
| Formulate a research hypothesis based on a specific question | o | o | o | o | o |

| | | | | | |
|---|---|---|---|---|---|
| Design an experiment or theoretical test of the hypothesis | o | o | o | o | o |
| Understand the importance of "controls" in research | o | o | o | o | o |
| Observe and collect data | o | o | o | o | o |
| Statistically analyze data | o | o | o | o | o |
| Interpret data by relating results to the original hypothesis | o | o | o | o | o |
| Reformulate your original research hypothesis (as appropriate) | o | o | o | o | o |
| Relate results to the "bigger picture" in your field | o | o | o | o | o |
| Orally communicate the results of research projects | o | o | o | o | o |
| Write a research paper for publication | o | o | o | o | o |
| Think independently | o | o | o | o | o |

ASSESSMENT OF TRAINING IN EXPERIMENTAL DESIGN

How would you rate your current level of **experience practicing** experimental design?

- No experience
- Minimal experience
- Some experience
- Moderate experience
- Extensive experience

Currently, how **uncomfortable or comfortable** are you with designing experiments?

- Extremely uncomfortable
- Moderately uncomfortable
- Slightly uncomfortable
- Neither comfortable or uncomfortable
- Slightly comfortable
- Moderately comfortable
- Extremely comfortable

**Additional Questions on Post-Course Survey in 2018**

In the past semester, which of the following experiences have contributed to your current level of **experience** practicing experimental design? Please select all that apply.

- ☐ Completing Coursework
- ☐ Participating in Laboratory Research
- ☐ Reading Scientific Literature
- ☐ Attending Research Seminars
- ☐ Giving Scientific Presentations
- ☐ Writing a Project Proposal
- ☐ Receiving Advice from Mentors
- ☐ Discussing Scientific Topics with Colleagues
- ☐ Other (please describe below)

From your list in the previous question, which experience was the most important in changing your level of **experience** practicing experimental design? Why?

In the past semester, which of the following experience have contributed to your current level of **comfort** in designing experiments? Please select all that apply.

- ☐ Completing Coursework
- ☐ Participating in Laboratory Research
- ☐ Reading Scientific Literature
- ☐ Attending Research Seminars
- ☐ Giving Scientific Presentations
- ☐ Writing a Project Proposal
- ☐ Receiving Advice from Mentors
- ☐ Discussing Scientific Topics with Colleagues
- ☐ Other (please describe below)

From your list in the previous question, which experience was the most important in changing your level of **comfort** in designing experiments? Why?

**Additional Demographic Questions at the end of Pre-Course Surveys**

YOUR PREVIOUS ACADEMIC AND PROFESSIONAL EXPERIENCE

Have you previously completed (or are you currently enrolled in) any of the following academic training programs listed below?

| | Yes | No |
|---|---|---|
| Completed a Bachelor's degree (A.B., B.A., B.S., B.Sc.) or equivalent | o | o |
| Currently enrolled in a Post-Baccalaureate (Post-bac), non-degree granting program | o | o |
| Completed a Post-Baccalaureate (Post-bac), non-degree granting program | o | o |
| Currently enrolled in a Master's degree (M.A., M.S., M.Sc.) or equivalent | o | o |
| Completed a Master's degree (M.A., M.S., M.Sc.) or equivalent | o | o |
| Currently enrolled in a Doctorate (Ph.D.) granting program | o | o |

Please list your undergraduate major(s) and minor(s) below (if not applicable, please write N/A).
Undergraduate Major #1:
Undergraduate Major #2:
Undergraduate Major #3:
Any additional Undergraduate Majors:
Undergraduate Minor #1:
Undergraduate Minor #2:
Undergraduate Minor #3:
Any additional Undergraduate Minors:

Master's Degree #1:
Master's Degree #2:
All additional Areas of Specialty for Master's:

Year in Ph.D. program
- o G1
- o G2
- o G3 or above

We would also like to get an approximate measure of your previous experience working in a research lab.

How many YEARS have you worked in a research lab? Solely for the purposes of normalizing our results, please consider yourself working in a lab for a given week as 5+ hours, on average, for that week. If you worked in a lab for 5+ hours per week for 12 weeks, you would estimate 0.25 and select 0.1-1.9 years below.
- o No previous experience
- o 0.1-1.9 years
- o 2-2.9 years
- o 3-3.9 years

- 4-4.9 years
- 5-5.9 years
- 6-6.9 years
- 7 or more years

**Additional Demographic Question at the end of Post-Course Surveys**

PERSONAL INFORMATION

With which gender do you identify?

- Female
- Male
- Transgender Female
- Transgender Male
- Trans
- Gender Variant/Non-conforming
- Not listed (please specify)
- Prefer not to answer