Perspectives

Context Matters: Volunteer Bias, Small Sample Size, and the Value of Comparison Groups in the Assessment of Research-Based Undergraduate Introductory Biology Lab Courses

Sara E. Brownell[1]*, Matthew J. Kloser[2], Tadashi Fukami[3], and Richard J. Shavelson[4]
[1]School of Life Sciences, Arizona State University, Tempe, AZ 85287,
[2]Institute for Educational Initiatives, University of Notre Dame, Notre Dame, IN 46556,
[3]Department of Biology, Stanford University, Stanford, CA 94305,
[4]Graduate School of Education, Stanford University, Stanford, CA 94305

The shift from cookbook to authentic research-based lab courses in undergraduate biology necessitates the need for evaluation and assessment of these novel courses. Although the biology education community has made progress in this area, it is important that we interpret the effectiveness of these courses with caution and remain mindful of inherent limitations to our study designs that may impact internal and external validity. The specific context of a research study can have a dramatic impact on the conclusions. We present a case study of our own three-year investigation of the impact of a research-based introductory lab course, highlighting how volunteer students, a lack of a comparison group, and small sample sizes can be limitations of a study design that can affect the interpretation of the effectiveness of a course.

"The committee recommends that project-based laboratories with discovery components replace traditional scripted "cookbook" laboratories to develop the capacity of students to tackle increasingly challenging projects with greater independence."
—Bio2010 (National Research Council (NRC), 2002, p. 75)

INTRODUCTION

Lab courses have been a staple of the undergraduate biology curriculum since its inception, but too often the labs have taken on a "cookbook" form where students follow a protocol, much like a recipe, to obtain a known answer (24, 34). Bio2010 (26) and Vision and Change (6) advocate shifting how we teach undergraduate laboratory courses from traditional "cookbook" labs to research-focused labs (which have been described a number of ways, including: inquiry-based; discovery; investigative; project-based; research-based; and course-based research experiences). These efforts seek to make labs more representative of authentic research, so that students spend more time "thinking like a biologist" rather than following a set of directions that often fail to engage deeper thinking.

In response to the call for reform, a significant number of research-based curricula have been developed, many of which have been formally evaluated and published. Research on these curricula report significant gains in student confidence in their abilities (4, 9, 10, 11, 14), their attitudes towards science and authentic research (13), and content knowledge using an array of pre-/post- surveys (3, 5) and restricted response summative assessment tests (e.g., Biology Field Test) (13). Based on these data, these biology education reforms look as though they have succeeded—reformed biology lab courses are measurably better.

While the biology education community has made progress in the development of novel research-based lab curricula, it is important that we interpret the effectiveness of these curricula with caution and remain mindful of inherent limitations to our study designs that may impact internal validity. What research design allows us to claim that these research-based labs are better? Looking at pre-course and post-course scores within one course? Comparing reformed courses to cookbook labs? Working with 10 students or 100 students?

Furthermore, on what basis could we claim that what we find in a local context generalizes to other contexts—diverse populations of students, different types of institutions, or variations in instructional methods? Unfortunately, the specific context of many recent studies is often not addressed, which can lead to biased generalizations threatening external validity. By citing the results of the study without limiting the conclusions to the context in which it was implemented, these potentially inaccurate generalizations can weaken the evidence-based foundation of undergraduate biology education reform. In addition, these generalizations can create the false assumption that continued investigation of research-based

---

*Corresponding author. Mailing address: School of Life Sciences, PO Box 874501, Arizona State University, Tempe, AZ 85287-4501. Phone: 480-965-0803. Fax: 480-965-6899. E-mail: sebbers@gmail.com.

lab courses is unnecessary because research has already supported their efficacy.

Although standards for reporting education research have been well documented and provide a framework of expectations for both quantitative and qualitative research in education (1, 2), some of these guidelines, particularly discussing the limits of a study's conclusions, seem to be underemphasized in the biology education research community. In this paper, we highlight some of the common limitations of educational research. As an example of the challenges of implementing one of these research-based lab courses, we will discuss our own data from a research-based lab course that highlights how very different conclusions can be drawn about the effectiveness of a study depending on the type of data used. As the biology education research community continues to use social science methodology to explore better ways to teach undergraduate biology, it is paramount that we acknowledge how the specific context of our research design influences our data and conclusions.

Threats to validity: volunteer bias, small sample size, and a lack of comparison groups

Logistical and institutional barriers that influence a study's design remain major challenges to conducting biology education research on research-based lab courses. Since randomized experimental design may not always be possible, the biology education research community must take extra precautions to address factors that can threaten the validity of reported conclusions (12, 28). This includes, but is not limited to, the bias of volunteer populations of students, small sample sizes, and not using a comparison group.

Volunteer bias. Double-blind randomized trials have long been considered the 'gold standard' of scientific research, including scientific educational research (8, 12, 23, 31). Randomization minimizes the chances of unobserved, systematic differences between two conditions biasing the results. However, some studies cannot avoid using volunteers who have some information about the intervention. Among other problems, this can result in researchers finding a significant impact resulting from an intervention even if none exists—the so-called Hawthorne effect, where participants show changes *because* they are being studied. In other scenarios, the bias might arise from self-selection if volunteers are aware of what will occur in the treatment condition (27). Volunteers may also produce a sample that differs from the comparison condition in terms of gender, level of self-confidence, willingness to take risks, or previous experience doing experiments (27). When recruiting volunteers to pilot a new classroom intervention, the study will almost certainly attract individuals more pre-disposed to the treatment than students who do not volunteer. For ethical or logistical reasons, the use of volunteers with some knowledge of the treatment sometimes cannot be avoided, thus making the careful interpretation of results extremely important.

Small sample sizes. Often, equipment, personnel, scheduling, and space are limiting factors that dictate how many students can participate in a reformed lab course, often resulting in reforms that focus on upper-level elective courses with a small number of students. Moreover, reforming courses requires extensive time from faculty, teaching assistants, and external evaluators to change the nature of teaching and learning and respond flexibly to obstacles that may arise. Thus, initial pilots might involve small numbers of students before scaling-up for hundreds of students. Clearly, advantages accrue for piloting new courses with small samples, including reducing cost and minimizing logistical problems. Reporting data from these pilot studies is important, as results from these studies provide an indicator of potential directions for improving teaching and learning. However, our reporting of results should make clear the ways in which the data might be impacted by the small sample size and what added affordances or obstacles will likely exist if the class is scaled up. Ideally, researchers should also publish data from the larger, scaled-up version of the course.

Lack of comparison groups. Traditional institutional structures and logistics make randomizing students into different versions of a science course very difficult, although all too often such designs are simply eliminated from consideration due to the perceived amount of effort it would take. Quasi-experimental methods without randomization provide an alternative for comparing student outcomes (8, 30). Publishing student achievement gains on a pre- and posttest in a pilot course provides helpful data about the likelihood that the course fosters student learning in this content area. While useful, if learning gains are not compared to the traditional course's outcomes, conclusions cannot be made about whether the new approach is just as beneficial as the traditional option. A comparison to the traditional course is critical to making the most informed decisions about whether the differences in performance are worth the costs of scaling-up new innovations. Generalizations made from the results of a study without a comparison group might wrongly attribute the learning gains to the specific reform intervention when, in fact, statistically similar gains are seen in other courses that have the same time on task, thus presenting a threat to external validity (23).

Case study illustrating the impact of volunteer bias and using a comparison group: a three-year evaluation of an introductory biology research-based lab course

The Department of Biology at Stanford University recently redesigned its introductory biology laboratory course. The traditional version of the course was a standard "cookbook" laboratory—students received detailed protocols that they followed step by step to achieve a known answer. The revised course was a research-based laboratory

that was aligned with the goals of Bio2010 and Vision and Change—students worked collaboratively on experiments with unknown answers and emphasis was placed on data interpretation and analysis, as opposed to getting the "right" answer. The revised lab course's content was rooted in ecology; students investigated the relationships of biotic and abiotic factors on yeast and bacteria that reside in a flowering plant called the monkeyflower. Students collected data, compiled it into a large database shared by all students in the class, posed novel hypotheses, and used statistical analysis to investigate correlational relationships. For more details about the course structure and design, see Kloser et al. (21) and Fukami (16). Data reported here include a combination of novel and previously published data; a description of research methodologies and statistical analyses can be found in Brownell et al. (7) and Kloser et al. (20).

In the first year, 20 volunteers took the "pilot" research-based course that was offered in two sections. Volunteers were recruited because the department did not believe it fair to require participation in an untested lab course. We compared the volunteers in the pilot lab to matched non-volunteers in the traditional lab. We checked that there were no differences in the two populations in gender distribution, class year, previous research experience, and self-reported GPA. Using Likert-scale surveys, we found that students in the pilot research-based lab course showed significant positive changes in attitudes towards authentic research, confidence in their abilities to do lab-related tasks, and interest in pursuing future research compared to students in the traditional lab (7).

Based on these encouraging results, the department agreed that no ethical problems would result from randomizing students into either course for a larger implementation study. The next year, we randomized students into the research-based lab or the traditional lab. We scaled up to 34 students in the research-based lab, while keeping the section size the same (approximately 10 students). For the evaluation, we first focused on pre- to post-course gains that students made in the research-based lab. Using the same set of Likert-scale surveys, we found that students in the research-based course showed significant gains in their attitudes towards authentic research, confidence in their ability to do lab-related tasks, and interest in future research (20). Students in the research-based lab also showed improvement on a performance assessment focused on data interpretation and designing an experiment.

However, when we compared the responses of students in the research-based lab to those of students in the traditional lab, we did *not* see the same differences observed previously (Fig. 1). Similar to the pilot, students in both conditions had similar gender distribution, class year, previous research experience, and self-reported GPA—there were no differences in these self-reported characteristics. Although we saw significant gains in student attitudes towards authentic research in the research-based course, there were no differences between the two groups
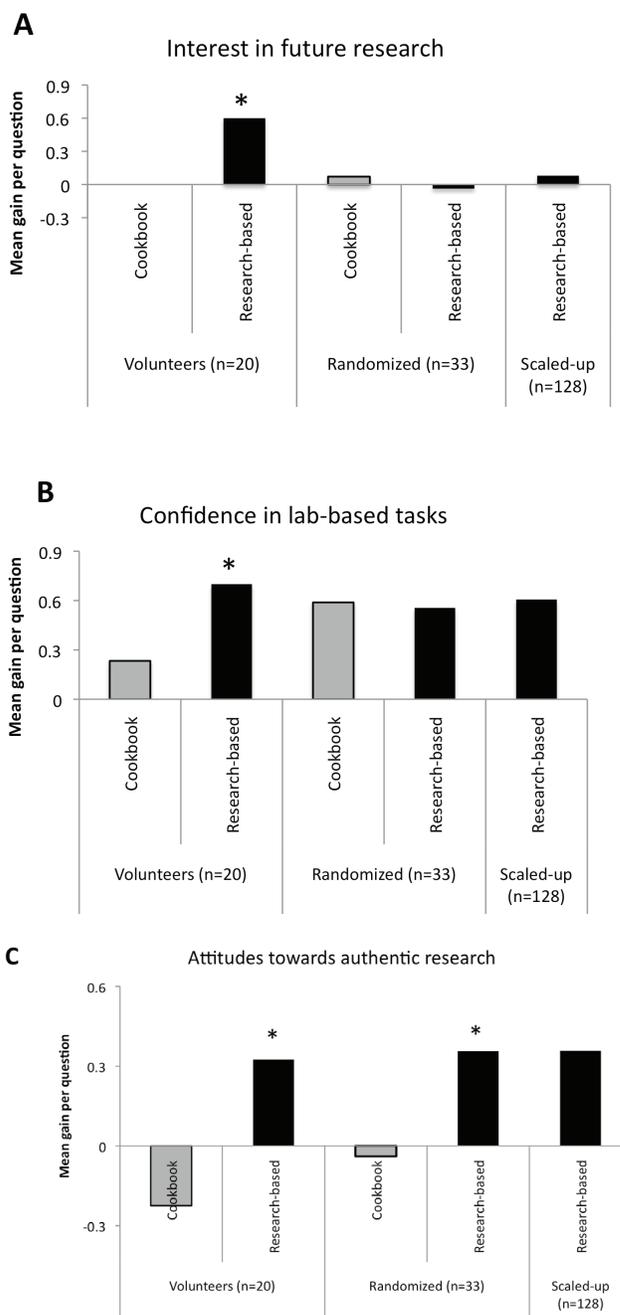

FIGURE 1. Likert-scale survey data from a three-year study of a research-based biology lab course. Students were asked a series of questions about (A) their interest in future research (2 questions), (B) their confidence in their ability to do lab-based tasks (6 questions), and (C) their attitudes towards authentic research (4 questions). Student scores on each question on the pre-course survey were subtracted from their scores on the post-course survey and averaged for that block of questions to get the main gain per question. Data shown are from three years that the course was offered to: volunteer students (n = 20 for the cookbook, n = 20 for research-based course), randomized students (n = 33 for cookbook, n = 33 for research-based course), and scaled-up research-based course students (n = 128). *$p < 0.05$ (Note: Data from Cookbook and Research-based Volunteers (7) and Research-based Randomized students (20) have previously been published.)

in student confidence in their ability to do lab-related tasks or student interest in pursuing future research. So why would the students in the research-based course in the randomized year not show higher gains in interest or confidence in ability compared with students in the traditional lab?

The notable difference between the pilot and randomized experiment was that students in the pilot course were volunteers. Might it be that students partial to authentic research and willing to take risks in a new course volunteered for the original pilot, thus skewing the study's results? While the initial study was important for our own internal evaluation and the results were interesting to the field as an initial indication of the impact of the treatment, the difference between the volunteer and non-volunteer group outcomes in our own work highlights the need to qualify findings and explicitly state limitations (7). We, along with the biology education research community, must carefully consider the limits of generalizing results from a volunteer population to *all* students in *all* universities because it could potentially have a negative impact on knowledge-building around issues of undergraduate lab reform.

In the third year, we scaled up from approximately 34 to 132 students. All students were enrolled in the research-based lab with a section size of approximately 20 students. We compared this larger group of non-volunteer students to the non-volunteer students in the traditional lab from the previous year and found results similar to the second year study (Fig. 1). Students in the research-based course compared with students in the traditional lab showed significant gains in their attitudes towards authentic research, but no differences in their interest in research or confidence in their ability to do lab-based tasks. Thus the difference in sample sizes (34 compared to 132) did not seem to have an impact, even though it was a potential limitation of our initial studies (7, 20).

We have compiled some of our previously published (and some unpublished) data to illustrate how important the context of a research study can be, particularly how the specific population of students and the inclusion of a comparison group can affect one's interpretations of the effectiveness of a course. To be clear, these are not new methodological findings, but rather, this case serves as an example from our own work that, in order for the field to move forward, we must be explicit about the limitations of our findings.

In our initial studies, if we had only reported the data from the volunteers in our research-based lab, we would have concluded that students in the research-based lab show increased interest in future research and confidence in their ability to do lab-related tasks (Fig. 2(A)). If we had compared these volunteers to matched non-volunteer students in the traditional lab, we would have concluded that students in the research-based course have greater increases in interest and confidence in their ability to do lab-related tasks than students taking a cookbook lab (Fig. 2(B)).

However, if we used the data from our second year of evaluation and reported data only from non-volunteers in the research-based lab, we would conclude that students only show gains in their confidence in their ability to do lab-related tasks, but *not* in their interest to pursue future research (Fig. 2(C)). These different conclusions likely stem from the participant population recruited—did these students volunteer or were they randomized without choice into the lab? If we added in the comparison data from randomized non-volunteers in the cookbook lab, we would conclude that students show gains in confidence in their ability to do lab-related tasks in *both* labs and do *not* show gains in their interest in doing future research in *both* labs (Fig. 2(D)). Using the comparison group demonstrates that students, regardless of what lab they are enrolled in, believe that their ability to do lab-based tasks improves—there is no difference between students in the two lab conditions. Thus, the conclusions based on these Likert-scale self-report surveys differ when using randomized non-volunteers with a comparison group from conclusions drawn from volunteers enrolled in a research-based course with no comparison group.

Although we do not see a difference in self-report interest or confidence in their ability to do lab-based tasks between the two groups when using randomized non-volunteers, we do see differences in students in the two labs in other measures: we consistently saw higher gains in students in the research-based course in their attitudes towards authentic research, regardless of whether students were volunteers or non-volunteers. We interpret this to mean that our authentic research-based course is effective at changing student attitudes about research in a positive way, independent of whether they chose to be in the course as volunteers or whether they were randomly assigned to it. We also saw that students in the research-based course showed gains in their experimental design and data interpretation skills. While we developed our own assessment of students' ability to design an experiment and interpret data, this assessment is limited to the ecological content studied in the research-based course and could not be used in the cookbook course. Recently published content-independent assessments for skills such as the experimental design ability tool (33) and scientific literacy test (18) could now be used to assess skills addressed or not addressed in the differently structured courses.

CONCLUSION

How biology education research addresses volunteer bias, sample size issues, and the presence or lack of comparison groups is important for determining the generalizability of research findings (12, 28). Other limitations must be considered as well. For example, even with a structured curriculum, the level and type of instruction is significantly impacted by an instructor's content knowledge (19, 36), pedagogical knowledge (25), pedagogical content knowledge (17, 22, 32), or personality (15, 29). Similarly, students' range
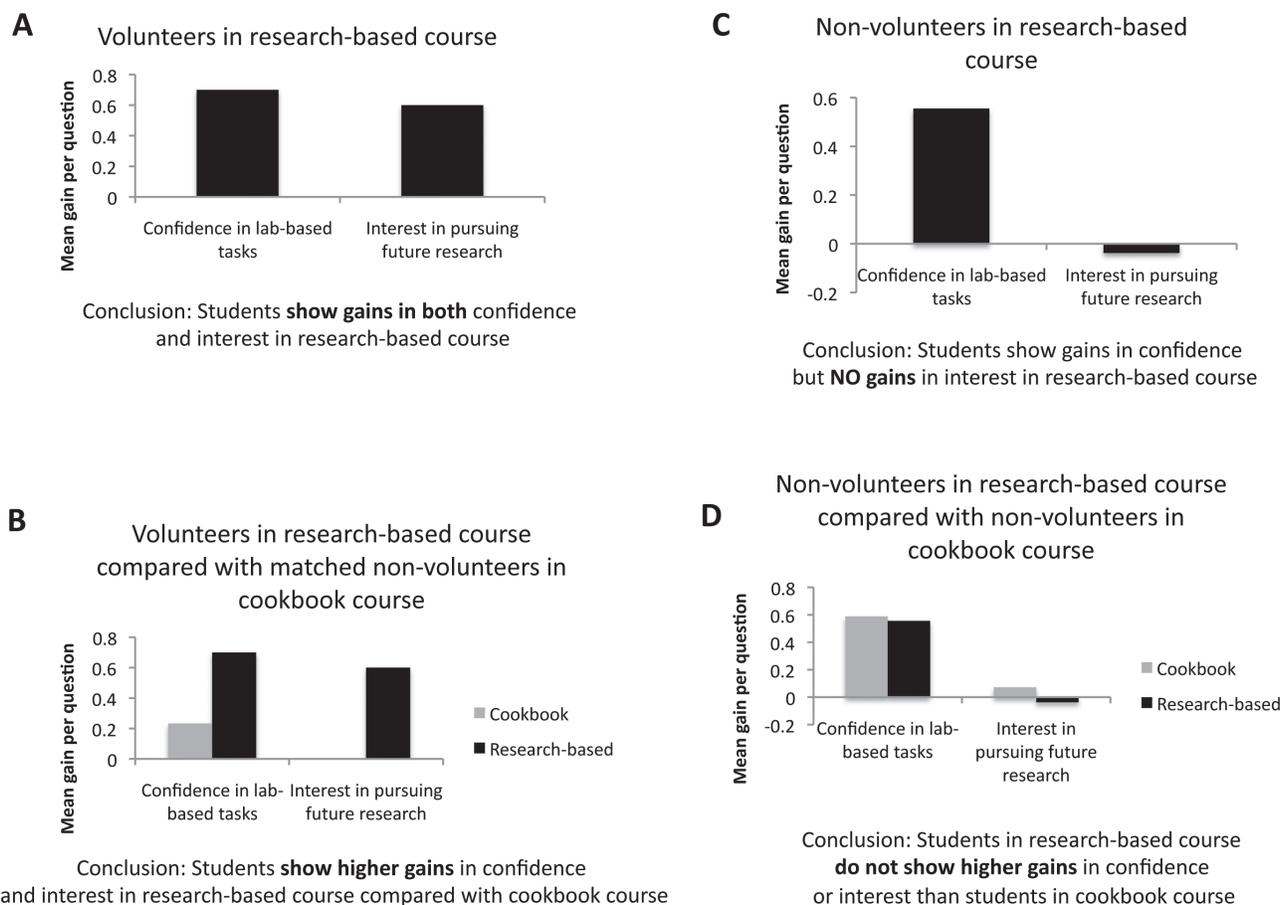
FIGURE 2. Conclusions about the effectiveness of the course differ based on which data are used. (A) The conclusion from only examining the data from the volunteers in the research-based course is that students show gains in both confidence in lab-based tasks and interest in pursuing future research. (B) The conclusion from comparing the volunteers in the research-based course with non-volunteers in the cookbook course is that students in the research-based course show higher gains than those in the cookbook course. (C) The conclusion from assessing non-volunteer students in the research-based course is that students show gains in confidence but no gains in interest. (D) The conclusion from comparing non-volunteers in the research-based course to non-volunteers in the cookbook course is that there are no differences between the students in interest or confidence in their ability to do lab-based tasks.

of general academic ability, extent of their prior knowledge and experiences, socioeconomic status, ethnicity, first generation status, and career aspirations can vary significantly within and among institutions (35). For example, conclusions drawn from data collected from a high-achieving student population may not apply to other student populations that have different learning needs. Making the problem more complicated is that some of these factors may appear to be at odds with each other. For example, if a specific number of students take a biology lab course each year, then taking volunteers or randomizing students into a treatment and comparison group will reduce the sample size within each condition, making any effect more difficult to statistically detect. However, without the comparison group, interpreting the impact of the new intervention will be limited. When dealing with a finite number of possible participants, it may be necessary to sacrifice a large sample size and qualify the interpreted data in light of this limitation. Such is the difficulty of conducting educational research, yet we must face these challenges and acknowledge these limitations in order to move the field forward.

While we focused on educational research, overextending conclusions based on limited or biased datasets is something that is present in all research science. For example, cost and labor constraints often limit the number of samples or subjects that are used in animal studies, large-scale metabolomic or transcriptomic experiments, or clinical trials. We should likely be asking about the appropriateness of generalizing from these conclusions as well.

The impetus for writing this article stemmed from a series of discussions at recent biology education conferences. On several occasions, when questions were raised regarding possible limitations of studies done on research-based lab courses, presenters dismissed these questions as irrelevant to the interpretation of the results. Perhaps these encounters are anecdotes, limited to only a few isolated public comments, and do not reflect the general consensus of the biology education community.

But whether anecdote or common occurrence, it is important for the field to remind itself about the need for qualifying our claims and generalizations so that we may better construct new knowledge in this important area of research. For the field to move forward, the biology education research community needs to be more aware of the potential limitations of its research and be cautious when generalizing specific work to other contexts. At the most basic level, all peer-reviewed biology education publications and presentations could be required to include an explicit limitations section. We must be more mindful of our specific recruitment methods, the demographics of our student population, and the caveats of comparing our data to other student groups. Given the complexities of *in situ* educational research we must continue to value exploratory studies that may not meet the rigors required in other scientific research, but in doing so we must also recognize the results as warrants for larger-scaled randomized trials and not final justification for a course design that 'works.' Then, we must convince members of our departments and institutions that larger-scaled, randomized trials are both possible and necessary to truly promote an evidence-based scholarship of teaching and learning. We hope that our case study exemplifies the importance of recognizing limitations in the interpretation of results and, perhaps more importantly, the generalization of conclusions. We want our foundation of research to be secure, even if that means less flashy conclusions in the short-term, but conclusions that in the long run will better prepare the next generation of biologists. Our research community and our students deserve nothing less.

ACKNOWLEDGMENTS

REFERENCES

1. **American Educational Research Association.** 2006. Standards for reporting on empirical social science research in AERA publications. Educ. Researcher **35**:33–40.
2. **American Psychological Association.** 2001. Publication manual of the American Psychological Association (5th ed.). Washington, DC.
3. **Aronson, B. D., and L. A. Silveira.** 2009. From genes to proteins to behavior: a laboratory project that enhances student understanding in cell and molecular biology. CBE Life Sci. Educ. **8**:291–308.
4. **Baumler, D. J., et al.** 2012. Using comparative genomics for inquiry-based learning to dissect virulence of *Escherichia coli* O157: H7 and *Yersinia pestis*. CBE Life Sci. Educ. **11**:81–93.
5. **Brame, C. J., W. M. Pruitt, and L. C. Robinson.** 2008. A molecular genetics laboratory course applying bioinformatics and cell biology in the context of original research. CBE Life Sci. Educ. **7**:410–421.
6. **Brewer, C. A., and D. Smith (ed).** 2011. Vision and change in undergraduate biology education: a call to action. American Association for the Advancement of Science, Washington, DC.
7. **Brownell S. E., M. J. Kloser, T. Fukami, and R. J. Shavelson.** 2012. Undergraduate biology lab courses: comparing the impact of traditionally based "cookbook" and authentic research-based courses on student lab experiences. J. Coll. Sci. Teach. **41**:18–27.
8. **Campbell, D. T., and J. C. Stanley.** 1963. Experimental and quasi-experimental designs for research, p. 1–37. Houghton-Mifflin, Boston, MA.
9. **Casem, M. L.** 2006. Student perspectives on curricular change: lessons from an undergraduate lower-division biology core. CBE Life Sci. Educ. **5**:65–75.
10. **Cox-Paulson, E. A., T. M. Grana, M. A. Harris, and J. M. Batzli.** 2012. Studying human disease genes in *Caenorhabditis elegans*: a molecular genetics laboratory project. CBE Life Sci. Educ. **11**:165–179.
11. **Cunningham, S. C., B. McNear, R. S. Pearlman, and S. E Kern.** 2006. Beverage-agarose gel electrophoresis: an inquiry-based laboratory exercise with virtual adaptation. CBE Life Sci. Educ. **5**:281–286.
12. **Derting T. L., K. S. Williams, J. L. Momsen, and T. P. Henkel.** 2011. Education research: set a high bar. Science **2**:1220–1221.
13. **Derting, T. L., and D. Ebert-May.** 2010. Learner-centered inquiry in undergraduate biology: positive relationships with long-term student achievement. CBE Life Sci. Educ. **9**:462–472.
14. **DiBartolomeis, S. M.** 2011. A semester-long project for teaching basic techniques in molecular biology such as restriction fragment length polymorphism analysis to undergraduate and graduate students. CBE Life Sci. Educ. **10**:95–110.
15. **Feldman, K. A.** 1986. The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: a review and synthesis. Res. Higher Educ. **24**:139–213.
16. **Fukami, T.** 2013. Integrating inquiry-based teaching with faculty research. Science **339**:1536–1537.
17. **Gess-Newsome, J.** 2002. Pedagogical content knowledge: an introduction and orientation, p. 3–17. *In* Examining pedagogical content knowledge. Springer, Netherlands.
18. **Gormally, C., P. Brickman, and M. Lutz.** 2012. Developing a test of scientific literacy skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. CBE Life Sci. Educ. **11**:364–377.

19. **Grossman, P. L., S. M. Wilson, and L. S. Shulman.** 1989. Teachers of substance: Subject matter knowledge for teaching, p. 23–36. *In* M. Reynolds (ed), The knowledge base for beginning teachers. Pergamon, New York.

20. **Kloser, M. J., S. E. Brownell, R. J. Shavelson, and T. Fukami.** 2013. Effects of a research-based ecology lab course: a study of nonvolunteer achievement, self-confidence, and perception of lab course purpose. J. Coll. Sci. Teach. **42**:72–81.

21. **Kloser, M. J., S. E. Brownell, N. R. Chiariello, and T. Fukami.** 2011. Integrating teaching and research in undergraduate biology laboratory education. PLoS biology **9**(11), e1001174.

22. **Magnusson, S., J. Krajcik, and H. Borko.** 1999. Nature, sources, and development of pedagogical content knowledge for science teaching, p. 95–132. *In* J. Gess-Newsome and N. Lederman (ed), Examining pedagogical content knowledge: The construct and its implications for science education. Kluwer Academic Publishers, Dordrecht, The Netherlands.

23. **Marczyk, G., D. DeMatteo, and D. Festinger.** 2005. Essentials of research design and methodology, p. 65–94. John Wiley & Sons, Hoboken, NJ.

24. **McComas, W.** 2005. Laboratory instruction in the service of science teaching and learning. Science Teacher **27**:24–29.

25. **Morine-Dershimer, G., and T. Kent.** 2002. The complex nature and sources of teachers' pedagogical knowledge, p. 21–50. *In* Examining pedagogical content knowledge. Springer, Netherlands.

26. **National Research Council (US). Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, & NetLibrary, Inc**. 2003. BIO 2010: transforming undergraduate education for future research biologists.

27. **Rosenthal, R., and R. L. Rosnow.** 1975. The volunteer subject, p. 121–133. John Wiley & Sons, New York.

28. **Ruiz-Primo, M. A., D. Briggs, H. Iverson, R. Talbot, and L. A. Shepard.** 2011. Impact of undergraduate science course innovations on learning. Science **331**:1269–1270.

29. **Rushton, S., J. Morgan, and M. Richard.** 2007. Teacher's Myers-Briggs personality profiles: identifying effective teacher personality traits. Teaching and Teacher Education **23**:432–441.

30. **Shadish, W. R., T. D. Cook, and D. T. Campbell.** 2002. Experimental and quasi-experimental designs for generalized causal inference, p. 1–81. Houghton Mifflin Company, New York.

31. **Shavelson, R. J., and L. Towne (ed).** 2002. Scientific research in education, p. 80–126. The National Academies Press, Washington, DC.

32. **Shulman, L. S.** 1986. Those who understand: knowledge growth in teaching. Educ. Res. **15**:4–14.

33. **Sirum, K., and J. Humburg.** 2011. The Experimental Design Ability Test (EDAT). Bioscene: J. Coll. Biol. Teach. **37**:8–16.

34. **Sundberg, M. D., J. E. Armstrong, and E. W. Wischusen.** 2005. A reappraisal of the status of introductory biology laboratory education in US colleges & universities. Am. Biol. Teach. **67**:525–529.

35. **Tanner, K., and D. Allen.** 2004. Approaches to biology teaching and learning: learning styles and the problem of instructional selection—engaging all students in science courses. Cell Biol. Educ. **3**:197–201.

36. **Wilson, S., L. Shulman and A. Richert.** 1987. "150 different ways of knowing": representations of knowledge in teaching, p. 104–123. *In* J. Calderhead (ed), Exploring teachers' thinking. Cassell, Eastbourne, UK.