Closing the Achievement Gap in STEM: A Two-Year Reform Effort at Brown University

Kyle F. Trenshaw[1], David M. Targan[1, 2], and James M. Valles[2]
[1]Science Center, [2]Department of Physics
Brown University

Undergraduate attrition from science, technology, engineering, and mathematics (STEM) fields remains a problem despite decades of research and education reform efforts to improve the STEM student experience. This study is one piece of a multifaceted STEM reform effort built on the hypothesis that more students will persist in learning environments that include evidence-based practices that expand subject comprehension and improve student satisfaction. At Brown University, four physical science departments worked together over two academic years to implement evidence-based practices in eight different large, introductory STEM courses. This study focuses on the subset of those courses in which both students' participation in voluntary collaborative learning experiences was tracked and students' pre-course preparation was assessed using course-specific cognitive measures. In these courses, the more often students participated in collaborative learning experiences, the more their performance improved compared to their pre-course preparation. Further, participation in collaborative learning experiences was found to disproportionately benefit the performance of female students and students from historically under-represented groups (HUGs). Based on these outcomes, this study suggests that implementing evidence-based practices such as collaborative learning in introductory STEM courses can help to close the achievement gap for women and HUGs.

Corresponding Author: Kyle F. Trensahw, kyle_trenshaw@brown.edu

Background and Significance

For years, the National Research Council and other organizations have been circulating reports about the ongoing crisis that is undergraduate science, technology, engineering, and mathematics (STEM) education in the United States. These reports describe the various problems with STEM education and potential solutions generally [1-5]; in regard to inclusivity for diverse populations [6-8]; and within the specific disciplinary contexts of engineering [9-10], chemistry [11], physics [12], and mathematics [13]. The message is clear: all students (although, disproportionately female students and students from historically under-represented groups (HUGs)) benefit from evidence-based pedagogical practices and these practices have been demonstrated to improve students' cognitive and affective outcomes in a variety of classroom settings and STEM disciplines.

Gender and racial disparity remain significant problems at all levels of STEM education despite decades of research and reform efforts to improve the STEM student experience. Data on gender from 35 Association of American Universities (AAU) institutions indicates that in 2012, 31% of assistant professors, 22% of associate professors, and just 13% of full professors in STEM were female, compared to 51%, 46%, and 29% in non-STEM respectively [14]. While the proportion of female professors in STEM is slowly increasing by one to two percentage points each year, the same cannot be said for STEM bachelor's degrees awarded to women, which have been hovering around 40% of awarded STEM degrees for almost a decade [14]. The situation is similarly grim when considering race and ethnicity, with less than 19% of STEM bachelor's degrees awarded to students from HUGs—defined as U.S. citizens identifying as African American, Black, Hispanic, Latino, Native American, or Pacific Islander—in 2012 compared to their representation in the U.S. population at 31% [15]. Although they may see representation of people of color from other countries or of Asian Americans, many HUG students will never see themselves reflected in their professors, or even their teaching assistants, because the number of doctoral degrees awarded to HUG students in the United States is still below 10%, and not every doctoral degree culminates in a faculty position [15].

These statistics have not gone unnoticed by STEM education researchers, who have shown that using evidence-based practices in the classroom improves student persistence in STEM fields. A recent study found that students who left chemistry had lower self-confidence and higher fear of failing, and experienced those feelings as more closely tied to their self-worth than their more persistent peers [16]. Many evidence-based practices are designed to specifically address affective outcomes like confidence and anxiety in STEM students and still others have been found to indirectly improve these outcomes [17]. As such, studies recommend exposing students to evidence-based practices as much as possible throughout STEM curricula to increase the likelihood that they will develop resilience in the face of adversity and persist in their chosen fields of study [17-18].

This study is one piece of a multifaceted STEM reform effort built on the hypothesis that more students will persist in learning environments that include evidence-based practices that expand subject comprehension and improve student satisfaction. In this study, we sought to answer the following research question: how does participation in voluntary collaborative learning experiences affect student performance on the final examination of the course when students' pre-course preparation is considered?

Context

At Brown University, four physical science departments, including the Division of Applied Mathematics, the Department of Chemistry, the School of Engineering, and the Department of Physics, collaborated on a proposal to become a project site for the AAU Undergraduate STEM Education Initiative. One of eight project-site grants was awarded to Brown and project work began during the summer of 2013. During the 2013-2014 and 2014-2015 academic years, instructors in the four departments implemented evidence-based practices in eight different large, introductory STEM courses. Because different evidence-based practices appealed to different instructors, and different instructors taught the same course across different semesters, each individual offering of each course is in many ways unique. However, because many instructors collaborated with one another between courses and even across disciplinary boundaries, there are also threads of similarity in the course implementations, one of which we describe in this study.

To answer our research question, we focus on a subset of the eight courses included in the project that share the following characteristics:

a) Lectures included some form of *active learning*, such as use of real-time student response systems, facilitated interaction between students, and/or hands-on demonstrations of scientific principles.
b) Collaborative learning experiences (CLEs) were offered to students outside of the course lecture, and were *voluntary* such that attendance did not contribute directly to students' grades in the course.
c) The CLE attendance of individual students was tracked over time.
d) Grading for the course was *not mandatory S/NC*[†] such that students' final examinations were valid measures of their overall course performance.
e) Students' pre-course preparation was assessed using a course-specific cognitive measure in the early weeks of the semester (see Method section below).

Four course implementations from the 2013-2014 and 2014-2015 academic years share all of these characteristics: the Spring 2014 offering of "Equilibrium, Rate, and Structure" (S2014 CHEM0330), both the Spring 2014 and Spring 2015 offerings of "Foundations of Electromagnetism and

[†] At Brown University, "mandatory S/NC" refers to a grading scheme where students can only receive a grade of satisfactory (S) or a grade of no credit (NC) on their transcripts, rather than a letter grade. Consequently, students may not perform to their best ability on the final examinations in these courses if they predict that the sum of the rest of their course performance will already yield a grade of S.

Modern Physics" (S2014 PHYS0060 and S2015 PHYS0060 respectively), and the Fall 2014 offering of "Foundations of Mechanics" (F2014 PHYS0050).

CHEM0330 is the introductory course for chemistry concentrators‡ and a required course in many other concentrations, most notably for students seeking to pursue a pre-med track. CLEs in S2014 CHEM0330 took the form of facilitated problem solving, in which students worked in small groups of three or four to solve both conceptual and quantitative chemistry problems. These problems were intentionally designed to be more complex than those students might see in textbooks or online and often required students to think deeply about the underlying chemistry concepts, rather than "plug-and-chug" through simplistic calculations. Instructors were available to facilitate discussion and monitor the pace of each group to avoid situations such as individual students working ahead or entire groups becoming frustrated and stalling out. Attendance at CLEs was *voluntary opt-in* because there were only enough seats to accommodate about one third of the total course enrollment. Students could voluntarily opt in to attending, but once they opted in, they were required to continue attending or their seat would be given to another interested student. This voluntary opt-in policy allowed instructors to assign students to groups and avoid some of the common difficulties with group work, such as students choosing their friends or isolated females or students of color in groups [19].

PHYS0050 (Foundations of Mechanics) and PHYS0060 (Foundations of Electromagnetism and Modern Physics) are a paired sequence of courses generally taken by students in non-physics STEM concentrations, such as engineering or geological sciences. All three PHYS offerings in this study were staffed by the same two instructors, one giving lectures and one organizing the CLEs. Thus, the pedagogy across the offerings was very similar. CLEs in the PHYS offerings were similar to those in S2014 CHEM0330 and included students working in small groups of three or four to solve physics problems with instructors available for support and guidance. However, the problems in PHYS CLEs tended to be more quantitative than conceptual, with the bulk chosen from the course textbook. Using textbook problems ensured that CLEs would help prepare students for their homework assignments, which were also largely made up of textbook problems. Aside from the homework-like textbook problems, each CLE also included one or two "challenge problems" which students could attempt if they completed the rest of the problems with time to spare. These challenge problems were more complex and conceptual, often relating to ongoing challenges in physics research or the physics behind real world phenomena. Attendance in CLEs was completely voluntary, so to incentivize students to attend, the instructors promised that one problem on each exam would be an exact copy of a problem students had seen during the CLEs. With two midterms and a final examination in each offering, students saw about 20 CLE problems prior to each exam. The instructors also promised that the problem chosen for the exam would be a problem that all students who attended the CLEs had completed during the allotted time; thus, students were asked to sign in at their tables and log the problems they completed on their sign-in sheet. Despite this incentive structure, attendance to CLEs varied and students could not be assigned groups. However, students were encouraged to find a group they felt they worked well with and to make plans to attend together on a regular basis. Consequently, many groups stayed consistent over time.

Method

The main evaluative technique in this study was the use of course-specific cognitive measures of students' pre-course preparation in the early weeks of each semester. The specific measures used in each course each semester are listed in Table 1.

Table 1. Cognitive measures of students' pre-course preparation by course

Course Offering	Disciplinary Pre-test	Mathematics Pre-test

‡ Concentrations at Brown University are synonymous with majors at many other institutions. Concentrators are students majoring, or "concentrating," in a particular discipline.

S2014 CHEM 0330	Chemistry Concept Inventory (CCI) [20]	University of Minnesota Mathematics Assessment for Second Semester Chemistry (UnMMASSC) [21]
S2014 PHYS 0060	-	internally developed instrument (unpublished)
F2014 PHYS 0050	Force Concept Inventory (FCI) [22]	internally developed instrument (unpublished)
S2015 PHYS 0060	Brief Electricity and Magnetism Assessment (BEMA) [23]	internally developed instrument (unpublished)

In S2014 CHEM0330, the pre-tests were administered as one assessment through an online course management system. While students were given homework points for completing the assessment, their performance on the pre-tests did not affect how many points they received. This completion grading of the pre-tests was used to discourage students from cheating to achieve a better score. In all PHYS course offerings, the pre-tests were administered on paper and in person during a scheduled course meeting. Students received no incentive to take the pre-tests other than the knowledge that doing so would give them insight into what types of concepts would be taught during the course and help the instructors tailor their course content to students' needs. Because of these differences in pre-test administration, S2014 CHEM0330 students are over-represented in the sample compared to all PHYS students, as seen in Table 2.

Table 2. Response rates of students who took both the pre-test and final examination by course

Course Offering	**Total Enrolled**	**Took Both Pre-test and Final**	**Response Rate (%)**
S2014 CHEM0330	227	217	96%
S2014 PHYS0060	64	42	66%
F2014 PHYS0050	121	92	76%
S2015 PHYS0060	79	41	52%
Total	**491**	**392**	**80%**

To check that our sample was representative of the total population of the four course offerings, we obtained gender and race/ethnicity information for the enrolled students through our institutional research office. Using this information, we compared the response rates of female, male, Asian, Black or African American, Hispanic or Latino, and White subsets of students in each course to the total response rate for that course, as seen in Table 3.

Table 3. Response rates of subsets of students who took both the pre-test and final examination by course

Sample Subset	**S2014 CHEM0330**	**S2014 PHYS0060**	**F2014 PHYS0050**	**S2015 PHYS0060**	**Total**
Female	96%	71%	79%	54%	83%
Male	95%	61%	73%	50%	77%
Asian	100%	71%	73%	**70%**	80%
Black or African American	100%	75%	82%	**33%**	86%
Hispanic or Latino	94%	75%	69%	57%	84%
White	95%	68%	80%	52%	80%
Total	**96%**	**66%**	**76%**	**52%**	**80%**

Overall, no subset response rate varied by more than six percentage points in comparison to the total response rate, with male students slightly under-represented and female, Black or African American, and Hispanic or Latino students slightly over-represented in the sample.

In individual courses, only S2015 PHYS0060 had any subset response rates that varied by more than ten percentage points from the total response rate for the course; the two subsets were Asian students, with 14 responding out of a total of 20, and Black or African American students, with 2 responding out of a total of 6. A low response rate increases the probability of a non-representative sample, and S2015 PHYS0060 had the lowest total response rate of the four course offerings, which could have contributed to the over-representation of Asian students and the under-representation of Black or African American students in the sample.

To analyze the entire sample of data together despite the disparate assessment methods and grading schemes, we used average-normalized scores (ANSs) for both the pre-tests and final examinations in each course. Final exam scores were used as a summative measure of student performance rather than final course grades because students in all four course offerings completed them on paper, in person, and individually, where final course grades included things like homework and laboratory assignments, which could be completed with peers and are more susceptible to cheating. ANSs for the pre-tests and final examinations scores were calculated using the following equation:

$$\text{ANS} = \frac{\text{individual student score}}{\text{average score for entire course sample}}$$

This calculation yielded a number that averaged to a value of 1 for the entire course sample. ANSs greater than 1 indicate that a student did comparatively better than their peers on that assessment and ANSs less than 1 indicate that a student did comparatively worse than their peers on that assessment.

ANSs for the pre-tests and final examinations were combined to create what we term improvement ratios (IRs) for each individual student using the following equation:

$$\text{IR} = \frac{\text{final exam ANS}}{\text{pre-test ANS}}$$

Thus, a students' IR is a measure of their improvement in course ranking between the pre-test and final examination. An IR of exactly 1 indicates no change in position relative to their peers, greater than 1 indicates an improvement in position, and less than 1 indicates a worsening in position. For example, if a student had performed a bit below average on the pre-test, with an ANS of 0.80, but had preformed a bit above average on the final examination, with an ANS of 1.2, their IR would be 1.5, indicating that they improved their position relative to their peers by the final examination. IRs for our sample ranged from 0.18 (a pre-test ANS of 1.0 and a final exam ANS of .18) to 4.0 (a pre-test ANS of 0.21 and a final exam ANS of 0.83).

To explore how IRs were affected by participating in CLEs, we correlated students' IRs with the percentage of total CLEs available during the semester (see Results section below). S2015 CHEM0330 included 12 CLEs, and all PHYS offerings included 11 CLEs. We used directional tests with a threshold value of $p < 0.05$ when checking for statistically significant correlations. Correlation effect sizes were defined using the following conventional scale [24]: small = 0.10-0.29, medium = 0.30-0.49, and large ≥ 0.50.

We want to acknowledge two sources of error that we cannot account for given the data available for this study. First, the validity of ANSs hinges on students giving their full effort on the assessments from which the original scores are derived. Because pre-tests were not graded in any of the course offerings in our sample, students' only incentive to put forth full effort was that the instructors requested that they do so. Further, in PHYS offerings, while students were all given the same amount of time to complete the pre-tests, they could also leave the classroom as soon as they had finished. In all four course offerings, students may have rushed through the pre-tests or selected random answers, despite both behaviors being explicitly discouraged by instructors. We cannot clean the data by selecting out for these behaviors because they are not captured in students' pre-test responses. Similarly, while none of the courses were mandatory S/NC, students at Brown can still opt to take a course S/NC rather than receive a letter grade. For these students, their final exam grades might not reflect their full effort. Second, some students are included multiple times in our sample because they took more than one of the four courses. While each course is a different experience and taking both PHYS0050 and PHYS0060 is standard, our sampling method could be seen as double-counting in terms of our total sample demographics. Because the pre-tests, final examinations, and overall populations are different for each course, we assert that including students multiple times is not strictly double-counting. Still, 29 students appear in the sample twice and 1 student appears three times. We cannot make claims about the complexities of how students might be affected by participating in CLEs in multiple courses, either positively or negatively, using a sample size of 30 students. We hope that in acknowledging these sources of error, the power and limitations of our results below can be better understood and referenced by others.

Results and Discussion

The Pearson correlation coefficient (PCC) between students' IRs and their CLE attendance was calculated for the entire sample and all gender- and race/ethnicity-defined subsets that contained at least 15 total students, with the exception of two race/ethnicity categories ("Non-Resident Alien" and "Two or More Races," both of which are unpredictably heterogeneous[§]). The results of these calculations can are summarized in Table 4.

Table 4. Correlation between students' improvement ratios and CLE attendance

Sample Subset	N	IR Increase Per 1/10 of CLEs Attended	PCC	Effect Size
Total	392	0.017	0.18*	Small
Female	212	0.021	0.20*	Small
Male	180	0.008	0.09	-
Asian	69	0.015	0.12	Small
Female	37	0.014	0.09	-
Male	32	-0.003	-0.04	-
Black or African American	42	0.021	0.31*	Medium
Female	23	0.012	0.17	Small
Male	19	0.035	0.46*	Medium
Hispanic or Latino	63	0.025	0.32*	Medium
Female	33	0.032	0.40*	Medium
Male	30	0.016	0.21	Small
White	152	0.012	0.16*	Small
Female	74	0.013	0.17	Small
Male	78	0.006	0.08	-

* Statistically significant based on a directional test with a threshold value of $p < 0.05$

In answer to our research question, participating in CLEs seems to improve students' performance on the final examination when their pre-course preparation is considered. CLE attendance was positively correlated to a small, statistically significant degree with IRs for all students, all female students, and all White students, and to a medium, statistically significant degree for all Black or African American students, all Hispanic or Latino students, male Black or African American students, and female Hispanic or Latino students. These results are consistent with previous studies where female students and students from HUGs disproportionately benefit from opportunities to make social connections and have their values affirmed in STEM contexts [25-26]. Further, for Black or African American students, males are often under-represented compared to females across STEM disciplines [27], so early positive experiences with CLEs could encourage more male Black or African American students to continue in STEM. Based on these results, implementing CLEs in introductory STEM courses could make the climate in STEM more welcoming and help close the achievement gap for students from HUGs.

While the IRs for the overall sample of students were positively correlated with their CLE attendance, when the results were disaggregated by the gender and race/ethnicity categories, attending CLEs resulted in null effects on students' IRs for 4 of the 15 subsets: all male students, female Asian students, male Asian students, and male White students. To better understand this apparent lack of benefit for some students, we considered how pre-course preparation might differ between the subsets. What we found was a flattening effect for students with proportionally higher pre-test scores compared to their peers. For example, male Asian students (the only subset with a negative correlation coefficient) began the course with the highest pre-test scores of all the subsets. Consequently, these students had the least room to improve on the final examination compared to their peers and had less varied IRs overall, which flattened their respective IR vs. CLE attendance curve. Even considering the subset of male Asian students, no statistically significant negative correlations were observed between students' CLE attendance and their IRs, and we still advocate the use of CLEs in introductory STEM courses based on our results.

The limitations of this study should be considered before attempting to generalize our findings to other contexts and populations. Brown University is a small, private institution with an acceptance rate of less than 10% annually. Thus, we cannot claim that our sample would be generalizable to a large, public institution with a higher acceptance rate. Further, our sample only includes introductory courses in chemistry and physics and over-represents the outcomes of chemistry students within the sample. While introductory STEM courses in disciplines such as engineering, mathematics, or computer science may be similar to the courses included in

[§] The authors want to acknowledge the inherent heterogeneity within all race/ethnicity categories used for federal reporting in the United States and how said categories can be used to marginalize and silence the voices of specific populations. We hope that in advocating for positive change in STEM education we are using the categories to their best advantage, rather than further marginalizing HUGs.

our sample, we cannot claim that implementing CLEs would have similar outcomes in such courses. However, because CHEM0330 and the PHYS0050/0060 pair serve different purposes within their respective departments and serve different populations of students at Brown, but yield similar results in terms of the benefits of CLEs, we hypothesize that implementing CLEs in other courses and disciplines would be more likely than not to similarly benefit students.

Conclusions

In our sample, the more often students participated in CLEs, the more their performance improved compared to their pre-course preparation. Further, participation in CLEs was found to disproportionately benefit the performance of female students and students from historically under-represented groups (HUGs). Based on these outcomes, this study suggests that implementing evidence-based practices such as collaborative learning in introductory STEM courses can help to close the achievement gap for females and HUGs.

To extend the findings of this study, we plan to 1) incorporate data from the 2015-2016 academic year (once available) to form a larger, more robust sample and 2) follow students' trajectories through Brown into the 2016-2017 and 2017-2018 academic years to see if experiencing evidence-based practices in introductory STEM courses makes students more likely to persist in STEM. We are also interested in students' attitudes about science, whether they complete a STEM degree or leave STEM altogether. How does completing a STEM course infused with evidence-based practices affect students' attitudes (interest, appreciation, etc.) about science, particularly if the course will be one of the only STEM experiences of their academic career? We hope to answer this and other questions with our future work.

To support instructors' interested in designing similar CLEs in their own STEM courses, some of our future work includes disseminating implementation guides about the full variety of CLEs that have been incorporated into introductory STEM courses at Brown. These guides will include detailed descriptions of the pedagogy and course structure, comments on the rationale for the design by both the reformers and the instructors involved, and a clear statement of the resources necessary for each implementation. We hope that these guides will provide a menu of possibilities, rather than a single "right" way, that instructors can use to find CLEs and evidence-based practices that fit their personal teaching style and the resources available to them at their institution.

Acknowledgements

Funding for this work was provided by an AAU Undergraduate STEM Education Initiative grant. We would like to thank all of the instructors of CHEM0330, PHYS0050, and PHYS0060 who committed their time and effort in implementing evidence-based practices in their classrooms and allowed us to evaluate their courses. We would also like to thank all the students who participated in the voluntary CLEs and took the pre-tests; we know how precious your time is in college, and we appreciate the portion of that time you gave to this work.

References

1. Heather B. Gonzalez and Jeffrey J. Kuenzi, "*Science, technology, engineering, and mathematics (STEM) education: A primer*," Washington, DC: Congressional Research Service, no. R42642 (August 2012).
2. National Research Council (NRC), "*Evaluating and improving undergraduate teaching in science, technology, engineering, and mathematics*," Washington, DC: National Academies Press, (2003).
3. National Research Council (NRC), "*Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*," Washington, DC: National Academies Press, (2012).
4. National Research Council (NRC), "*Reaching stutents: What research says about effective instruction in undergraduate science and engineering*," Washington, DC: National Academies Press, (2015).
5. President's Council of Advisors on Science and Technology, "*Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and*

mathematics," Washington, DC: Executive Office of the President, (2012).
6. National Research Council (NRC), "*Expanding underrepresented minority participation: America's science and technology talent at the crossroads*," Washington, DC: National Academies Press, (2011).
7. National Research Council (NRC), "*Colloquy on minority males in science, technology, engineering, and mathematics*," Washington, DC: National Academies Press, (2012).
8. National Research Council (NRC), "*Surmounting the barriers: Ethnic diversity in engineering education: Summary of a workshop*," Washington, DC: National Academies Press, (2014).
9. National Research Council (NRC), "*Infusing real world experiences into engineering education*," Washington, DC: National Academies Press, (2012).
10. National Research Council (NRC), "*Educating engineers: Preparing 21st century leaders in the context of new modes of learning: Summary of a forum*," Washington, DC: National Academies Press, (2013).
11. National Research Council (NRC), "*Undergraduate chemistry education: A workshop summary*," Washington, DC: National Academies Press, (2014).
12. National Research Council (NRC), "*Adapting to a changing world—Challenges and opportunities in undergraduate physics education*," Washington, DC: National Academies Press, (2013).
13. National Research Council (NRC), "*The Mathematical Sciences in 2025*," Washington, DC: National Academies Press, (2013).
14. Rebecca Carr, "*Women in the academic pipeline for science, technology, engineering, and math: Nationally & at Association of American Universities Data Exchange (AAUDE) instituitions*," Nashville, TN: AAUDE, (2013).
15. National Center for Science and Engineering Statistics, "*Women, minorities, and persons with disabilities in science and engineering: 2015*," Arlington, VA: National Science Foundation, no. nsf15311 (2015).
16. Randi Shedlosky-Shoemaker and Jessica M. Fautch, "Who leaves, who stays? Psychological predictors of undergraduate chemistry students' persistence," *Journal of Chemical Education*, 92, no. 3 (January 2015): 408-414.
17. Brian A. Couch, Tanya L. Brown, Tyler J. Schelpat, Mark J. Graham, and Jennifer K. Knight, "Scientific Teaching: Defining a taxonomy of observable practices," *CBE-Life Sciences Education*, 14 (March 2015) 1-12.
18. Scott E. Lewis, "Investigating the longitudinal impact of a successful reform in general chemistry on student enrollment and academic performance," *Journal of Chemical Education*, 91, no. 12 (September 2014): 2037-2044.
19. Barbara Oakley, Richard M. Felder, Rebecca Brent, and Imad Elhajj, "Turning student groups into effective teams," *Journal of Student-Centered Learning*, 2, no. 1 (2004): 9-34.
20. Douglas R. Mulford and William R. Robinson, "An inventory for alternate conceptions among first-semester general chemistry students," *Journal of Chemical Education*, 79, no. 6 (June, 2002): 739-744.
21. Doreen Geller Leopold and Barbara Edgar, "Degree of mathematics fluency and success in second-semester introductory chemistry," *Journal of Chemical Education*, 85, no. 5 (May 2008): 724-731.
22. Ibrahim Abou Halloun and David Hestenes, "Common sense concepts about motion," *American Journal of Physics*, 53, no. 11 (November 1985): 1043-1055.
23. Lin Ding, Ruth Chabay, Bruce Sherwood, and Robert Beichner, "Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment," *Physics Education Research*, 2 (March 2006): 1-7.
24. Jacob Cohen, "A power primer," *Psychological Bulletin*, 112, no. 1 (July 1992): 155-159.
25. Donna M. Hall, Amanda J. Curtin-Soydan, and Dorian A. Canelas, "The Science

Advancement through Group Engagement Program: Leveling the Playing Field and Increasing Retention in Science," *Journal of Chemical Education*, 91 (2014): 37-47.

26. Akira Miyake, Lauren E. Kost-Smith, Noah D. Finkelstein, Steven J. Pollock, Geoffrey L. Cohen, and Tiffany A. Ito, "Reducing the Gender Acheivement Gap in College Science: A Classroom Study of Values Affirmation," *Science*, 330 (November 2010): 1234-1237.
27. Valerie C. Lundy-Wagner, "Is It Really a Man's World? Black Men in Science, Technology, Engineering, and Mathematics at Historically Black Colleges and Universities," *The Journal of Negro Education*, 82, no. 2 (2013): 157-168.