A *Vision and Change* Reform of Introductory Biology Shifts Faculty Perceptions and Use of Active Learning

Anna Jo Auerbach[†][*] and Elisabeth Schussler[‡]

[†]Department of Genetics, University of Georgia, Athens, GA 30602; [‡]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996

ABSTRACT

Increasing faculty use of active-learning (AL) pedagogies in college classrooms is a persistent challenge in biology education. A large research-intensive university implemented changes to its biology majors' two-course introductory sequence as outlined by the *Vision and Change in Undergraduate Biology Education* final report. One goal of the curricular reform was to integrate core biological concepts and competencies into the courses using AL pedagogical approaches. The purpose of this study was to observe the instructional practices used by faculty ($N = 10$) throughout the 3-year process of reform to determine whether the use of AL strategies (including student collaboration) increased, given that it can maximize student learning gains. Instructors participated in yearly interviews to track any change in their perceptions of AL instruction. Instructors increased their average use of AL by 12% (group AL by 8%) of total class time throughout the 3-year study. Interviews revealed that instructors shifted their definitions of AL and talked more about how to assess student learning over the 3 years of the project. Collaboration, feedback, and time may have been important factors in the reform, suggesting that small shifts over time can accumulate into real change in the classroom.

INTRODUCTION

Traditional lecture-style teaching of undergraduate science courses does not enhance student learning to the same degree as student-centered active learning (AL) (Freeman *et al.*, 2014). While lecture is an efficient way to disseminate a large amount of information to a large number of students, this teacher-centered strategy promotes passive student learning (National Research Council, 2000) while stifling student motivation and enthusiasm (Weimer, 2002). As evidence of this, many students who leave science majors say it was because they felt uninspired by their science classes and perceived the instructors as not caring (Seymour and Hewitt, 1997). Roughly 60% of science, technology, engineering, and mathematics (STEM) undergraduate students, including 58% of life sciences students, switch their majors or drop out from the university; this percent for life sciences students increases to 80% for women and students from minority groups (President's Council of Advisors on Science and Technology, 2012).

AL is a student-centered pedagogical approach that engages student thinking through the use of class activities that require students to reflect upon and often explicitly discuss their ideas and their application (Collins and O'Brien, 2003; Smith *et al.*, 2009). Studies of introductory biology courses have found that the use of AL improves student performance, including that of underrepresented students (Michael, 2006; Freeman *et al.,* 2007; Armbruster *et al.*, 2009). Increasing the use of student-centered practices is a main recommendation of the *Vision and Change in Undergraduate Biology Education* final report (American Association for the Advancement of Science [AAAS], 2011). Freeman *et al.* (2014) suggested that it is time to stop researching the benefits of AL, which have clearly been shown, and instead begin implementing it.

Various instructional methods can be used to incorporate AL in the classroom, with common practices including clicker questions, verbal questions, and written activities (Couch *et al.*, 2015). Many instructors use student response systems or "clickers" that allow large classrooms of students to select the answer for a projected problem or question using a handheld personal response device. Instructors also ask questions out loud, called verbal questions, to produce responses either through student volunteers or calling on students to respond. Even when a single student provides an answer, verbal questions engage students who do not answer the question (Obenland *et al.*, 2012). Writing is another form of eliciting responses from students. Instructors can use classroom activities such as figure interpretations, model drawings, and phylogenetic trees and ask students to respond in writing on notebook paper, note cards, or preprinted worksheets. AL in the classroom can take any form so long as the students are explicitly engaged in answering questions, solving problems, and discussing solutions and reasoning about the material with their peers while receiving regular feedback from the teacher (Wieman, 2014).

AL techniques can be used to engage students individually or in a collaborative context. While both are successful strategies for increasing student learning, combining peer discussion with instructor explanation of the answer maximizes student learning gains resulting from the use of clickers in biology classes (Smith *et al.*, 2009, 2011). As such, providing opportunities for student discussion is seen to be critical for a student-centered course (AAAS, 2011). Couch *et al.* (2015) suggest that peer interaction is one component of best practices for teaching science, specifically allowing students to refine their knowledge through activities such as worksheets that require small-group discussion. Eddy *et al.* (2015) found that students explaining their answers to peers, hearing other students describe their logic, and participating in activities that require group work are all elements of best practices in AL. Research in physics has suggested that, although classroom discussion is important for student understanding (Turpen and Finkelstein, 2010), how discussion is employed in the classroom varies by instructor, with some professors promoting much higher levels of student discussion than others (Turpen and Finkelstein, 2009).

Increasing faculty use of AL pedagogies in college classrooms is an ongoing challenge for biology education, despite some instructors' willingness to seek professional development opportunities. Changing instructor beliefs about teaching is seen by many as a necessary prerequisite to changing practice (Pajares, 1992), but the interventions most effective in changing faculty beliefs are still being investigated. It is known that a one-time workshop is not enough to invoke meaningful change in faculty use of evidence-based practices (Sunal *et al.*, 2001). Even after more extensive training and familiarity with best practices literature, implementation of AL pedagogies can be nonexistent or disappointingly low (Kember, 1997; Gess-Newsome *et al.*, 2003; Dancy and Henderson, 2010; Ebert-May *et al.*, 2011). Thus, although a change in beliefs is believed to be necessary to change practice, a change in beliefs does not guarantee changes in practice. This disconnect has been probed in college faculty, with a lack of time to prepare new teaching techniques being reported as the biggest barrier to the use of AL practices (Dancy and Henderson, 2010).

Gormally *et al.* (2014) highlight the importance of supportive instructional feedback as faculty attempt to implement AL strategies in their classrooms, and studies suggest the need for faculty to form communities of practice to work together on change (Lave and Wenger, 1991). Faculty are more likely to put effort into pedagogical reform if they understand the goal, are committed to it, and believe in its success (Kluger and DeNisi, 1996). Given this, a process that incorporates the collaborative development of instructional goals among the faculty implementing the change and providing instructional feedback during implementation may be beneficial to enacting pedagogical change. In this paper, we describe research that followed faculty involved in a curricular reform community of practice to see how their teaching practices and perceptions of their teaching practices changed over the 3 years of a programmatic curricular reform. For this study, we did not focus on causal mechanisms of change, but rather documented the changes that occurred and considered conditions that may have made a difference for this group of instructors.

The study took place over a period of 3 consecutive years as the instructors of an introductory biology sequence at a southeastern public research university implemented curricular changes guided by the *Vision and Change* final report (AAAS, 2011). The research built on previous research examining AL use in introductory biology at the same institution (Auerbach and Schussler, 2016). One of the goals of the curricular reform was to encourage faculty to use AL as a means of engaging students with the new curriculum. The purpose of this study was to track faculty perceptions and use of AL instruction in the introductory biology majors' courses throughout the 3-year curricular reform process using classroom observations and interviews. These combined data (quantitative and qualitative) answered the following research questions:

1. How is the amount of time faculty devote to AL practices changing over the curricular reform?
2. How are faculty perceptions of their teaching practices changing over the curricular reform?
3. Are changes in faculty perceptions of their teaching practices related to changes in the amount of time they devote to AL practices?

METHODS

Participants

The potential participants for this study included instructors ($N = 15$) who taught one of the two introductory majors' biology courses (Biodiversity and Cell Biology) between Fall 2012 and Spring 2015. Instructors were not included in the data analysis if they did not teach at least twice over the 3 years of the curricular reform period, because the intent of the study was to track instructional change over time. This reduced the number of instructor participants to 10; five instructors taught the Cell Biology course and five taught the Biodiversity course (Table 1). All 10 instructors taught at least once a year for 2 years, with four instructors teaching all 3 years. All of the instructors held a PhD, half of the instructors were tenure-track faculty, and the other half were full-time permanent lecturers. Most of the instructors ($N = 7$) had taught the introductory course for less than 5 years, two had 5–10 years of experience with the course, and one had more than 10 years of experience with the course;

TABLE 1. Summary of instructor pseudonyms, course taught, experience with course, and the years they were observed and interviewed for this study

Pseudonym	Course	Teaching experience in course	Years observed/ interviewed
Albert	Biodiversity	<5 years	Year 1
			Year 3
Bruce	Biodiversity	10+ years	Year 1
			Year 2
			Year 3
Celine	Biodiversity	<5 years	Year 1
			Year 2
			Year 3
David	Biodiversity	<5 years	Year 1
			Year 2
Erin	Biodiversity	<5 years	Year 1
			Year 2
			Year 3
Frank	Cell Biology	5–10 years	Year 1
			Year 3
Gail	Cell Biology	<5 years	Year 1
			Year 2
Harold	Cell Biology	<5 years	Year 1
			Year 2
			Year 3
Ingrid	Cell Biology	<5 years	Year 2
			Year 3
Juanita	Cell Biology	5–10 years	Year 2
			Year 3

only one instructor who participated in this study was new to either course at the beginning of the study (Table 1). We received informed consent from all participating instructors for both observations and interviews before collecting data. All data-collection procedures were reviewed and approved by the University of Tennessee Review Board for Human Subjects.

Courses and Process of Change

The Biodiversity and Cell Biology courses were offered in multiple lecture class sections in both Fall and Spring semesters. There were eight total lecture sections observed each year, with four representing Cell Biology and four representing Biodiversity. Each lecture section contained 150–225 students, depending on the time it was taught. There were honors lecture sections for each course that were not included in this study, and the Summer offerings of each course were also not included in the study. Biodiversity is the first course of the sequence and covers topics such as ecology, evolution, and the diversity of life. The Cell Biology course covers topics such as macromolecules, cell structure, metabolism, and DNA structure and function.

The curricular changes undertaken were guided by the *Vision and Change in Undergraduate Biology Education* (AAAS, 2011) final report. Select faculty members across the three biology departments served by the introductory courses formed a task force in 2010 to consider ways to improve the courses. After the publication of the *Vision and Change* final report (AAAS, 2011), the faculty decided to adopt the concepts and competencies from the report as the new unified learning objectives for the two courses. A decision was also made to remove

the labs appended to each course and instead add a weekly 50-minute small-group discussion led by teaching assistants focused on primary literature and biological literacy in an AL context. This resulted in a proposal to switch from a traditional two-course, 8-credit introductory sequence to a three-course, 8-credit sequence with two lecture/discussion courses (3 credits each) and an independent 2-credit lab/discussion starting in Fall 2014. This research project focused only on the lecture portions of the 4-credit courses that existed through Spring 2014, and then the first delivery of the two newly created 3-credit lecture/discussion courses in the 2014–2015 academic year. The lecture content in the courses remained the same, but the structural changes to the new courses in 2014–2015 resulted in 50 minutes less lecture time per week than in the original courses (150 minutes per week in the old courses vs. 100 minutes per week in the new courses).

The curricular changes were phased in over 3 years of reform to ease the transition to the new course structure. During the first year (2012–2013), four faculty members piloted the *Vision and Change* concepts and competencies within their lecture courses. During the second year (2013–2014), all faculty used the *Vision and Change* concepts and competencies in an AL context in their courses, but the course structure remained as a 4-credit lecture/lab format. During the third year (2014–2015), the new course structure with 3-credit lecture/discussions and 2-credit lab/discussion were implemented.

Throughout the curricular reform process, faculty met as a group to discuss the reform, participate in professional development sessions, and share resources. Instructors worked together and jointly developed the implementation guidelines for the revised lecture courses. On average, these meetings occurred once a month and lasted at least 1 hour each. These meetings included faculty review of primary literature on topics such as backward design and student-centered learning. Some meetings were spent learning strategies such as how to create a voice-over PowerPoint or how to use learning catalytics, an interactive student response tool that allows for more than just the multiple-choice questions used with clickers. Faculty also discussed course data (such as instructor AL use collected for this study) and AL pedagogies throughout the process. The meeting topics were decided by the group based on what they felt they needed for implementation that semester, or sometimes even on a meeting-by-meeting basis.

There were no course or sequence requirements for how AL or the new curricula were to be implemented, beyond what was decided upon by the group. Course learning objectives, major topics, total course points, and textbook materials were the same; however, topic sequence, specific activities and assignments, and book readings, for example, were allowed to vary by instructor. Faculty typically designed their own PowerPoints, activities, and exams for their courses. While faculty were encouraged to use AL, they were not given specific directives about the type or amount that had to be used. The only program message was to try to increase the use of AL (of whatever type they felt comfortable using) and to try to implement it with student peer discussion as much as possible.

At the end of each semester of observations, the instructors had access to their own AL data and the compiled data for the program. All data were shared anonymously. Some of the instructors also participated in communities of faculty,

postdoctoral associates, graduate students, and undergraduates who worked together to design the discussion curricula; however, the meetings with instructors about changes to the lecture courses were separate from the meetings about the new discussion curricula.

Data Collection

Classroom observations were used to sample instructional practices within the introductory courses over the 3 years of the reform process. Observations were undertaken using an observation protocol created via an iterative process of classroom observation and classroom event categorization by the first author (A.J.A.) in early Fall 2012 (Auerbach and Schussler, 2016). Existing observation protocols were also tested, but at the time data collection began, no other protocol existed to capture the data considered central to this study (the Classroom Observation Protocol for Undergraduate STEM [COPUS; Smith *et al.*, 2013] had not yet been published). After the publication of COPUS, the authors discussed switching to that instrument, but the switch would not have resulted in improved data for the study, so the decision was made to keep the data-collection process the same.

To collect data, freehand notes were used to capture the classroom events relevant to this study; they were used to log the duration of time spent on each event and, in the case of AL events, whether they were conducted as an individual or group activity. The standard that we used to identify AL in the faculty classrooms was any question by the faculty member to the students in which an electronic, verbal, or written response was expected (even a response by a single student to a verbal question), thus inferring that thinking was expected. All classroom observations for this study were done in person, not video recorded. To confirm the reliability of the created observation protocol, three external evaluators were trained in the use of the protocol and observed at least two of the same classes as the first author (A.J.A.) each semester. Observers classified all classroom events into the same categories, and the duration of the events in all observations matched exactly or within 1 minute of each other.

The first author (A.J.A.) conducted observations of each consenting instructor once per month over the full semester of instruction for 3 academic years. The observation lasted for the entire length of the chosen class, which was 50 or 75 minutes, depending on the lecture day (50 minutes for courses taught on Mondays, Wednesdays, and Fridays vs. 75 minutes for courses meeting on Tuesdays and Thursdays). The observations were unannounced, and the observer typically sat in the back of the class to be as unobtrusive as possible.

There were seven categories of classroom events that were recorded throughout the study: clicker questions, instructor verbal questions, activities, student questions, lecture, video, and no class. These categories were classified as either AL or non-AL. Classroom events were classified as AL if the instructor asked the students to engage with the material in some manner (i.e., answer a clicker question, discuss a concept, draw a phylogenetic tree). AL was also broken into subcategories of individual (students work alone) and group (students work together) AL. The classroom events of clickers, verbal questions, and activities were considered to be instructor-led AL practices and were further subcategorized as individual AL or group AL. Any clicker or verbal questions that were not content based (e.g., feedback about a topic or asking about class logistics) were not counted in the AL total. Student questions were placed into a separate category and not considered AL, because in these cases, the instructors were not explicitly asking students to be engaged in their learning, and this study was focused solely on instructor-directed AL. Lecture, video, and no class (class beginning late or dismissed early) were classified as non-AL events because they were not explicitly asking students to engage in thinking.

Five of the AL categories were created before testing of the protocol, based on previous classroom observations. The activities and video categories emerged as new data-collection categories during class observations and resulted in revision of the observation protocol. Activities included the use of paper, such as students being asked to analyze data and turn the work in on paper, or the use of worksheets to guide discussion and responses. Activity topics included students creating hypotheses, predicting future phenotypes, creating population growth curves, drawing a food web, graphing data, and creating phylogenetic trees. In summary, the only classroom events that counted toward AL in a class were individual or group conceptual clicker questions, individual or group instructor conceptual (verbal) questions, and individual or group class activities.

The observer recorded the length and frequency of all classroom events for the entire scheduled class period. For questions such as clicker questions and verbal questions, the total amount of time recorded reflected the time it took to pose the question, the time it took students to generate an answer, any redirecting on the instructor's part, and additional student answers. At no point was a judgment ever recorded about the quality of the classroom event, for example, whether certain clicker questions were more effective at fostering conceptual learning or whether particular activities seemed to provoke more student discussion. The data therefore reflect only the time that an instructor used AL and not the relative quality of the AL.

Each instructor was interviewed at the end of the semester he or she taught. Research has shown that teaching beliefs can influence teaching practices, though changes in beliefs do not do not necessarily translate into changes in teaching practice (Kember, 1997; Gess-Newsome *et al.*, 2003). However, the potential impacts of instructor perceptions of their teaching on teaching practices mean it is important to follow teacher thinking throughout a process of attempting to change teaching practices. The same questions (see the Supplemental Material) were asked at each interview to identify changes in perception and use of AL over time. The interview questions were designed based on a priori categories (course planning, AL implementation, AL definition, AL types, and change in teaching) to track changes over time in the way faculty thought about, planned for, and implemented AL within their classrooms. For example, faculty were asked what AL means to them and what types of AL they use in their classroom. Follow-up questions were often used to clarify or extend participant responses. The interviews were either held in the participant's or first author's (A.J.A.) office, depending on the participant's preference. The first author performed all interviews. The interviews were audio recorded with permission; they were begun only after participants received all information about the study and signed an informed consent form. The interviews ranged from 18 to 36 minutes each. Instructors were interviewed twice if they taught

2 of the 3 years and three times if they taught every year, resulting in a total of 24 interviews with the 10 participants over the course of the project.

Data Analysis

Observations. While 15 instructors were observed, only the 10 core instructors were included in data analyses. During the first year of observations, the number of observations per instructor varied from three to five. For each instructor to have equal representation and for each year to be compared over time, it was determined that each instructor would have three observations in the Fall (one each in September, October, and November) and three in the Spring (one each in February, March, and April). Thus, for the first year of data collection, one observation was selected each month for each semester an instructor taught to standardize the data analysis (Table 1). Eight of the 10 instructors taught each year; thus, a total of 24 observations were analyzed each year for a total of 72 observations over the course of reform.

Observational data were used to calculate the proportion of class time each instructor spent on each classroom event per observation and was recorded into a spreadsheet for analysis. The percentage of time spent on each classroom event was calculated as a function of the entire class period time (50 or 75 minutes) and then averaged for the three observations for each instructor. These values ($N = 24$ per year) were then averaged across instructors to determine the overall percentage of class time spent on each classroom event per year for the program. The values that represented the three AL events (clickers, verbal questions, and activities) were also added together for each instructor to produce the total proportion of class time spent in AL overall (ALO). These values were also broken into individual and group AL events based on the implementation (i.e., students answer a clicker individually vs. after discussing the question with classmates) to produce a measure of group AL overall (GALO) for each instructor. These values were then averaged across instructors to determine the overall percent of class time spent on each classroom event (clickers, verbal questions, and activities), ALO, and GALO over each of the 3 academic years. The data were assessed for coding errors and missing data, then standardized scores were created to check for outliers. A Shapiro-Wilk test was used to assess the data for nonnormality, as the sample size was less than 300 (Tabachnick and Fidell, 2001). Appropriate skewness and kurtosis were also confirmed (Westfall and Henning, 2013).

A one-way repeated-measures analysis of variance (ANOVA) using the observation data was performed with the total amount of AL (ALO) as the dependent variable and year as the independent variable to determine whether instructor use of AL was changing across the reform. Separate one-way repeated-measures ANOVAs were performed to determine whether changes occurred across time in use of group AL (GALO) or classroom event type used (i.e., clickers, verbal questions, and activities). For analyses that produced significant results, post hoc comparisons using the Bonferroni correction were performed to determine which years were different, and effect sizes were calculated in SPSS using Cohen's *d* (IBM Corporation, 2011; Middlemis Maher, 2013).

Interviews. Of the 10 instructors included in the study overall, only eight of the 10 taught each academic year of the project.

Because the intention was to analyze interview responses for potential changes on a yearly basis, the interview analyses were carried out on only the eight instructors who taught each year, rather than all 10 instructors each year. The 24 interviews were transcribed fully into a word-processing program for analysis. The interviews were then analyzed in two successive cycles using two independent researchers, one being the primary analyst and the other independently confirming the themes identified by the first.

In the first round of analysis, interviews were analyzed using the a priori categories that were developed and reflected in the interview questions. These categories were used as a guide to examine the responses and identify themes within each category in order to search for change across time (Kvale and Brinkmann, 2009). For this analysis, all responses relating to each category were highlighted. Responses for each category were read and notes were made about potential themes. The responses were then reread, and initial sorting into themes was done to determine whether they reflected the data. The themes were finalized when it was apparent that they represented the majority of the responses (Saldaña, 2013). These themes were then used to create a summary of each category by tallying the number of instructors whose responses could be placed in each theme. For example, under the category of course planning, instructors reported two themes of behavior: using previous course materials as a guide for course planning or developing learning objectives for the course to serve as the guiding structure.

After this first round by the primary analyst, a co-researcher (E.S.) in a subsequent round also assessed the interviews. The co-researcher identified participant responses using the a priori categories and themes created by the primary researcher. The co-researcher also made notes of possible a posteriori categories and themes that were not planned for or expected but emerged from the data nonetheless. Once the co-researcher sorted the participant responses under each category by the themes, the two researchers met to compare results and reach consensus. At this time, some themes were retitled or compacted, but there were no significant disagreements about the presence of the created themes. However, there were two categories, in addition to the a priori categories, that the researchers felt warranted further exploration (assessment and professional development). Each researcher independently searched the interviews for participant responses that related to the two new categories. Then they each assigned the responses to a theme within each a posteriori category. Finally, they met again to compare results and reach consensus. Assessment was added as a category, but it was decided that there was not enough about professional development to warrant a category addition. Summaries for each category and theme were then compared across the years using comparative analysis (Saldaña, 2013).

Aligning Perceptions with Practices. Two analyses were conducted to link the data about instructor practice to the data about instructor teaching perceptions. The first was to compare faculty estimates of time spent in ALO from their interviews with their actual classroom use of ALO, to see whether these two variables increased in alignment over time. An increase in alignment between these two variables would suggest that data sharing and feedback about their actual practice were being used by

the faculty member to inform their perceptions of their practice. To do this, we performed a regression analysis using the average time spent in ALO reported by each instructor in his or her interview to predict the actual amount of time spent in the individual classrooms on AL each year. The second analysis was to align each instructor's classroom use of ALO and GALO to shifts in instructional perceptions articulated in his or her interview, to see whether shifts in these perceptions could be related to shifts in practice. To do this, we compared the number and type of shifts that occurred in themes within a category for each instructor each year with the instructor's use of ALO and GALO each year. We use quotations from the transcripts and pseudonyms to refer to participants when presenting the data.

RESULTS

Observations

Over the 3 years of reform, there was an overall significant effect of ALO use when comparing lecture observations over time ($p = 0.026$; Figure 1). Instructors averaged anywhere from 26% to 39% of class time in ALO across the 3 years of reform (Table 2). The results of a Bonferroni post hoc test showed no significant results for paired-year comparisons; however, the use of ALO from the first to third year of implementation showed a large difference in effect size ($d = 0.88$; Table 3). ALO did not decrease from year 2 to year 3, despite a loss of one-third of instructional time. There were no significant differences when comparing the use of GALO or any of the types of AL events (i.e., clickers, verbal questions, and activities) in class
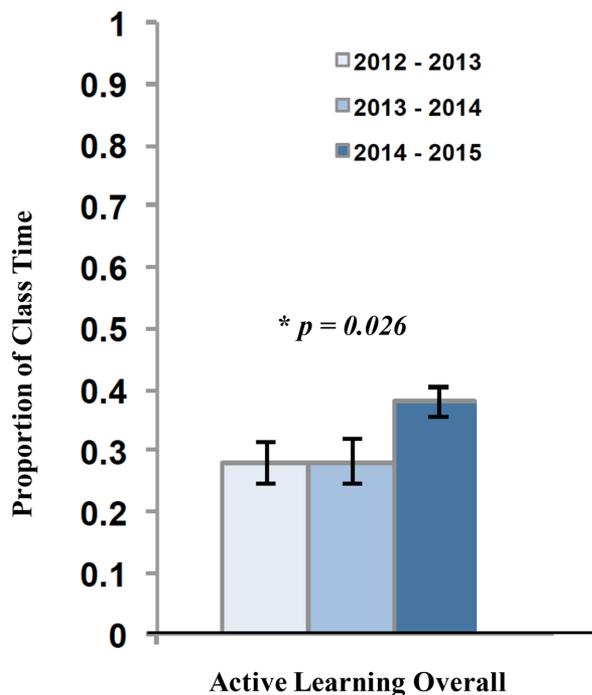
FIGURE 1. Proportion of class time spent in active learning (ALO) in programmatic introductory biology lecture classes across 3 years of reform (2012–2013, 2013–2014, 2014–2015) showed an overall significant effect, $F(2, 46) = 3.95$, $p = 0.026$, partial $\eta^2 = 0.147$, power = 0.682.

TABLE 2. Average proportion of total class time spent in different types of AL in introductory biology courses over 3 academic years[a]

	2012–2013	2013–2014	2014–2015
Active learning (ALO)	0.26	0.27	0.39*
Group active learning (GALO)	0.13	0.15	0.21
Clicker questions	0.13	0.11	0.17
Verbal questions	0.08	0.07	0.10
Activities	0.05	0.09	0.12

[a]The average proportion of class time is reported as the mean of all instructors ($N = 10$) for each academic year. ALO was the total of all types of AL, while GALO was only when students were allowed to discuss clicker questions, verbal questions, or work together on activities. Asterisk (*) indicates a significant increase of AL type use by year, $F(2, 46) = 3.95$, $p = 0.026$, partial $\eta^2 = 0.147$, power = 0.682.

across the 3 years. GALO typically comprised 13–21% of the total class time, and clickers were always the most-used average AL practice among the instructors (Table 2).

Interviews

For the a priori categories probed by the interview questions (course planning, AL implementation, AL definition, AL type, and change in teaching) the identified themes were tallied over time to identify changes in instructor perceptions and use of AL (Table 4). Themes within each category are italicized in the descriptions below (see the Supplemental Material for additional supporting quotes for themes).

Course Planning. Within this category, two themes were identified: "*instructor used previous course materials as a guide*" and "*instructor used learning objectives as a guide*." For instance, some instructors reported using old PowerPoint presentations or the textbook as a starting point for their course planning. Other instructors talked about learning objectives for the course or course content and how these were used to shape the course planning. As one instructor said,

> I start with specific learning objectives. What do I want them to do? How can I achieve that? [instructor used learning objectives as a guide]

The number of faculty using previous course materials as a guide for course planning decreased over time. In the first year, two instructors were using learning objectives to plan their course (Table 4). In the second and third years, those same two instructors continued to use learning objectives and two additional instructors also started using learning objectives. Thus, while use of course materials decreased, the use of learning objectives to guide course planning increased within the group of instructors.

TABLE 3. Effect size comparisons by year[a]

Year	1	2	3
1	—	0.06	0.81
2	0.06	—	0.77
3	0.81	0.77	—

[a]Effect size (Cohen's *d*) difference of ALO (AL overall) reported by comparing years. A measure of 0.2 indicates a small effect size; 0.5 is medium; and 0.8 is large.

TABLE 4. Summary of interview categories and themes[a]

Category	Theme	Year 1	Year 2	Year 3
Course planning	Instructor used available course materials as guide	5	5	4
	Instructor created learning objectives	2	3	4
AL implementation	AL planned after lecture is set	7	7	5
	Lecture planned around AL	1	0	2
	AL planned simultaneously with lecture content	0	1	1
AL definition	Students engaged, involved, thinking about content	7	8	7
	Students interacting with each other	0	4	5
	Students as knowledge constructors	1	1	3
AL type	Instructor reported using group AL	6	8	8
	Instructor did not report using group AL	2	0	0
	Instructor reported using clickers	8	8	7
	Instructor reported using verbal questions	6	7	7
	Instructor reported using activities	5	6	8
Change in teaching	Change in overall approach	2	5	5
	Change in time pressure	2	3	1
	Change in pedagogical strategy	2	0	2
	No change	3	1	1
Assessment		1	3	6

[a]Frequency of themes for each interview category across 3 years of reform for instructors (*N* = 10). A maximum of eight instructors taught each academic year, so for each column *N* = 8. Themes within each category were identified by qualitative analysis of instructor interviews. All categories were created a priori with the exception of "Assessment," which was created a posteriori.

AL Implementation. There were three themes identified for when the planning for AL implementation occurred: "*planned after lecture is set*," "*planned simultaneously with lecture content*," "*planned lecture around AL*." Instructors described different ways they planned for AL in their classrooms. Some instructors only thought about AL after the lecture was set. Others talked about planning their AL in concert with the content and planning their AL learning first and then considering how to plan their lectures. For example,

> The one thing I think I centered more about the activities I had. I thought about a set of activities that I liked that had worked well in the past. And so I basically designed around this. [planned lecture around AL]

Throughout the reform process, the number of instructors who shifted their planning for AL implementation from "*planned after lecture is set*" decreased from seven instructors to five. One instructor shifted to "*plan simultaneously with lecture content*," and the other began to "*plan lecture around AL.*"

AL Definition. There were three themes within this category: "*students engaged, involved, thinking about content*," "*students interacting with each other*," and "*students as knowledge constructors*." Instructors could be classified into more than one theme in this category. If they talked about students being engaged with the content, they were counted under that theme. If instructors instead or additionally talked about student collaboration as a part of what AL means, it was counted as a second theme. The last theme represented instructors talking about AL as a process where students construct their own knowledge. For example,

> Practicing!! Giving them an opportunity to practice the things I want them to be able to do on the exam, which then are also the things I think are important to learn in the class so anything that gets them to think, consider, working together mainly just to think, to build their understanding. [students engaged, involved, thinking about content, students interacting with each other, students as knowledge constructors]

The way instructors defined or thought of AL also changed throughout the reform process (Table 4). In the first year, most instructors (7) described AL as "*students engaged, involved, thinking about content*." Only one instructor defined AL using the "*students as knowledge constructors*" theme. In the second year, all instructors used the theme of "*students engaged, involved, thinking about content*" in their definition of AL. Some instructors (4) also began to add "*students interacting with each other*." And by the third year, most instructors (7) defined AL as "*students engaged, involved, thinking about content*." Five instructors added "*students interacting with each other*" and three added "*students as knowledge constructors*" theme to their definitions.

AL Type. There were five themes within the category of AL type: "*instructor used group AL*," "*instructor did not report using group AL*," "*instructor used clickers*," "*instructor used verbal questions*," and "*instructor used activities*." The type of AL that instructors described implementing in their lectures changed across the 3 years (clickers, verbal questions, and activities), as did the manner in which they asked students to work on those AL events (individually or in groups). For example,

> Open-ended questions that lead to discussion, think-pair-share, clicker questions, handouts. [used group AL and used clickers and used verbal questions and used activities]

In the first year, most of the instructors reported that they (6) "*used group AL*" (Table 4). By the second year, all instructors reported using GAL, and that continued throughout the third year. All instructors (8) "*used clickers*" during the first year, and by the third year, one instructor had ceased using them. Most instructors (6) "*used verbal questions*" during the first year of reform. An additional instructor began using verbal questions in the second year and continued into the third year of reform. Some instructors (5) "*used activities*" during the first year of reform. In the second year, another instructor began using activities, and by the third year all faculty were using activities to implement AL into their classroom.

Change in Teaching. There were four themes under change in teaching, for which instructors reported whether the curricular reform process impacted their teaching in terms of their "*overall approach*," "*time pressure*," "*pedagogical strategies*," or that the curricular reform had "*no change*" on their teaching.

Instructors thought about how the reform had impacted their overall teaching approach, such as the way they organized their courses or creativity in delivering their courses. Instructors expressed concerns about how a reduction in time for lecture would change the way they taught. Instructors also reported that the curricular reform would not change their teaching and indicated that it was because their teaching was already in a process of reform. Instructors commented about the pedagogical choices they used in the classroom, such as trying to lecture less and be more purposeful with the practice they gave students in class. During the third year of reform, one instructor began thinking about how pedagogical choices he made can affect different groups of students in class. For example,

> I think you know if we talk about awhile ago I was not doing any of this, learning, active learning. I was not thinking too much about it. I think a lot more about what I would like, I'm more purposeful in the way I design my class. I'm thinking more about what do I want the students to know so this backward design, I'm doing this a lot more.

As the reform progressed, more faculty expressed that the curricular reform was resulting in change of their overall approach to teaching. The first and third years were when faculty reported the most changes to their use of pedagogical strategies. Although several faculty were concerned about upcoming pressures to losing class time in the second year of reform, by the third year (when faculty class hours were reduced) only one instructor was still expressing concern.

Assessment. Unlike the other categories identified, assessment was a category that was identified a posteriori and emerged from the data. After noticing instructors mentioning assessment in their interviews, all responses related to assessment were marked and reread. Assessment did not have any themes. When instructors mentioned assessment, it was always their reflections about what to test, how to test, and using assessment itself as an instructional strategy. For example,

> I'm still struggling with assessment in the course. I'm happier with my assessment this year. [My teaching assistant] helped

me to revise how I ask question so that they are scenario based and so now I think that the assessment is much more in line with the activities that we do in class.

During the first year of reform, only one instructor reflected on assessment when thinking about his or her teaching. During the second year, three instructors were thinking about assessment in their classrooms. And by the third year, six instructors were considering assessment when describing their AL practices.

Aligning Perceptions with Practices

For data collected during the first year of reform, a nonsignificant regression analysis revealed that there was no relationship between faculty estimates of time spent on AL and actual use of AL. A similar pattern was found for the second year of reform (see the Supplemental Material). However, a significant regression analysis using the third year of data showed faculty a significant relationship between their estimate of their AL use and their actual AL use, ($p = 0.001$; Figure 2).

For comparison of individual instructor changes in their teaching perceptions (from the interview themes) with changes in their AL practices (from observations of ALO and GALO used in classes), the results of the study were aligned over time for each instructor (Table 5). All but two of the instructors increased their average use of ALO and GALO when comparing their first and last course deliveries over the reform period, although in some cases there was some decrease in AL use from the first to second or second to third year (Bruce and Celine). In contrast, David decreased his use of ALO and GALO from year 1 to year 2, and did not teach in year 3. Interestingly, David also had a change in teaching perception that reverted to a practice less consistent with best practice over the reform period; he stopped using learning objectives to plan for AL in year 2. Erin's average use of GALO was slightly lower in year 3 than year 1, despite an increased use of GALO in year 2. However, these changes were
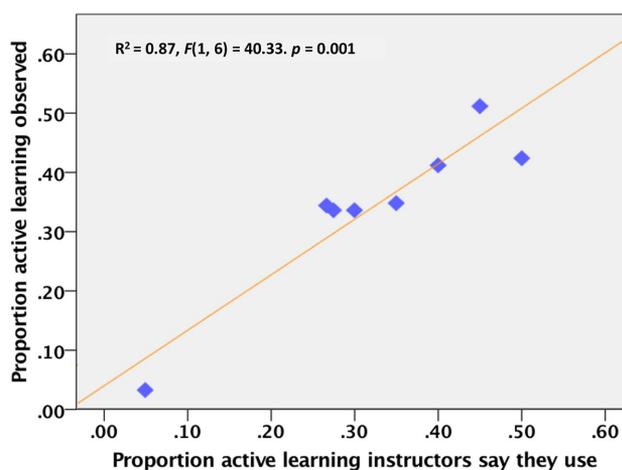
FIGURE 2. Proportion of total amount of active learning that instructors reported using in their classrooms (from the interviews) used to predict the actual proportion of time they spent using active learning in their classes (from the observations) for the third year of the reform. This year had a significant relationship between program faculty reported and actual use of active learning.

TABLE 5. Summary of ALO practices by instructor by year[a]

Instructor	Data	Year 1	Year 2	Year 3
Albert	ALO	0.36	—	0.38
	GALO	0.17	—	0.21
Bruce	ALO	0.30	0.30	0.53
	GALO	0.10	0.03	0.18
Celine	ALO	0.23	0.50	0.38
	GALO	0.06	0.44	0.14
David	ALO	0.36	0.08	—
	GALO	0.15	0.03	—
Erin	ALO	0.39	0.41	0.45
	GALO	0.36	0.38	0.35
Frank	ALO	0.35	—	0.44
	GALO	0.21	—	0.22
Gail	ALO	0.05	0.14	—
	GALO	0.00	0.07	—
Harold	ALO	0.06	0.13	0.13
	GALO	0.00	0.00	0.13
Ingrid	ALO	—	0.38	0.39
	GALO	—	0.23	0.28
Juanita	ALO	—	0.18	0.37
	GALO	—	0.05	0.18

[a]Means (M) of observed AL proportion (ALO) and group AL proportion (GALO) of total class time by individual instructor by year ($N = 10$). A dash mark in the ALO or GALO row indicates the instructor did not teach that year.

TABLE 6. Summary of beliefs and practices alignment[a]

Instructor	Year 2 shifts	Year 3 shifts
Albert		Course planning Teaching style
Bruce	Assessment AL implementation	Assessment AL implementation
Celine	Teaching style	Assessment Teaching style
David	AL definition Course planning (−)	
Erin	AL definition Assessment	Assessment AL implementation AL definition
Frank		AL definition Assessment AL type
Gail	AL definition AL type	
Harold	AL type	AL definition AL type
Ingrid		AL definition
Juanita		AL type Teaching style

[a]Shifts in interview themes by individual instructor by year ($N = 10$). The appearance of an interview category in a particular year indicates that the instructor shifted themes within that category toward a belief more consistent with AL, such as adding GAL, planning the lecture around AL, or adding to his or her definitions of AL. However, David's shift in course planning in year 2 was toward a belief less consistent with AL. This is highlighted with a (−) after that category.

not accompanied by any stated reversion in teaching practices consistent with AL.

For the instructors who showed positive changes in use of ALO or GALO over time, course planning was not a common theme aligned with changes in practice. Only Albert shifted his perception of course planning over time (Table 6). Similarly, how instructors planned for using AL (AL implementation) shifted for only two of the instructors (Bruce and Erin) and was also accompanied by other shifts in perceptions of their practice. In contrast, half of the instructors had shifts in the category of "AL definition" toward themes more consistent with AL (students interacting and constructing knowledge) and also were seen to increase their use of ALO and GALO over the course of the reform. Similarly, four of the instructors started discussing the category of "assessment" in their interviews and also experienced increases in their use of GALO and ALO. Three and four instructors, respectively, also self-reported the reform changing their teaching style and self-reported changing their use of AL types.

DISCUSSION

Instructor beliefs about teaching and learning are hypothesized to shape the practices they actually implement in the classroom (Pajares, 1992). However, what instructors say about their teaching may not match with what actually occurs (Ebert-May *et al.*, 2011), revealing a disconnect between the perception and reality of instruction. This study provided evidence that some changes in instructor's perceptions of their teaching practices occurred at the same time as increases in use of AL when instructors were engaged in a curricular reform, particularly in how they defined AL and increased their thinking about

assessment. Although this study was not designed to investigate the underlying causality of these changes over time, we note that these observed changes happened as instructors engaged in a community of practice were given opportunities to 1) reflect on how AL is defined and how assessment informs teaching practice, 2) participate in reflection facilitated by reviewing data on teaching practices within a collaborative learning community, and 3) be given sufficient time (3 years) to integrate these new ideas. These features align with previous work suggesting that communities of practice are important for faculty pedagogical change (Lave and Wenger, 1991).

Instructor ALO significantly increased across the 3 years of curricular reform, with the most change occurring in year 3, when the faculty spent an average of 39% of class time in AL exercises. Previous observational research of faculty members who teach large-enrollment STEM courses has shown instructors to range in their use of AL strategies, with some instructors devoting as little as 7% of class time to AL (Smith *et al.*, 2014). Our research suggests that reforming instruction in a meaningful way can be accomplished by making small programmatic changes that accumulate over time to create the change, thus reducing barriers such as time for planning (Dancy and Henderson, 2010). Indeed, the average use of ALO in the lecture classes increased by 12% from the first year to the third year of reform, despite the loss of one-third of the class time. This suggests that if you embed the changes in perceptions or practices into the program early, they may persist despite the addition of new barriers such as loss of class time. This may be due to comfort level with the changes after 3 years, or perhaps opportunities to

practice the use of AL. Future studies should investigate whether these may be factors in helping faculty adopt AL pedagogies.

Although faculty in this study were collectively able to increase their use of ALO and maintain this increase when losing instructional time, the first 2 years of this study showed extremely similar and sometimes identical numbers. The shifts in use of ALO started slowly and then gained momentum in the third year. The third year is also when faculty were more likely to correctly predict their use of AL. This may be due to several factors. First, the largest barrier to change is reported to be time (Dancy and Henderson, 2010). It is possible that the reduction in contact time for instructors the third year could have opened up additional planning time for them to integrate AL. The curricular changes in year 3 also explicitly assigned graduate teaching assistants as discussion leaders and lecture assistants to each course, which may have provided additional grading support for AL assignments and increases faculty reflection about assessment in year 3. Second, instructors need time to think about and process change. In the same way that students need time to process and construct new information, processing time may be an important factor in achieving pedagogical shifts in faculty instruction. Part of this may be because shifts in practices require shifts in how instructors define AL, and much like misconceptions, it may be hard to dislodge old teaching beliefs. Third, having a supportive learning community of instructors may be instrumental in persevering with change, as faculty may want to abandon efforts when change does not happen instantly (Henderson and Darcy, 2007). However, it is important to remember that time to allow for change and a supportive group may not be enough to shift instructional practices for all instructor groups. Ebert-May et al. (2011) evaluated a 3-year professional development program in which teams of faculty met three to six times during the program and found the majority of faculty to still use lecture-based, teacher-centered pedagogies after participating in the program despite the faculty reporting the use of student-centered approaches. It may be that the nature of this reform, with instructors teaching the same courses at the same university and meeting regularly, is an important part of how to support AL implementation.

One limitation of this study was that the quality of the AL that occurred in the classroom was not assessed. This is a necessary next step to better understand the conditions under which AL implementation can improve student outcomes (Andrews et al., 2011). Another limitation was the number of observations. The specific AL techniques used by an instructor can fluctuate over different parts of the course, so three classroom visits may not accurately capture the true nature of AL use in each class. However, three observations were used in similar observational studies (Sawada et al., 2002; Eddy et al., 2015).

The interviews provided evidence for how instructional change among the faculty was achieved. There were shifts over the course of the reform in how faculty defined, planned for, and implemented AL in their classrooms; however, the alignment of teaching perceptions and teaching practices for individual instructors suggests that changes in AL definition and consideration of assessment may be more impactful on practices than changes in planning for AL. Interestingly, the conversations about assessment within the group were supported by conversations about learning objectives, but it appears that considerations of assessment were more aligned with changes in practice than the use of learning objectives. Faculty also changed how they conceptualized AL across the years. Almost all instructors saw AL as engaging or involving the students with the content in the first year of reform. During the second year, instructors built onto that definition by saying that AL is interactive. This was reflected in their classrooms, as all instructors reported using and were observed to use group AL that year. In the third year, instructors also started to see AL as a process in which students are constructors of knowledge. This may be why more instructors self-reported using activities in their classes in year 3, because activities typically provide more extended opportunities for group work and knowledge construction.

Part of the shift in faculty practices and perceptions of teaching may be attributed to the use of feedback (Kluger and DeNisi, 1996) and collaboration (Sunal et al., 2001) throughout the reform process, factors that seem to assist faculty when dealing with barriers to change (Gormally et al., 2014). The AL observation results were shared at the end of each semester so that each instructor was aware of his or her own use of AL and the use for the program as a whole. The feedback was likely responsible for instructors changing their perceptions of their AL use to become more aligned with actual AL use, as evidenced in year 3. It should be noted that instructors did not merely change their responses when asked on average how much time they spent in AL to match their observations from the prior year. Rather, they increased their estimates of time spent each year and showed an actual increase in the time they spent in AL in the third year. Instructors then used this information as part of their discussions about programmatic goals for AL; increasing the use of group AL, for example, was discussed as a change the group wanted to make over time. Through this feedback and discussion, instructors were also introduced to the variety of ways AL was being implemented in the classroom when they shared ideas with each other. As such, it is recommended that institutions fostering reform should host regular meetings with the instructors where they are asked to discuss and make decisions about implementation of their courses; this may be key to the long-needed switch from advocating for AL to implementing AL (Freeman et al., 2014; Gormally et al., 2014).

Another factor that may be key to change is this faculty learning community was led by a biology education researcher. This allowed for proper navigation of faculty through the biology education literature and effective practices. The monthly meetings instructors participated in included discussion of best practices from the literature, including learning objectives, backward design (Wiggins and McTighe, 1998), and using collaboration in classroom activities. The leader also provided a link to contacts within the field of biology education, which allowed for securing professional development opportunities for the faculty. Three instructors commented during the interviews on the value of these meetings and the collaborative community context in which they occurred. Frank said, "It's allowed me to interact with the other instructors of the course and see what kinds of activities they do." Celine talked about the impact on her teaching, "I think having the data on what I was doing in my classroom presented to me and then being able to compare myself with other people that are teaching the

same course, I think it led to me using more, asking students to talk to one another more and I think I got ideas for activities that I may not have otherwise done." Ingrid also shared about the impact in her classroom, "I think a lot more about what I would like, I'm more purposeful in the way I design my class. I'm thinking more about what do I want the students to know so this backward design, I'm doing this a lot more."

It may be important to highlight the factors that we believe did not drive change. There were no meaningful data on student learning changes over the reform period, because there were no common exams across these courses and no national instrument was identified to capture these changes. Therefore, instructors were presumably not motivated by data showing increased student conceptual learning. We did have data showing increased student gains in scientific literacy (Auerbach and Schussler, 2017), but these data were not completed until after year 3. We also do not believe that change was driven by discussions of enjoyment of teaching within the community or changes in student relationships. These were not conversations that occurred as part of the community of practice, and these topics were not mentioned in any of the faculty interviews. The community of practice focused almost exclusively on logistics of curricular changes and teaching practices. Motivation changes were not an explicit part of the community process, so any changes in motivation that drove faculty practice were implicit and not observable as part of the project.

Instructional change to increase the use of AL in large introductory biology programs is possible, though this study suggests that it may require time for the changes to take effect. A recommendation for departments that wish to align their instruction with best practices is to do it collaboratively and to give faculty some autonomy to make decisions about implementation. Faculty should meet regularly and talk about their classroom practices, share ideas and resources, discuss relevant literature, and conduct structured observations of their practice in order to provide feedback to inform their teaching perceptions. Through community approaches such as these, introductory biology may move closer to the types of practices that can help all students achieve success.

ACKNOWLEDGMENTS

REFERENCES

American Association for the Advancement of Science (2011). *Vision and change in undergraduate biology education: A call to action. Final report*. Washington, DC.

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education*, *10*, 394–405.

Armbruster, P., Patel, M., Johnson, E., & Weiss, M. (2009). Active learning and student-centered pedagogy improves student attitudes and performances in introductory biology. *CBE—Life Sciences Education*, *8*, 203–213.

Auerbach, A. J., & Schussler, E. E. (2016). Instructor use of group active learning in an introductory biology sequence. *Journal of College Science Teaching*, *45*, 67–74.

Auerbach, A. J., & Schussler, E. E. (2017). Curriculum alignment with Vision and Change improves student scientific literacy. *CBE—Life Sciences Education*, *16*(2), ar29.

Collins, J. W. III., & O'Brien, N. P. (Eds.). (2003). *The Greenwood dictionary of education*. Westport, CT: Greenwood.

Couch, B. A., Brown, T. L., Schelpat, T. J., Graham, M. J., & Knight, J. K. (2015). Scientific teaching: Defining a taxonomy of observable practices. *CBE—Life Sciences Education*, *14*, ar9.

Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, *78*, 1056–1063.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *Bioscience*, *61*, 550–558.

Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE—Life Sciences Education*, *14*, ar23.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Oloroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, *111*, 8410–8415.

Freeman, S., O'Conner, E. O., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., & Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. *CBE—Life Sciences Education*, *6*, 132–139.

Gess-Newsome, J., Southerland, S. A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal*, *40*(3). doi:10.3102/00028312040003731

Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about teaching in higher ed: Neglected opportunities to promote change. *CBE—Life Sciences Education*, *13*, 187–199.

Henderson, C., & Darcy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics–Physics Education Research*, *3*, 020102.

IBM Corporation. (2011). *IBM statistics for windows, Version 22.0*. Armonk, NY.

Kember, D. (1997). A reconceptualization of the research into university academics' conceptions of teaching. *Learn Instagram*, *7*(3), 255–275.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*, Thousand Oaks, CA: Sage.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education*, *30*, 159–167.

Middlemis Maher, J., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE—Life Sciences Education*, *12*, 345–351.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Obenland, C. A., Munson, A. H., & Hutchinson, J. S. (2012). Silent students' participation in a large active learning science classroom. *Journal of College Science Teaching*, *42*(2), 90–98.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*, 307–332.

President's Council of Advisors on Science and Technology.(2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: U.S. Government Office of Science and Technology.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. Los Angeles, CA: Sage.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics

class-rooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, *102,* 245–253.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview.

Smith, M. K., Jones, F. H. M., Gilbert, S., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12,* 618–627.

Smith, M. K., Vinson, E. L., Smith, J. A., Lewin, J. D., & Stetzer, M. R. (2014). A campus-wide study of STEM courses: New perspectives on teaching practices and perceptions. *CBE—Life Sciences Education*, *13,* 624–635.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, *323,* 122–124.

Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE—Life Sciences Education*, *10,* 55–63.

Sunal, D. W., Hodges, J., Sunal, C. S., Whitaker, K. W., Freeman, M., Edwards, L., … Odell, M. (2001). Teaching science in higher education: Faculty professional development and barriers to change. *School Science and Mathematics*, *101,* 246–257.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*, (4th ed.). Needham Heights, MA: Allyn & Bacon.

Turpen, C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: Variations in physics' professors implementation of peer instruction. *Physical Review Special Topics–Physics Education Research*, *5,* 1–18.

Turpen, C., & Finkelstein, N. D. (2010). The construction of different classroom norms during peer instruction: Students perceive differences. *Physical Review Special Topics–Physics Education Research*, *6,* 1–22.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco, CA: Jossey-Bass.

Westfall, P. H., & Henning, K. S. S. (2013). *Understanding advanced statistical methods*. Boca Raton, FL: CRC.

Wieman, C. (2014). Large-scale comparison of teaching methods sends clear message. *Proceedings of the National Academy of Sciences*, *111*(13), 8319–8320.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.