"What Will I Experience in My College STEM Courses?" An Investigation of Student Predictions about Instructional Practices in Introductory Courses

Clara L. Meaders,[†]* Emma S. Toth,[‡§] A. Kelly Lane,[∥¶] J. Kenny Shuman,[¶] Brian A. Couch,[∥] Marilyne Stains,[¶] MacKenzie R. Stetzer,[‡#] Erin Vinson,[‡] and Michelle K. Smith[†]

[†]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; [‡]Maine Center for Research in STEM Education, [§]School of Biology and Ecology, and [#]Department of Physics and Astronomy, University of Maine, Orono, ME 04469; [∥]School of Biological Sciences and [¶]Department of Chemistry, University of Nebraska–Lincoln, Lincoln, NE 68588

ABSTRACT

The instructional practices used in introductory college courses often differ dramatically from those used in high school courses, and dissatisfaction with these practices is cited by students as a prominent reason for leaving science, technology, engineering, and mathematics (STEM) majors. To better characterize the transition to college course work, we investigated the extent to which incoming expectations of course activities differ based on student demographic characteristics, as well as how these expectations align with what students will experience. We surveyed more than 1500 undergraduate students in large introductory STEM courses at three research-intensive institutions during the first week of classes about their expectations regarding how class time would be spent in their courses. We found that first-generation and first-semester students predict less lecture than their peers and that class size had the largest effect on student predictions. We also collected classroom observation data from the courses and found that students generally underpredicted the amount of lecture observed in class. This misalignment between student predictions and experiences, especially for first-generation and first-semester college students and students enrolled in large- and medium-size classes, has implications for instructors and universities as they design curricula for introductory STEM courses with explicit retention goals.

INTRODUCTION

Recent national reports have cited ongoing issues in undergraduate science, technology, engineering, and mathematics (STEM) education. Approximately half of first-year undergraduate students who start in STEM fields graduate with a STEM bachelor's degree 6 years later, and most of this loss of students occurs between the first and second year of college (Seymour and Hewitt, 1997; Chen, 2013; Eagan *et al.*, 2014). Although they have similar rates of STEM aspiration, students from underrepresented minority (URM) and first-generation backgrounds leave STEM majors by either switching majors or leaving college at higher rates than their classmates (Engle and Tinto, 2008; Chen, 2013; Cataldi *et al.*, 2018). This unequal attrition is a concern, because it leads to a systematic underrepresentation of certain populations within STEM majors, directly contributing to underrepresentation at graduate and professional school levels and within the STEM workforce (National Center for Science and Engineering Statistics, 2017).

Multiple studies have investigated the characteristics of students who leave STEM majors in college and the reasons why (Seymour and Hewitt, 1997;

Alting and Walser, 2007; Chang *et al.*, 2008; Shaw and Barbuti, 2010; Sithole *et al.*, 2017). These factors include demographic characteristics (including gender, URM or first-generation status, and income), family background, pre-college academic preparation (including STEM courses, and in particular math courses taken in high school), current college experiences (including STEM courses taken during the first year of college), institutional context, campus climate, and institutional support (Griffith, 2010; Chen, 2013). A multivariate analysis of these factors in a study that tracked 16,700 first-year students at multiple institutions nationwide found that students who switched out of STEM majors completed fewer STEM courses during their first year, took introductory instead of advanced math courses during their first year, and had lower grades in their STEM courses than their peers (Chen, 2013). Other common reasons students cited for leaving STEM majors include a lack or loss of interest in STEM disciplines or seeing other majors as offering a better education (Seymour and Hewitt, 1997).

Several of the factors that contribute to STEM retention relate to how faculty teach and the messages they send. Notably, more than 90% of students who left STEM majors mentioned concern about the poor quality of teaching in their introductory college courses (Seymour and Hewitt, 1997). Moreover, large introductory STEM courses have a reputation of being "weed-out" courses that focus on lecture (Mervis, 2011). Students in passive lecture courses report general dissatisfaction with the classroom environment, lack of structure, and impersonal nature of their courses (Cooper and Robinson, 2000).

Several national reports have called for universities to move away from lecture and incorporate active learning into undergraduate STEM courses (American Association for the Advancement of Science, 2011; President's Council of Advisors on Science and Technology, 2012). Active-learning practices such as peer instruction with clicker questions or small-group work have been shown to be more effective for student learning and engagement than solely lecturing in college classrooms (Haak *et al.*, 2011; Smith *et al.*, 2011; Freeman *et al.*, 2014). Furthermore, incorporating active learning in the classroom has been found to increase equitable outcomes for URM and first-generation students and decrease student failure rates (Stephens *et al.*, 2012; Eddy and Hogan, 2014; Ballen *et al.*, 2017; Gavassa *et al.*, 2019). In addition to improving student learning outcomes, active-learning approaches positively impact student retention. For example, peer instruction (Mazur, 1997) in physics courses resulted in both increased student learning outcomes and persistence of students in STEM majors when compared with traditional lecturing (Crouch and Mazur, 2002; Watkins and Mazur, 2019).

Despite the call for incorporating active learning into classrooms, lecture is still the predominant instructional practice in college STEM courses regardless of course level or class size (Akiha *et al.*, 2018; Stains *et al.*, 2018). This focus on lecture in college classrooms differs significantly from instructional practices used in high school classes (Akiha *et al.*, 2018). A study characterizing how class time was spent in 480 middle school, high school, first-year college, and advanced college classrooms found that the median percent of 2-minute class intervals spent lecturing shifted from 32% in high school to 80% in first-year college classes (Akiha *et al.*, 2018). Students in middle and high school courses spent more time working individually or in groups on in-class activities. The shift from high school to first-year college is the most dramatic instructional transition students experience between starting middle school and finishing college.

Because of this large shift in instructional practices, first-year student expectations may impact their course experiences and subsequent retention, particularly if those expectations are inaccurate. One framework in higher education that guides the use of student expectation data is service quality, which is borrowed from commercial enterprises (Hill, 1995; Sultan and Wong, 2010). Through this lens, student survey and interview data can be used to promote increased alignment between instructors and students and can inform shifts from instructor-led to student-centered teaching. Specifically, these data can help instructors manage student expectations within course constraints and provide insights on how to improve the classroom experience (Sander *et al.*, 2000).

To begin to explore how student expectations affect course experiences, students in two sections of a large-enrollment biology course at one institution were surveyed about how they predicted class time would be used (Brown *et al.*, 2017). Compared with than their returning student peers, first-year students expected more class time to be spent completing activities and working in small groups as opposed to listening to lecture, suggesting that students come in with different expectations for the instructional practices that they will experience. Brown *et al.*, 2017 viewed the discrepancies between student predictions and class practices through the lens of expectancy violation theory (Burgoon, 1978) and proposed that these discrepancies can negatively impact a student's experience within a course. Another study of student expectations across 14 sections of math and humanities courses at one university found that first-year students were more likely than returning students to both value and expect group work in their large-enrollment courses (Messineo *et al.*, 2007). These student expectations are important, because they reveal underlying differences in student conceptions of the college experience and reflect the degree of adjustment students will need to make during a course.

Although students have cited instructional practices as a reason for leaving, the intersection of course expectations, perceptions, and instructional practices has not been investigated in depth across STEM disciplines. In this study, we aimed to expand our understanding of the expectations students have about the instructional styles they will encounter in their college courses beyond biology to include a larger set of STEM disciplines. We built upon previous work about student expectations by surveying students across 10 STEM disciplines at three universities and by using linear mixed modeling to account for demographic differences within the student body. We also collected observation data about teaching practices from the classes in which these students were enrolled using the Classroom Observation Protocol for Undergraduate STEM, or COPUS (Smith *et al.*, 2013). Specifically, we asked, 1) What types of instructional practices do students predict for their college STEM courses? 2) Do those predictions vary by student demographics or course characteristics? 3) To what extent do the learning environments provided in introductory STEM courses align with student expectations? To address these research questions, we surveyed students during the first week

of class in the Fall semester to measure their predictions about instructional practices and midway through the semester to assess their perceptions of the instructional practices used. The answers to these questions help provide a more complete understanding of how students transition from high school to college STEM classes and have implications for how faculty in introductory STEM courses choose to structure their courses and communicate about expectations during the first days of class.

METHODS

Development of the Survey Instrument

To develop our first-week and midsemester surveys, we 1) wrote questions asking students about how often instructional practices such as lecture would occur, 2) gave a pilot version of the survey to 2540 students at two research-intensive universities during the Fall 2017 semester, 3) analyzed the pilot data to determine the types of information the survey was eliciting, 4) met with a focus group of undergraduate students to explore how the questions were being interpreted, and 5) made further revisions based on feedback from several discipline-based education research groups that included undergraduates, graduate students, postdocs, and faculty. More information about survey development can be found in Supplemental Appendix S1.

Data Collection

During Fall 2018, we distributed revised first-week and midsemester surveys to students at the two universities from the pilot survey as well as at a third research-intensive university. Survey items are included in Supplemental Appendix S2.

We identified students to participate in our study by reaching out to instructors who taught large introductory STEM courses and who were simultaneously participating in a topic-based faculty learning community (Cox, 2004) focused on the high school to first year of college instructional transition during the 2018–2019 academic year. The courses were taught by a total of 20 individual or groups of instructors and ranged in size from 20 to 565 students. The instructors distributed surveys by email or through links posted on course management systems in a total of 22 courses. The first-week predictions survey occurred during the first week of the semester, and the midsemester perceptions survey occurred between weeks 6 and 8. The course subjects in our study broadly covered all of STEM and included biology, chemistry, computer science, earth science, ecology and environmental science, economics, engineering, forestry, mathematics, physics, and statistics.

Students completed the surveys online, outside class. The surveys were open for 1 week, starting after the end of the first class period. Some faculty gave extra credit or participation credit points as an incentive. For the first-week survey, we received 2436 student survey responses. The total course enrollment was 3916 students, resulting in an average response rate of 62%. For the midsemester survey, we received 1671 responses, resulting in an average response rate of 42%. We removed responses from the data set if students 1) did not agree to the consent form, 2) reported being under 18 years old, or 3) left more than 50% of the content questions blank (excluding optional demographic questions). If students responded to the survey for the same class more than once, we kept only their first responses.

At the end of the survey, we included several demographic questions (including first-generation student status, URM status, English language spoken at home, gender, and first semester on a college campus; Supplemental Appendix S3). These demographic questions were included based on a literature search to identify predictor variables that affect student expectations or that were included in other national studies related to college courses (Chen, 2013; Yee, 2016; Brown *et al.*, 2017). We also included two questions asking about international student status and transfer student status to account for students who either were unfamiliar with the U.S. education system or were at a new university despite being returning students.

We matched students who submitted responses for both the first and second survey by full name and student ID. For matched students who answered both surveys, we removed those who changed their answer for demographic questions from the first-week to the midsemester survey from the data set. If a matched student left a demographic question blank on one survey, but answered it on the other survey, the answer was filled in to match on both surveys. We removed responses if the student selected "prefer not to answer," "other," or left any of the demographic questions blank on both surveys, as these responses were not informative to the research questions. This processing yielded a data set with 1638 responses for the first-week survey and 1269 responses to the midsemester survey, with 829 students who responded to both surveys. To examine how representative the survey population was to each course as a whole, we obtained course-level percentages of gender and URM status from each university's registrar and compared the percent of male, female, and URM students who responded to our survey with the total percent of each group within each course (Supplemental Table S1). Additional student demographic information is shown in Table 1, and course characteristic information is shown in Table 2.

Student Predictions about Class Time

We used mixed-effects model regression analyses to determine whether a particular set of demographic variables could be used to explain the variation in student responses to a first-week survey question asking about the percent of class time that would be spent with the instructor lecturing (Supplemental Appendixes S2 and S3). We chose to focus on lecture, as this instructional teaching practice is prevalent across college courses (Stains *et al.*, 2018). We used the R base package for testing mixed-effects models (RStudio Team, 2015) and the ggplot2 package for graphics (Wickham, 2016).

Our first step was to classify predictor variables as random or fixed effects. We followed the recommendations outlined in Theobald (2018) for classifying predictor variables as random or fixed effects and for model selection. We treated gender as a binary, categorical predictor, as the majority of students (99.7%) self-identified as either male or female. All other demographic responses were binary predictors, except for course size, which was a three-level categorical predictor (small, medium, and large based on course size designations in Freeman *et al.*, 2014). We included gender, first-generation status, URM status, transfer student status, international student status, English language spoken at home, and course size as fixed effects. We were interested in the effects of these specific demographic variables

TABLE 1. Demographic characteristics of the student responses for the first-week (*n* = 1638 students) and midsemester (*n* = 1269 students) surveys, with total numbers within each group and percent out of the total number of responses reported

Student variables	First-week survey	Midsemester survey
College experience		
First-semester	779 (48%)	647 (51%)
Returning student	859 (52%)	622 (49%)
English spoken at home		
English spoken at home as a child	1483 (91%)	1172 (92%)
English not spoken at home as a child	155 (9%)	97 (8%)
First-generation status		
First-generation	443 (27%)	379 (30%)
Continuing generation	1195 (73%)	890 (70%)
Gender		
Male	798 (49%)	652 (51%)
Female	840 (51%)	617 (49%)
International student		
Domestic	1546 (94%)	1216 (96%)
International	92 (6%)	53 (4%)
Transfer student		
Nontransfer	1466 (89%)	1153 (91%)
Transfer	172 (11%)	116 (9%)
URM status		
URM	281 (17%)	192 (15%)
Non-URM	1357 (83%)	1077 (85%)

and whether students had predictions that differed about large introductory courses. Although university type has been associated with student retention, for example, private versus public

TABLE 2. Course characteristics of the student responses for the first-week (*n* = 1638 students) and midsemester (*n* = 1269 students) surveys, with total numbers within each group and percent out of the total number of responses included

Course variables	First-week survey	Midsemester survey
Course size		
Small (<50 students): 3 sections	45 (3%)	47 (4%)
Medium (51–110 students): 6 sections	219 (13%)	227 (18%)
Large (>110 students): 13 sections	1374 (84%)	995 (78%)
Subject		
Biology	563 (34%)	381 (30%)
Chemistry	191 (12%)	187 (15%)
Computer science	159 (10%)	116 (9%)
Earth science	47 (3%)	26 (2%)
Economics	113 (7%)	24 (2%)
Engineering	17 (1%)	11 (1%)
Forestry	38 (2%)	36 (3%)
Math	66 (4%)	90 (7%)
Physics	214 (13%)	241 (19%)
Statistics	230 (14%)	157 (12%)
University		
1	878 (54%)	513 (40%)
2	574 (35%)	597 (47%)
3	186 (11%)	159 (13%)

or selective versus nonselective (Chen, 2013), the effect of each university was not a primary research question. Consequently, we treated instructor and university as random effects to account for clustering of student responses within a course or university.

After classifying predictors as fixed or random, we measured associations between potential categorical predictors using the Goodman Kruskal package in R (Pearson, 2016). We did not include subject as a factor in our models, because subject was highly correlated with course size, instructor, and university, and some disciplines were represented by multiple courses in our data set, but other subjects had only one course. The random effect of instructor was correlated with university, so instructor was nested within university.

Our first-week survey data set included 79 responses from students who were enrolled in more than one STEM course participating in our study. Before performing model selection, we tested whether students who were enrolled in multiple courses were more likely to submit similar predictions about lecture. We hypothesized that students would have prior expectations about what college courses would be like regardless of course and, consequently, students enrolled in multiple courses would submit similar predictions for each of those courses. We analyzed the responses from students enrolled in multiple courses and used a one-way analysis of variance (ANOVA) to test whether there were significant differences in predictions based on the individual students (Supplemental Table S2). The results from the one-way ANOVA indicated that students enrolled in multiple classes were more likely to predict similar amounts of lecture for each of their classes. To account for clustering of student responses, we included a random effect for individual students.

Once we had determined which predictor variables to test, we used model selection to identify which variables to include in a final linear mixed-effects regression model using the R package lme4 (Bates *et al.*, 2015). We performed model selection using Akaike's information criterion (AIC) and a Bayesian information criterion (BIC) to decide which random effects to include in our model and then to determine which combination of fixed effects best explained our data (Supplemental Appendix S4). As instructor and course were highly correlated, we did not include them together in models. We chose to include instructor and student ID as random effects in our final model and not course for the following reasons: 1) models with instructor and student ID had the lowest AIC and BIC values, 2) some courses had multiple lecture sections that were taught by different instructors participating in our study, and 3) instructors may vary in how they introduce the class on the first day. In addition, a model with instructor nested within university as random effects as well as student ID had an equivalent AIC value as a model with only instructor and student ID. An ANOVA between the two models showed that the models had equivalent fits (Supplemental Appendix S4). We chose to use the simplest model, and included only instructor and student ID as random effects.

To select fixed effects, we used the R package MuMIn, which also uses AIC values to select the best-fitting model among all combinations of variables (Barton, 2019). We also used MuMIn to calculate conditional and marginal coefficients of determination (R^2 values). We compared the best-fitting mixed-effects

model with a null model that included only the random effects (Supplemental Appendix S4). The best-fitting model was further analyzed for significance using the R package lmerTest and by calculating variance inflation factor values using the car package (Fox and Weisberg, 2011; Kuznetsova *et al.*, 2017).

Although we had accounted for the clustering of responses from a single student by including student as a random effect (Supplemental Table S2), we also performed model selection using a data set in which we randomly removed responses so that we only had one response per student (Supplemental Appendix S5). This process allowed us to confirm that the predictors identified from our analyses with the complete data set were significant regardless of how many times students were represented. The best-fitting model using either the full (multiple-response-per-student) data set or the one response per student (single-response-per-student) data set output the same significant predictor variables, indicating that including multiple responses from students does not change the overall pattern of student predictions. Because both the single-response-per-student and multiple-response-per-student sets showed similar patterns, we chose to keep responses from students enrolled in multiple classes. This decision allowed us to maintain a more representative pool of students within each course.

Classroom Observations

Because we are interested in the transition from high school to college, we observed the first four to five class periods each instructor taught using the COPUS (Smith *et al.*, 2013). Although previous studies using classroom observations have used two to three class periods either over 1 week of a course or from different points in the semester (Smith *et al.*, 2014; Lewin *et al.*, 2016; Durham *et al.*, 2018; Pelletreau *et al.*, 2018), four class periods have been found to reflect variation in an instructor's teaching practices (Stains *et al.*, 2018). For courses with one instructor, the class periods were all at the beginning of the semester, excluding the first day of class, which typically includes a higher level of the administration code. For courses cotaught by multiple instructors, the class periods were spaced throughout the semester but included the first four or five periods of the participating instructor's teaching. For courses in which where the instructor taught multiple sections of the same course, only one lecture section per instructor was observed. In total, we conducted 108 class period observations.

The COPUS consists of coding 25 instructor and student behaviors at 2-minute time intervals over the course of a class meeting. There are various ways to explore COPUS observation data (Smith *et al.*, 2013; Stains *et al.*, 2018; Erdmann and Stains, 2019). We used the COPUS analyzer tool at COPUSprofiles.org to incorporate all 25 student and instructor behaviors into an aggregated cluster between 1 and 7, which can be further coded into either didactic, interactive lecture, or student-centered teaching practices. We used R to visualize the differences between the classroom COPUS profiles in a heat map.

The percent of 2-minute intervals that included lecture for each course was determined by counting the 2-minute intervals in which lecture was marked and then by dividing that number by the total number of 2-minute intervals in a class session. For example, if instructor lecturing (Lec) was coded 25 times during a 50-minute lecture, then 100% of the possible 2-minute

time intervals contained lecture. We then determined the average percent of 2-minute intervals that students experienced within any given course that included lecture. These calculations slightly overestimate the amount of time that an instructor spends on any one activity, because the instructor may have also included other activities within the space of any of the 2-minute intervals. To explore the impact of this possible overestimation, we identified the total number of 2-minute intervals that included lecture, and then examined the co-occurrence of lecture with active learning or other codes (Supplemental Appendix S6).

Correlation Analysis

We plotted the average percent of 2-minute intervals coded as lecture with the average percent of in-class time instructors spent lecturing reported by individual students within that course. A regression line was fitted in R to calculate the linear relationship between observed and reported lecture. For visualization of any differences in reporting between different demographic groups, within each course, the student survey data were disaggregated by the demographic variable and plotted against observed time.

Open-Response Question Analysis

The first-week predictions survey included a short-answer question asking students to describe what information or experiences they used to predict the percent of instructional approaches that would be used in the courses. We used inductive coding to analyze student responses. First, coauthors A.K.L. and J.K.S. read through 100 responses from the question at one institution and together developed a list of ideas appearing in those responses. The list served as the initial codebook. Second, these two coders tested the initial codebook on responses from another institution, further refining the ideas. This second stage continued with sets of responses from each institution until the coders felt that no further refinements were needed, because no new ideas were emerging. Third, each coder independently coded 48 responses, which included responses from all three institutions. Once each code had a percent agreement of 90% or greater, a single coder coded the remaining responses. Any responses that could not be readily coded using the established codebook were marked as "other" and reviewed by both coders to confirm that these responses could not be coded with any established code. After coding was complete, we reviewed the codes and illustrative responses, which revealed three themes described in the *Results*.

RESULTS

What Types of Instructional Practices Do Students Predict for Their College STEM Courses?

To explore how students predicted class time would be spent in their introductory college STEM courses, we analyzed data from the first-week survey (Supplemental Appendix S2). A multipart question asked students to predict the percent of class time the instructor would lecture to students, ask students to work alone, ask students to work in groups, or ask students to do other things. The responses were required to add to 100% for students to submit their responses and move to the next question. The survey data showed variation in responses, but overall, students predicted that the majority of class time would be spent listening
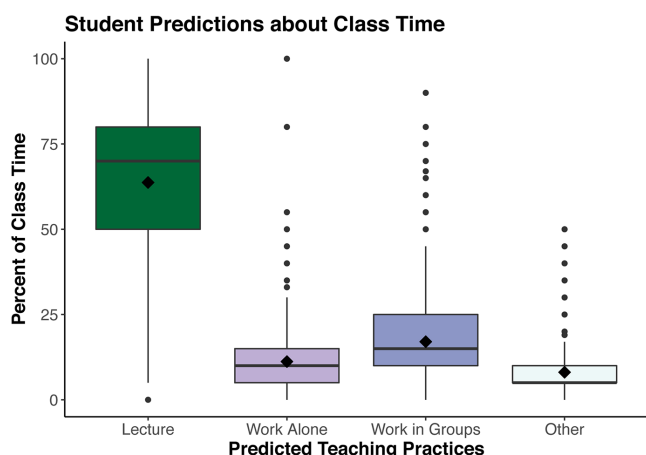
Student Predictions about Class Time

FIGURE 1. Box plot of the percent of in-class time that students predicted would be dedicated to lecture, working alone, working in groups, or other activities. The boxes represent the interquartile range (IQR) of responses for each category. Lines within each box represent the median, and diamonds represent the mean response. Whiskers represent 1.5 times the IQR. Dots represent outliers.

to lecture (Figure 1). Specifically, the mean predictions were that the instructor would spend 64% (SD ±21) of class time lecturing to students, 11% (SD ±9) of class time asking students to work individually, 17% (SD ±13.5) of class time asking students to work in groups, and 8% (SD ±7) of class time on "other."

Next, we determined what information students were using to make their predictions. Students were asked, "What experiences or information did you use to make your predictions about how class time will be spent (for example, experiences or information you received before or during the semester)?" Analysis of the open-ended question responses revealed a variety of sources students used to inform their predictions, and these sources could be divided into three major themes: first-hand experiences with the course or instructor, course characteristics, and information acquired outside the course (Table 3). These responses provide insights into experiences and preconceptions that shape student predictions.

Do Student Predictions about Instructional Practices Vary by Student Demographics or Course Characteristics?

Given the range of responses for student predictions about the percent of class time that would be spent listening to lecture, we wanted to determine whether certain demographic variables predicted student responses. We included each of the demographic variables (Table 1) as a fixed effect in a linear mixed-effects model, along with course size. The best-fitting model includes first-generation student status, first-semester on a college campus status, and course size as significant predictors and explains 56% of the variation in the data (Table 4).

According to the model, a first-generation, first-semester student in a large course (specified by the model intercept) predicted that ~64.29% of in-class time would be dedicated to lecture (Table 4). The model also provided estimates based on demographic variables or course size relative to the intercept. For example, continuing-generation students predicted that ~3.33% more class time would be spent listening to lecture compared with their first-generation student peers. Similarly, returning stu-

dents predicted ~4.40% more lecture than their first-semester student peers. Adding the estimates together (64.29% + 3.33% + 4.40% = 72.02%) yields the percent of lecture predicted by continuing-generation students who were returning students in large-enrollment courses. Course size also was a significant factor, and the negative values associated with medium and small classes indicate that students predicted that 20.62% (medium) and 19.41% (small) less class time would be spent listening to lecture compared with students in large-enrollment classes.

We visualized the variation in student predictions about in-class time that would be spent on lecture in their college STEM courses with box plots (Figure 2). These box plots were consistent with the results from our model and showed that first-generation students predicted less time would be spent listening to lecture compared with their continuing-generation peers (Figure 2A) and that first-semester students on a college campus also predicted that less class time would be spent listening to lecture compared with students who were returning to college (Figure 2B). First-generation college students predicted a mean of 59% (SD ±21) of class time would be dedicated to lecture, while continuing-generation college students predicted that 65% (SD ±20) of class time would be dedicated to lecture. The mean percent of class time students predicted would be dedicated to lecture was 61% (SD ±19) for first-semester students and 66% (SD ±22) for returning students. These differences in predictions for first-generation and first-semester students are slightly higher than the predicted differences from our linear mixed model, which accounts for both first-generation student status and first-semester status together, along with course size. To determine whether the broader differences in overall student predictions were consistent across courses, we investigated patterns of differences in student predictions at the course level. We found that, in the majority of courses, first-generation and first-semester students predicted less lecture than their peers (Supplemental Appendix S7), indicating that these demographic factors impact student predictions across courses.

On the first-week survey, course size was also a predictor in the best-fitting linear regression model (Table 4). Students in large-enrollment courses (*n* = 13 courses) predicted a mean of 67% (SD ±19) of class time dedicated to lecture, while students in medium (*n* = 6 courses) and small courses (*n* = 3 courses) predicted a mean of 47% (SD ±21) and 48% (SD ±22), respectively (Figure 2C).

To What Extent Do the Learning Environments Provided in Introductory STEM Courses Align with Student Expectations?

Analysis of the first-week survey data led to questions about how student predictions compared with the amount of lecture that actually occurred in their college STEM courses. We observed the first four to five class periods taught by each instructor and analyzed the relative frequency of the "lecture" code. We calculated the mean percent of 2-minute intervals that included lecture across all of the courses. This observed number (74%) is higher than the average student predictions about the percent of class time that the instructor would lecture to the students (Figure 2D). Notably, the predictions of first-generation and first-semester students on the first-week survey are most distant from the observed average of 2-minute intervals with lecture within each course.

TABLE 3. Open-response question analysis of a survey question asking students about what experiences or information they used to make their predictions about how class time would be spent

Theme	Experience or information	Responses containing a particular code	Example quotation[a]
Firsthand experiences with course and/or instructor	First day(s)—nonspecific	26%	I also used my experience from this morning's [course number] class.
	First day(s)—activities in class	8%	So far [the class] has mostly been lecture with some student involvement.
	First day(s)—instructor's description of instructional practices	13%	I based it on what the instructor described during our first lecture.
	Syllabus/course website	11%	The online class website provides details on what will be covered during class.
	Interacting with the instructor outside of class time	1%	My teacher gave a presentation [at a student orientation event], which included how class time would be typically spent.
Course characteristics	Based on the subject or content of the course	11%	I inferred that due to this class being about software engineering [time] would be spent working in groups and working on coding.
	Based on class size or structure	11%	Mostly because the course is called lecture. There are far too many students to be trying to split into groups etc. I assume that is what lab time is to be used for.
Information acquired outside course	Nonspecific prior knowledge/ experience and assumptions	14%	This is what other classes usually are set up like.
	Experience in previous classes— in general	13%	I based my predictions on past experiences with college classes.
	Talking to individuals who are not in the course	10%	My older sisters have told me a lot about college, and I made predictions based off what they said.
	Experience in high school classes	8%	In the past, with high school classes, many lectures involved student interaction so I feel as though this may also be the case.
Miscellaneous	I do not know/complete guess	2%	I guessed.
	Off-topic	2%	—
	Other	4%	—
	No response	6%	—

[a]Example quotations are included to provide context for how codes were generated.

Students enrolled in large courses predicted more lecture than students in medium and small courses (Figure 2C). The classroom observation data also followed this trend, with the observed percent of class time involving lecture increasing as course size increased (Figure 2E). The observed number of 2-minute intervals with lecture in the large and medium classes was higher than the amount of lecture that students predicted (Figure 2E). For the small courses, the mean predicted amount

TABLE 4. Estimated coefficients for variables from the best-fitting linear mixed-effects model that examines how different predictors impact the percent of in-class time students expect the instructor to spend lecturing

Predictors	Estimate	SE	t value	p value[a]	2.5% Confidence interval	97.5% Confidence interval
(Mean intercept)	64.29	3.04	21.14	5.97 e–15***	58.45	70.16
Course size						
Medium course	−20.62	4.87	−4.24	0.00046***	−30.03	−11.28
Small course	−19.41	6.19	−3.14	0.0035**	−31.28	−7.53
Continuing generation	3.33	1.06	3.15	0.0017**	1.26	5.41
Returning student	4.40	0.96	4.57	5.33 e–06***	2.50	6.28
Random effects						
Instructor + Student ID Intraclass Correlation Coefficient (ICC) = 0.48 Observations: 1638 students $R^2m = 0.14/R^2c = 0.56$						

The intercept represents a first-generation, first-semester student in a large course. The t value reported is the (regression coefficient)/(standard error).
[a]Statistical significance is indicated by **, $p < 0.01$; and ***, $p < 0.001$.
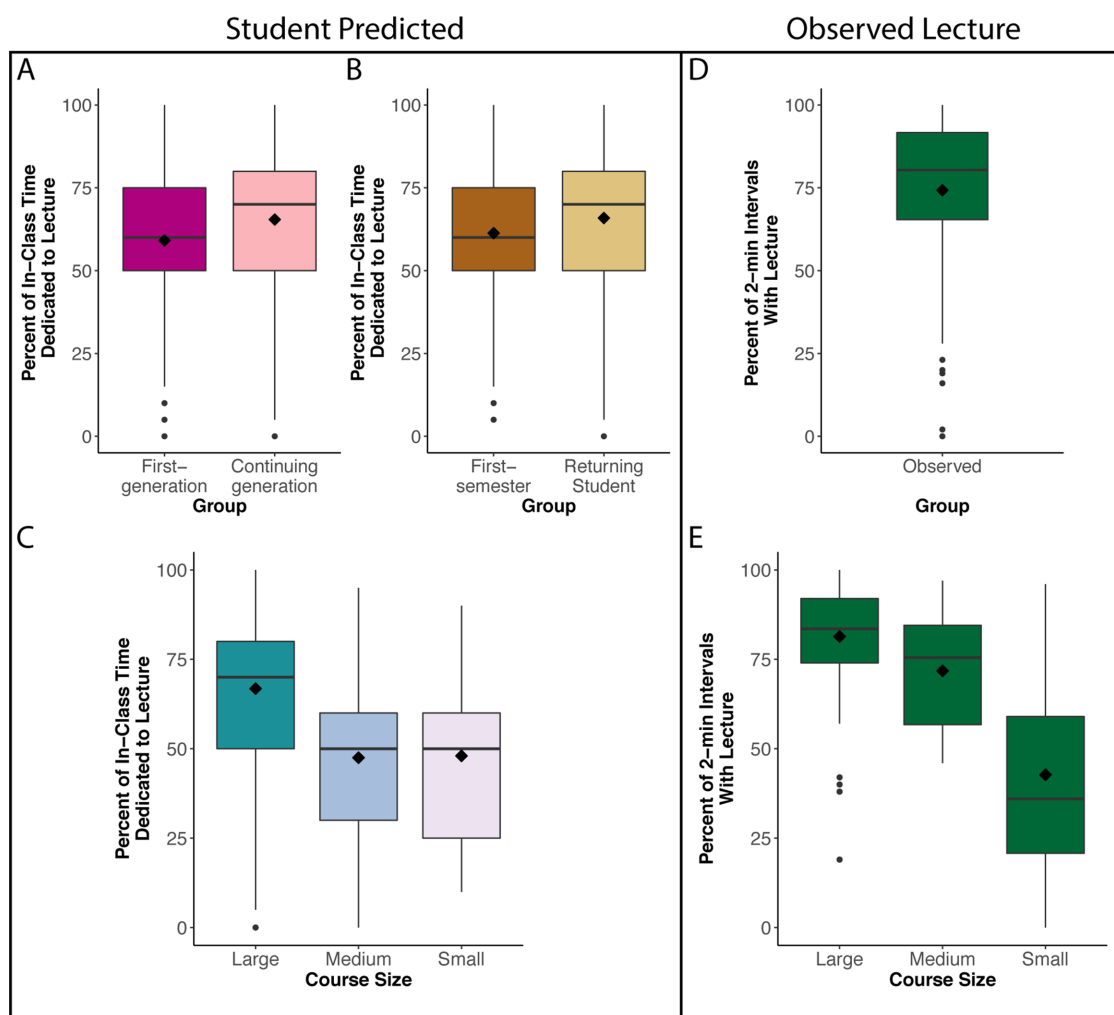
FIGURE 2. Box plots of student-predicted lecture disaggregated by variables identified in linear mixed-effects models and box plots of COPUS observations. Student predictions of in-class time that would be dedicated to lecture disaggregated by (A) first-generation or continuing-generation status, (B) first-semester on a college campus or returning student status, and (C) course size. Average percent of 2-minute intervals that contained lecture for each COPUS observed class period (D) for full sample and (E) disaggregated by course size. The boxes represent the interquartile range of responses for each category. Lines within each box represent the median, and diamonds represent the mean response. Whiskers represent the largest and smallest values within 1.5 times the IQR. Dots represent outliers.

of lecture matched more closely to the observed. Owing to the limited number of small courses in our study and the smaller number of student responses due to smaller course size ($n = 3$ courses and 43 student responses), the differences seen in small classes may not be representative of student predictions across small courses. However, the trend we observed of more lecture in larger classes and less lecture in smaller classrooms matches with observations across other North American institutions (Stains *et al.*, 2018).

When the percentage of 2-minute time intervals was averaged across courses, the overall average was 74% (Figure 2D). We identified that, while the majority of 2-minute intervals consisted of only lecture or lecture-related codes, there were 2-minute intervals that had lecture and active learning or other activities coded, which could have resulted in an overestimation of how much time instructors spent lecturing (Supplemental Appendix S6). Therefore, we selected a random sample of fifty 2-minute time intervals that included lecture as well as active

learning or other activities and timed the seconds dedicated to lecture. The timing analysis suggests that, overall, ~68.5% of total in-class time was dedicated to lecture. This corrected total amount of time is higher than the mean amount of lecture predicted by students overall (64%) and, notably, is also higher than the amount predicted by first-generation (59%) and first-semester students (61%).

Because the lecture code can co-occur with other COPUS codes and instructional styles may differ across class periods, we also used COPUSprofiles.org to gain a more holistic view of the instructional practices. This profile program aggregates the observation data into seven different types of instructional practices spanning the range from didactic to student-centered for each class period (Stains *et al.*, 2018). Our results show that instructors in large courses used primarily didactic instructional practices, with some interactive lecture and a few student-centered class periods (Figure 3). In contrast, instructors from medium courses represented a range of teaching styles, with no
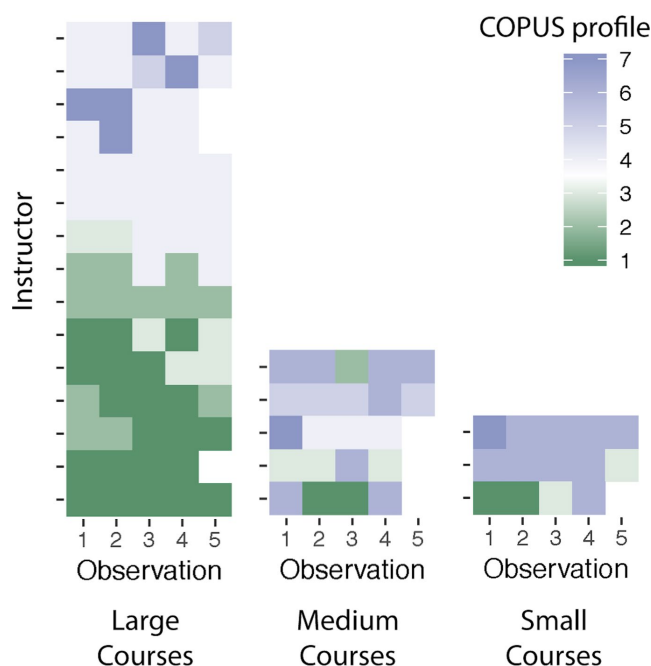
FIGURE 3. COPUS profiles of each individual instructor's first four or five class periods, disaggregated by course size. Each row includes observations from an individual instructor, arranged chronologically from left to right. COPUS profiles represent seven types of instructional styles, indicated by 1 to 7 on the heat map, and range from majority lecture to student centered. Clusters 1 and 2 are "didactic" and are primarily lecture based: Cluster 1 (dark green) has no student involvement except questions to and from students, while cluster 2 (green) sometimes incorporates clicker questions. Clusters 3 (light green) and 4 (lightest blue) are categorized as "interactive lecture" and include either other group activities or clicker questions in groups, respectively. Clusters 5, 6, and 7 (light blue, blue, dark blue) are "student-centered" class periods, with cluster 5 representing regular usage of group work, cluster 7 slightly less usage, and cluster 6 including group worksheets and one-on-one assistance from the instructors. Instructors are organized from most didactic to most student centered on the *y*-axis. Blank spaces indicate when only four observations of a particular instructor occurred.

instructor using solely didactic practices. Although we only had three instructors from small courses in our study, 10 out of 14 of the observed class periods featured student-centered practices. The aggregate COPUS profiles are consistent with our analysis of the "lecture" code alone, indicating that our focus on the percent of 2-minute intervals that included lecture reflects broader student experiences within a course. Our results are also similar to previous studies that showed that instructors are generally consistent in their instructional patterns across multiple class periods (Pelletreau *et al.*, 2018) and that any variation tends to be between just two of the broad COPUS profiles teaching styles (Stains *et al.*, 2018).

Is Variation in Student Predictions about How Class Time Will Be Spent Due to Differences in How Students Perceive Time?

Although we found differences between the predictions of students about the percent of class time dedicated to lecture

(Figure 1), a confounding question is whether students are accurate in reporting the amount of lecture they receive. To answer this question, we gave students a survey at the midsemester time point and asked them to report the percent of class time during which the instructor lectured on a typical day (Supplemental Appendix S2).

To visualize how closely student perceptions at the midsemester point matched COPUS observation data from within each course, we plotted the data on scatter plots and fitted simple linear regressions (Figure 4A). The slope of the regression line (0.94) indicates that, at the midsemester point, students were accurate in reporting the amount of in-class time dedicated to lecture.

We also plotted individual student predictions about lecture from the first-week survey and compared them with the amount of time the instructor spent lecturing using COPUS observation data (Figure 4B). Although there was a positive relationship between student predictions and COPUS observations, the slope of the regression (0.48) is lower than at the midsemester point, likely because the students have less experience in the class. Taken together, the comparison of student responses and COPUS observations indicates that differences between student predictions about lecture on the first-week survey are due to differences in expectations and not due to differences in how students experience time.

In addition, one explanation for the differences between the predictions of different demographic groups (Table 4 and Figure 2) is that first-generation and first-semester students perceive class time and teaching practices in different ways. To address this question, we disaggregated the data by first-generation and first-semester student status and performed additional correlation analyses between the observed percent of class time and midsemester survey data (Supplemental Appendix S8). Disaggregating students by demographic group had borderline to no significant effect on the strong correlation between perceptions and COPUS observations. Taken together, these data suggest that first-generation and first-semester students perceive the amount of lecture in similar ways as their classmates and lend further credence to the finding that these groups report different predictions on the first-week survey (Figure 2).

Discussion and Implications

To investigate student experiences of the high school to first-year college instructional transition, we surveyed more than 1500 students from three research-intensive universities in courses across 10 STEM disciplines. Our results showed that student predictions about the instructional practices in introductory STEM courses differed modestly based on certain demographic variables, including first-semester and first-generation student status, and by course size (Table 4 and Figure 2). Although the estimates from students with the two significant demographic variables (first-generation and first-semester student status) were lower than the estimates from students enrolled in different course sizes, these demographic differences are notable when we consider that none of the other demographic predictors were identified in the best-fitting model and also that the trends were consistent at the course level (Supplemental Appendix S7).
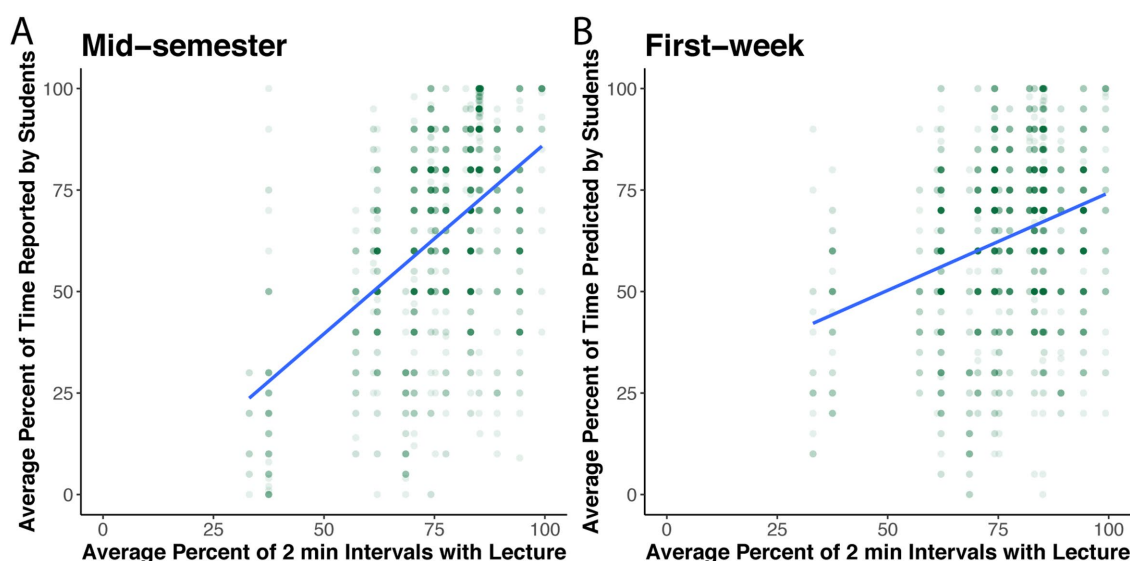
FIGURE 4. Scatter plots of in-class time spent lecturing reported by students (*y*-axis) compared with the average observed percent of 2-min intervals that contained lecture for that course (*x*-axis). Each dot represents perceptions or predictions from one student, increased opacity indicates that several students reported similar percents of time. (A) Midsemester perceptions, regression line $0.94x - 7.3$, R^2 0.26. (B) First-week predictions, regression line $0.48x + 26.26$, R^2 0.08.

Student Predictions about the In-Class Time of Their Introductory College STEM Courses

There was a wide range of student predictions for the percent of class time that would be dedicated to lecture (Figure 1) and what experiences students used to inform their predictions, including the first days of class (Table 3). However, students generally predicted that the majority of their in-class time would be spent listening to lecture, but still underpredicted the amount of lecture they would actually experience (Figure 2). This underprediction of the amount of lecture was more pronounced for first-semester and first-generation students. These results are consistent with previous studies of first-semester students who predicted more active learning in a course than their upper-class peers (Messineo *et al.*, 2007; Brown *et al.*, 2017).

Differences in student predictions is of interest to educators and researchers, as first-generation students leave STEM majors at higher rates than their peers, and students overall leave STEM majors at high rates after their first year of introductory courses (Chen, 2013; Eagan *et al.*, 2014). The predictions from first-semester and first-generation students reveal that students from various backgrounds may perceive college courses differently, which could be indicative of larger disconnects in their college experiences. For example, most first-semester students likely recently experienced more active learning in their high school courses (Akiha *et al.*, 2018), and they may expect their college courses to be similar to their high school courses (Lowe and Cook, 2003). Similarly, first-generation students may have less overall familiarity with the norms of college, in particular of faculty expectations (Collier and Morgan, 2008), or with the study skills needed to succeed (Horowitz, 2019). First-generation students may require more time than their peers to understand faculty expectations within courses and may also continue to expect that their college courses will be more similar to their high school courses.

There are at least two ways instructors can address the disconnect between student predictions and student experiences: 1) talk more explicitly to students about the instructional practices in the course and 2) change instructional practices in introductory courses to include less lecture and more active learning. These two responses are aligned with the service quality framework recommendations of using student feedback to inform how instructors can better manage expectations and improve instruction (Hill, 1995). Furthermore, both of these responses do not require instructors to explicitly focus on particular demographic groups in class, but rather use approaches that could be beneficial to all students in a variety of course sizes.

Talking to Students about Course Instructional Practices

Although instructors may not know what assumptions individual students are making about how class time will be spent, they can be fairly confident that there is a mix of student expectations in their classroom on the first day, with some students expecting their college courses to use instructional practices similar to those used in their high school courses. Encouraging instructors to be explicit and deliberate in their explanations and activities during the first day of class can set the tone for what instructional practices students can expect during the semester (Gaffney and Whitaker, 2015). Our data reveal that students use within-course experiences such as syllabi, course websites, or instructor talk and actions during the first day of class to inform their expectations (Table 3). Moreover, aligning student expectations early on is recommended as a way to increase student buy-in and student engagement with active-learning practices (Brazeal *et al.*, 2016; Brazeal and Couch, 2017; Cavanagh *et al.*, 2016; Tharayil *et al.*, 2018). For example, instructors can describe that they are using active-learning techniques such as clicker questions with peer discussion or small-group discussion activities because these techniques have been shown to decrease failure rate and

increase learning (Freeman *et al.*, 2014; Ballen *et al.*, 2017). In addition, faculty can show summary data (Wieman, 2014) and/or assign articles about the benefits of active learning written for a general audience for their first week reading (Bajak, 2014; Wieman, 2014). Notably, professional development programs and general advice articles (Lang, 2019) give recommendations for how to structure the first day of class to set student expectations, but more research is needed on what instructors do during the first day, how often introductory instructors mention changes from high school to college, and how the information emphasized impacts students.

Changing Instructional Practices in Introductory Courses to Include Less Lecture and More Active Learning

In addition to being explicit about their instructional practices, introductory STEM course instructors could also increase the amount of active learning. Our work, along with that of Brown *et al.* (2017) and Messineo *et al.* (2007), suggests that introductory students would be amenable to more active learning in their courses. The addition of more active learning would both better align with the expectations of first-semester and first-generation students (Figure 2) and would likely improve student pass rates and learning outcomes (Freeman *et al.*, 2014), particularly for first-generation and URM students (Eddy and Hogan, 2014; Ballen *et al.*, 2017). Beyond increasing course performance, active learning can promote inclusive teaching practices and has been shown to increase student self-efficacy (Ballen *et al.*, 2017), which is another known factor that contributes to STEM retention (Lent *et al.*, 2008; Sawtelle *et al.*, 2012). Furthermore, using active learning promotes student–student interactions, which provide specific benefits for first-generation students, who often interact less with their teaching teams, academic support staff, and fellow classmates (Yee, 2016).

Student predictions of more active learning than they are experiencing are notable, because the discipline-based education literature has focused on how student resistance to active learning is a barrier (Seidel and Tanner, 2013). Instructors often expect students to resist active learning and cite this perceived student resistance as a reason for not incorporating more individual and group work into their classrooms (Felder and Brent, 1996; Michael, 2007). One lens that has been used to investigate the interactions between student predictions, student perceptions, and instructional practices is expectancy violation theory (Burgoon, 1978). This theory was originally developed in studies of psychology about personal space and suggests that, when an event differs from what was predicted, expectations are violated, which may impact one's experience. In the discipline-based education research literature, it has been used as a framework to examine the implications of student expectations (Gaffney *et al.*, 2010; Brown *et al.*, 2017) and has most often been discussed in relation to student resistance to active-learning practices (Gaffney *et al.*, 2010; Seidel and Tanner, 2013; Keeley, 2014). For example, instructors who wish to add more active learning to their courses often express concern about violating their students' expectations of what a "typical" college course should look like (i.e., predominantly lecture). Our survey data suggest that faculty in introductory courses are violating student expectations by lecturing more than is expected, and these results could be used to encourage additional course transformations.

The Impact of Course Size on Student Expectations

The variable that had the largest impact on student predictions was course size (Table 4), with students in the large class sizes predicting and experiencing the most lecture (Figures 2 and 3). These results align with a national trend that students in larger classes experience more lecture (Stains *et al.*, 2018). The number of students in large courses can seem like a barrier to implementing in-class problem solving or active-learning strategies (Michael, 2007). However, students cite that instructors can make large courses feel like small ones by using instructional practices such as group work (Cash *et al.*, 2017). Furthermore, there are large STEM courses that have successfully implemented individual and small-group work in their classrooms for up to 1000 students (Exeter *et al.*, 2010; Deslauriers *et al.*, 2011; Ballen *et al.*, 2017). Resources such as Allen and Tanner (2005) provide strategies for faculty interested in implementing active learning into large courses. In addition, faculty professional development programs can focus on providing support for using active learning in large-enrollment classes, and universities can work to highlight large-enrollment active-learning courses to attract potential new students.

Future Directions

Previous research about retention in STEM revealed that students who leave STEM majors are often dissatisfied with the instructional practices they experienced (Seymour and Hewitt, 1997). As instructors make changes to their courses to more explicitly set and better meet student expectations, investigations that include student interviews could provide additional insights into student perceptions of instructional practices, in particular for large introductory courses. Interviews with students who have large differences between how much lecture they predict and receive could reveal the extent to which discrepancies in expectations impact personal experiences in a course. Furthermore, as more courses incorporate active learning, it will be interesting to explore whether these changes impact student expectations. Future longitudinal studies are needed to track how student predictions about instructional practices correlate with retention in STEM majors. Specifically, it is important to explore whether or not students whose predictions are more closely aligned with teaching practices are more likely to persist in STEM than those whose predictions are less aligned. Additionally, it is important to conduct similar studies of student expectations and instructional practices at primarily undergraduate institutions and community colleges. The prevalence of lecture at research-intensive institutions has been documented on a national level (Stains *et al.*, 2018), and yet we identified a disconnect between student expectations and experiences. Given the consistency of our results across three institutions, we would predict similar findings at primarily undergraduate institutions and community colleges, but the potential for smaller class sizes and emphasis on student–faculty connections could reveal variations in student expectations and/or faculty teaching practices.

The type of survey data that we collected, which ranges across multiple universities and disciplines, is useful for identifying general trends in student thinking. However, it is also important to consider how these data can be used by faculty to promote change for individual introductory STEM course instructors. While many faculty care about student thinking and

want to explore data (Hora *et al.*, 2017), few have personal or institutional systems that support their attempts. Rather, most faculty deal with data on their own without collaboration from colleagues or experts (Hora *et al.*, 2017). Therefore, one approach to addressing these factors and sharing data with faculty is through faculty learning communities (FLCs), which are small groups of faculty who meet regularly over the course of a year to discuss and reflect on a common goal (Cox, 2004). Topic-based FLCs that address a specific issue or concern, such as the transition from high school to college STEM courses, enable faculty to work together to devise solutions. Furthermore, FLCs can inform instructional practices and improve student learning (Pelletreau *et al.*, 2018).

CONCLUSION

Our survey- and observation-based study of student predictions and experiences within introductory STEM courses shows that first-generation and first-semester college student predictions are less well aligned with actual teaching practices than those of their peers. In addition, course size has a large influence on student predictions of the amount of lecture time, and students in medium and large STEM courses underpredict the amount of class time they will spend listening to lecture. Encouraging introductory STEM instructors to talk more explicitly to students about the instructional practices in the course and use active learning could help students more successfully navigate the transition between high school and college and pursue their intended STEM majors.

ACKNOWLEDGMENTS

REFERENCES

Akiha, K., Brigham, E., Couch, B. A., Lewin, J., Stains, M., Stetzer, M. R., ... & Smith, M. K. (2018). What types of instructional shifts do students experience? Investigating active learning in science, technology, engineering, and math classes across key transition points from middle school to the university level. *Frontiers in Education*, *2*, 1–18. https://doi.org/10.3389/feduc.2017.00068

Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex. *Cell Biology Education*, *4*(4), 262–268. https://doi.org/10.1187/cbe.05-08-0113

Alting, A., & Walser, A. (2007). Retention and persistence of undergraduate engineering students: "What happens after the first year?" *ASEE*, *2007*, 9.

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

Bajak, A. (2014, May 12). Lectures aren't just boring, they're ineffective, too, study finds. Retrieved from sciencemag.org/news/2014/05/lectures-arent-just-boring-theyre-ineffective-too-study-finds

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education*, *16*(4), ar56. https://doi.org/10.1187/cbe.16-12-0344

Barton, K. (2019). MuMIn: Multimodel inference (Version 1.43.6). Retrieved March 18, 2019, from https://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brazeal, K. R., Brown, T. L., & Couch, B. A. (2016). Characterizing student perceptions of and buy-in toward common formative assessment techniques. *CBE—Life Sciences Education*, *15*(4), ar73. https://doi.org/10.1187/cbe.16-03-0133

Brazeal, K. R., & Couch, B. A. (2017). Student buy-in toward formative assessments: The influence of student factors and importance for course success. *Journal of Microbiology & Biology Education*, *18*(1), 1–10. https://doi.org/10.1128/jmbe.v18i1.1235

Brown, T. L., Brazeal, K. R., & Couch, B. A. (2017). First-year and non-first-year student expectations regarding in-class and out-of-class learning activities in introductory biology. *Journal of Microbiology & Biology Education*, *18*(1), 1–9. https://doi.org/10.1128/jmbe.v18i1.1241

Burgoon, J. K. (1978). A communication model of personal space violations: Explication and an initial test. *Human Communication Research*, *4*(2), 129–142. https://doi.org/10.1111/j.1468-2958.1978.tb00603.x

Cash, C. B., Letargo, J., Graether, S. P., & Jacobs, S. R. (2017). An analysis of the perceptions and resources of large university classes. *CBE—Life Sciences Education*, *16*(2), ar33. https://doi.org/10.1187/cbe.16-01-0004

Cataldi, E. F., Bennett, C. T., Chen, X., & Simone, S. A. (2018). Stats in brief first-generation students (college access, persistence, and postbachelor's outcomes) (pp. 1–31). National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2018/2018421.pdf

Cavanagh, A. J., Aragón, O. R., Chen, X., Couch, B., Durham, M., Bobrownicki, A., ... & Graham, M. J. (2016). Student buy-in to active learning in a college science course. *CBE—Life Sciences Education*, *15*(4), ar76. https://doi.org/10.1187/cbe.16-07-0212

Chang, M. J., Cerna, O., Han, J., & Sàenz, V. (2008). The contradictory roles of institutional status in retaining underrepresented minorities in biomedical and behavioral science majors. *Review of Higher Education: Journal of the Association for the Study of Higher Education*, *31*(4), 433–464. http://dx.doi.org/10.1353/rhe.0.0011

Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields* (NCES 2014-001). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Collier, P. J., & Morgan, D. L. (2008). "Is that paper really due today?": Differences in first-generation and traditional college students' understandings of faculty expectations. *Higher Education*, *55*(4), 425–446. https://doi.org/10.1007/s10734-007-9065-5

Cooper, J. L., & Robinson, P. (2000). The argument for making large classes seem small. *New Directions for Teaching and Learning*, *2000*(81), 5–16. https://doi.org/10.1002/tl.8101

Cox, M. D. (2004). Introduction to faculty learning communities. *New Directions for Teaching and Learning 2004*(97), 5–23. doi:10.1002/tl.129

Crouch, C. H., & Mazur, E. (2002). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977. https://doi.org/10.1119/1.1374249

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*(6031), 862–864. https://doi.org/10.1126/science.1201783

Durham, M. F., Knight, J. K., Bremers, E. K., DeFreece, J. D., Paine, A. R., & Couch, B. A. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *International Journal of STEM Education*, *5*(1), 1–15. https://doi.org/10.1186/s40594-018-0128-1

Eagan, K., Hurtado, H., Figueroa, T., & Hughes, B. (2014, April 2015). Examining STEM Pathways among Students Who Begin College at Four-Year Institutions. Washington, DC: National Academy of Sciences. (Commissioned paper prepared for the Committee on Barriers and Opportunities in Completing 2- and 4-Year STEM Degrees).

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education*, *13*(3), 453–468. https://doi.org/10.1187/cbe.14-03-0050

Engle, J., & Tinto, V. (2008). *Moving beyond access: College success for low-income, first-generation students*. Washington, DC: Pell Institute.

Erdmann, R. M., & Stains, M. (2019). Classroom as genome: Using the tools of genomics and bioinformatics to illuminate classroom observation data. *CBE—Life Sciences Education*, *18*(1), es1. https://doi.org/10.1187/cbe.18-07-0116

Exeter, D. J., Ameratunga, S., Ratima, M., Morton, S., Dickson, M., Hsu, D., & Jackson, R. (2010). Student engagement in very large classes: The teachers' perspective. *Studies in Higher Education*, *35*(7), 761–775. https://doi.org/10.1080/03075070903545058

Felder, R. M., & Brent, R. (1996). Navigating the bumpy road to student-centered instruction. *College Teaching*, *44*(2), 43–47. https://doi.org/10.1080/87567555.1996.9933425

Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, 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*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gaffney, J. D. H., Gaffney, A. L. H., & Beichner, R. J. (2010). Do they see it coming? Using expectancy violation to gauge the success of pedagogical reforms. *Physical Review Special Topics–Physics Education Research*, *6*(1), 1–16. https://doi.org/10.1103/PhysRevSTPER.6.010102

Gaffney, J. D. H., & Whitaker, J. T. (2015). Making the most of your first day of class. *Physics Teacher*, *53*(3), 137–139. https://doi.org/10.1119/1.4908079

Gavassa, S., Benabentos, R., Kravec, M., Collins, T., & Eddy, S. (2019). Closing the achievement gap in a large introductory course by balancing reduced in-person contact with increased course structure. *CBE—Life Sciences Education*, *18*(1), ar8. https://doi.org/10.1187/cbe.18-08-0153

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, *29*(6), 911–922. https://doi.org/10.1016/j.econedurev.2010.06.010

Haak, D. C., HilleRisLambers, J., Pietre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213–1216. https://doi.org/10.1126/science.1204820

Hill, F. (1995). Managing service quality in higher education: The role of the student as primary consumer. *Quality Assurance in Education*, *3*(3), 10–21.

Hora, M. T., Bouwma-Gearhart, J., & Park, H. J. (2017). Data driven decision-making in the era of accountability: Fostering faculty data cultures for learning. *Review of Higher Education*, *40*(3), 391–426. https://doi.org/10.1353/rhe.2017.0013

Horowitz, G. (2019). *Teaching STEM to first generation college students: A guidebook for faculty & future faculty*. Charlotte, NC: Information Age Publishing.

Keeley, S. M. (2014). Coping with student resistance to critical thinking: What the psychotherapy literature can tell us. *College Teaching*, *43*(4), 140–145.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lang, J. M. (2019). How to teach a good first day of class. Retrieved May 2, 2019, from www.chronicle.com/interactives/advice-firstday

Lent, R. W., Sheu, H. Bin, Singley, D., Schmidt, J. A., Schmidt, L. C., & Gloster, C. S. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of Vocational Behavior*, *73*(2), 328–335. https://doi.org/10.1016/j.jvb.2008.07.005

Lewin, J. D., Vinson, E. L., Stetzer, M. R., & Smith, M. K. (2016). A campus-wide investigation of clicker implementation: The status of peer discussion in STEM classes, *CBE—Life Sciences Education*, *15*, ar6. https://doi.org/10.1187/cbe.15-10-0224

Lowe, H., & Cook, A. (2003). Mind the gap: Are students prepared for higher education? *Journal of Further and Higher Education*, *27*(1), 53–76.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

Mervis, J. (2011). Undergraduate science: Weed-out courses hamper diversity. *Science*, *334*(6061), 1333. https://doi.org/10.1126/science.334.6061.1333

Messineo, M., Gaither, G., Bott, J., & Ritchey, K. (2007). Inexperienced versus experienced students' expectations for active learning in large classes. *College Teaching*, *55*(3), 125–133. https://doi.org/10.3200/CTCH.55.3.125-133

Michael, J. (2007). Faculty perceptions about barriers to active learning. *College Teaching*, *55*(2), 42–47, DOI: 10.3200/CTCH.55.2.42-47.

National Center for Science and Engineering Statistics. (2017). *Women, minorities, and persons with disabilities in science and engineering* (Special Report NSF 17-310). Arlington, VA. Retrieved from www.nsf.gov/statistics/wmpd/

Pearson, R. (2016). *Goodman Kruskal: Association analysis for categorical variables*. Retrieved May 2, 2019, from https://cran.r-project.org/package=GoodmanKruskal

Pelletreau, K. N., Knight, J. K., Lemons, P. P., Mccourt, J. S., Merrill, J. E., Nehm, R. H., … & Smith, M. K. (2018). A faculty professional development model that improves student learning, encourages active-learning instructional practices, and works for faculty at multiple institutions. *CBE—Life Sciences Education*, *17*(2), es5. https://doi.org/10.1187/cbe.17-12-0260

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 (pp. 36–38).

RStudio Team. (2015). *RStudio: Integrated development for R*. Boston, MA: RStudio, Inc. www.rstudio.com/

Sander, P., Stevenson, K., King, M., & Coates, D. (2000). University students' expectations of teaching. *Studies in Higher Education*, *25*(3), 309–323. https://doi.org/10.1080/03075070050193433

Sawtelle, V., Brewe, E., & Kramer, L. H. (2012). Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of Research in Science Teaching*, *49*(9), 1096–1121. https://doi.org/10.1002/tea.21050

Seidel, S. B., & Tanner, K. D. (2013). "What if students revolt?"—Considering student resistance: Origins, options, and opportunities for investigation. *CBE—Life Sciences Education*, *12*(4), 586–595. https://doi.org/10.1187/cbe-13-09-0190

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

Shaw, E. J., & Barbuti, S. (2010). Patterns of persistence in intended college major with a focus on STEM majors. *NACADA Journal*, *30*(2), 19–34. https://doi.org/10.12930/0271-9517-30.2.19

Sithole, A., Chiyaka, E. T., McCarthy, P., Mupinga, D. M., Bucklein, B. K., & Kibirige, J. (2017). Student attraction, persistence and retention in STEM programs: Successes and continuing challenges. *Higher Education Studies*, *7*(1), 46. https://doi.org/10.5539/hes.v7n1p46

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & 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*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

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*(4), 624–635. https://doi.org/10.1187/cbe.14-06-0108

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*(1), 55–63. https://doi.org/10.1187/cbe.10-08-0101

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., … & Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, *102*(6), 1178–1197. https://doi.org/10.1037/a0027143

Sultan, P., & Wong, H. Y. (2010). Service quality in higher education—a review and research agenda. *International Journal of Quality and Service Sciences*, *2*(2), 259–272. https://doi.org/10.1108/17566691011057393

Tharayil, S., Borrego, M., Prince, M., Nguyen, K. A., Shekhar, P., Finelli, C. J., & Waters, C. (2018). Strategies to mitigate student resistance to active learning. *International Journal of STEM Education*, *5*(7). https://doi.org/10.1186/s40594-018-0102-y

Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE—Life Sciences Education*, *17*(3), rm2. https://doi.org/10.1187/cbe.17-12-0280

Watkins, B. J., & Mazur, E. (2019). Retaining students in science, technology, engineering, and mathematics (STEM) majors, *42*(5), 36–41.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. http://ggplot2.org

Wieman, C. E. (2014). Large-scale comparison of science teaching methods sends clear message, *Proceedings of the National Academy of Sciences USA*, *111*(23), 8319–8320. https://doi.org/10.1073/pnas.1407304111

Yee, A. (2016). The unwritten rules of engagement: Social class differences in undergraduates' academic strategies. *Journal of Higher Education*, *87*(6), 831–858. https://doi.org/10.1080/00221546.2016.11780889