Look Who's Talking: Teaching and Discourse Practices across Discipline, Position, Experience, and Class Size in STEM College Classrooms

JOURJINA SUBIH ALKHOURI, CRISTINE DONHAM, TÉA S. PUSEY, ADRIANA SIGNORINI, ALEXANDER H. STIVERS, AND PETRA KRANZFELDER

Students are more likely to learn in college science, technology, engineering, and math (STEM) classrooms when instructors use teacher discourse moves (TDMs) that encourage student engagement and learning. However, although teaching practices are well studied, TDMs are not well understood in college STEM classrooms. In STEM courses at a minority-serving institution (MSI; n = 74), we used two classroom observation protocols to investigate teaching practices and TDMs across disciplines, instructor types, years of teaching experience, and class size. We found that instructors guide students in active learning activities, but they use authoritative discourse approaches. In addition, chemistry instructors presented more than biology instructors. Also, teaching faculty had relatively high dialogic, interactive discourse, and neither years of faculty teaching experience nor class size had an impact on teaching practices or TDMs. Our results have implications for targeted teaching professional development efforts across instructor and course characteristics to improve STEM education at MSIs.

Keywords: undergraduate, graduate, teacher discourse moves, COPUS, CDOP

The Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith et al. 2013) and the Classroom Discourse Observation Protocol (CDOP; Kranzfelder et al. 2019a) are two classroom observation tools that allow researchers to assess teaching and discourse practices. A previous study combining COPUS and CDOP results showed that it is possible to create a classroom environment with high student-centered, evidence-based teaching practices (EBTPs) that encourage student learning, but with low *dialogic, interactive* discourse (Kranzfelder et al. 2020). This indicates that even when instructors are engaging in active learning teaching practices, they are still instructing with teacher-centered discourse practices, where they are dominating classroom conversations. However, this previous work only examined biology instructors' classroom teaching and discourse practices in mostly introductory undergraduate biology classes at a research-intensive, predominantly white institution (PWI). Therefore, building on that work in biology classrooms, we wanted to expand our understanding of EBTPs by examining teaching and discourse practices across a range of instructor and course characteristics, such as in different science, technology, engineering, and mathematics (STEM) disciplines, instructor types, years of teaching experience, and class size at a research-intensive, minority-serving institution (MSI).

Classroom discourse

Instructors play a key role in facilitating student engagement through enacted classroom discourse, or the verbal instructor–student and student–student interactions used to construct scientific knowledge and ideas (Cazden 2001, Mortimer and Scott 2003b, Truxaw and DeFranco 2008, Michaels and O'Connor 2015, Wei et al. 2018). Students are more likely to learn in college STEM classrooms when they are encouraged to analyze and challenge questions and work collaboratively in small groups to answer instructors'

https://doi.org/10.1093/biosci/biab077

questions (Ebert-May et al. 1997, Knight and Wood 2005, Gray et al. 2010). Therefore, instructors facilitate student engagement through deliberate actions taken to mediate, participate in, or influence classroom discourse (Krussel et al. 2004, Mercer 2010, Oliveira 2010, Knight et al. 2013). One type of classroom discourse, *teacher discourse moves* (TDMs), are the conversational strategies used by instructors to support student understanding of content knowledge (Warfa et al. 2014, Kranzfelder et al. 2020) and they have been found to foster student learning by engaging students in a deeper understanding of the scientific ideas (Osborne 2010, Berland and Reiser 2011, Kuhn et al. 2017). In contrast, Seidel and colleagues (2015) coined the term *instructor talk* to describe the non-content-related conversational language used by instructors. An example of *instructor talk* would be when an instructor gives instructions for classroom activities or justifications for active learning use (Seidel et al. 2015). Although this type of discourse facilitates overall learning in the classroom, it is different from the content-related discourse that we refer to as TDMs.

Prior work assessing TDMs in primary and secondary STEM classrooms found that the initiate–response–evaluate (IRE) discourse pattern that is focused on fixed communication, was the prevailing form of dialogue between instructors and students (Sinclair and Coulthard 1975, Howe and Abedin 2013). An example of IRE would be an instructor asking a yes or no question (initiate), receiving a yes or no answer from the student (response), and confirming that answer as either correct or incorrect (evaluate). However, the less frequently occurring initiate–response–feedback (IRF) discourse pattern creates opportunities for student–instructor dialogue by generating collaborative discussions. In contrast to IRE, an example of IRF would be an instructor asking a question (initiate), receiving an answer from the student (response), and then prompting the student for follow-up dialogue (feedback). IRF discourse approaches are more effective than IRE in promoting student discussions as they create opportunities for students to develop critical reasoning and argumentation; with the IRF discourse approach, students are asked to think beyond whether their answer is correct or incorrect but, rather, spend more time reasoning through and supporting their answers with evidence (Duschl and Osborne 2002, Jiménez-Aleixandre and Erduran 2007, Duschl 2008). More recently in undergraduate biology classrooms, Kranzfelder and colleagues (2020) found that the less engaging IRE discourse pattern was the most dominant form even when instructors were teaching with student-centered, active learning strategies. For example, an instructor could be moving around and guiding small group and whole class discussions (i.e., student-centered, active learning activity) but providing information to the student by making analogies or connects to students' personal experiences (i.e., teacher-centered, authoritative discourse practice). Such studies are of vital importance as they broaden our understanding of teaching and discourse patterns currently employed in college STEM classrooms.

Classroom observation protocols

Classroom observation protocols are tools that help us measure college STEM teaching practices, especially EBTPs, including active learning (AAAS 2012, Williams et al. 2015). In contrast to self-report surveys or interviews that are predisposed to biases (Ebert-May et al. 2011, Mitchell and Martin 2018, van der Lans 2018), well-developed, reliable classroom observation protocols provide a more objective way of documenting teaching practices in real time or via audio or video recordings (AAAS 2012). Many classroom observation protocols have been developed to look at college STEM teaching practices, including the PORTAAL (Practical Observation Rubric to Assess Active Learning; Eddy et al. 2015), the DART (Decibel Analysis for Research in Teaching; Owens et al. 2017), the RTOP (Reformed Teaching Observation Protocol; Sawada et al. 2002), the COPUS (Smith et al. 2013), and the CDOP (Kranzfelder et al. 2019a). There are differences in how each of these classroom observation protocols measure STEM teaching practices, but a combination of two classroom observation protocols, COPUS (Smith et al. 2013) and CDOP (Kranzfelder et al. 2019a), has been found to provide a holistic view into college STEM classrooms (Kranzfelder et al. 2020).

COPUS is a popular protocol for measuring traditional lecturing versus active learning at department-wide (Kranzfelder et al. 2019b, Reisner et al. 2020), institution-wide (Smith et al. 2014, Lund and Stains 2015, Lund et al. 2015, Lewin et al. 2016, Akiha et al. 2018, Meaders et al. 2019, Tomkin et al. 2019, Denaro et al. 2021), and multiple-institution scales (Stains et al. 2018, Borda et al. 2020, Lane et al. 2021) for education research, faculty teaching professional development (PD), and tenure and promotion purposes. In contrast, CDOP is a new protocol for measuring discourse practices, particularly TDMs, in STEM classrooms with both traditional lecturing and active learning (Kranzfelder et al. 2019a). Kranzfelder and colleagues (2020) showed that even when instructors mostly implemented student-centered, active learning teaching practices, they were not always paired with student-centered TDMs. However, that study had limitations as it only examined the classroom practices of biology instructors teaching in mostly introductory undergraduate biology classes. Therefore, it is essential to investigate teaching practices using COPUS and discourse practices using CDOP across different instructor and course characteristics, including STEM disciplines, instructor types, years of faculty teaching experience, and class size, to expand on previous research and broaden our understanding of what is happening in college STEM classrooms.

Instructor and course characteristics that might affect teaching and discourse practices

Prior studies have found differences in teaching practices as measured by COPUS across STEM disciplines (Lund et al. 2015, Eagan 2016, Stains et al. 2018). First, Lund and colleagues (2015) found that chemistry instructors *lectured* disproportionately more than biology instructors, whereas biology instructors implemented more *peer instruction*, and mathematics instructors used more *collaborative learning*. More recently, Eagan (2016) found that mathematics and engineering instructors consistently used fewer electronic quizzes with immediate feedback and student inquiry to drive learning compared to biology instructors. Finally, Stains and colleagues (2018) found that mathematics instructors used the most *student-centered* instructional style (i.e., instructor used group work strategies consistently), biology instructors used the most *interactive lecture* instructional style (i.e., instructor used some group work strategies), and chemistry instructors used the most *didactic* instructional style (i.e., instructor spent 80% or more of class time lecturing). These studies suggest that different STEM disciplines have different cultures of implementing student-centered EBTPs in their courses.

Instructors' academic positions or instructor types have been shown to influence teaching practices (Bush et al. 2020, Harlow et al. 2020, Xu and Solanki 2020). For example, in the University of California system, there are three main instructor types: tenure-track research faculty, tenure-track teaching faculty (also known as *lecturer with potential security of employment*), and non-tenure-track lecturer (also known as *contingent faculty*, *part-time*, or a *unit-18 lecturer*). Each of these instructor types includes widely different expectations for research, teaching, service, and opportunities for teaching PD (Harlow et al. 2020, Xu and Solanki 2020). For example, tenure-track research faculty are primarily evaluated on the success of their research programs (Brownell and Tanner 2012), and their teaching is generally not an important area for advancement (Figlio et al. 2015). In contrast, tenure-track teaching faculty are expected to spend more time preparing for their classroom instruction and to be more knowledgeable about student-centered EBTPs (Harlow et al. 2020). Finally, lecturers are the predominant instructor type in higher education with teaching expectations but not research or service (Murray 2019). Also, when comparing tenure-track teaching faculty to lecturers, tenure-track teaching faculty tend to have more opportunities for teaching PD and a smaller teaching load than that of lecturers who often teach up to five courses per semester (Adu and Okeke 2014, Murray 2019). It has been well documented that in years subsequent to discipline-based PD there is an improvement in student performance outcomes (Huberman 1994, Horn 2010, Council 2012, Kennedy 2016, Manduca et al. 2017), because PD promotes opportunities for faculty to learn about alternative approaches to teaching (Mizell 2010).

Two more instructor and course characteristics that might affect teaching and discourse practices are years of faculty teaching experience and class size (Dancy and Henderson 2010, Budd et al. 2013). First, it has been shown that novice teachers hold simplistic views of teaching and learning (Putnam and Borko 1997) and have teaching anxiety that diminishes with teaching experience (Keavney and Sinclair 1978), suggesting that they are most likely not incorporating EBTPs into their teaching repertoire. In addition, with teaching experience comes a better understanding of classroom management, which can increase opportunities for involvement and improve communication between instructor and students (Berger et al. 2018). Also, Lund and colleagues (2015) found that more experienced faculty members (i.e., more than 6 years of teaching experience) are in general more interested in implementing and integrate more student-centered EBTPs in their classrooms. Second, the number of students enrolled in a class (i.e., class size) has often been cited as a barrier to implementation of student-centered EBTPs (Gess-Newsome et al. 2003, Henderson and Dancy 2007, Hora 2012). For example, Smith and colleagues (2014) found a significant positive correlation between the percentage of *presenting* information as measured by COPUS and class size (Pearson's $r = .401$, $p < .05$), indicating that instructors who teach larger class sizes tend to *present* information more often. However, Lund and colleagues (2015) found no differences in implementation of student-centered pedagogies across small (1–25 students), medium (25–100 students), and large (more than 100 students) class sizes. In contrast, Stains and colleagues (2018) found that courses with small class sizes do not necessarily implement more student-centered strategies, and Akiha and colleagues (2018) reported that even in small class sizes (30 and below), instructors continue to *present* information, indicating that class size did not affect teaching practices of the instructors in their study context. The previous work described above suggests that class size may or may not influence the implementation of student-centered EBTPs.

Therefore, we wanted to better understand the dynamics of teaching and discourse practices in *all* STEM classrooms at a research-intensive MSI. Specifically, we asked the following three questions: How do teaching practices correlate with discourse practices? Are there differences among STEM instructors with regards to teaching practices and discourse practices? And are there differences in teaching and discourse practices across various instructor and course characteristics, including STEM disciplines, instructor types, years of faculty teaching experience, and class size?

Institution, instructor, and course characteristics

We compared 35 instructors teaching 74 in-person class sessions in undergraduate and graduate STEM courses, including biology, chemistry, mathematics, physics, and engineering, at a mid-size, public, research-intensive university designated as an MSI. Table 1 shows the characteristics of the instructors and their courses. The possible instructor type categories are tenure-track or tenured research faculty (referred to as *research faculty* hereafter), tenure-track or tenured teaching faculty (referred to as *teaching faculty* hereafter), and non-tenure-track contingent faculty (referred to as *lecturers* hereafter). The years of teaching experience is based on the number of years of faculty

Table 1. Demographic characteristics of instructors (n = 35) and their courses (n = 74 class sessions).

Characteristics	n	Percentage
Years of teaching experience		
0–5	14	40.0
6–10	8	22.9
11+	13	37.1
Instructor type		
Research faculty	14	40.0
Teaching faculty	7	20.0
Lecturers	14	40.0
STEM discipline of instructor		
Biology	16	45.7
Molecular and Cellular Biology	(12)	
Life and Environmental Sciences	(2)	
Quantitative Systems Biology	(2)	
Chemistry	9	25.7
Mathematics	4	11.4
Other STEM	6	27.1
Physics	(4)	
Engineering	(2)	
Class size (class sessions)		
Small (≤60 students)	24	32.4
Medium (61–100 students)	6	8.1
Large (>100 students)	43	59.5
Class level (class sessions)		
Lower division	55	74.3
Upper division	14	18.9
Graduate	5	6.8

Note: Some instructors taught more than one course, but demographics and class sessions are included per instructor. Parentheses indicate the number in the subcategory.

teaching experience at the institution of study. The years of teaching experience ranged from 0 to more than 10 years, with 40% of all instructors having more than 10 years of teaching experience. Participating instructors varied across STEM departments, with the majority being in biology, followed by chemistry, mathematics, and other STEM (engineering and physics). In addition, courses were mostly taught by a sole instructor (i.e., not coteaching or team teaching), and the class sizes ranged from 4 to 292 students, with the mean class size being 110 students. Instructors taught mainly lower-division courses that were designated for majors (table 1). Class sessions ranged from 38 to 82 minutes, avoiding class sessions in which the entire meeting time was dedicated to exams, student presentations, or special group project work. However, we included class sessions in which quizzes were given because these are a regular part of the daily or weekly class sessions and only took up to 15 minutes in a 75-minute class session (or 20% of the time).

Instructor recruitment

We sent out an initial recruitment email to research and teaching faculty through faculty department email list serves and individual emails to lecturers in the departments of biology, chemistry, physics, and mathematics. We additionally sent out individual emails to teaching faculty in engineering. This initial email included the purpose of our study, procedures, benefits, IRB approval, potential dissemination of results, classroom observation scheduling information, and contact information for questions. We invited instructors to participate in our study who met the following selection criteria: They taught either an undergraduate or a graduate STEM course, they taught the lecture component of the course—not laboratory or discussion—or they taught the course in-person between two academic years (Fall 2018, Spring 2019, Fall 2019, and Spring 2020 semesters before the COVID-19 global pandemic). Initially, 41 instructors consented to participate in the study; however, two were excluded because of classroom observation scheduling conflicts, two were excluded because of either being a lab or discussion component of the course, and two were excluded as they did not teach in-person in the Spring 2020 semester because of the transition to remote instruction during the COVID-19 global pandemic. We are unable to give the participation rate because the total number of instructors in the email list serves is unknown. The study was classified by the UC Merced Institutional Review Board as exempt (protocol ID no. UCM2020-3).

COPUS data collection

We used COPUS (Smith et al. 2013) to quantify teaching practices observed across instructors and compared them across STEM disciplines, instructor types, years of faculty teaching experience, and class size. COPUS documents teaching practices in 2-minute intervals throughout a class session using 12 individual instructor codes categorized into four collapsed instructor codes adapted from (Smith et al. 2014) and (Kranzfelder et al. 2019b): *presenting, guiding, administering*, and *other*. Individual instructor codes include teaching practices, such as *lecturing*, *posing a question, answering questions*, and *moving and guiding* (Smith et al. 2013). We followed the code descriptions outlined by Smith and colleagues (2013), with the exception that *one-on-one discussions* were coded by observers when the instructor was helping one student or a small group and not paying attention to the rest of the class. Also, *whole-class discussion* was coded when students were leading a discussion, such as an in-class debate or Socratic seminar (supplemental tables S1 and S2).

The live COPUS observations were conducted by 14 Students Assessing Teaching And Learning (SATAL) undergraduate student interns working for the Center for Engaged Teaching and Learning at the institution of study. SATAL interns support faculty and staff's PD by observing their teaching and learning through live COPUS observations, class interviews, and focus groups and provide instructors with actionable feedback (Signorini and Pohan 2019). SATAL interns were trained to conduct COPUS observations in 3 hours by three of the authors (JA, AMS, and PK) according to the training outlined in Smith and colleagues (2013) until moderate interrater reliability (IRR) was established between all coders (κ = .55, 95% confidence interval [CI] = .55–.56; supplemental table S5). Fleiss' kappa statistics of .01–.20 indicate no to slight agreement, .21–.40 indicate fair agreement, .41–.60 indicate moderate agreement, .61–.80 indicate substantial agreement, and .81–1.00 indicate almost perfect agreement (Fleiss 1971). At minimum, two SATAL interns were present in the classroom for each of the live observations. In addition to reaching a moderate IRR during training, SATAL interns would meet for up to 30-minutes after each classroom observation to discuss their codes and resolve any coding disagreements until reaching 100% consensus. By having both a moderate kappa score and consensus building after each classroom observation, the data collected by SATAL interns were considered reliable.

CDOP data collection

We used CDOP (Kranzfelder et al. 2019a) to quantify the discourse practices observed across instructors and compared them across STEM disciplines, instructor types, years of faculty teaching experience, and class size. CDOP documents discourse practices, specifically TDMs, in 2-minute intervals throughout a class session using 17 individual instructor codes such as *sharing*, *real-worlding*, *checking in*, *contextualizing*, and *requesting* into four collapsed instructor codes as is described in the present article (supplemental tables S3 and S4; Mortimer and Scott 2003a, Kranzfelder et al. 2019b):

Authoritative, noninteractive is classroom discourse in which the instructor focuses on their point of view with no student participation opportunities (e.g., lecturing).

Authoritative, interactive is classroom discourse in which the instructor is the main participant but leads students through a question-and-answer routine to consolidate their point of view (e.g., lecturing with IRE-type questions).

Dialogic, interactive is classroom discourse in which both the instructor and students participate. Here, the instructor listens and responds to student discourse, and students benefit from the teacher's guidance (e.g., whole-class discussion with IRF-type questions).

Other was used when a TDM was observed, but no identifiable codes fit.

For CDOP analysis, we collected audio recordings for each of the instructors using either a Sony HDR camcorder with a microphone or a Swivl with a remote marker and an Apple iPad. We listened to audio recordings while using the CDOP to quantify the TDMs used by instructors. One coder (JA) was trained for 3 hours by the corresponding author (PK), whereas two coders (CD and AHS) were trained by the first author (JA) according to the training outlined in Kranzfelder and colleagues (2019a) until substantial IRR was established between all four coders (κ = .79, 95% CI .72–.86; table S6). Over several months, three coders (JA, CD, and AHS) independently coded all of the audio recordings as first coders, whereas two coders (PK and JA) served as second coders for over 25% of the recordings (κ = .83, supplemental table S7). If the average kappa score was below .6, then the coders met to discuss the discrepancies until 100% consensus was reached.

Data analyses

Following Lewin and colleagues (2016), Meaders and colleagues (2019), and Kranzfelder and colleagues (2020), we analyzed the COPUS and CDOP individual codes using the percentage of 2-minute time intervals to determine and compare the frequency of a particular code. We divided the number of 2-minute time intervals marked for each code (e.g., *sharing*) by the total number of 2-minute time intervals for that class session. For example, if *sharing* was marked 20 of the 2-minute time intervals out of a possible 30 2-minute time intervals (i.e., 60-minute class session), then 20/30 or 66.7% of the possible 2-minute time intervals contained *sharing*. This calculation overestimates the amount of time an instructor spends on any one behavior as the behavior is counted for the entire 2-minute time interval even if the instructor only spends 10 seconds on it.

Similar to Smith and colleagues (2014), Lewin and colleagues (2016), and Kranzfelder and colleagues (2020), we also analyzed the COPUS and CDOP collapsed data using the percentage of codes to get a more holistic view of multiple codes and compare across broad teaching and discourse practices. In addition, we analyzed COPUS and CDOP collapsed data by the percentage of codes to determine differences across STEM disciplines, instructor types, years of faculty teaching experience, and class sizes. More specifically, we added the total number of times each code was marked and divided it by the total number of codes. For example, if *sharing* was marked 20 times and there were 50 codes in total, then *sharing* would correspond to 20/50 or 40% of the total codes. This calculation slightly underestimates the amount of time an instructor spends on any one behavior as it counts the behavior relative to all other behaviors.

We categorized our data to quantify how teaching and discourse practices differed among instructors' STEM discipline, instructor type, years of teaching experience, and class size. We made categories on the basis of samples with at least 10 class sessions for all four variables. First, we divided the STEM disciplines into four categories: biology (i.e., molecular and cellular biology, quantitative and systems biology, and life and environmental sciences), chemistry, mathematics, and other STEM (engineering and physics). We grouped instructors who taught life sciences courses into biology and grouped engineering and physics into other STEM. Second, we divided instructor types into three categories following categorization from Xu and Solanki (2020): research faculty, teaching faculty, and lecturers. Third, following Lund and colleagues (2015), we divided the years of faculty teaching

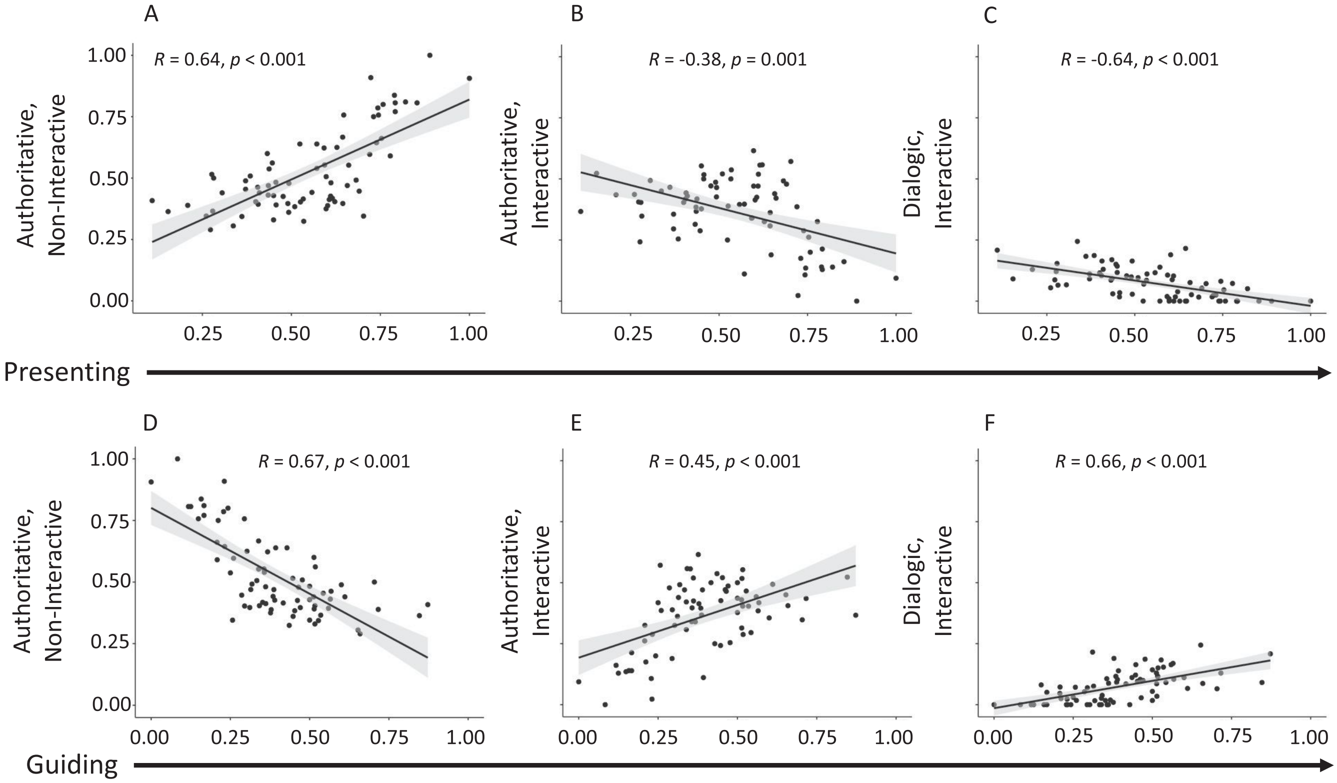

Figure 1. Three discourse approaches (i.e.,* authoritative, noninteractive*;* authoritative, interactive*; and* dialogic, interactive*) percentage of codes on the* y*-axis, in response to teaching practices (i.e., presenting and guiding) percentage of codes on the* x*-axis. Scatter plots and best-fit lines are shown with Spearman's correlation coefficient (ρ) and* p*-value for (a)* presenting *as a function of* authoritative, noninteractive*; (b)* presenting *as a function of* authoritative, interactive*; (c)* presenting *as a function of* dialogic, interactive*; (d)* guiding *as a function of* authoritative, noninteractive*; (e)* guiding *as a function of* authoritative, interactive*; and (f)* guiding *as a function of* dialogic, interactive. *The shaded area shows the 95% confidence intervals for the best-fit line.

experience on the basis of the number of years teaching as the instructor of record at this institution of study into three categories: 0–5 years, 6–10 years, and 11 or more years. In contrast to Lund and colleagues (2015), we determined the range of the categories to ensure that there was a large enough sample size for each of the three categories. Fourth, we divided the class size (or the number of students per class) into two categories: small (up to 60 students) and medium (61–100 students) together and large (more than 100 students).

Statistical analyses

To determine whether there were relationships between teaching and discourse practices across instructors, we calculated nonparametric Spearman's rank correlation tests. More specifically, we correlated two COPUS collapsed instructor codes (*presenting* and *guiding*) to three CDOP collapsed codes (*authoritative, noninteractive*; *authoritative, interactive*; and *dialogic, interactive*). We explored the relationships of *presenting* and *guiding* to the three discourse approaches as these teaching practices create opportunities for conversations between instructors and students around content.

To determine whether there were differences in the teaching and discourse practices across instructors, we calculated nonparametric Friedman tests as it does not assume a normal distribution, and post hoc pairwise comparisons using Wilcoxon signed-rank tests with Bonferroni corrections. In addition, we used Kendall's *W* test for calculating effect size, which uses the Cohen's interpretation guidelines of .1–.3 for a small effect, .3–.5 for a moderate effect, and greater than .5 for a large effect (Cohen 1988, Tomczak and Tomczak 2014).

To determine whether there were differences between instructional and discourse practices by STEM discipline of the course, instructor type, years of teaching experience, and class size, first, we calculated a nonparametric aligned ranks transformation ANOVA (Wobbrock et al. 2011) with the ARTool package in R (Kay and Wobbrock 2020). Second, we calculated post hoc pairwise comparisons with Bonferroni corrections. And finally, we calculated the partial eta-squared measure (η_p^2) for calculating effect size, which uses .01–.06 to indicate a small effect, .06–.14 to indicate a moderate effect, and greater than.14 to indicate a large effect (Cohen 1988, Tomczak and Tomczak 2014). All statistical analyses were conducted using the R statistical software

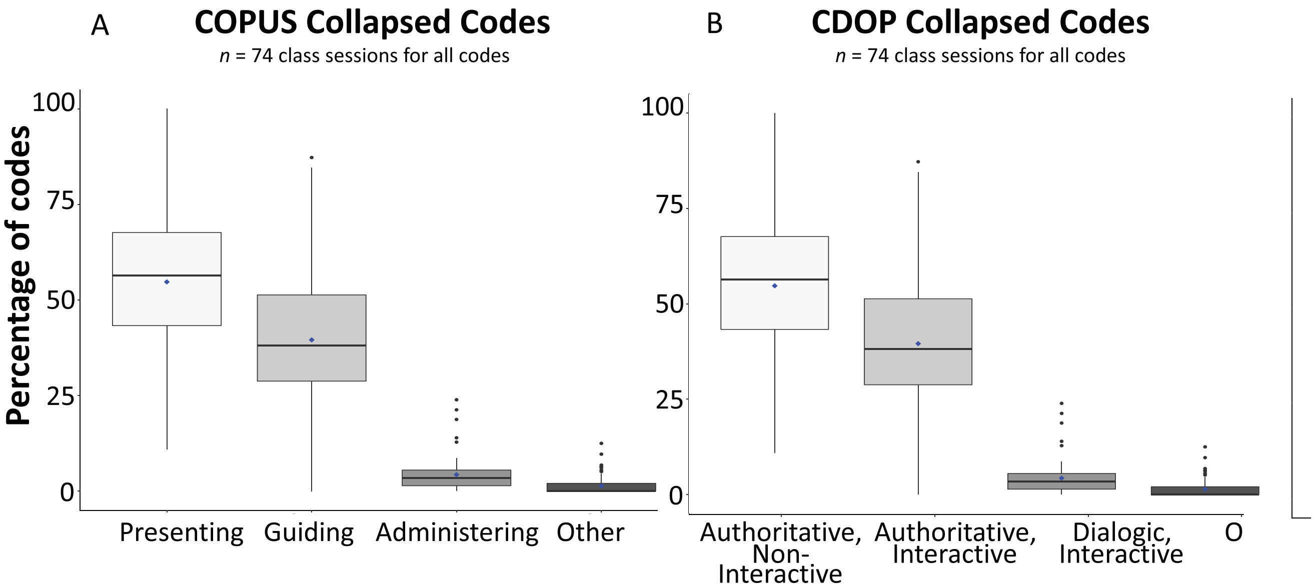

Figure 2. Box-and-whisker plots showing the percentage of codes that instructors spent on different teaching practices (a) and discourse practices (b) across 74 STEM class sessions. The boxes represent the interquartile range (IQR) of practices for each collapsed code, whiskers represent the largest and smallest values within 1.5 times the IQR, lines within each box represent the median, the blue diamond represents the mean, and the black dot represents the outliers.

(R Core Team 2020) and the significance threshold (α) was set at .05 for all tests. As was suggested by Wasserstein and colleagues (2019), we presented the actual p-value unless it is less than .001, which we presented as $p < .001$.

Correlations between teaching (COPUS) and discourse practices (CDOP) used by STEM instructors

We correlated two COPUS collapsed codes to three CDOP collapsed codes and found significant associations between all six pairs of variables ($p < .001$; figure 1, supplemental tables S12 and S13). We found that *presenting* positively correlated with *authoritative, noninteractive* ($\rho = .64$; figure 1a), but negatively correlated with *authoritative, interactive* ($\rho = -.38$, figure 1b) and dialogic*, interactive* ($\rho = -.64$, figure 1c). In contrast, *guiding* negatively correlated with *authoritative, noninteractive* ($\rho = -.67$, figure 1d), but positively correlated with *authoritative, interactive* ($\rho = .45$, figure 1e) and dialogic*, interactive* ($\rho = .66$, figure 1f). This suggests that *presenting* teaching practices and *authoritative, noninteractive* discourse practices were commonly implemented together, whereas *guiding* and *authoritative, interactive* and *dialogic, interactive* were commonly implemented together (figure 1).

Broad teaching practices used by STEM instructors (COPUS)

We used COPUS collapsed codes to quantify broad teaching practices of our STEM instructors and found that they were mainly *presenting* information to students (e.g., *lecturing*), but also *guiding* students in active learning activities (e.g., moving around and facilitating small group or whole-class discussion; $\chi^2(3) = 189$, $p < .001$, $W = .85$). More specifically, STEM instructors were spending significantly more of their class time *presenting* information to students (mean [M] = 55%, range of 11%–100% across all class sessions) than *guiding* students in active learning activities (M = 40%, range of 0%–87% across all class sessions). Finally, STEM instructors were spending significantly less class time *administering* (M = 4%) and *other* teaching practices (M = 1%; figure 2a, supplemental table S8).

Broad discourse practices used by STEM instructors (CDOP)

We used CDOP collapsed codes to quantify the broad discourse practices of our STEM instructors and found that they were mainly using *authoritative* discourse approaches (i.e., only lecturing or lecturing with IRE-type questions) and spent significantly less time on *dialogic* discourse approaches (i.e., the instructor asks students to talk about content; $\chi^2(3) = 175$, $p < .001$, $W = .79$). For example, *authoritative* discourse practices were eleven times more likely to occur than dialogic ones. More specifically, STEM instructors spent significantly more of their class time using *authoritative, noninteractive* discourse practices (M = 53%, range of 29%–100% across all class sessions) compared to *authoritative, interactive* discourse practices (M = 36%, range of 0%–62% across all class sessions), *dialogic, interactive* discourse practices (M = 7%, range of 0%–24% across all class sessions), and *other* (i.e., *no content discourse*) discourse practices (M = 4%, range of 0%–22% across all class sessions; figure 2b, supplemental table S10).

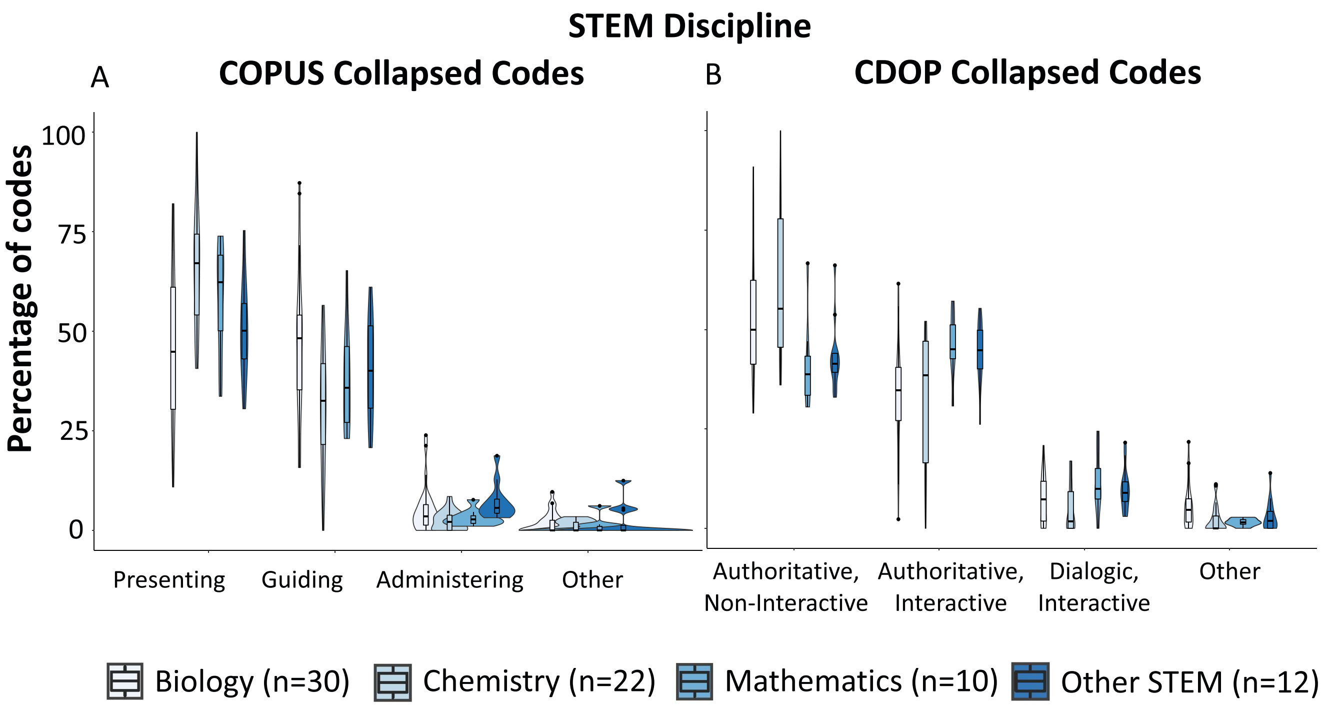

Figure 3. Violin and box-and-whiskers plots show the percentage of codes that instructors spent on different teaching practices (a) and discourse practices (b) across STEM disciplines, including biology, chemistry, mathematics, and other STEM. The violin represents the density of the code frequency. The boxes represent the interquartile range (IQR) of practices for each collapsed code, whiskers represent the largest and smallest values within 1.5 times the IQR, lines within each box represent the median, and the black dot represents the outliers.

Teaching (COPUS) and discourse (CDOP) practices across STEM disciplines

We found significant differences in the teaching practices across STEM disciplines ($F(9,280) = 4.85$, $p < .001$, $\eta_p^2 = .13$). Looking at individual STEM disciplines and COPUS codes, the average percentage use of the two different teaching practices and their ranges for the different disciplines were, for *presenting*, biology, $M = 47\%$, range = 11%–82%; chemistry, $M = 66\%$, range = 41%–100%; mathematics, $M = 58\%$, range = 34%–74%; other STEM, $M = 51\%$, range = 31%–75%, and for *guiding*, biology, $M = 46\%$, range = 16%–87%; chemistry, $M = 31\%$, range = 0%–57%; mathematics, $M = 38\%$, range = 23%–65%; and other STEM, $M = 41\%$, range = 21%–61%.

Overall, we found that chemistry instructors used significantly more *presenting* than biology instructors ($p = -.005$), whereas they used significantly less *guiding* than biology instructors ($p = .04$; figure 3a, supplemental tables S14 and S15).

Similarly, we found significant differences in discourse practices across STEM disciplines ($F(9,280) = 3.25$, $p < .001$, $\eta_p^2 = .09$). The average percentage use of the three different discourse practices, with ranges in parentheses, by different disciplines were, for *authoritative, noninteractive*, biology, $M = 51\%$, range = 29%–84%; chemistry, $M = 61\%$, range = 36%–100%; mathematics, $M = 50\%$, range = 31%–91%; and other STEM, $M = 43.7\%$, range = 33%–66%; for *authoritative, interactive*, biology, $M = 36\%$, range = 11%–62%; chemistry, $M = 32\%$, range = 0%–52%; mathematics, $M = 38\%$, range = 2%–57%; and other STEM, $M = 43\%$, range = 26%–55%; and for *dialogic, interactive*, biology, $M = 7\%$, range = 0%–22%; chemistry, $M = 5\%$, range = 0%–17%; mathematics, $M = 10\%$, range = 0%–24%; and other STEM, $M = 10\%$, range = 3%–22%.

Overall, we found that chemistry instructors used significantly more *authoritative, noninteractive* discourse than other STEM disciplines ($p < .001$; figure 3b, supplemental tables S16, and S17).

Teaching (COPUS) and discourse (CDOP) practices across instructor types

We found significant differences in the teaching practices across instructor types ($F(6,284) = 2.48$, $p = .02$, $\eta_p^2 = .05$), but significances were lost after Bonferroni corrections (figure 4a, supplemental tables S18 and S19). The average percentage use of the two different teaching practices and their ranges by different faculty types—research faculty, teaching faculty, and lecturers—were, for *presenting*, research faculty, $M = 54\%$, range = 15%–82%; teaching faculty, $M = 46\%$, range = 11%–69%; and lecturers, $M = 60\%$, range = 28%–100% and for *guiding*, research faculty, $M = 42\%$, range = 15%–84.6%; teaching faculty, $M = 47\%$, range = 29%–87%; and lecturers, $M = 34\%$, range = 0%–70%.

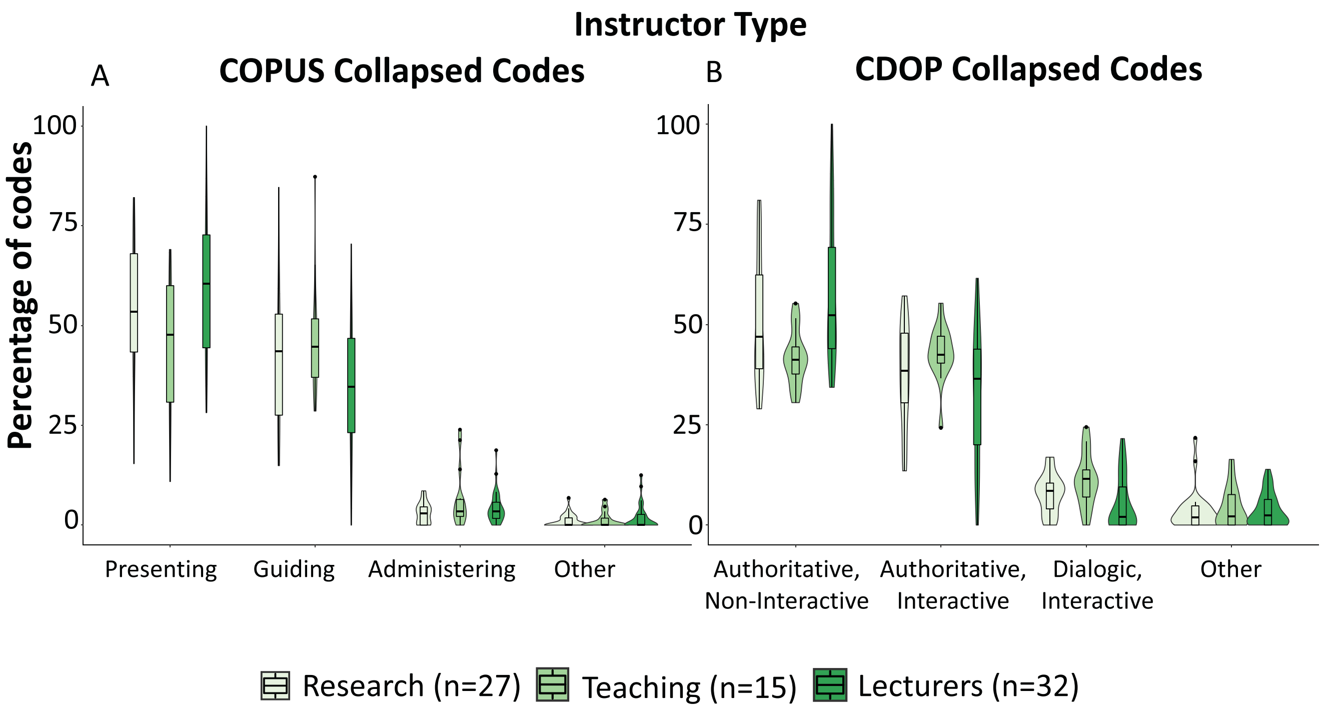

Figure 4. Violin and box-and-whisker plots showing the percentage of codes used by instructor types for teaching practices (a) and discourse practices (b). The violin represents the density of the code frequency. The boxes represent the interquartile range (IQR) of practices for each collapsed code, whiskers represent the largest and smallest values within 1.5 times the IQR, lines within each box represent the median, and the black dot represents the outliers.

Similarly, we found significant differences between instructor type and discourse practices on the percentage of codes ($F(6,284) = 5.554$, $p < .001$, $\eta_p^2 = .11$). The average percentage use of the three different discourse practices and their ranges by different instructor types—research faculty, teaching faculty, and lecturers—were, for *authoritative, non-interactive*, research faculty, M = 52%, range = 29%–81%; teaching faculty, M = 42%, range = 31%–55%; and lecturers, M = 58%, range = 34%–100%; for *authoritative, interactive*, research faculty, M = 37%, range = 13%–57%; teaching faculty, M = 43%, range = 24%–55%; and lecturers, M = 33%, range = 0%–62%; and for *dialogic, interactive*, research faculty, M = 8%, range = 0%–17; teaching faculty, M = 11%, range = 0%–24%; and lecturers, M = 5%, range = 0%–22%.

Overall, we found that teaching faculty used significantly less *authoritative, noninteractive* than lecturers (p = .004; figure 4B, supplemental tables S20 and S21).

Teaching (COPUS) and discourse (CDOP) practices across years of faculty teaching experience

We did not find significant differences in the teaching practices across years of faculty teaching experience ($F(6,284) = 0.76$, $p = .6$, $\eta_p^2 = .05$ (figure 5a, supplemental tables S22 and S23). Similarly, for CDOP, we did not find significant differences in the discourse practices across years of faculty teaching experience ($F(6,284) = 1.06$, $p = .38$, $\eta_p^2 = .02$; figure 5b, supplemental tables S24 and S25).

Teaching (COPUS) and discourse (CDOP) practices across class size

We did not find significant differences in the teaching practices across class size ($F(3,288) = 0.11$, $p = .95$, $\eta_p^2 < .001$; figure 6a, supplemental tables S26 and S27). Similarly, we did not find significant difference in the discourse practices across class size ($F(3,288) = 0.43$, $p = .73$, $\eta_p^2 < .001$; figure 6b, supplemental tables S28 and S29).

Variation in teaching and discourse practices across STEM instructors

The way instructors guide students' engagement can foster student learning with a deeper understanding of scientific ideas (Osborne 2010, Berland and Reiser 2011, Kuhn et al. 2017), and classroom observations can help us understand how instructors are implementing these active engagement practices (Williams et al. 2015). Prior studies have investigated STEM teaching practices across different instructor and course characteristics, such as STEM discipline, course level, class size, classroom physical layout, and faculty teaching experience (Lund et al. 2015, Akiha et al. 2018, Stains et al. 2018). However, discourse practices were only investigated on biology instructors teaching in mostly introductory undergraduate biology classes at a PWI (Kranzfelder et al. 2020). Therefore, for this study, we investigated at an MSI how teaching practices correlate with discourse practices, which teaching and discourse

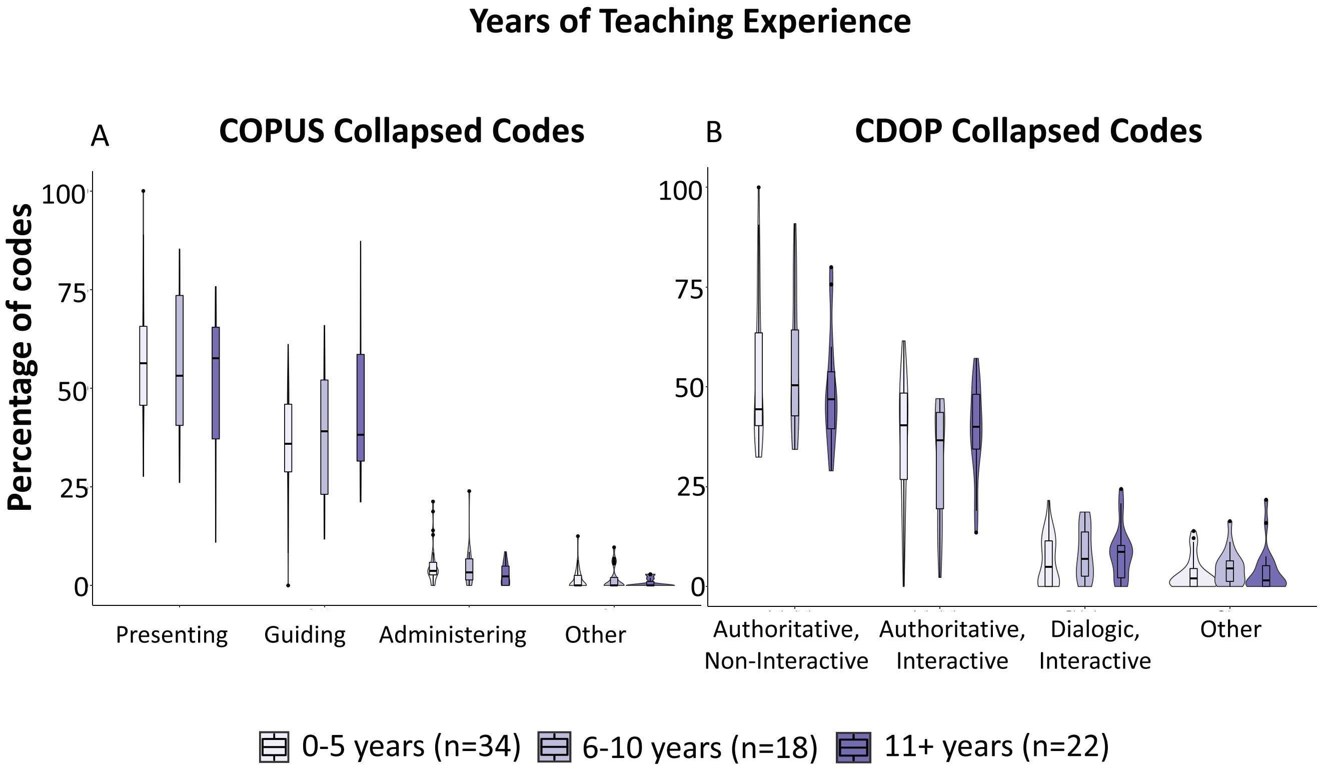

Figure 5. Violin and box-and-whisker plots showing the percentage of codes by instructors' years of teaching experience for teaching practices (a) and discourse practices (b). The violin represents the density of the code frequency. The boxes represent the interquartile range (IQR) of practices for each collapsed code, whiskers represent the largest and smallest values within 1.5 times the IQR, lines within each box represent the median, and the black dot represents the outliers.

practices are used across instructors, and how teaching and discourse practices vary across several instructor and course characteristics. Our findings suggest that instructors teaching in college STEM classrooms are mostly using teacher-centered, *authoritative* discourse practices with differences across disciplines and instructor types at a research-intensive MSI.

Presenting is associated with *authoritative*, *noninteractive* discourse, whereas guiding is associated with *dialogic*, *interactive* discourse. First, we correlated teaching and discourse practices. We found that *presenting* and *authoritative, noninteractive* were positively correlated to each other, whereas *guiding* was positively correlated to both *interactive* discourse practices. This indicates that when STEM instructors used teacher-centered pedagogies, such as lecturing or showing a video, they are most likely the dominant voice being heard in the classroom (i.e., *authoritative*). For example, when an instructor is *presenting* content material by mainly lecturing, they dominate the conversations and discuss only their point of view, employing the *authoritative, noninteractive* discourse approach. This magnifies the issue of inclusion in our STEM classrooms, because students traditionally underrepresented in the sciences may not voice their misconceptions or questions when an instructor dominates the conversation. In contrast, students of privileged ethnicities tend to voice their misconceptions and questions regardless of an instructors' teaching style (Ochoa and Pineda 2008). In addition, Myers and Rocca (2000) discuss how a "dominant and contentious" communication style leaves students with a negative impression and can adversely affect the student experiences. Conversely, when an instructor is *guiding* students in mainly active learning activities, then they are most likely providing opportunities for the students' point of view and voice to be heard in the classroom and creating opportunities for students to develop their content ideas (i.e., *dialogic, interactive* discourse practices). Fassinger (1996) conducted a study at a Midwestern liberal arts college and found that students' perceptions and peer dynamics influence their participation; however, instructors play a key role in allowing such participation and student discussions by either controlling the activities and conversations (similar to *presenting* in an *authoritative* manner) or involving students in the learning process (similar to *guiding* in a *dialogic* manner). Therefore, promoting both student-centered teaching practices (i.e., *guiding*) and student-centered discourse practices (i.e., *dialogic, interactive*) can promote more student involvement and create equitable and inclusive learning environments that serve all students.

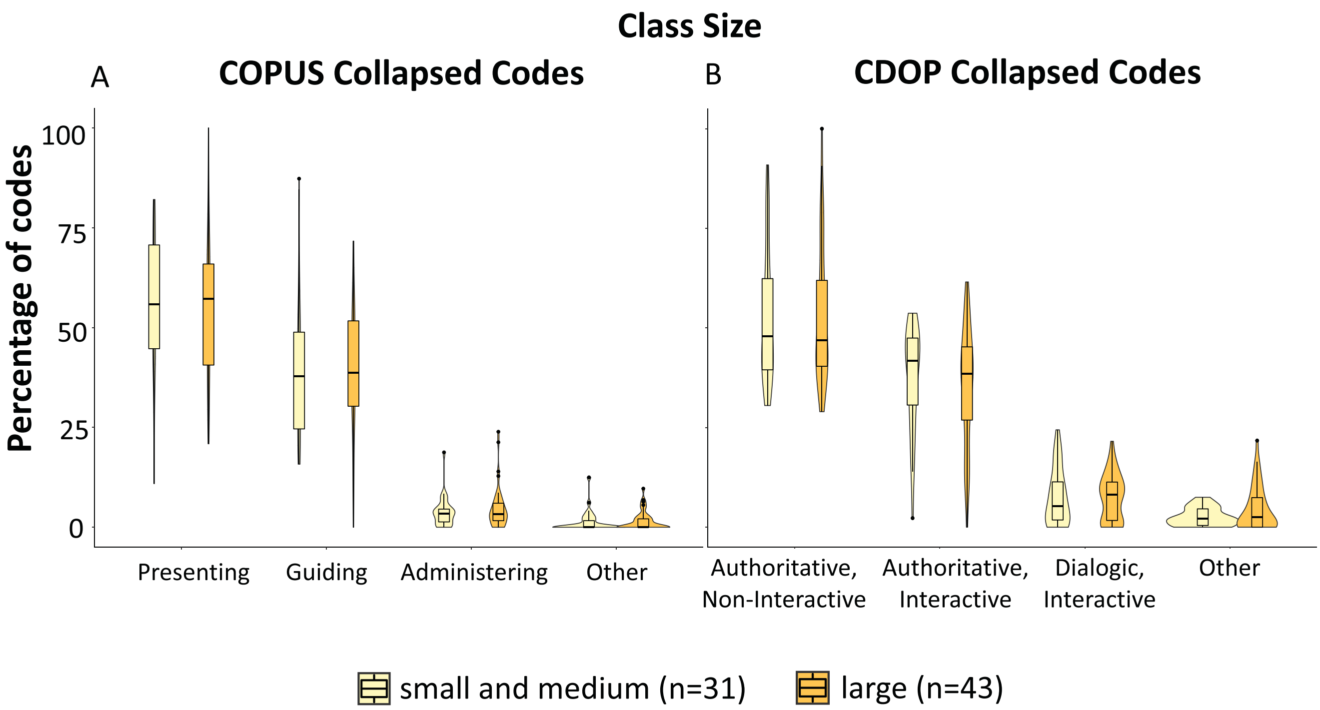

Figure 6. Violin and box-and-whisker plots showing the percentage of codes used by instructors with respect to teaching practices (a) and discourse practices (b) in varying class sizes. The violin represents the density of the code frequency. The boxes represent the interquartile range (IQR) of practices for each collapsed code, whiskers represent the largest and smallest values within 1.5 times the IQR, lines within each box represent the median, and the black dot represents the outliers.

Instructors used mostly *presenting* and *authoritative*, *noninteractive* practices in their college STEM classes. Second, we examined teaching and discourse practices across all STEM instructors. We found that instructors across all STEM disciplines primarily used teacher-centered teaching practices, such as *presenting* information to students, and teacher-centered discourse practices, such as activating prior course concepts or knowledge (*linking*) or priming future course concepts or knowledge (*forecasting*). This trend is prevalent despite evidence suggesting that active learning teaching practices (Freeman et al. 2014) and engaging students in dialogic discourse can promote student learning (Duschl and Osborne 2002, Krussel et al. 2004, Jiménez-Aleixandre and Erduran 2007, Duschl 2008, Freeman et al. 2014). In addition, student-centered pedagogies, such as *guiding* teaching practices and *dialogic, interactive* discourse practices, could have the opportunity to narrow the achievement gap for underrepresented students in STEM fields (Gavassa et al. 2019, Theobald et al. 2020). However, prior studies suggest that implementing student-centered EBTPs continues to remain low (Henderson et al. 2012b) and college STEM classes are still largely being taught using traditional lecturing, not active learning (Stains et al. 2018). Therefore, our findings were consistent with previous studies showing that teacher-centered discourse patterns are the most prevalent in both K–12 classrooms (Sinclair and Coulthard 1975, Howe and Abedin 2013) and college biology classrooms (Kranzfelder et al. 2020).

***Presenting* and *authoritative*, *noninteractive* dominated teaching practices and TDMs across STEM disciplines.** Third, we expanded our understanding of teaching and discourse practices across a range of instructor and course characteristics, including different STEM disciplines, instructor types, years of faculty teaching experience, and class size. We found differences in teaching and discourse practices across STEM disciplines, including biology, chemistry, and other STEM, similar to other studies (Grossman and Stodolsky 1995, Breslyn and McGinnis 2012, Freeman et al. 2014). When we analyzed the teaching and discourse practices across these disciplines, we found that chemistry instructors *presented* more than biology instructors and employed *authoritative, noninteractive* discourse more than other STEM instructors. Looking at the average use of teaching and discourse practices, we found that although biology instructors spent almost half of their class session *guiding* students in active learning activities, their discourse was mostly *authoritative*, not *dialogic*. Our findings are supported by recent studies showing that chemistry instructors lectured more than biology instructors who implemented more peer instruction and collaborative learning (Lund and Stains 2015, Lund et al. 2015) and student-centered instructional styles (Stains et al. 2018). In addition, it has been shown that biology instructors tend to implement more inquiry-based learning (Edelson et al. 1999, Spronken-Smith and Walker 2010) and team-based learning (Michaelsen and Sweet 2008, Leupen et al. 2020), which can promote student scientific investigation and student learning (Breslyn and McGinnis 2012). In addition, some studies have found that chemistry instructors spend more time focusing on content knowledge and student misconceptions and less time on instructional delivery and discourse (Thiele and Treagust 1994, Van Driel et al. 2002, Breslyn and McGinnis 2012, Lund and Stains 2015). For example, Lund and Stains (2015) found that chemistry instructors were somewhat more likely to believe that "teaching with new instructional methods will limit content coverage." Also, these patterns may be due to chemistry instructors employing the same teaching techniques that they received while they were students (Galbraith and Shedd 1990). Our findings of differences across STEM disciplines suggest that PD should be tailored to the specific needs (i.e., either more training on student-centered teaching practices or discourse practices) of the discipline or department.

Teaching faculty used less *authoritative*, *noninteractive* discourse than lecturers. We found that teaching faculty used less *authoritative, noninteractive* discourse practices than lecturers. However, we did not find significant differences in *authoritative, interactive* or *dialogic, interactive* discourse practices between instructor types. These findings are not surprising on the basis of the roles and expectations of the three studied instructor types. Xu and Solanki (2020) recently described teaching faculty as tending to have more teaching PD opportunities, lighter teaching loads, and more consistency in courses taught from one term to the next when compared to lecturers; therefore, teaching faculty might have more time and opportunities to learn about and implement student-centered practices. Generally, lecturers have relatively low compensation, minimal benefits, limited participation in departmental decisions, and lack of job security, leading to low supports and incentives for PD to improve their teaching skills and practices (Umbach and Wawrzynski 2005, Bettinger and Long 2010, Xu and Solanki 2020). Taken together, we conclude that although we did not find significant differences across instructor types with teaching practices after performing Bonferroni corrections, we see that on average, teaching and research faculty *guide* their students through active learning activities, and teaching faculty tend to involve students in the conversations, especially using *authoritative, interactive* discourse. From our findings, we suggest providing equitable institutional PD supports, incentives, and opportunities to all three instructor types, which may increase the implementation of student-centered EBTP in the classroom.

Years of faculty teaching experience did not affect teaching or discourse practices. We found that instructors' years of faculty teaching experience did not affect teaching or discourse practices. On the basis of findings in prior studies (Keavney and Sinclair 1978, Hoy and Spero 2005, Lund et al. 2015, Berger et al. 2018), this was somewhat surprising to us as we expected the instructors with the most teaching experience (i.e., 11 or more years) to employ more student-centered teaching and discourse practices. Therefore, we expected more experienced instructors might increase student participation as a result of having greater confidence in their knowledge, skills, and practices, and that could have allowed them to provide more opportunities for student involvement. A possible explanation to why our results do not reflect what has been observed in other studies could be due to lack of buy-in (Patrick et al. 2016), professional identity of the instructors (Brownell and Tanner 2012), or perceived student resistance to active learning strategies (Finelli and Borrego 2020). Moreover, other studies have found other resource and time barriers to implementing active learning, such as lack of time for preparations of class material and in-class active learning activities, lack of technology that supports in-class active learning, lack of training, lack of incentives, and lack of administrative support (Henderson et al. 2010, Anderson et al. 2011, Patrick 2020). Implementing active learning in STEM classrooms requires buy-in, resources, and time from instructors; therefore, if they are not supported in their implementation of active learning, then they are less likely to implement it regardless of how long they have taught at the institution. Despite the lack of significant differences, instructors in our study had a wide range of years of teaching experience within each category, but they are all predisposed to their own beliefs, knowledge, and skills. For example, two faculty with 6 years of teaching experience might have different pedagogical beliefs, knowledge, and skills and, therefore, may implement active learning to varying degrees. Our findings suggest that instructors of varying years of teaching experience may benefit from more PD opportunities and being incentivized by their departments to participate in these opportunities (e.g., teaching awards), potentially leading to more implementation of student-centered EBTPs.

Class size did not affect teaching or discourse practices. We found that neither teaching practices nor discourse practices differed across class sizes. This is in contrast to previous studies that have cited class size as a barrier to faculty's implementation of student-centered EBTPs (Gess-Newsome et al. 2003, Henderson and Dancy 2007, Hora 2012, Smith et al. 2014, Lund and Stains 2015, Lund et al. 2015, Shadle et al. 2017, Akiha et al. 2018, Stains et al. 2018). For example, Lund and colleagues (2015) found statistically significant differences in instructional styles on the basis of class size and Lund and Stains (2015) found that 100% of their biologists self-reported class size dictating their teaching methods. Therefore, although it is promising that we did not find differences across class sizes, there needs to be a shift in faculty perception about the influence of class size on implementation of EBTPs to allow for more active learning opportunities in large class sizes.

Recommendations for college STEM administrators and instructors. First, we recommend that institutions provide department- or discipline-specific teaching PD on student-centered teaching or discourse practices. It has been well documented that PD brings multilayered improvements in instructional practices when it is department- or discipline-specific (i.e., biology and chemistry), and not just "good teaching remedies," such as implementing active learning techniques for engaging students (Henderson et al. 2012a). Our findings, similar to Lund and Stains (2015), highlighted differences across STEM disciplines, suggesting the importance of not treating all departments and disciplines identically when reform efforts or training occurs at an institution. Some Centers for Teaching and Learning, such as our own institution, offer opportunities for classroom observations using COPUS to help instructors visualize their teaching practices. However, in addition to COPUS peer observations, we recommend using CDOP to help instructors assess their instructional discourse by providing a baseline for instructors to reflect on their TDMs. We suggest that faculty discuss their CDOP results with each other to reflect on their patterns of questioning students and encourage one another to incorporate more *dialogic, interactive* discourse moves. A simple way of moving toward *dialogic, interactive* discourse in the classroom is to ask students to evaluate each other's ideas (e.g., *challenging*) as this encourages students to think about concepts and *challenge* each other's answers. This is aligned with most institution's mission to support student academic success by allowing them to be involved in the learning process.

Second, we recommend that college instructors across all STEM fields take advantage of institutional pedagogical PD to learn how to apply EBTPs in their classrooms. More specifically, we recommend that departments incentivize these pedagogical trainings to improve instructional and discourse practices. We make this suggestion as college STEM instructors may be more willing to participate in such pedagogical trainings if measures of teaching effectiveness, such as teaching practices based on COPUS observations, are evaluated as part of the tenure and promotion process (Henderson et al. 2011, Brownell and Tanner 2012, Stains et al. 2018, Kranzfelder et al. 2019b). Also, STEM departments can affect faculty's beliefs and motivations and promote changes to teaching culture by valuing both contributions to teaching and research equally during evaluation (Herman et al. 2018).

Third, we encourage faculty to create faculty learning communities (FLCs) or communities of practice (COP) to adopt a new belief system that values teaching and to establish long-term collaborations between faculty supporting each other in the use of active learning (Wenger 1998, Kezar et al. 2017, Herman et al. 2018, Tomkin et al. 2019). FLCs or COP are usually attended by those faculty interested in advancing their pedagogical skills as participation is voluntary and no certifications are awarded (Weaver et al. 2016). Recently, Tomkin and colleagues (2019) found that COP are particularly effective when they consist of small, disciplinary teams working on the same courses and all using EBTPs. Therefore, we recommend that chemistry instructors teaching large enrollment introductory chemistry courses work together in their pedagogical reform efforts.

Limitations and future directions

We acknowledge that although our study aimed to propel college STEM education forward into research-based practices, we have limitations that we hope to address in future studies. These limitations include limited generalizability due to only studying STEM instructors at one, not multiple institutions, a limited ability to measure differences in student learning across different discourse approaches using only classroom observation data, a limited ability to measure the impacts of instructor and course characteristics on practices with small sample size, and a limited ability to measure how PD effects teaching and discourse practices. First, this study was performed at only one higher education institution. All instructors shared the same resources and were under the same leadership; therefore, expectations were uniform. It may be beneficial to conduct a similar study across multiple institutions to paint a more detailed picture of instructional and discourse practices in higher education. Second, the classroom observation protocols employed in our study documented the presence or absence of teaching or discourse practices and do not touch on cognitive student engagement or student performance outcomes. Although student-centered teaching practices, such as implementing active learning, are associated with improved student performance outcomes (Freeman et al. 2014), we did not measure these outcomes. In the future, we would like to collect student learning gains through concept inventories, such as GenBio-MAPS for general biology (Couch et al. 2019), to further investigate the impact of different discourse approaches on student learning. For example, does *dialogic, interactive* discourse lead to improved student learning gains? Third, with a larger sample size across various institutions, we would like to measure the impacts of different instructor and course characteristics on teaching or discourse practices. An interesting observation we saw was that biology instructors implemented *presenting* teaching practices with a wide range (10.9%–82.1%). In the future, we would like to study which instructor and course characteristics, such as gender, teaching experience, instructor type, and class size, are more likely to implement active learning pedagogies broadly, and within disciplines. Fourth, we did not study the effects of discipline- or department-specific PD on instructional and discourse practices. It has been shown that PD improves instructional practices (Henderson et al. 2012a); therefore, it would be interesting to investigate access to PD and the effects of PD across various variables and how that affects teaching and discourse practices. For example, would a non-tenure-track chemistry lecturer with over 10 years of teaching experience have different teaching or discourse practices if they had more access to teaching PD? Taken together, our work shines a light on teaching and discourse practices in college

STEM classrooms, and we hope that both STEM educators and STEM education researchers focus not just on *what* instructors and students are doing in classrooms, but *who* and *how* instructors and students are talking about science content in these classrooms.

Acknowledgments

We would like to thank all the instructors who welcomed us into their classrooms. We would also like to thank all the student interns for their contributions to COPUS and CDOP data collection: Sandy Dorantes, Hoa Nguyen, Leslie Bautista, Guadalupe Covarrubias-Oregel, Jesus Lopez, Sara Patino, Andrew Perez, Matias Lopez, Andrea Presas, Abrian Villalobos, Monica Ramos, Christian Urbina, Ayo Babalola, and Gurpinder Bahia. Special thanks to members of the Schuchardt and Warfa research teams at the University of Minnesota, Kranzfelder Lab, especially Namitha Bhat, and the Discipline-Based Education Research Journal Club at the University of California Merced, especially Erik Menke, for their valuable input in brainstorming, critiquing, revising the study questions and direction of this study, and feedback that critically improved the manuscript. This study was funded by Howard Hughes Medical Institute award no. GT11066, National Science Foundation Hispanic-Serving Institution award no. 1832538, and start-up funding from the Department of Molecular and Cellular Biology at the University of California, Merced, for PK.

Supplemental material

Supplemental data are available at *BIOSCI* online.

References cited

Adu E, Okeke C. 2014. Factors affecting lecturers' participation in continuing professional development (CPD). Journal of Sociology and Social Anthropology 5: 271–281.

Akiha K, Brigham E, Couch BA, Lewin J, Stains M, Stetzer MR, Vinson EL, Smith MK. 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: 68.

[AAAS] American Association for the Advancement of Science. 2012. Describing and Measuring Undergraduate STEM Teaching Practices. AAAS.

Anderson WA, et al. 2011. Changing the culture of science education at research universities. Science 331: 152–153.

Berger J-L, Girardet C, Vaudroz C, Crahay M. 2018. Teaching Experience, teachers' beliefs, and self-reported classroom management practices: A coherent network. SAGE open 8: 2158244017754119.

Berland LK, Reiser BJ. 2011. Classroom communities' adaptations of the practice of scientific argumentation. Science Education 95: 191–216.

Bettinger EP, Long BT. 2010. Does cheaper mean better? The impact of using adjunct instructors on student outcomes. Review of Economics and Statistics 92: 598–613.

Borda E, Schumacher E, Hanley D, Geary E, Warren S, Ipsen C, Stredicke L. 2020. Initial implementation of active learning strategies in large, lecture STEM courses: Lessons learned from a multi-institutional, interdisciplinary STEM faculty development program. International Journal of STEM Education 7: 4.

Breslyn W, McGinnis JR. 2012. A comparison of exemplary biology, chemistry, earth science, and physics teachers' conceptions and enactment of inquiry. Science Education 96: 48–77.

Brownell SE, Tanner KD. 2012. Barriers to faculty pedagogical change: Lack of training, time, incentives, and…tensions with professional identity? CBE—Life Sciences Education 11: 339–346.

Budd D, Van der Hoeven Kraft K, McConnell D, Vislova T. 2013. Characterizing teaching in introductory geology courses: Measuring classroom practices. Journal of Geoscience Education 61: 461–475.

Bush SD, Stevens MT, Tanner KD, Williams KS. 2020. Disciplinary bias, money matters, and persistence: Deans' perspectives on science faculty with education specialties (SFES). CBE—Life Sciences Education 19: 34.

Cazden CB. 2001. Classroom Discourse: The Language of Teaching and Learning. Heinemann.

Cohen J. 1988. Statistical Power Analysis for the Behavioral Sciences. Taylor and Francis.

Couch BA, Wright CD, Freeman S, Knight JK, Semsar K, Smith MK, Summers MM, Zheng Y, Crowe AJ, Brownell SE. 2019. GenBio-MAPS: A programmatic assessment to measure student understanding of vision and change core concepts across general biology programs. CBE—Life Sciences Education 18: 1.

Council NR. 2012. Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. National Academies Press.

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

Denaro K, Sato B, Harlow A, Aebersold A, Verma M. 2021. Comparison of cluster analysis methodologies for characterization of Classroom Observation Protocol for Undergraduate STEM (COPUS) data. CBE—Life Sciences Education 20: 3.

Duschl RA. 2008. Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. Review of Research in Education 32: 268–291.

Duschl RA, Osborne J. 2002. Supporting and Promoting Argumentation Discourse in Science Education. Taylor and Francis.

Eagan K. 2016. Becoming More Student-Centered? An Examination Of Faculty Teaching Practices across STEM and Non-STEM Disciplines between 2004 and 2014. Alfred P. Sloan Foundation.

Ebert-May D, Brewer C, Allred S. 1997. Innovation in large lectures: Teaching for active learning. BioScience 47: 601–607.

Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE. 2011. What we say is not what we do: Effective evaluation of faculty professional development programs. BioScience 61: 550–558.

Eddy SL, Converse M, Wenderoth MP. 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: 23.

Edelson DC, Gordin DN, Pea RD. 1999. Addressing the challenges of inquiry-based learning through technology and curriculum design. Journal of the Learning Sciences 8: 391–450.

Fassinger PA. 1996. Professors' and students' perceptions of why students participate in class. Teaching Sociology 24: 25–33.

Figlio DN, Schapiro MO, Soter KB. 2015. Are tenure track professors better teachers? Review of Economics and Statistics 97: 715–724.

Finelli CJ, Borrego M. 2020. Evidence-based strategies to reduce student resistance to active learning. Pages 943–952 in Mintzes JJ, Walter E, eds. Active Learning in College Science: The Case for Evidence-Based Practice. Springer.

Fleiss JL. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin 76: 378.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences 111: 8410–8415.

Galbraith MW, Shedd PE. 1990. Building skills and proficiencies of the community college instructor of adult learners. Community College Review 18: 6–14.

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: 8.
Gess-Newsome J, Southerland SA, 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: 731–767.
Gray K, Steer D, McConnell D, Owens K. 2010. Using a student-manipulated model to enhance student learning in a large lecture class. Journal of College Science Teaching 40: 86–95.
Grossman PL, Stodolsky SS. 1995. Content as context: The role of school subjects in secondary school teaching. Educational Researcher 24: 5–23.
Harlow A, Lo SM, Saichaie K, Sato BK. 2020. Characterizing the University of California's tenure-track teaching position from the faculty and administrator perspectives. PLOS ONE 15: e0227633.
Henderson C, Beach A, Finkelstein N. 2011. Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. Journal of Research in Science Teaching 48: 952–984.
Henderson C, Beach AL, Finkelstein N. 2012a. Four categories of change strategies for transforming undergraduate instruction. Pages 223–245 in Tynjälä P, Stenström M-L, Saarnivaara M, eds. Transitions and Transformations in Learning and Education. Springer.
Henderson C, Dancy M, Niewiadomska-Bugaj M. 2012b. Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? Physical Review Special Topics: Physics Education Research 8: 020104.
Henderson C, Dancy MH. 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.
Henderson C, Finkelstein N, Beach A. 2010. Beyond dissemination in college science teaching: An introduction to four core change strategies. Journal of College Science Teaching 39: 18–25.
Herman GL, Greene JC, Hahn LD, Mestre JP, Tomkin JH, West M. 2018. Changing the Teaching Culture in Introductory STEM Courses at a Large Research University. Journal of College Science Teaching 47: 32–38.
Hora MT. 2012. Organizational factors and instructional decision-making: A cognitive perspective. Review of Higher Education 35: 207–235.
Horn IS. 2010. Teaching replays, teaching rehearsals, and re-visions of practice: Learning from colleagues in a mathematics teacher community. Teachers College Record 112: 225–259.
Howe C, Abedin M. 2013. Classroom dialogue: A systematic review across four decades of research. Cambridge Journal of Education 43: 325–356.
Hoy AW, Spero RB. 2005. Changes in teacher efficacy during the early years of teaching: A comparison of four measures. Teaching and Teacher Education 21: 343–356.
Huberman AM. 1994. Lives of Teachers. Teachers College Press.
Jiménez-Aleixandre MP, Erduran S. 2007. Argumentation in Science Education: An Overview. Pages 3–27 in Erduran S, Jiménez-Aleixandre MP, eds. Argumentation in Science Education: Perspectives from Classroom-Based Research. Springer.
Kay M, Wobbrock J. 2020. ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. University of Washington.
Keavney G, Sinclair KE. 1978. Teacher concerns and teacher anxiety: A neglected topic of classroom research. Review of Educational Research 48: 273–290.
Kennedy MM. 2016. How does professional development improve teaching? Review of Educational Research 86: 945–980.
Kezar A, Gehrke S, Bernstein-Sierra S. 2017. Designing for success in STEM communities of practice: Philosophy and personal interactions. Review of Higher Education 40: 217–244.
Knight JK, Wise SB, Southard KM. 2013. Understanding clicker discussions: Student reasoning and the impact of instructional cues. CBE—Life Sciences Education 12: 645–654.
Knight JK, Wood WB. 2005. Teaching more by lecturing less. Cell Biology Education 4: 298–310.
Kranzfelder P, Bankers-Fulbright JL, García-Ojeda ME, Melloy M, Mohammed S, Warfa A-RM. 2019a. The Classroom Discourse Observation Protocol (CDOP): A quantitative method for characterizing teacher discourse moves in undergraduate STEM learning environments. PLOS ONE 14: e0219019.
Kranzfelder P, Bankers-Fulbright JL, García-Ojeda ME, Melloy M, Mohammed S, Warfa A-RM . 2020. Undergraduate biology instructors still use mostly teacher-centered discourse even when teaching with active learning strategies. BioScience 70: 901–913.
Kranzfelder P, Lo AT, Melloy MP, Walker LE, Warfa A-RM. 2019b. Instructional practices in reformed undergraduate STEM learning environments: A study of instructor and student behaviors in biology courses. International Journal of Science Education 41: 1944–1961.
Krussel L, Edwards B, Springer G. 2004. The teacher's discourse moves: A framework for analyzing discourse in mathematics classrooms. School Science and Mathematics 104: 307–312.
Kuhn D, Arvidsson TS, Lesperance R, Corprew R. 2017. Can engaging in science practices promote deep understanding of them? Science Education 101: 232–250.
Lane AK, Meaders CL, Shuman JK, Stetzer MR, Vinson EL, Couch BA, Smith MK, Stains M. 2021. Making a first impression: Exploring what instructors do and say on the first day of introductory STEM courses. CBE—Life Sciences Education 20: 7.
Leupen SM, Kephart KL, Hodges LC. 2020. Factors influencing quality of team discussion: Discourse analysis in an undergraduate team-based learning biology course. CBE—Life Sciences Education 19: 7.
Lewin JD, Vinson EL, Stetzer MR, Smith MK. 2016. A campus-wide investigation of clicker implementation: The status of peer discussion in STEM Classes. CBE—Life Sciences Education 15: 6.
Lund, Pilarz M, Velasco JB, Chakraverty D, Rosploch K, Undersander M, Stains M. 2015. The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. CBE—Life Sciences Education 14: 1–12.
Lund TJ, Stains M. 2015. The importance of context: An exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. International Journal of STEM education 2: 1–21.
Manduca CA, Iverson ER, Luxenberg M, Macdonald RH, McConnell DA, Mogk DW, Tewksbury BJ. 2017. Improving undergraduate STEM education: The efficacy of discipline-based professional development. Science Advances 3: e1600193.
Meaders CL, Toth ES, Lane AK, Shuman JK, Couch BA, Stains M, Stetzer MR, Vinson E, Smith MK. 2019. "What will I experience in my college STEM courses?" An investigation of student predictions about instructional practices in introductory courses. CBE—Life Sciences Education 18: 60.
Mercer N. 2010. The analysis of classroom talk: Methods and methodologies. British Journal of Educational Psychology 80: 1–14.
Michaels S, O'Connor C. 2015. Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. Pages 347–362 in Resnik LB, Asterhan CSC, Clarke SN, eds. Socializing Intelligence through Talk and Dialogue. American Educational Research Association.
Michaelsen LK, Sweet M. 2008. The essential elements of team-based learning. New Directions for Teaching and Learning 2008: 7–27.
Mitchell KM, Martin J. 2018. Gender bias in student evaluations. Political Science and Politics 51: 648–652.
Mizell H. 2010. Why Professional Development Matters. ERIC.
Mortimer E, Scott P. 2003a. Meaning Making in Secondary Science Classrooms. McGraw-Hill Education.
Mortimer EF, Scott PH. 2003b. Meaning Making in Secondary Science Classrooms. Open University Press.
Murray DS. 2019. The precarious new faculty majority: Communication and instruction research and contingent labor in higher education. Communication Education 68: 235–245.

Myers SA, Rocca KA. 2000. The relationship between perceived instructor communicator style, argumentativeness, and verbal aggressiveness. Communication Research Reports 17: 1–12.

Ochoa GL, Pineda D. 2008. Deconstructing power, privilege, and silence in the classroom. Radical History Review 2008: 45–62.

Oliveira AW. 2010. Improving teacher questioning in science inquiry discussions through professional development. Journal of Research in Science Teaching 47: 422–453.

Osborne J. 2010. Arguing to learn in science: The role of collaborative, critical discourse. Science 328: 463–466.

Owens MT, Seidel SB, Wong M, Bejines TE, Lietz S, Perez JR, Sit S, Subedar Z-S, Acker GN, Akana SF. 2017. Classroom sound can be used to classify teaching practices in college science courses. Proceedings of the National Academy of Sciences 114: 3085–3090.

Patrick LE. 2020. Faculty and student perceptions of active learning. Pages 889–907 in Mintzes JJ, Walter EM, eds. Active Learning in College Science: The Case for Evidence-Based Practice. Springer Nature.

Patrick LE, Howel LA, Wischusen W. 2016. Perceptions of active learning between faculty and undergraduates: Differing views among departments. Journal of STEM Education: Innovations and Research 17: 55–63.

Putnam RT, Borko H. 1997. Teacher learning: Implications of new views of cognition. Pages 1223–1296 in Biddle BJ, Good TL, Goodson I, eds. International Handbook of Teachers and Teaching. Springer.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Reisner BA, Pate CL, Kinkaid MM, Paunovic DM, Pratt JM, Stewart JL, Raker JR, Bentley AK, Lin S, Smith SR. 2020. I've been given COPUS (Classroom Observation Protocol for Undergraduate STEM) data on my chemistry class… now what? Journal of Chemical Education 97: 1181–1189.

Sawada D, Piburn MD, Judson E, Turley J, Falconer K, Benford R, Bloom I. 2002. Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. School Science and Mathematics 102: 245–253.

Seidel SB, Reggi AL, Schinske JN, Burrus LW, Tanner KD, Tomanek D. 2015. Beyond the biology: A systematic investigation of noncontent instructor talk in an introductory biology course. CBE—Life Sciences Education 14: 43.

Shadle SE, Marker A, Earl B. 2017. Faculty drivers and barriers: Laying the groundwork for undergraduate STEM education reform in academic departments. International Journal of STEM Education 4: 8.

Signorini A, Pohan C. 2019. Exploring the Impact of the Students Assessing Teaching and Learning Program. International Journal for Students as Partners 3: 139–148.

Sinclair JM, Coulthard M. 1975. Towards an Analysis of Discourse: The English Used by Teachers and Pupils. Oxford University Press.

Smith MK, Jones FHM, Gilbert SL, Wieman CE. 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 MK, Vinson EL, Smith JA, Lewin JD, Stetzer MR. 2014. A campus-wide study of stem courses: New perspectives on teaching practices and perceptions. CBE—Life Sciences Education 13: 624–635.

Spronken-Smith R, Walker R. 2010. Can inquiry-based learning strengthen the links between teaching and disciplinary research? Studies in Higher Education 35: 723–740.

Stains M, et al. 2018. Anatomy of STEM teaching in North American universities. Science 359: 1468–1470.

Theobald EJ, Hill MJ, Tran E, Agrawal S, Arroyo EN, Behling S, Chambwe N, Cintrón DL, Cooper JD, Dunster G. 2020. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proceedings of the National Academy of Sciences 117: 6476–6483.

Thiele RB, Treagust DF. 1994. An interpretive examination of high school chemistry teachers' analogical explanations. Journal of Research in Science Teaching 31: 227–242.

Tomczak M, Tomczak E. 2014. The need to report effect size estimates revisited: An overview of some recommended measures of effect size. Trends in Sport Sciences 21: 19–25.

Tomkin JH, Beilstein SO, Morphew JW, Herman GL. 2019. Evidence that communities of practice are associated with active learning in large STEM lectures. International Journal of STEM Education 6: 1.

Truxaw MP, DeFranco TC. 2008. Mapping mathematics classroom discourse and its implications for models of teaching. Journal for Research in Mathematics Education 39: 489–525.

Umbach PD, Wawrzynski MR. 2005. Faculty do matter: The role of college faculty in student learning and engagement. Research in Higher Education 46: 153–184.

van der Lans RM. 2018. On the "association between two things": The case of student surveys and classroom observations of teaching quality. Educational Assessment, Evaluation, and Accountability 30: 347–366.

Van Driel JH, Jong OD, Verloop N. 2002. The development of preservice chemistry teachers' pedagogical content knowledge. Science education 86: 572–590.

Warfa A-RM, Roehrig GH, Schneider JL, Nyachwaya J. 2014. Role of teacher-initiated discourses in students' development of representational fluency in chemistry: A case study. Journal of Chemical Education 91: 784–792.

Wasserstein RL, Schirm AL, Lazar NA. 2019. Moving to a world beyond "$p < 0.05$." American Statistician 73: 1–19.

Weaver GC, Burgess WD, Childress AL, Slakey L. 2016. Transforming Insitutions: 21st Century Undergraduate STEM Education. Purdue University Press.

Wei L, Murphy PK, Firetto CM. 2018. How can teachers facilitate productive small-group talk? An integrated taxonomy of teacher discourse moves. Elementary School Journal 118: 578–609.

Wenger E. 1998. Communities of practice: Learning as a social system. Systems Thinker 9: 2–3.

Williams CT, Walter EM, Henderson C, Beach AL. 2015. Describing undergraduate STEM teaching practices: A comparison of instructor self-report instruments. International Journal of STEM Education 2: 18.

Wobbrock JO, Findlater L, Gergle D, Higgins JJ. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. Pages 143–146 in Grinter R, Rodden T, eds. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery.

Xu D, Solanki S. 2020. Tenure-track appointment for teaching-oriented faculty? The impact of teaching and research faculty on student outcomes. Educational Evaluation and Policy Analysis 42: 66–86.

Jourjina Subih Alkhouri, Cristine Donham, and Petra Kranzfelder (pkranzfelder@ucmerced.edu) are affiliated with the Department of Quantitative and Systems Biology; Téa S. Pusey is affiliated with the Department of Sociology; Alexander H. Stivers and Petra Kranzfelder are affiliated with the Department of Molecular and Cellular Biology; and Adriana Signorini is affiliated with the Center for Engaged Teaching and Learning, all at the University of California, Merced, in Merced, California, in the United States.