RESEARCH   **Open Access**

Evidence that communities of practice are associated with active learning in large STEM lectures

Jonathan H. Tomkin[1]* ⓘ, Shereen O. Beilstein[2], Jason W. Morphew[2] and Geoffrey L. Herman[3]

Abstract

Background: With the increased attention on the need to retain students within STEM majors, it is important for STEM instructors to adopt evidence-based instructional practices that are student-centric and employ active learning techniques. However, traditional approaches for increasing student-centric, active learning practices such as workshops, seminars, and department or college mandates have been either inefficient or ineffective at motivating institutional change. This is particularly true for introductory STEM courses with large enrollments. One promising approach is to develop and support instructors in forming communities of practice around reforming introductory and foundational STEM courses. By engaging instructors within these communities of practice, and connecting the communities with each other, instructors may be more likely to experience an epistemological shift that will lead to adoption of active learning practices.

To explore whether participating in a community of practice is related to the use of student-centered, active learning techniques, 25 undergraduate foundational STEM courses whose instructors were members of a community of practice were observed using the Classroom Observation Protocol for Undergraduate STEM Courses. The results were compared to a sample of 35 undergraduate foundational STEM courses whose instructors were not members of a community of practice.

Results: Instructors who were members of a community of practice were much more likely to employ student-centric practices, such as asking questions, following up, and engaging in discussion, and much less likely to use instructor-centered practices, such as lecturing. In addition, students in these classes were more likely to be actively engaged in problem-solving activities rather than passively listening. We found that student-centric, active learning practices correlated with students attending and actively participating in class, an effect that is stronger for courses taught by instructors who were members of a community of practice.

Conclusion: Communities of practice are a potentially effective mechanism for enhancing student learning and retention by increasing the use of active learning practices by STEM instructors. These communities are particularly effective when they consist of small, disciplinary teams working on the same course(s) and are linked with other individuals or groups that use evidence-based instructional practices.

Keywords: Evidence-based instruction, Communities of practice, Instructional change

---

*Correspondence: tomkin@illinois.edu
[1]School of Earth, Society, and the Environment, University of Illinois at Urbana–Champaign, Green St, Urbana 61801, IL, USA
Full list of author information is available at the end of the article

Introduction

There is a prevalent need to reform science, technology, engineering, and mathematics (STEM) lecture sections (Austin 2011; Handelsman et al. 2004; Singer et al. 2012). University STEM instructors traditionally emphasize content, believing that lecturing effectively transmits the content directly from instructor to student. A considerable body of research (Freeman et al. 2014; Handelsman et al. 2004) suggests that this view is mistaken, and that student-centric approaches where students actively participate in lectures lead to more learning. Despite the vast evidence, it is hard to get faculty to adopt active learning approaches as faculty often demonstrate a "resistance to change" (Henderson et al. 2011; Kegan and Lahey 2009). In addition to a lack of training and time (Brownell and Tanner 2012), the lack of adoption of active learning techniques is often the result of instructors associating their teaching method with an implicit world-view—a set of beliefs about how learning occurs. From this perspective, change requires instructors not only to learn and apply new technical skills, but also to adopt a new belief system—specifically the belief that instruction utilizing active learning techniques is more beneficial for student learning than traditional instruction utilizing passive learning techniques, such as lecturing. A successful campaign to increase the use of active learning should therefore focus on convincing faculty to change their epistemology about teaching and learning (making a so-called adaptive change (Kegan and Lahey 2009)).

The usual methods of academic communication are not effective at altering pre-existing beliefs connected with teaching and learning. Academic articles and seminars often do not lead to long-term lasting changes in teaching practices (Henderson et al.). Standard methods of bureaucratic intervention (e.g., "top-down" mandates, teaching workshops, and institutional centers for teaching and learning) are similarly ineffective because these methods do not change pre-existing belief structures, resulting in little to no adoption of evidence-based instructional practices (Kezar et al. 2015). This suggests that exposing instructors to the literature is not sufficient to affect large-scale change and that a different approach is required to promote the widespread adoption of active learning in lectures across an institution (Kezar et al. 2015). Collaborative teaching models are, however, thought to be a channel through which new beliefs about teaching can be cultivated (Gehrke and Kezar 2017). One potential established collaborative approach is to encourage faculty to create communities of practice (CoPs) (Kezar et al. 2017; Wenger 1998).

In prior work, we described a reform effort that stressed a "bottom-up," collaborative approach (Ma et al.). Faculty members who supported the use of active learning techniques were encouraged to mentor other faculty members who wanted to improve some aspect of their course, whether focused on content or pedagogy. These CoPs were intended to persist, so as to encourage long-term collaboration, with the hope that (a) mutual trust between peers would lead to the spread of constructive beliefs about active learning, as has been suggested in prior research (Kezar et al. 2017; Villachia et al. 2013), and (b) continuing communication between experts and practicing faculty inside the CoPs would lead to faster and more efficient implementation of active learning techniques. In this study, we use the Classroom Observation Protocol for Undergraduate STEM courses (COPUS) to compare the teaching practices of faculty who participated in CoPs with faculty who did not participate in CoPs.

Background

Reforming lectures

Calls for reform in STEM education have emphasized the need to increase the quality and number of graduating STEM students (Austin 2011; Brewer and Smith 2011; Fairweather 2009; President's Council of Advisors on Science and Technology 2012). There are also persistent calls to increase the number of STEM graduates to meet growing workforce demands. For example, the president's office of technology has called for annual increases of 34% of STEM graduates (President's Council of Advisors on Science and Technology 2012). However, a major barrier to achieving this goal lies in the poor retention rates of students in STEM majors (Seymour 2002). In addition, industry and government stakeholders have increased pressure on STEM programs to produce "T-shaped" students: students not only with technical depth but also with a breadth of professional skills such as communication, teamwork, ethics, creativity, and desires and skills for lifelong learning (Jamieson and Lohman 2012; Sheppard et al. 2014).

Growing evidence suggests that traditional, lecture-centric instruction fails to achieve these goals. The traditional, lecture-centric approach, which emphasizes passive leaning, has been linked to less learning for, and lower retention rates of, STEM students relative to more active approaches (Freeman et al. 2014; Henderson et al. 2011). Further, if students merely passively listen to their instructors, they lack opportunities to practice the professional skills asked for by industry and government stakeholders. It is not surprising then that lecture-centric instruction fails to achieve both the depth and breadth required to produce "T-shaped" students (Austin 2011; Freeman et al. 2014). In contrast, student-centric and active learning approaches have been shown to improve learning gains for students (Freeman et al. 2014) and have been shown to lead to higher retention of STEM students relative to traditional approaches (Henderson et al. 2011).

Because efforts that focus on changing the teaching practices of individual instructors have been generally ineffective at promoting large-scale change, there has been a growing interest in community-based methods. In particular, CoPs are being explored because they have been shown as an effective mechanism for spreading knowledge within organizations (Kezar et al. 2015; Lea 2005).

Communities of practice

CoPs can be formal or informal structures within or across organizations and are defined primarily by a common interest in, and commitment to, a domain of knowledge (Wenger et al. 2002; Cox 2005). CoPs go beyond mere common-interest groups as members are primarily practitioners whose participation is focused on improving shared knowledge and practices within the domain (Cox 2005). The community is characterized by regular meetings in which members participate in joint activities and discussions whose purpose is to share information that allows members to learn from each other (Cox 2005). Thus, a primary criterion for classifying a group of faculty members in our study as a CoP was whether they met on a regular basis to discuss how they would teach particular large-enrollment STEM courses (Ma et al.).

Effective CoPs focus less on formal structures or roles and more on capitalizing on the community's own agency and energy (Ianquinto et al. 2011). Consequently, CoPs grow and emerge in response to changes in membership, emerging interests, and evolving goals. CoPs often have a regular rhythm or cycles of activities that maintain steady engagement with the community (Ianquinto et al. 2011; Wenger et al. 2002). This growth is facilitated by open, collaborative dialogue within and outside the CoP, helping members learn from each other, but avoiding "group think" by inviting new and challenging ideas into the community. In our context, the CoPs were formed around the common challenge of improving student outcomes amid growing course enrollments. The regular cycles of delivering courses each term and meeting weekly were intended to promote critical collaborative discussions. By providing mentors, we created avenues for ideas to spread across CoPs.

CoPs can provide an environment for challenging faculty whose resistant identities and beliefs deter the adoption of new teaching approaches (Borrego et al. 2013). By surrounding faculty with respected colleagues, CoPs can help faculty believe that adopting new practices will be accepted by their valued peers (Wenger et al. 2002). Beyond addressing belief structures, CoPs generate a sustainable, ongoing context that offers faculty needed training. Collaborations with colleagues provide a forum for feedback on teaching practices and exposure to new ideas about teaching, ultimately leading to learning and improved practice (Lave and Wenger 1991; Wenger 1998; Wenger et al. 2002). Through this process, CoPs foster mutual trust and reflective engagement (Wenger 1998; Wenger et al. 2002), and in such a context, assumptions and beliefs about teaching can change implicitly (Davenport and Prusak 1970; Hildreth and Kimble 2002). This trust and community can fuel faculty members' intrinsic motivations to learn about effective teaching, circumventing a lack of motivation derived from extrinsic rewards (Herman et al. 2017; Ryan and Deci 2000). This type of learning can decrease the learning curve for novices, reduce creation of redundant resources or reenactments of failures, and promote creativity (Lesser and Storck 2001).

Classroom Observation Protocol for Undergraduate STEM

The Classroom Observation Protocol for Undergraduate STEM courses (COPUS) (Smith et al. 2013) is a flexible tool that can capture a variety of passive and active instructional practices in diverse courses that vary by content and enrollment. The protocol accounts for both the instructor and their students' behaviors during each 2-min interval for the entire class, using 25 individual codes that range from practices that are passive, such as lecturing for the instructor and listening for students, to those that are active, such as asking clicker questions for instructors and working in groups for students (see Table 1 for a complete list of the individual COPUS codes). The method is non-judgmental in that there is no estimate of instructor quality or effectiveness; however, because the method records the absence or presence of specific classroom behaviors, it is well-suited to document the presence and frequency of instructors' use of active learning practices. In addition, multiple behaviors can be coded within each 2-min segment, leading to an accurate representation of each class.

The original iteration of COPUS can capture a complex array of classroom behaviors; however, it can become unwieldy for comparison and analysis (Smith et al. 2014). To provide a more streamlined way of looking at the data, an updated version of COPUS (Smith et al. 2014) condenses the original 25 codes into four categories for instructor behaviors and four categories for student behaviors (see Table 1). The two most important instructor behaviors from the collapsed categories for discriminating between active and passive learning practices are guiding and presenting, respectively. Instructor behaviors such as posing and answering questions, asking and following up on clicker questions, among others, map onto the collapsed category of guiding, which we contend reflects instructors' use of active learning practices. Whereas instructor practices such as lecturing, writing on the board, or showing a demonstration or video map onto the collapsed category of presenting, which

Table 1 List of individual COPUS codes (from Smith et al. 2013) by their collapsed categories (from Smith et al. 2014)

	Collapsed categories	Individual codes	A/P
Instructor is:	Presenting	Lecturing	P
		Real-time writing	–
		Showing or conducting demo/video	P
	Guiding	Follow-up on clicker question/activity	A
		Posing non-clicker question to students (non-rhetorical)	A
		Asking clicker question (entire time, not just when first asked)	A
		Listening to and answering student questions to entire class	–
		Moving through class guiding ongoing student work	–
		One-on-one extended discussion with individual students	–
	Administration	Administration (assign homework, return tests, etc.)	–
	Other	Waiting (instructor late, working on fixing technical problems)	–
		Other	–
Students are:	Receiving	Listening to instructor	P
	Working	Individual thinking/problem solving	–
		Discussing clicker question in groups of students	A
		Working in groups on worksheet activity	A
		Other assigned group activity	–
		Making a prediction about a demo or experiment	A
		Test or quiz	–
	Talking to class	Student answering question posed by instructor	–
		Student asks question	–
		Students engaged in whole-class discussion	A
		Students presenting to entire class	A
	Other	Waiting (instructor late, working on fixing technical problems)	–
		Other	–

The far-right column lists predictions for which code will be systematically more common in active (A) or passive (P) lectures (see far-right column)

we contend reflects instructors' use of passive learning practices.

For student behaviors, three collapsed categories—working, talking, and receiving—can be used for discriminating between active and passive learning practices. The first category, working, aligns readily with active learning practices. In class observations, students are considered to be working if they engage in student-centric behaviors such as individual thinking and problem solving, discussing clicker questions in a group, working in groups on worksheets or other activities, or making predictions. Conversely, the last category, receiving, aligns readily with passive learning practices. Students are considered to be receiving if they are observed listening to either the instructor or another student speak. The collapsed category of talking falls somewhere in between. Students considered to be Talking if they engage in discursive practices such as asking or answering a question posed by the instructor or contributing to a whole-class discussion. Considering whether talking can be classified as active learning is contingent on the unit of analysis. For an individual student who is answering a question, for example, talking could be classified as an active learning event. This would not generally be true for the class as a whole, however, as the other students would be passively listening to the question and answer.

Two large, multi-institution studies—conducted by Lund and colleagues (Lund et al. 2015), whose project included 73 instructors and 269 lectures, and Stains and colleagues (Stains et al. 2018), whose project included 548 instructors and 2008 lectures—applied cluster analysis to COPUS data to determine if there are consistent styles of lectures. Although the two research groups arrived at their clusters independently, their findings have strikingly similar characteristics. Lund and colleagues (Lund et al. 2015) described four broad categories of lecturing behavior: "lecturing," "socratic," "peer instruction," and "collaborative learning." Whereas, Stains and colleagues (Stains et al. 2018) described three broad categories: "didactic," "interactive lecture," and "student-centered." Lecturing and socratic (Lund et al. 2015) map neatly onto didactic (Stains et al. 2018)—all types sharing the trait that instructors lectured in at least 80% of the 2-min periods and that the amount of student work was negligible, averaging 10% or less. Similarly, peer instruction Lund et al. 2015 and interactive lecture (Stains et al. 2018) have concordant instructor behaviors; lecturing, on average, between 55% and 76% of the periods in the former case and in roughly 75% of the periods in the latter, and with students engaged in group work (with student response systems or without), averaging between a quarter to a half of the periods observed. Collaborative learning is associated with relatively low levels of lecturing (50% of periods or less) and high levels of student work (around

50%). The student-centered clusters broadly share these collaborative learning traits.

The coherent lecturing styles described by these two large studies suggest that COPUS is well-suited to capture the instructor and student behaviors that are diagnostic of both traditional lecturing and active learning lecturing. The consistency of results also supports the claim that COPUS is a reliable measure.

Context of current study

Two intertwined programs—the College of Engineering-funded Strategic Instructional Innovations Program (SIIP) and an NSF-funded Widening Implementation and Demonstration of Evidence-based Reform (WIDER)—were conceived in 2012 at a large Midwestern University (Herman et al. 2018). This was a large-scale intervention: between the two programs, 28 departments and 231 faculty were involved. Though the scope of these two programs differed (SIIP focused solely on engineering courses and faculty, while WIDER was STEM inclusive), these initiatives were both created with the goal of transforming instructional practices in large lecture courses using the same model of change. These programs used CoPs as a way to create an environment of support for faculty who did not intentionally use evidence-based instructional practices prior the advent of these programs (Herman et al. 2018).

The CoPs consisted of core groups of three or more instructors responsible for the delivery of gateway, or foundational, STEM courses. The composition of these groups varied across the different departments, and included lab staff, faculty, and graduate assistants, all of whom were directly involved in the delivery of the course. CoPs met on weekly basis to discuss both short- and long-term goals. For example, the group may spend part of a meeting helping a member troubleshoot her use of peer instruction in preparation for the next class. They could then use the remaining time to plan data collection for an ongoing research project and evaluate the success of the changes made to that course. CoP meetings also included an outside faculty member with some experience implementing active learning techniques as well as an external evaluator. Many of these outside members mentored multiple CoPs and would often mention successful practices that other CoPs were using. Sometimes these suggestions would lead to members of different CoPs meeting to share ideas.

The main goal shared by CoP members was the improvement of the courses they teach. For each CoP group, these were either a single foundational course or a connected sequence of courses. CoP members collectively identified and implemented their planned reforms. The reforms chosen by each CoP was not prescribed but rather emerged from the many discussions within

the CoP, which included reviewing scholarly articles and evaluating suggestions from the outside faculty mentor. The mentors encouraged the CoP members to "teach the way you do research," so as to discourage individuals from proposing teaching methods based on hunches or personal experience without regard for evidence on effectiveness. Since this is a research-intensive institution, and research faculty appreciate how STEM research is conducted, this approach helped CoP members understand the importance of using evidence-based instructional practices (such as active learning) in their courses.

Method
Evaluating communities of practice

As the lecture section is standard practice for the large STEM courses affiliated with SIIP and WIDER, we examined the characteristics of the lecture section to determine if instructors involved with CoPs adopted active learning techniques. Given that the COPUS method can capture a variety of activities instructors utilize and the corresponding student behaviors these activities elicit (Smith et al. 2013), COPUS was used to address our central research question: how do the lectures of instructors from CoPs vary in relation to their peers who were not part of CoPs? Although traditional lecture sections typically involve long periods of actual lecturing from the instructor (accompanied by students passively listening), reformed lecture periods can involve significant student activity. Student-centric lectures, therefore, should have less frequent lecturing from the instructor and more periods when the instructor is guiding students as they engage in collaborative problem-solving activities. Additionally, most university faculty teach in isolation, maintaining sole jurisdiction of their courses and rarely talking to other faculty (Spalter-Roth et al. 2010; Tanner and Allen 2006). While most faculty practice teaching, few can be considered to be members of communities focused on improving teaching practice and are therefore not part of a CoP. Thus, we hypothesize that CoPs encourage the adoption of evidence-based instructional practices, and so predict that CoP involvement correlates with an increase in time spent using active learning techniques in lecture sections, and a reduction in time spent traditionally lecturing.

Instructors were externally rated by a project evaluation rubric (Herman et al. 2018) to determine whether they were functioning as a CoP. These evaluations were made independently of the observations and data used in the current study. The rubric evaluated teams across five dimensions (administrative support, collaborative development and ownership, faculty outcomes, student outcomes, and sustainability/trajectory). In this study, we compare the teaching practices of instructors who were verified as participating in a CoP with those who were not part of a CoP. We will refer to these two groups as

CoP instructors and non-CoP instructors throughout the paper.

Classroom observations

Sixty undergraduate STEM lecture classes were observed across 14 departments at a large Midwestern University. We selected the classes to observe using purposive sampling techniques (Teddlie and Yu 2007) with the objectives of (1) collecting classroom observations that were representative of the STEM courses offered; (2) conducting observations on non-test, typical class days; and (3) comparing these classroom observations using the COPUS protocol to record the presence and frequency of active learning practices utilized by CoP and non-CoP instructors.

All of the observed classes were required for STEM majors, and most (81.67%) of the observed classes were "foundational." Foundational courses in our sample meet the following criteria: they are introductory; required for STEM majors; serve as prerequisites for upper-level, discipline-specific courses; and are taken by students in their first or second year.

We observed at least one lecture for each of the following foundational courses: Calculus I, II, and III; Chemistry 0, I and II; Physics 0, I and II; Integrative Biology I; Molecular and Cellular Biology I; Mechanical Engineering I, II, and III; Electrical and Computer Engineering I and II; Computer Science I, II and III; Civil and Environmental Engineering I; Materials Science and Engineering I; and General Engineering I.

In addition to foundational courses, a further 11 upper-level STEM courses were observed as part of the reform. Four of these courses were taught by CoP instructors, and seven were taught by non-CoP instructors. These courses were taught in Civil and Environmental Engineering, Mechanical Engineering, Integrative Biology, Computer Science, Industrial Engineering, Electrical and Computer Engineering, and Materials Science and Engineering.

Comparing classroom observations by CoP

The instructors included in the current study's sample all teach at a large Midwestern University. Twenty-five of the classes observed were taught by CoP instructors, and 35 of the classes observed were taught by non-CoP instructors as determined by (Herman et al. 2018). All instructors were aware that they were being observed for the study. Observation times were arranged in advance with instructors on non-test,"business-as-usual" class days, with consent obtained prior to the observation.

Most of the courses (83.3%) observed in this sample were 50 min long. However, 10 courses (16.7%) lasted 80 min. For the longer courses, observations were concluded after 50 min to ensure durational consistency across the entire sample. The percentage of courses observed that were 80 min long were balanced across CoP (16.0%) and non-CoP (17.1%) instructors. Course enrollment for the observed classes varied, but was generally large with a mean enrollment of 208.8, median of 187, and range from 33 to 692 students.

Instructors of the courses included in the dataset were observed once—with three exceptions. In two instances, a scheduling error led to two separate observations of the same instructor teaching the same course. In the other, two observations were conducted within the same week for the same course because the instructor utilized two class formats—traditional and flipped—throughout the semester. For all three of these instances, data from the observations were averaged and collapsed into one entry. Also, it should be noted that the first and last authors were each observed once as part of this study.

Approximately 33% of observations were conducted by two different observers, and substantial inter-rater reliability was achieved (Cohen's kappa = .89) (Cohen 1960; Landis and Koch 1977). Discrepancies between the two observers were discussed and reconciled after the observation, and the reconciled results are used here. Because of our interest in examining the presence of active and passive teaching practices, we treated the collapsed categories in a binary fashion, creating totals for the collapsed categories that reflect the presence or absence of instructor and student behaviors, instead of the density with which these behaviors occurred. For example, an instructor could employ multiple guiding techniques in one 2-min interval (e.g., asking a clicker question and following up on clicker question). Thus, totals for the condensed categories were calculated as follows: if instructors employed several guiding techniques during a 2-min interval, they would receive only one instance of guiding for that interval. Codes for student behavior were treated in the same fashion. For each 2-min interval, it was noted once whether students were engaged in any working, talking, and receiving behaviors. Totals of the condensed instructor (i.e., guiding, presenting, and other) as well as student (i.e., working, receiving, talking, and other) behaviors were obtained by summing the number of 2-min intervals where a certain behavior occurred for the duration of the class.

Analytical methods

To determine if our hypothesis that CoPs encourage evidence-based instruction is supported, we analyzed the class observation data in several ways. We expected the observations to yield quantifiable differences in passive and active teaching practices between groups of instructors by CoP affiliation; these expected differences are listed in Table 1. This analysis was complicated by the non-normal distribution of much of the data and by the

lack of random CoP assignment. These issues are discussed in more detail below.

Collapsed COPUS categories

We initially analyzed the data using Smith and colleague's (Smith et al. 2013) collapsed COPUS categories (Table 1). For instructors, the two collapsed categories of interest were presenting and guiding. The former is associated with lecturing, while the latter corresponds to the use of student-centric activities (e.g., posing questions for individuals or groups, using clickers, and following up on questions). As such, we expected to see CoP instructors to enact significantly more guiding practices than their counterparts.

For students, the three categories of interest were receiving (e.g., listening to the lecture or another student talking), talking (e.g., answering a question the instructor posed), and working (e.g., considering a question individually or as a group, and completing worksheets). Again, for this dimension, we expected to see students in classrooms taught by CoP instructors engaging in significantly higher frequencies of working as opposed to listening and talking when compared to their peers from classes taught by non-CoP instructors.

Shapiro-Wilk tests indicated that the observed distribution for instructor guiding was normal ($W = 0.96$, $p = .06$); however, the remaining variables (instructor presenting, instructor other, student working, student receiving, student talking, and student other) were not normally distributed ($p < .01$). Therefore, the group differences were analyzed using non-parametric Kruskal-Wallis tests. Standardized effect sizes were calculated using Hedge's *g*, which is interpreted similarly to Cohen's *d*. Hedge's *g* is used here because Cohen's *d* tends to overestimate the standardized effect size, particularly for small samples (Lakens 2013). Although there is no consensus concerning the interpretation of standardized effect sizes (Pek and Flora 2017), Hattie (Hattie 2008) conducted a synthesis of over 800 meta-analyses and found that for educational studies, the mean effect size was 0.4 and suggested that effect sizes of 0.2, 0.4, and 0.6 be considered small, medium, and large, respectively. Another measure of the magnitude of the effects can be obtained using the difference in the mean counts for each group to determine the difference in the number of 2-min intervals in which a particular behavior is observed.

An important limitation in the study's design is that CoP affiliation could not be randomly assigned. To control for instructor and course characteristics, we conducted seven multiple regressions with the counts of the seven collapsed student and instructor behaviors described above (i.e., instructor guiding, instructor presenting, instructor other, student working, student receiving, student talking, student other) as the dependent variable, and CoP membership, instructor rank (non-tenure track, assistant professor, and associate or full professor), course enrollment, and course type (foundational or non-foundational) as the independent variables.

Given the relative normality of the instructor guiding distribution, a multiple linear regression was conducted. However, because the distribution of the remaining collapsed student and instructor behaviors are highly skewed count data, multiple regression techniques designed for normally distributed continuous data would lead to biased estimates of regression coefficients and their standard errors (Coxe et al. 2009; Gardner et al. 1995). The simplest distribution for skewed count data is the Poisson distribution. Poisson regression models count data where the possible values of the dependent variable are non-negative integers. Poisson distributions are truncated at zero, skewed to the right when the mean is small, become more Gaussian as the mean increases, and are characterized by a single parameter that describes the mean, variance, and skew (Agresti 1996). One student behavior (student receive) and one instructor behavior (instructor present) occurred very frequently in most of the observed classrooms. Because the distributions of these two codes were truncated at the maximum of 25 and highly negatively skewed, these variables were modeled using the number of 2-min segments where the behaviors were not observed, allowing for the use of a Poisson distribution. The results then can be interpreted by reversing the direction of an effect. For example, if instructors from CoPs are found to have fewer 2-min segments where student receiving was *not* observed, we could therefore conclude that CoP instructors have more 2-min segments where student receiving was observed.

The COPUS protocol records the presence or absence of instructor and student behaviors for each 2-min segment of class. For some collapsed categories, there were several instances where a behavior never occurred during an observation resulting in a count of zero. Poisson regression can underestimate the standard error of regression parameter when there are more zeros than expected under a Poisson distribution. To account for the presence of more zeros than would be expected under a Poisson distribution in the distributions for some of the collapsed categories, we used zero-inflated Poisson (ZIP) regression (Agresti 2010; Gardner et al. 1995). The presence of excessive zeros in the distributions was assessed using Vuong's test (Vuong 1989). The results of the Vuong tests indicate that a ZIP regression model fits better for student working, student other, and the absence of student receiving. A ZIP regression model did not fit the data significantly better for student talking, instructor other, and the absence of instructor presenting. Therefore, these variables were modeled using Poisson regression.

In summary, the relationship between CoP membership and student and instructor behaviors were modeled using seven multiple regressions with the counts of the collapsed student and instructor behaviors as the dependent variable, and CoP membership, instructor rank, course enrollment, and course type (foundational or nonfoundational) as the independent variables. A multiple linear regression was used to model the counts for instructor guiding. A Poisson log-linear model was used to model the counts for student talking and instructor other. A ZIP log-linear model was used to model the counts for student working, student other, the absence of student receiving, and the absence of instructor presenting. In addition, to analyze the relationship between instructor and student behavior for both CoP and non-CoP instructors, we calculated Pearson correlations between student working and instructor guiding, and between student talking and instructor guiding for each group. We present the coefficient of determination (i.e., the square of the Pearson correlation) in the "Results" section because this quantity represents the percentage of variance in one variable that is predictable from the other variable.

Detailed COPUS categories

To examine what specific behavior underlies the observed differences between CoP and non-CoP instructors, we analyzed the original COPUS protocol with all 25 individual codes. Shapiro-Wilk tests indicated that the observed distributions for all of the counts for the detailed COPUS codes were not normally distributed (all had $p < .001$), so group differences were compared using non-parametric Kruskal-Wallis tests. Standardized effect sizes were calculated using Hedge's *g*, while another measure of the magnitude of the effects can be obtained using the difference in the mean counts.

Results

Collapsed COPUS codes

The collapsed COPUS categories (Fig. 1, Table 2) show that CoP instructors and their students engage in significantly more active practices than non-CoP instructors and their students. CoP instructors spent much more time guiding (occurring, on average, in 15 2-min periods rather than in 10: $g = 0.97$, 95% CI $[0.41, 1.49]$) and much less time presenting (occurring, on average, in 19 2-min periods, rather than in 23: $g = -1.40$, 95% CI $[-1.99, -0.83]$). Students in these COP instructor-led lectures spent much more time working (occurring, on average, in eight 2-min periods rather than in two: $g = 1.33$, 95% CI $[0.78, 1.91]$) and less time receiving ($g = -0.86$, 95% CI $[-1.41, -0.33]$). The amount of time students spent talking was similar between the two groups ($g = -0.39$, 95% CI $[-0.91, 0.12]$).

Figure 2 shows the frequency distribution of the number of student working periods, for both COP instructors (blue bars), and for the non-CoP instructors (green bars). Fifteen of the 35 lectures led by instructors who were not part of a COP had zero instances of students doing any individual COPUS code associated with working, while only 1 of the 25 lectures led by a COP instructor exhibited this pattern. Conversely, only 1 of the 35 lectures led by non-CoP instructors had 10 or more 2-min segments that contained student working, while eight of the 25 CoP instructor-led lectures did.

Collapsed COPUS codes: multiple regressions

The differences between the CoP and non-CoP groups are robust when controlling for other factors. A summary of the multiple regression analyses is presented below with 95% confidence intervals in brackets. Complete results can be found in the supplementary file for this article.
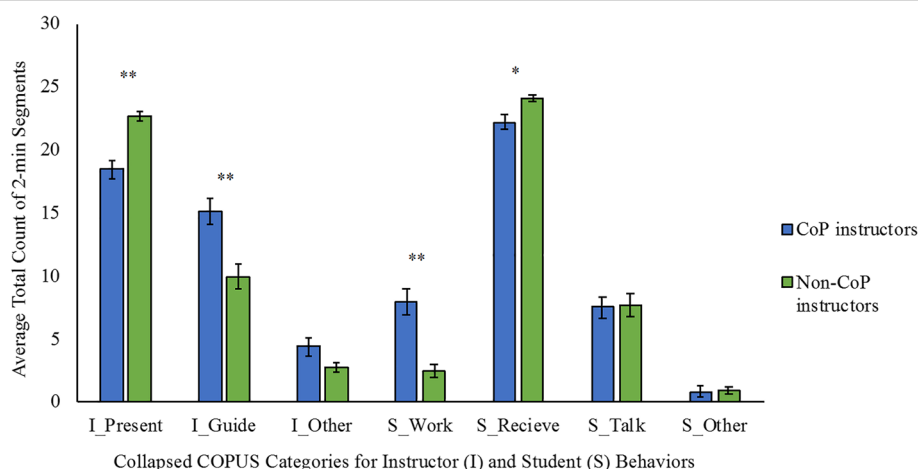

Fig. 1 Student and instructor behavior frequency. Comparison of mean counts for collapsed COPUS categories by CoP ($n = 25$) and non-CoP ($n = 35$) instructors. Significant differences using Kruskal-Wallis tests: *p < .05 , **p < .001

Table 2 Comparisons of central tendency and variation of the total counts of 2-min intervals where behavior, as indicated by the collapsed categories for student and instructor codes, was observed by CoP (*n* = 25) and non-CoP (*n* = 35) instructors

Variable	CoP		Non-CoP		Kruskal-Wallis		Effect size	
	Mean (median)	SD	Mean (median)	SD	$\chi^2(1)$	*p*	Hedge's *g*	[95% CI]
Student behavior								
Work	7.92 (7)	5.04	2.45 (1)	3.17	19.41	< .001	1.33	[0.78, 1.91]
Receive	22.18 (23)	2.94	24.07 (25)	1.36	9.33	.002	− 0.86	[−1.41, −0.33]
Talk	7.50 (8)	4.30	7.66 (5)	5.50	0.01	.93	0.39	[− 0.12, 0.91]
Other	0.84 (0)	2.21	0.90 (0)	1.60	0.86	.35	− 0.03	[− 0.55, 0.48]
Instructor behavior								
Present	18.44 (19)	3.78	22.67 (23)	2.26	19.71	< .001	− 1.40	[−1.99, − 0.83]
Guide	15.14 (16)	5.20	9.94 (10)	5.63	12.20	< .001	0.94	[0.41, 1.49]
Other	4.38 (4)	3.61	2.74 (2)	1.99	2.53	.11	0.33	[− 0.18, 0.85]

Student work

The ZIP regression model indicates that the number of segments when students were working are 2.2 [1.6, 2.9] times greater for CoP instructors while controlling for other factors (*p* < .001). In addition, CoP instructors were less likely to have no 2-min segments where students were working (*p* = .01). The number of segments when students were working are 1.6 [1.04, 2.4] times greater for foundational courses (*p* = .01). Instructor rank was related to the number of segments when students were working. As compared to non-tenure track faculty, the number of segments when students were working were 1.6 [1.1, 2.4] times lower for assistant professors (*p* = .03); however, there was no difference for associate or full professors (*p* = .63). Course size was not related to the number of segments when students were working (*p* = .11).

Student receive

The ZIP regression model indicates that the number of segments when students were *not* receiving is 2.2 [1.3, 3.8] times greater for CoP instructors while controlling for other factors (*p* < .01). In addition, CoP instructors were less likely to have no segments where students were *not* receiving—or, in other words, were less likely to have students receiving in every observed segment (*p* = .04). The number of segments when students were *not* receiving is 3.2 [1.3, 5.5] times lower for foundational courses (*p* < .01). Instructor rank was related to the number of segments when students were receiving. As compared to non-tenure track faculty, the number of segments when students were *not* receiving was 2.7 [1.5,6.8] times lower for associate or full professors (*p* = .03); however, there was no difference for assistant professors (*p* = .51). In other words, the number of segments when students were
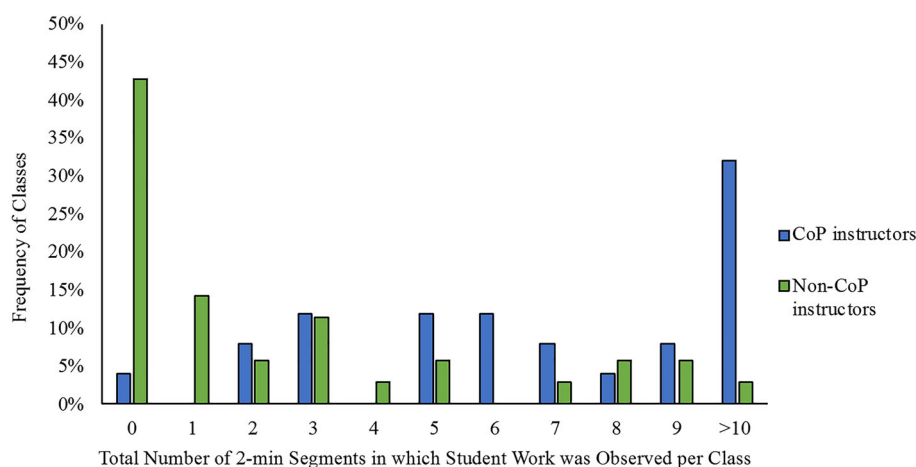
Fig. 2 Student working histogram. Distribution of the frequency of classes in which the collapsed COPUS category of student work was observed, indicated by the total number of 2-min segments, for CoP (*n* = 25) and non-CoP (*n* = 35) instructors

receiving was higher for assistant professors as compared to non-tenure track faculty. Course size was not related to the number of segments when students were receiving ($p = .79$).

Student talk

The Poisson model indicates that the number of segments when students were talking is not different for CoP instructors while controlling for other factors ($p = .52$). However, CoP instructors were less likely to have no segments where students were talking ($p = .01$). Instructor rank was not related to the number of segments when students were talking. The number of segments when students were talking was 1.6 [1.1, 2.2] times lower for assistant professors ($p < .01$); however, there was no difference for associate or full professors ($p = .33$). The number of segments when students were talking is not different for foundational courses ($p = .07$). Course size was not related to the number of segments when students were talking ($p = .52$).

Student other

The ZIP regression model indicates that the number of segments when students were engaged in other behaviors, which describe those behaviors not related to instruction such as waiting for a technical problem to be resolved, is not different for CoP instructors while controlling for other factors ($p = .51$). In addition, CoP instructors were not different in their likelihood to have no segments where students were engaged in other behaviors ($p = .32$). There was no difference between assistant professors ($p = .65$), or associate or full professors ($p = .42$) and non-tenure track faculty; the number of segments when students were engaged in other behaviors. The number of segments when students were engaged in other behaviors is not different for foundational courses ($p = .70$), and course size was not related to the number of segments when students were engaged in other behaviors ($p = .11$).

Instructor guide

The multiple linear regression model indicates that CoP instructors have 6.4 [3.5, 9.3] more segments when instructors were guiding while controlling for other factors ($p < .001$). Foundational courses have 3.8 [0.1, 7.4] more segments when instructors were guiding than non-foundational courses ($p = .04$). Instructor rank was related to the number of segments when instructors were guiding. As compared to non-tenure track faculty, assistant professors have 5.0 [0.6, 9.3] fewer segments where instructors were guiding ($p = .03$); however, there was no difference for associate or full professors ($p = .70$). Course size was not related to the number of segments when instructors were guiding ($p = .88$).

Instructor present

The Poisson model indicates that the number of segments when instructors were *not* presenting is 2.8 [2.1, 3.8] times greater for CoP instructors while controlling for other factors ($p < .001$). The number of segments when instructors were *not* presenting is 1.6 [1.1, 2.2] times lower for foundational courses ($p < .01$). In other words, instructors in foundational courses do more presenting. There was no difference between assistant professors ($p = .40$), or associate or full professors ($p = .11$) and non-tenure track faculty.
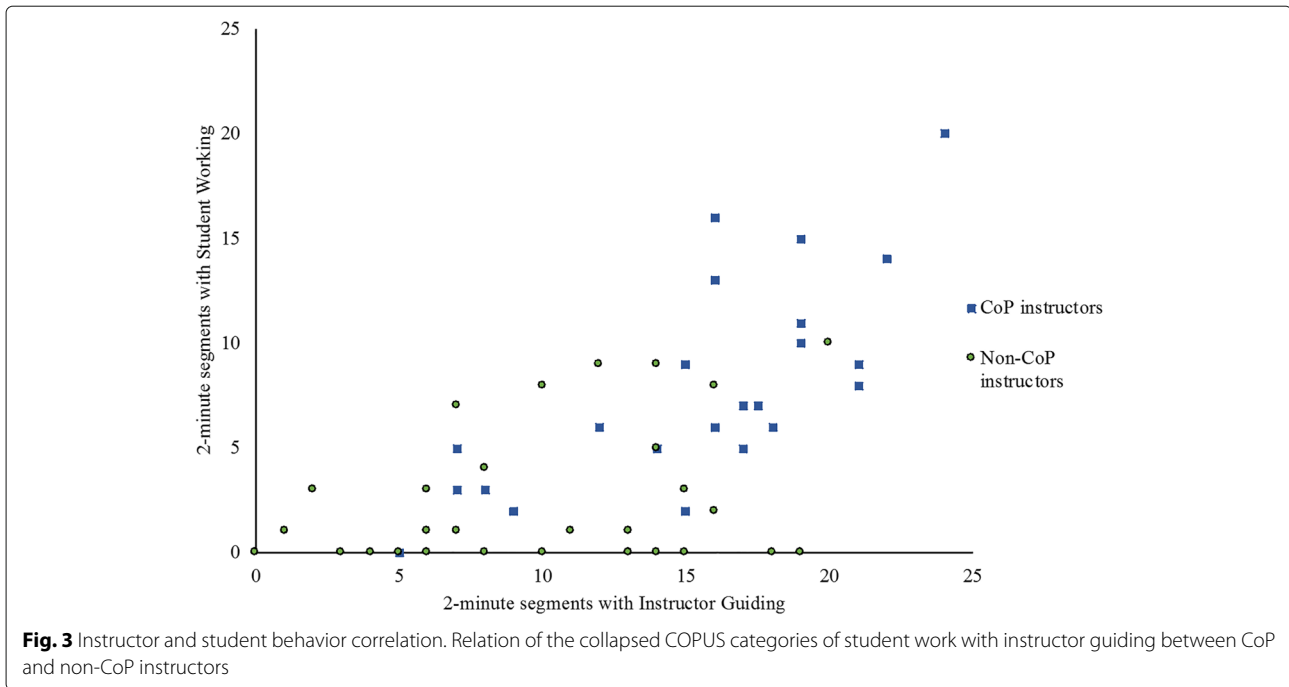
Instructor other

The Poisson model indicates that the number of segments when instructors were engaged in other activities (e.g., turning on lights or troubleshooting technology) is 1.5 [1.1, 2.0] times greater for CoP instructors while controlling for other factors ($p < .001$). The number of segments when instructors were engaged in other activities was 1.7 [1.2, 2.5] times lower for foundational courses ($p < .01$). As compared to non-tenure track faculty, the number of segments when instructors were engaged in other activities was 1.7 [1.1, 2.5] times higher for assistant professors ($p=.02$); however, there was no difference for associate or full professors ($p = .94$). Courses with larger enrollments were more likely to have instructors engaged in other activities ($p < .001$).

Relationship between instructor and student behaviors by communities of practice

For CoP instructors, the correlation between instructor guiding and student working was moderate, with $r^2 = 0.54$. For non-CoP instructors, the correlation between instructor guiding and student working was $r^2 = 0.14$. Instructor guiding and student working were non-normally distributed for non-CoP instructors, however, so we did not conduct significance tests for these correlations. This data is shown in Fig. 3. For CoP instructors, the correlation between instructor guiding and student talking was $r^2 = 0.04$. For non-CoP instructors, it was $r^2 = 0.76$.

Detailed COPUS codes

The descriptive statistics for the detailed COPUS codes from classroom observations are enumerated on Table 3 (instructors) and Table 4 (students), and are grouped by CoP affiliation. Table 3 shows that the single most common instructor activity was lecturing for both CoP and non-CoP instructors. Table 4 shows that the single most common student activity was listening. The maximum number of 2-min observations over a 50-min period is 25, and so the listening values shown on Table 4 are close to the possible maximum. For non-CoP instructors, the median value was at the maximum of 25. In other

Fig. 3 Instructor and student behavior correlation. Relation of the collapsed COPUS categories of student work with instructor guiding between CoP and non-CoP instructors

words, for most of the classes taught by non-CoP instructors, every single 2-min period included time when the students were passively listening.

Although there are similarities between the two groups in the instructor and student behaviors, there are also significant differences. As is shown in Table 3, CoP instructors were far more likely to ask clicker questions ($g = 1.32$, 95% CI [0.75, 1.89]), follow-up on those questions (i.e., clicker follow-up, $g = 1.08$, 95% CI [0.53, 1.64]), and

spend time moving through class guiding student work (i.e., moving and guiding, $g = 0.72$, 95% CI [0.18, 1.25]). Conversely, these instructors were much less likely to be lecturing ($g = -1.48$, 95% CI [$-2.06, -0.90$]) or real-time writing ($g = -0.75$, 95% CI [$-1.28, -0.21$]).

Student behaviors also show significant differences between the CoP and non-CoP instructors (Table 4). Students in classes taught by CoP instructors were much more likely to be actively working in groups on clicker

Table 3 Comparisons of central tendency and variation of the total counts of 2-min intervals where instructor behavior, as indicated by the original list of COPUS codes (from Smith et al. 2013), was observed by CoP ($n = 25$) and non-CoP ($n = 35$) instructors

Variable	CoP		Non-CoP		Kruskal-Wallis		Effect size	
	Mean (median)	SD	Mean (median)	SD	$\chi^2(1)$	p	Hedge's g	[95% CI]
Lecturing	16.20 (18)	5.02	21.81 (22)	2.47	23.63	< .001	1.48	[$-2.06, -0.90$]
Real-time Writing	8.38 (7)	7.56	14.66 (16)	8.77	8.36	.004	-0.75	[$-1.28, -0.21$]
Demo/video	3.30 (1)	4.46	1.41 (0)	2.39	2.59	.108	0.55	[0.02, 1.07]
Clicker follow-up	6.82 (6)	4.99	2.43 (2)	3.12	12.83	< .001	1.08	[0.53, 1.64]
Posing question	6.38 (6)	3.92	6.09 (4)	4.46	0.31	.577	0.07	[$-0.45, 0.59$]
Asking clicker question	6.30 (6)	5.32	1.10 (0)	2.39	19.09	< .001	1.32	[0.75, 1.89]
Answering question	3.26 (3)	2.83	3.32 (2)	3.91	0.31	.575	-0.02	[$-0.54, 0.50$]
Moving and guiding	2.58 (1)	4.39	0.40 (0)	1.33	11.47	< .001	0.72	[0.18, 1.25]
One-on-one discussion	2.04 (0)	4.42	0.40 (0)	1.09	4.93	.026	0.55	[0.02, 1.07]
Administration	1.98 (1)	2.04	2.01 (2)	1.53	0.41	.521	-0.19	[$-0.71, 0.33$]
Waiting	2.16 (1)	2.25	0.87 (0)	1.61	6.88	.009	0.67	[0.14, 1.20]
Other	0.52 (0)	2.02	0.17 (0)	0.57	0.02	.893	0.25	[$-0.27, 0.77$]

Table 4 Comparisons of central tendency and variation of the total counts of 2-min intervals where student behavior, as indicated by the original list of COPUS codes (from (Smith et al. 2013)), was observed by CoP (n = 25) and non-CoP (n = 35) instructors

Variable	CoP		Non-CoP		Kruskal-Wallis		Effect size	
	Mean (median)	SD	Mean (median)	SD	$\chi^2(1)$	p	Hedge's g	[95% CI]
Listening	22.18 (23)	2.94	24.07 (25)	1.36	9.33	.002	− 0.86	[− 1.40, − 0.32]
Ind prob solving	0.54 (0)	1.22	0.84 (0)	1.70	0.26	.610	− 0.19	[− 0.71,0.32]
Group clicker	5.62 (5)	4.72	0.67 (0)	2.23	25.26	< .001	1.40	[0.83, 1.98]
Group working	0.28 (0)	0.98	0.60 (0)	2.05	0.03	.863	− 0.19	[− 0.71, 0.33]
Group other	1.20 (0)	2.80	0.17 (0)	0.57	4.43	.035	0.55	[0.02, 1.08]
Prediction	0.60 (0)	1.58	0.20 (0)	0.68	1.64	.200	0.35	[− 0.18, 0.87]
Test/quiz	0.28 (0)	1.40	0.00 (0)	0.00	1.40	.237	0.31	[− 0.21, 0.83]
Answer question	3.26 (3)	2.83	3.33 (2)	3.91	0.31	.575	− 0.02	[− 0.54, 0.50]
Ask question	2.28 (2)	2.21	3.23 (2)	3.76	0.41	.524	− 0.29	[− 0.81, 0.23]
Class discussion	0.28 (0)	0.89	0.31 (0)	1.16	0.14	.708	− 0.03	[− 0.55, 0.49]
Presentation	0.12 (0)	0.60	0.00 (0)	0.00	1.40	.237	0.31	[− 0.21, 0.83]
Waiting	0.72 (0)	1.86	0.81 (0)	1.60	0.33	.564	− 0.05	[− 0.57, 0.47]
Other	0.12 (0)	0.44	0.03 (0)	0.17	0.83	.361	0.29	[− 0.23, 0.81]

questions (i.e., group clicker, $g = 1.40$, 95% CI [0.83, 1.98]) and other group problems (i.e., group other, $g = 0.55$, 95% CI [0.02, 1.08]) and spent significantly less time passively listening ($g = − 0.86$, 95% CI [− 1.40, − 0.32]).

As part of the lecture observation, a count was made of the number of students attending. Student attendance ranged from 20 to 600 students, with a mean of 136.5 and a median of 112.5. The size of the class, as measured by student attendance, was not statistically different between the CoP and non-CoP instructors, $t(58) = 0.79, p = 0.43$. The mean and one standard deviation of the observed number of students attending the lectures were 151.5 ± 99.7 for CoP instructors, and 129.0 ± 116.0 otherwise. CoP instructor-led courses had higher rates of student attendance, however, as measured by comparing student attendance to course enrollment, with an average of 77% of the enrolled students attending, versus 61% for non-CoP, instructor-led courses. This difference was significant: $t(57) = 3.05, p = .003, g = 0.79$, 95% CI [0.25, 1.33].

Discussion

Getting faculty to adopt evidence-based instructional practices, such as active learning in lectures, is a challenge. If faculty are most likely to learn from one another, then bringing instructors together into voluntary learning communities—CoPs—might be a way to encourage such a change. But are these CoPs associated with the use of evidence-based instructional practices? In this study, we found that they are.

The CoP instructor-led classes showed much higher levels of active learning than those led by non-CoP instructors. Using COPUS as an observational tool, we found that CoP instructors employed significantly more evidence-based instructional practices that we predicted would correlate with active, or reformed, lecturing (Table 1), even when controlling for instructor and class characteristics. Also as predicted, non-CoP instructors were observed using significantly more passive techniques that map onto a traditional-lecturing style (Table 1). This was found to be true for both the individual COPUS codes (Tables 3 and 4) and collapsed categories (Table 2).

Interestingly, results from the collapsed COPUS categories suggest that instructors' use of active learning techniques—as captured by instructor guiding—does not guarantee increases in active learning for students—as captured by student working. Although student working does increase with instructor guiding (Fig. 3), this relationship was much stronger in the CoP instructor classes ($r^2 = 0.54$) than in the non-CoP instructor classes (($r^2 = 0.14$). In other words, the time that CoP instructors spent guiding was more readily transformed into student working than it was for non-CoP instructors. The relationship between instructor guiding and student talking demonstrated the opposite pattern. When non-CoP instructors spent more time guiding, their students were more likely to be recorded as talking ($r^2 = 0.76$). Conversely, for CoP instructors, there was essentially no correlation between instructor guiding and student talking ($r^2 = 0.04$).

This shows that when CoP instructors were engaged in guiding students in the classroom, they did so in the context of engaging students in active work. When non-CoP instructors were engaged in guiding students, they

focused on asking or answering questions of individual students. This approach led to individual students spending time talking with the instructor while the rest of the class listened, which is not an active learning technique. Such differences in instructional approaches across the CoPs can serve as a mechanism to increase instructor understanding about the importance of active learning and, ultimately, to transition the nature of large enrollment STEM lectures from instructor- to student-centric.

The quantitative differences between the CoP and non-CoP instructors' classes are consistent with what would be observed if the two groups differed in their lecturing styles. As discussed earlier, two previous studies (Lund et al. 2015; Stains et al .2018) used cluster analysis of COPUS observations to classify lecturing styles. The average values for each of the two groups in our study naturally separate according to the the clusters found in these studies. The non-CoP instructors lectured in 87% of the 2-min periods and answered student questions in 13% of the periods, while students did group work in just under 10% of periods. These are within one standard deviation of the averages reported by (Stains et al. 2018) for the didactic model and inside the interquartile range of the lecturing model reported by (Lund et al. 2015). The other COPUS observations are also consistent with these clusters, and no others. As the students in non-CoP instructor-led courses spent on 13% of the periods asking questions, Lund and colleagues (Lund et al. 2015) would classify these as lecturing rather than socratic classes, for example.

CoP instructors, on the other hand, lectured in 65% of the 2-min periods, performed follow-up activities in 27% of the periods, and students did group work in 32% of periods. These are within one standard deviation of the average values of the peer instruction (Stains et al. 2018) cluster and inside the interquartile range of the interactive lecture cluster (Lund et al. 2015). The observations are not consistent with a classification in any of the other clusters.

Given the average instructor and student behaviors observed in our sample, we consider traditional, passive-lecture styles to be consistent with the lecturing and didactic modes of instruction found previously. In contrast, active learning lecture styles are consistent with peer instruction and interactive lecture modes as they make less use of direct lecture and require students to spend less time passively listening and more time working and collaborating. CoP instructors used, on average, an active learning lecture style, while non-CoP instructors used, on average, a traditional lecture style.

The impact of CoP-based reforms on active learning in large STEM lectures

The difference in student working between the Cop and non-Cop instructors was both large and significant, with a standard effect size of 1.3. Other measures of instructor

and student activity showed similar differences: CoP-led classes had more active practices and fewer passive practices than non-CoP-led classes. This suggests that the reform was an effective way in which to increase the amount of active-learning used in STEM lectures.

We predict that a replication of a CoP-based reform would result in more student activity in large lectures, but we do not expect the overall effect size of an intervention to be as large as we found in this study, for two reasons.

Firstly, as this was a voluntary reform, not all of the instructors invited to take part formed CoPs. Measures of an institution-wide reform effort would need to include those instructors who did not choose to participate when calculating the overall impact.

Secondly, as participation in the CoPs was voluntary, the variable of interest (CoP vs. non-CoP) was not randomly assigned. This means that we cannot simply make the causal claim that CoP membership was responsible for the differences between the two groups. It is very likely that some of the observed correlation was the consequence of unobserved instructor characteristics. For example, an instructor who is interested in pedagogy is more likely to both (1) use active learning in lectures, and (2) participate in an education-focused CoP. This is the "correlation versus causation" problem and implies that at least some fraction of the observed difference was not actually caused by the reform.

There is evidence that the relationship was at least partly causal, however, and that the CoPs did increase the time spent in instructor guiding and student working. We can say this as we know that the CoPs were instrumental in the adoption of student response technologies (iClickers). CoP instructors used this technology significantly more than non-CoP instructors (in a way, we described as being active-learning; see "clicker follow-up" and "asking clicker question" in Table 3, and "group clicker" on Table 4). In a 2015 survey of the CoP instructors in the study (Herman and Mena 2015), 83% of program participants noted that they began to use iClickers as a consequence of the reforms. A subsequent study (Ma et al.) mapped the spread of instructional technology across departments and found that CoP mentors were important in transferring their use; that this transfer occurred between CoPs; and that this transfer coincided with the reform effort. We can therefore say that the CoPs either started or accelerated the adoption of a technology that significantly increased instructor guiding and student working scores in this study.

Future directions
Because this study shows that instructor participation in CoPs is associated with increased use of active learning techniques in STEM lectures, we propose that CoPs should be used as a mechanism to enhance

STEM students' learning experiences. A promising—but challenging—area of future research is to therefore investigate how and why CoPs form and flourish. Although the authors of this paper are working to understand this question, the diverse nature of STEM faculty and college settings implies that multi-institutional data and a variety of approaches are required to make definitive recommendations. From the results of this study and those of a previous network analysis (Ma et al.), we suggest that the CoPs should consist of small, disciplinary teams working on the same course or set of related courses, and that they should be linked to other individuals or groups that use active learning techniques.

Given that there is an effort to create CoPs, the subsequent persistence of CoPs appear to depend on idiosyncratic properties of the individuals in each department, as the group characteristics of the CoP and non-CoP instructors do not systematically differ. The individual and departmental characteristics that are important in the adoption and spread of evidence-based instructional practices are not known and are a worthy area of future study. Although CoPs cluster by department (see Fig. 2 from (Ma et al.)), the type of department was not correlated to whether or not a CoP would form given the opportunity. Just one of each of the following pairs of departments had a functioning CoP as a consequence of the reform program: Integrative Biology/Molecular and Cellular Biology, Civil and Environmental Engineering/Geology, Physics/Mathematics, and Electrical Engineering/Mechanical Engineering. It is not obvious that these paired departments would attract STEM instructors with consistently differing views on pedagogy.

Conclusion

We found that educational communities of practice (CoPs) were strongly associated with active learning in undergraduate STEM lectures at a large university. As measured by the Classroom Observation Protocol for Undergraduate STEM (COPUS) instructors who were members of a CoP spent less time lecturing to, and more time guiding the learning of, their students. CoP instructors were also more effective than other instructors at translating this guiding time into student activity. As a consequence of these differences, the students in CoP instructor-led lectures spent less time passively listening, and more time actively working, than they did in lectures led by instructors who were not part of a CoP. These differences were large and statistically significant, with the presence of CoPs associated with standardized effect sizes of -0.9 for a measure of student passivity, and 1.3 for a measure of student activity. We encourage further efforts to form CoPs, as our results suggest this is a promising method with which to reform undergraduate STEM lectures.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1] School of Earth, Society, and the Environment, University of Illinois at Urbana–Champaign, Green St, Urbana 61801, IL, USA. [2] Department of Educational Psychology, University of Illinois at Urbana–Champaign, 1310 S 6th St, Champaign 61820, IL, USA. [3] Department of Computer Science, University of Illinois at Urbana–Champaign, 201 N. Goodwin Ave, Urbana 61801, IL, USA.

References
Agresti, A. (2010). *Analysis of Ordinal Categorical Data, vol. 656*. New York: Wiley.
Agresti, A. (1996). *An Introduction to Categorical Data Analysis, vol. 135*. New York: Wiley.
Brewer, CA, & Smith, D (2011). *Vision and change in undergraduate biology education: a call to action*. Washington, DC: American Association for the Advancement of Science.
Austin, A (2011). *Promoting evidence-based change in undergraduate science education*. Washington, DC: National Academies National Research Council.
Borrego, M., Froyd, J.E., Henderson, C., Cutler, S., Prince, M. (2013). Influence of engineering instructors' teaching and learning beliefs on pedagogies in engineering science courses. *International Journal of Engineering Education*, *29*(6), 1456–1471.
Brownell, S., & Tanner, K. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and tensions with professional identity. *CBE-Life Sciences Education*, *11*(4), 339–346.
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.
Cox, A (2005). What are communities of practice? A comparative review of four seminal works. *Journal of Information Science*, *31*(6), 527–540.
Coxe, S, West, S.G., Aiken, L.S. (2009). The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121–136.

Davenport, T., & Prusak, L. (1970). L.(2000). Working knowledge: How organizations manage what they know.

Fairweather, J. (2009). *Linking evidence and promising practices in science, technology, engineering and mathematics (STEM) undergraduate education*: National Academies National Research Council Board of Science Education. http://www.nsf.gov/attachments/117803/public/Xc--Linking_Evidence--Fairweather.pdf.

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*, *111*(23), 8410–8415.

Gardner, W., Mulvey, E.P., Shaw, E.C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, *118*(3), 392.

Gehrke, S., & Kezar, A. (2017). The roles of STEM faculty communities of practice in institutional and departmental reform in higher education. *American Educational Research Journal*, *54*(5), 803–833.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S.M., Wood, W.B. (2004). Scientific teaching. *Science*, *304*(5670), 521–522.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 Meta-analyses relating to achievement*. London: Routledge.

Henderson, C., Beach, A., Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytical review of the literature. *Journal of Research in Science Teaching*, *48*, 952–984.

Henderson, C., Dancy, M., Niewiadomska-Bugaj, M. The 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*(2), 020104.

Herman, G.L., Goldberg, D.E., Trenshaw, K.F., Somerville, M., Stolk, J. (2017). The intrinsic-motivation course design method. *International Journal of Engineering Education*, *33*(2A), 558–574.

Herman, G.L., Greene, J.C., Hahn, L.D., Mestre, J.P., Tomkin, J.H., West, M. (2018). Changing the teaching culture in introductory stem courses at a large research university. *Journal of College Science Teaching*, *47*(6), 32–38.

Herman, G.L., & Mena, I.B. (2015). Tracking the spread of research-based instructional strategies, In *Frontiers in Education Conference (FIE), 2015 IEEE* (pp. 1–5). Los Angeles: IEEE.

Hildreth, P.M., & Kimble, C. (2002). The duality of knowledge.

Ianquinto, B., Ison, R., Faggian, R. (2011). Creating communities of practice: Scoping purposeful design. *Journal of Knowledge Management*, *15*(1), 4–21.

Jamieson, L., & Lohman, J. (2012). *Innovation with impact: Creating a culture for scholarly and systematic innovation in engineering education*. Washington, DC: American Society for Engineering Education.

Kegan, R., & Lahey, L.L. (2009). Immunity to change: How to overcome it and unlock the potential in yourself and your organization.

Kezar, A, Gehrke, S, Bernstein-Sierra, S (2017). Designing for success in STEM communities of practice: Philosophy and personal interactions. *The Review of Higher Education*, *40*(2), 217–244.

Kezar, A., Gehrke, S., Elrod, S. (2015). Implicit theories of change as a barrier to change on college campuses: an examination of stem reform. *The Review of Higher Education*, *38*(4), 479–506.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, *4*, 863.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation, vol. 521423740*. Cambridge: Cambridge University Press.

Lea, M. (2005). Communities of practice in higher education. In D Barton & K Tusting (Eds.), *Beyond Communities of Practice: Language, Power and Social Context* (pp. 180–197). New York: Cambridge University Press.

Lesser, E.L., & Storck, J. (2001). Communities of practice and organizational performance. *IBM Systems Journal*, *40*(4), 831–841.

Lund, T.J., Pilarz, M., Velasco, J.B., 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*(2), 18.

Ma, S., Herman, G.L., West, M., Tomkin, J., Mestre, J. Studying STEM faculty communities of practice through social network analysis. *The Journal of Higher Education*, *in submission*.

Pek, J., & Flora, D.B. (2017). Reporting effect sizes in original psychological research: A discussion and tutorial.

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: Executive Office of the President. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf.

Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.

Seymour, E (2002). Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education*, *86*(1), 79–105.

Sheppard, S.D., Antonio, A.L., Brunhaver, S.R., Gilmartin, S.K. (2014). Studying the career pathways of engineers. In A Johri & B Olds (Eds.), *Cambridge Handbook of Engineering Education Research*. New York: Cambridge University Press.

Singer, S.R., Nielsen, N.R., Schweingruber, H.A. (2012). *Discipline-based education research*. Washington, DC: The National Academies.

Smith, M.K., Jones, F.H., 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.

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.

Spalter-Roth, R., Mayorova, O., Scelza, J., Vooren, N.V. (2010). *Teaching alone? sociology faculty and the availability of social networks*. Washington, DC: American Sociological Society Research Briefs.

Stains, M., Harshman, J., Barker, M., Chasteen, S., Cole, R., DeChenne-Peters, S., Eagan, M., Esson, J., Knight, J., Laski, F., et al. (2018). Anatomy of stem teaching in north american universities. *Science*, *359*(6383), 1468–1470.

Tanner, K., & Allen, D. (2006). Approaches to biology teaching and learning: On integrating pedagogical training into the graduate experiences of future science faculty. *CBE-Life Sciences Education*, *5*(1), 1–6.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, *1*(1), 77–100.

Villachia, S., Marker, A.W., Plumlee, D., Huglin, L., Chegash, A. (2013). The arrows in our backs: Lessons learned trying to change the engineering curriculum, In *120th American Society for Engineering Education Annual Conference & Exposition, Atlanta, GA*.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307–333.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems Thinker*, *9*(5), 2–3.

Wenger, E., McDermott, R.A., Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Cambridge: Harvard Business Press.