

## Article

# A Campus-Wide Investigation of Clicker Implementation: The Status of Peer Discussion in STEM Classes

Justin D. Lewin,<sup>\*†</sup> Erin L. Vinson,<sup>\*†</sup> MacKenzie R. Stetzer,<sup>†‡</sup> and Michelle K. Smith<sup>\*†</sup>

<sup>\*</sup>School of Biology and Ecology, University of Maine, Orono, ME 04469-5751; <sup>†</sup>Maine Center for Research in STEM Education, University of Maine, Orono, ME 04469-5727; <sup>‡</sup>Department of Physics and Astronomy, University of Maine, Orono, ME 04469-5709

Submitted October 26, 2015; Revised December 19, 2015; Accepted January 5, 2016  
Monitoring Editor: Peggy Brickman

At the University of Maine, middle and high school teachers observed more than 250 university science, technology, engineering, and mathematics classes and collected information on the nature of instruction, including how clickers were being used. Comparisons of classes taught with ( $n = 80$ ) and without ( $n = 184$ ) clickers show that, while instructional behaviors differ, the use of clickers alone does not significantly impact the time instructors spend lecturing. One possible explanation stems from the observation of three distinct modes of clicker use: peer discussion, in which students had the opportunity to talk with one another during clicker questions; individual thinking, in which no peer discussion was observed; and alternative collaboration, in which students had time for discussion, but it was not paired with clicker questions. Investigation of these modes revealed differences in the range of behaviors, the amount of time instructors lecture, and how challenging the clicker questions were to answer. Because instructors can vary their instructional style from one clicker question to the next, we also explored differences in how individual instructors incorporated peer discussion during clicker questions. These findings provide new insights into the range of clicker implementation at a campus-wide level and how such findings can be used to inform targeted professional development for faculty.

## INTRODUCTION

A number of national reports informed by emerging education research have advocated for active-engagement instruction in postsecondary science, technology, engineering, and mathematics (STEM) courses (American Association for the Advancement of Science, 2010; President's Council of Advisors on Science and Technology [PCAST], 2012; Singer *et al.*, 2012). Moreover, a recent comprehensive meta-analysis of 225 science education research articles indicates that students learn more in and are less likely to drop out of STEM courses

that use these active-engagement instructional approaches (Freeman *et al.*, 2014). One such instructional approach involves instructors posing multiple-choice conceptual questions, fostering peer discussion about these questions among the students, and asking students to indicate their answers via personal response systems or clickers.

Clickers are electronic voting devices that allow instructors to obtain real-time student responses to multiple-choice questions in order to assess student thinking and to inform instruction (e.g., Mazur, 1997; Caldwell, 2007; Smith *et al.*, 2011). A recent nationwide survey found that 86% of U.S. college faculty members are familiar with clickers and 12% of faculty members have adopted clickers in their own classrooms (FTI Consulting, 2015). In addition, a study of student engagement in a large-enrollment undergraduate science class found the use of clicker questions and the follow-up to clicker questions to be the most engaging of all in-class activities observed, as measured by the Behavioral Engagement Related to Instruction protocol (Lane and Harris, 2015).

It is often suggested that clicker questions be supported by an instructional strategy known as peer instruction (Mazur, 1997). In a high-fidelity enactment of peer instruction, 1) the instructor poses a multiple-choice conceptual

CBE Life Sci Educ March 1, 2016 15:ar6

DOI:10.1187/cbe.15-10-0224

Address correspondence to: Michelle K. Smith (michelle.k.smith@maine.edu).

© 2016 J. D. Lewin *et al.* CBE—Life Sciences Education © 2016 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

question; 2) students are given time to think; 3) students determine their individual answers and vote; 4) if there is variation in the student answers, neighboring students discuss their answers with one another; 5) students vote again after peer discussion; and 6) the instructor explains the correct answer to the whole class, often displaying a histogram of all student responses and soliciting explanations from students for incorrect and correct answers. Peer instruction encourages student interactions during lecture and breaks the monotony of passive listening while also offering an opportunity for the instructor to walk around the room and interact with students. These interactions allow the instructor to gauge the level of student understanding and thus gain insight into incorrect lines of reasoning.

Multiple studies have shown that the peer discussion portion of peer instruction increases student performance on clicker questions (Smith *et al.*, 2009; Porter *et al.*, 2011; Knight *et al.*, 2015; Barth-Cohen *et al.*, 2016) and that peer discussion produces higher performance outcomes when compared with other tasks such as quiet reflection (Lasry *et al.*, 2009). Furthermore, the largest gains in student performance occur when peer discussion is immediately followed by an instructor explanation (Smith *et al.*, 2011). Notably, students have more positive attitudes about the utility of clickers when faculty encourage peer discussion and are successfully able to create opportunities for students to discuss the multiple-choice questions (Keller *et al.*, 2007).

Although the learning benefits of using clicker questions with peer discussion have been documented through the use of carefully designed protocols, it has also been noted that faculty often change and modify research-based pedagogies and tools, such as clickers, in their classrooms (Henderson and Dancy, 2007; National Research Council [NRC], 2013). For example, survey results from faculty in multiple disciplines showed that 15% of faculty who use clickers reported that they did not allow or did not encourage peer discussion during clicker questions (Keller *et al.*, 2007). In another observation-based study of undergraduate physics classrooms, researchers found that none of the faculty had students record their individual answers before talking with peers (Turpen and Finkelstein, 2009), a component of peer instruction that is advocated by researchers (Mazur, 1997; Smith *et al.*, 2011).

In this study, we used classroom observation data from 21 different STEM departments to explore the spectrum of instructional practices associated with clickers. Specifically, we asked: 1) Are there differences in instructional behaviors in classes that are taught with and without clicker questions? 2) In classes that use clickers, is there variation in how click-

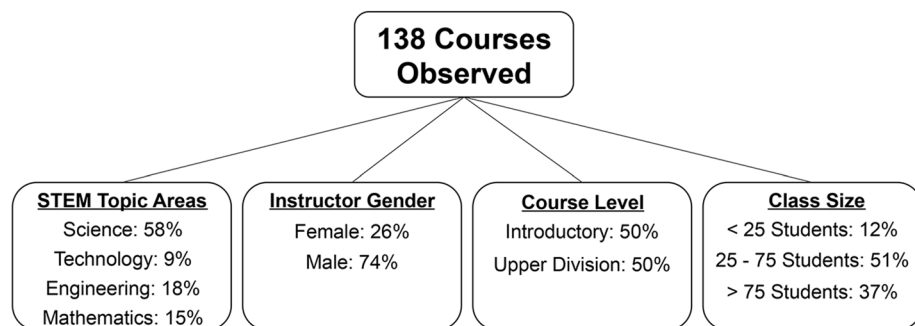
ers are implemented? 3) How do individual instructors vary their implementation of clicker questions? The answers to all three questions are critical for identifying common-use cases on which to focus future research and for optimizing faculty professional development so that it may better support the effective implementation of clickers.

## METHODS

For this investigation, University of Maine STEM instructors were sent emails asking them if they would allow middle and high school teachers to visit their classrooms and collect observation data. The instructors were receptive, with 74% agreeing to allow the teachers to observe their courses. Faculty who declined typically cited reasons such as giving an exam, canceling class, or having a guest lecturer present on the day of the proposed observation.

Observations were conducted in both February and April during the Spring 2014 and Spring 2015 semesters and in November during the Fall 2014 semester. Altogether, 270 class sessions were observed. These observations represented 119 instructors who taught 138 courses in 21 different departments (biology and ecology; chemical and biological engineering; chemistry; civil and environmental engineering; computer sciences; earth sciences; ecology and environmental sciences; economics; electrical and computer engineering; food and agriculture; forest resources; marine science; mathematics and statistics; mechanical engineering; molecular and biomedical science; new media; nursing; physics and astronomy; plant, soil, and environmental science; psychology; and wildlife, fisheries, and conservation biology). On average there were 12.6 (SE  $\pm$  2.5) class sessions observed per department. Demographic information about the types of courses and instructors observed is included in Figure 1. Data from 97 observations from Spring 2014 were reported in an earlier study (Smith *et al.*, 2014).

All faculty members who agreed to be observed were given a human subjects consent form. Approval to evaluate teacher observation data of classrooms (exempt status, protocol no. 2013-02-06) was granted by the institutional review board at the University of Maine. Because of the delicate nature of sharing observation data with other faculty members and administrators, the consent form explained that the data would only be presented in aggregate and would not be subdivided according to variables such as department. Faculty members were given access to observation data from their own course(s) upon request after we collected observation



**Figure 1.** Demographic information about the courses observed at the University of Maine. The STEM designation is based on the course title; the description “Introductory” vs. “Upper Division” is based on the order in which the course is taken in the major.

and survey data for this study. In total, 68% of the observed faculty members requested their data and met with a professional development coordinator to discuss the results.

### Selection and Training of Middle and High School Teachers

Thirty-eight teachers from the state of Maine conducted the classroom observations. To record instructional behaviors in the classroom, the middle and high school teachers were trained to use the Classroom Observation Protocol for Undergraduate STEM (COPUS) according to the training procedure outlined by Smith *et al.* (2013). COPUS was adapted from the Teaching Dimensions Observation Protocol (Hora *et al.*, 2013; Hora and Ferrare, 2014). Briefly, at the beginning of the training, the 25 COPUS codes and code descriptions (Table 1)

**Table 1.** COPUS instrument codes used to describe instructor and student behaviors in class and a description of the collapsed COPUS codes used to compare class sessions

	COPUS codes	Collapsed codes
What the instructor is doing	Lec—lecturing or presenting information RtW—real-time writing D/V—showing or conducting a demo, experiment, simulation, etc.	Presenting
	FIUp—follow-up/feedback on clicker question or activity to class PQ—posing nonclicker question to students (nonrhetorical) CQ—asking clicker question (entire time, not just when first asked) AnQ—listening to and answering student questions to entire class MG—moving through class guiding ongoing student work 1o1—one-on-one extended discussion with individual students	Guiding
	Adm—administration (assign homework, return tests, etc.)	Administration
	W—waiting (instructor late, working on fixing AV problems, etc.) O—other	Other
	L—listening to instructor	Receiving
	Ind—individual thinking/problem solving CG—discussing clicker question in groups of two or more students WG—working in groups on worksheet activity OG—other assigned group activity Prd—making a prediction about a demo or experiment TQ—test or quiz	Students working
	AnQ—student answering question posed by instructor SQ—student asks question WC—students engaged in whole-class discussion SP—students presenting to entire class	Students talking to class
	W—waiting (instructor late, working on fixing AV problems, etc.) O—other	Other
What the students are doing		

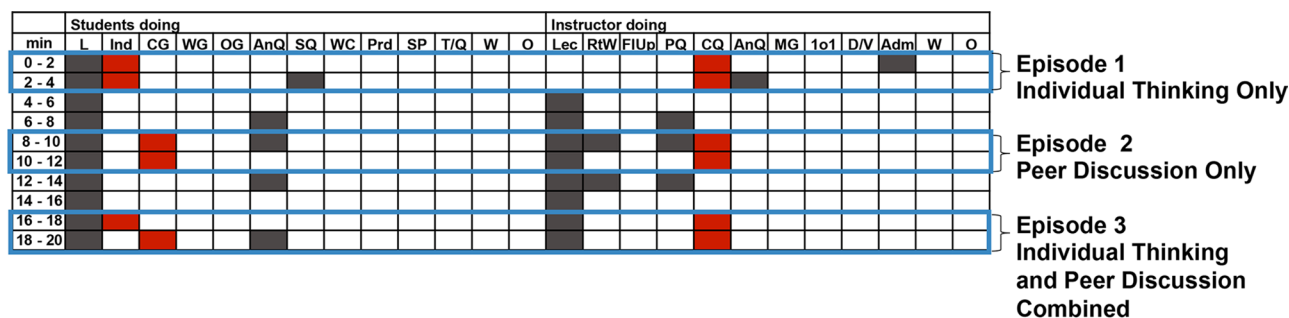
were discussed with the teachers (sample COPUS protocol sheets can be found in Smith *et al.*, 2013, and at [www.cwsei.ubc.ca/resources/COPUS.htm](http://www.cwsei.ubc.ca/resources/COPUS.htm)). The teachers then practiced coding videos of classrooms and discussed codes that were not unanimously selected. In total, the training took approximately 2 h. Teachers conducted observations in pairs, and we calculated Cohen's kappa scores to measure interrater reliability (Landis and Koch, 1977) for every in-class observation; more details are included in the next section.

### Analyzing COPUS Data

Pairs of teachers observed classes and were instructed to record their COPUS results independently. Cohen's kappa interrater scores were calculated for each observation pair to establish coder reliability. The mean Cohen's kappa was 0.89 ( $SE \pm 0.006$ ) for all observations. Because some of the paired observations had low interrater reliability scores, we removed the six lowest paired observations, which all had Cohen's kappa interrater scores below 0.650. The average Cohen's kappa interrater score for the remaining 264 observations was 0.91 ( $SE \pm 0.005$ ), indicating strong agreement among paired observers (Landis and Koch, 1977). With strong reliability between coders, only codes that both observers marked during each 2-min interval were included in the data set analyzed for this study.

To determine the relative abundance of each COPUS code, we added the total number of times each code was marked and divided by the total number of codes, resulting in a percent of code. For example, if both observers marked instructor lecture (Lec) during the same 24 time intervals in a 50-min class period and marked 29 total instructor codes for the duration of the class, then 24/29 or 82.8% of the instructor codes correspond to lecture. Because it was difficult to visually represent and compare 25 COPUS codes in 264 different class sessions, we also collapsed the codes into four categories that describe what the instructor is doing and four categories that describe what the students are doing, as reported in Smith *et al.* (2014) and shown in Table 1. The percentage of each collapsed code was determined by adding the percent code of each individual code within the collapsed category. This collapsed-code representation is advantageous because it allows for a holistic view of multiple COPUS codes at the same time and facilitates comparisons across broad instructional approaches.

However, when trying to determine and compare the frequency of a single code, such as instructor lecturing (Lec) or student listening (L), percent-code calculations can be difficult to interpret, because multiple COPUS codes can be marked at the same time, which in turn can impact the denominator of the calculation (Lund *et al.*, 2015). In particular, some codes are often marked together, such as instructor real-time writing (RtW) and instructor lecturing (Lec). Therefore, we also compared class sessions by calculating the percentage of 2-min time intervals in which specific codes, such as instructor lecturing (Lec) or student listening (L), were observed. The percentage of 2-min time intervals was determined by counting the number of 2-min time intervals in which each code was marked and then dividing that by the total number of time intervals that were coded. For example, if instructor lecture (Lec) was marked during 24 two-minute time intervals out of the possible 25 two-minute time intervals then 24/25 or 96.0% of the possible 2-min time intervals contained lecture.



**Figure 2.** Example excerpt of COPUS codes from a clicker class session with three clicker episodes. Abbreviated COPUS codes, described in Table 1, are along the top; the 2-min time intervals are along the left side. Student and instructor codes of interest are shaded red, with blue boxes surrounding each clicker episode.

### Analyzing Instances of Clicker Use

For this study, we were particularly interested in class sessions that used clicker questions. At the University of Maine, instructors largely started implementing clickers in their classrooms in 2005 (Strukov, 2008), and the Faculty Development Center estimates that currently more than 5000 students are enrolled in courses that use clickers each academic year. To determine which class sessions used clickers, we looked for instances in which the instructor clicker question (CQ) code was marked by both observers during a single 2-min interval. The CQ code was identified in 80 of the 264 class sessions observed.

To find out more about how the clicker questions were used in each of the 80 class sessions, we looked for blocks of 2-min time intervals with individual or consecutive instructor clicker question (CQ) codes, and we called these “clicker episodes” (Figure 2). In total, 181 clicker episodes were observed, and the duration of each episode was determined by counting the number of consecutive 2-min time intervals marked with clicker question (CQ) codes. Overall, clicker episodes had a mean duration of 2.4 (SE  $\pm$  0.11) 2-min time intervals; therefore, the average clicker episode was less than 5 min ( $2 \times 2.4 = 4.8$ ) in duration. To determine how clicker questions were used during these episodes, we examined student behaviors during the same time intervals. Only two student behaviors were selected along with instructor CQ: individual thinking (Ind) and clicker group discussion (CG). Thus, there were three possible combinations of student behaviors during clicker episodes: Individual Thinking Only, Peer Discussion Only, and Individual Thinking and Peer Discussion Combined. For example, Figure 2 shows a class session with three clicker episodes, indicated by three segments of time with instructor clicker question (CQ) codes that are separated by one or more 2-min time intervals. The first clicker episode is solely characterized by individual thinking (Ind) student codes and is classified as Individual Thinking Only. The second clicker episode has clicker group work (CG) only and is classified as Peer Discussion Only. The third clicker episode has individual thinking (Ind) student codes followed by clicker group discussion (CG) codes, and is classified as Individual Thinking and Peer Discussion Combined.

### Clicker Use in Class Sessions

Upon analysis of all clicker episodes, two broad class session modes were identified, those with peer discussion during

clicker questions and those without peer discussion during clicker questions (Figure 3). Peer Discussion class sessions had at least one clicker episode with peer discussion, indicated by a clicker group discussion (CG) code with a corresponding instructor clicker question (CQ) code (Figure 4A). These class sessions may have also included student individual (Ind) or group work (WG, OG) codes, but the presence of at least one student CG code defined this mode.

The class sessions without peer discussion during clicker questions were further classified into two modes, Individual Thinking and Alternative Collaboration (Figure 3). Individual Thinking class sessions had no clicker episodes with peer discussion, and were thus characterized by the presence of instructor clicker question (CQ) codes paired with student individual thinking (Ind) codes (Figure 4B). In the Individual Thinking class sessions, students never discussed class material in groups. Alternative Collaboration class sessions had no clicker episodes with peer discussion, and therefore individual thinking (Ind) was the only student code that coincided with instructor clicker question (CQ) codes. However, Alternative Collaboration class sessions included worksheet-based group work (WG) or other group work (OG) at another point in the class period (Figure 4C). While students in Alternative Collaboration class sessions voted on clicker questions as individuals, just like students in Individual Thinking class sessions, observer notes suggested that the clicker questions were often tied to the group activities. Thus, students in some Alternative Collaboration class sessions had the opportunity to discuss relevant question material with peers, just not in the context of the clicker questions themselves.

To determine whether there is variation in how individual instructors use clickers at different times, we looked at clicker data for instructors who were observed teaching with at least two clicker episodes ( $n = 25$  instructors). Specifically, we examined the percentage of clicker questions that had Individual Thinking Only, Peer Discussion Only, and Individual Thinking and Peer Discussion Combined. The percent was calculated by dividing the number of episodes in each category by the sum of all episodes for that instructor.

### Observer Feedback

To collect additional information about class sessions and to give the middle and high school teachers opportunities to reflect on the instruction, we developed a feedback survey for



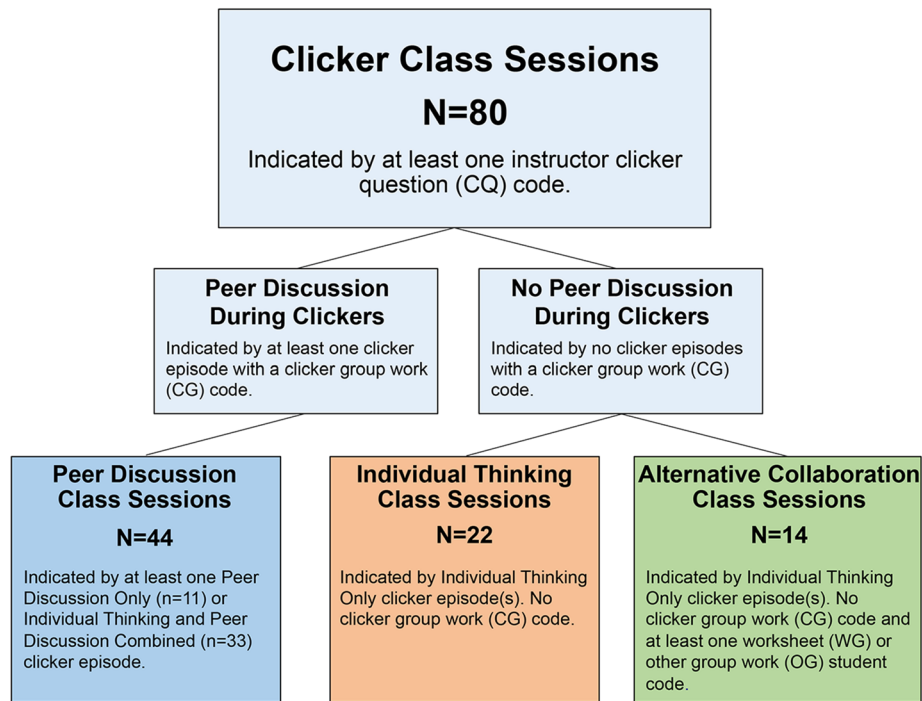


Figure 3. Description of the three distinct modes of clicker class sessions: Peer Discussion, Individual Thinking, and Alternative Collaboration. These modes were identified based on the presence of an instructor clicker question (CQ) code and four student codes: individual thinking (Ind), clicker group discussion (CG), worksheet group work (WG), and other group work (OG).

the teachers to complete after each observation. The survey was developed during Summer 2014, piloted in Fall 2014, revised based on teacher interviews and written feedback, and implemented in Spring 2015. Observers completed the sur-

vey for each observation in pairs, discussing their reasoning for each answer. Discussions were audio recorded to monitor the usefulness of the survey and to ascertain whether or not additional clarification was needed on any items.

### A. Peer Discussion Class Session

	Students doing													Instructor doing												
min	L	Ind	CG	WG	OG	AnQ	SQ	WC	Prd	SP	T/Q	W	O	Lec	RtW	FIUp	PQ	CQ	AnQ	MG	1o1	D/V	Adm	W	O	
0 - 2																										
2 - 4																										
4 - 6																										
6 - 8																										
8 - 10																										

### B. Individual Thinking Class Session

	Students doing													Instructor doing												
min	L	Ind	CG	WG	OG	AnQ	SQ	WC	Prd	SP	T/Q	W	O	Lec	RtW	FIUp	PQ	CQ	AnQ	MG	1o1	D/V	Adm	W	O	
0 - 2																										
2 - 4																										
4 - 6																										
6 - 8																										
8 - 10																										

### C. Alternative Collaboration Class Session

	Students doing													Instructor doing												
min	L	Ind	CG	WG	OG	AnQ	SQ	WC	Prd	SP	T/Q	W	O	Lec	RtW	FIUp	PQ	CQ	AnQ	MG	1o1	D/V	Adm	W	O	
0 - 2																										
2 - 4																										
4 - 6																										
6 - 8																										
8 - 10																										

Figure 4. Example excerpt of COPUS codes for the three class modes: (A) Peer Discussion class sessions, which include the presence of at least one student clicker group discussion (CG) code. (B) Individual Thinking class sessions, in which students did not have the opportunity to talk in groups, so only the student individual thinking (Ind) code was selected. (C) Alternative Collaboration class sessions, in which students discussed material in groups (student codes worksheet group work [WG] and/or other group work [OG]) but did not have peer discussion during the clicker question. Abbreviated COPUS codes, described in Table 1, are listed along the top; the 2-min time intervals are indicated along the left side. Student and instructor codes of interest are shaded red.

A section of this survey included a question about how challenging the clicker questions were for students. In general, the clicker questions were:

1. Challenging for students—the class vote was often split.
2. Easy for students—the majority of the students answered correctly.
3. Cannot determine—the instructor did not talk about the class voting results.

To help teachers answer this question, during training sessions, we watched training videos of instructors using clicker questions and showed teachers how the clicker system worked in detail by asking them a few clicker questions. We also demonstrated how the clicker results could be displayed to students and how the teachers could learn about how challenging the questions are for students based on the voting results. Because this survey was not fully implemented until Spring 2015, feedback is provided for 41 of the 80 clicker class sessions. Seventeen middle and high school teachers provided feedback on clicker class sessions.

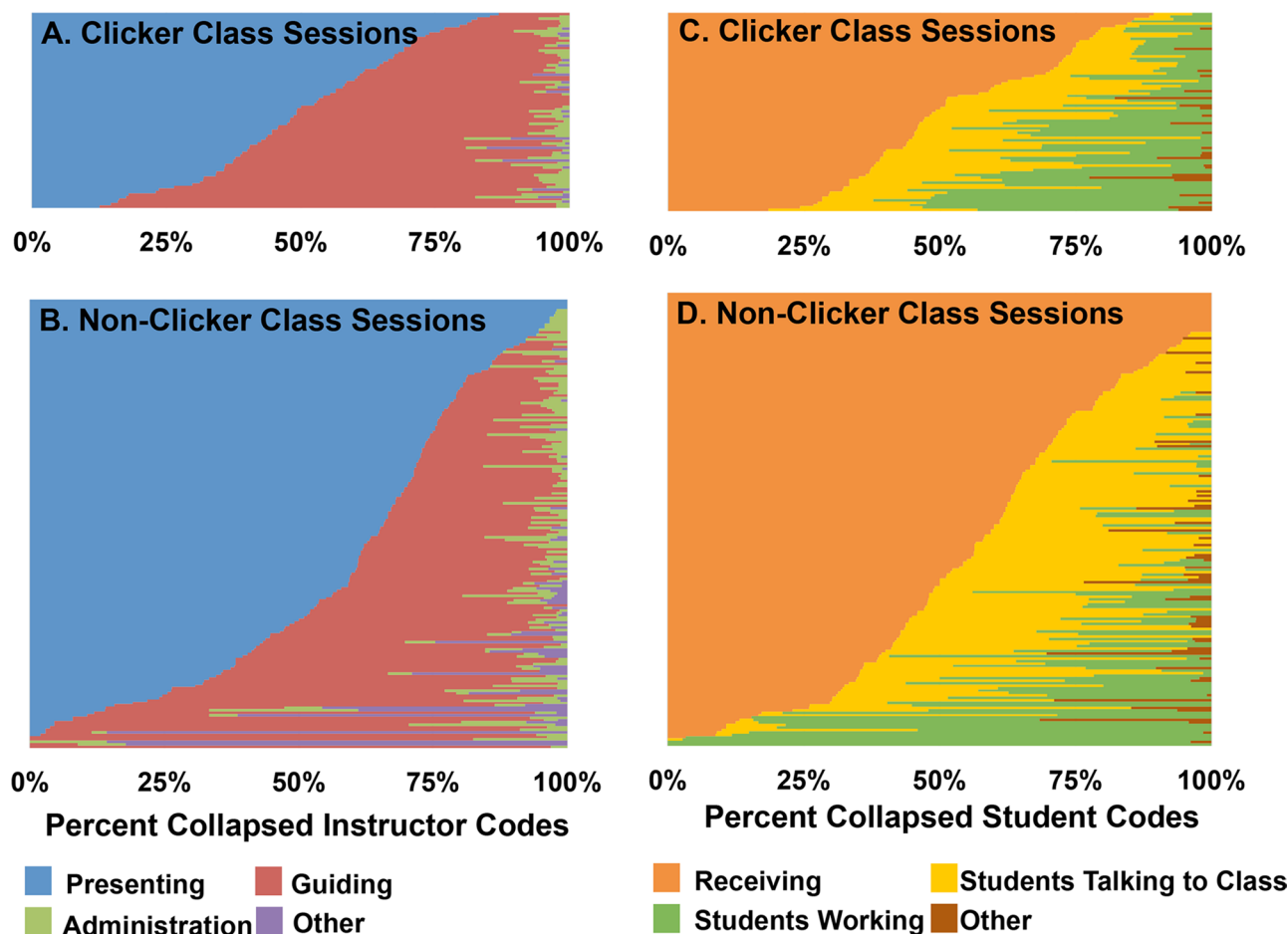
All statistical analyses were performed using SPSS (IBM, Armonk, NY).

## RESULTS

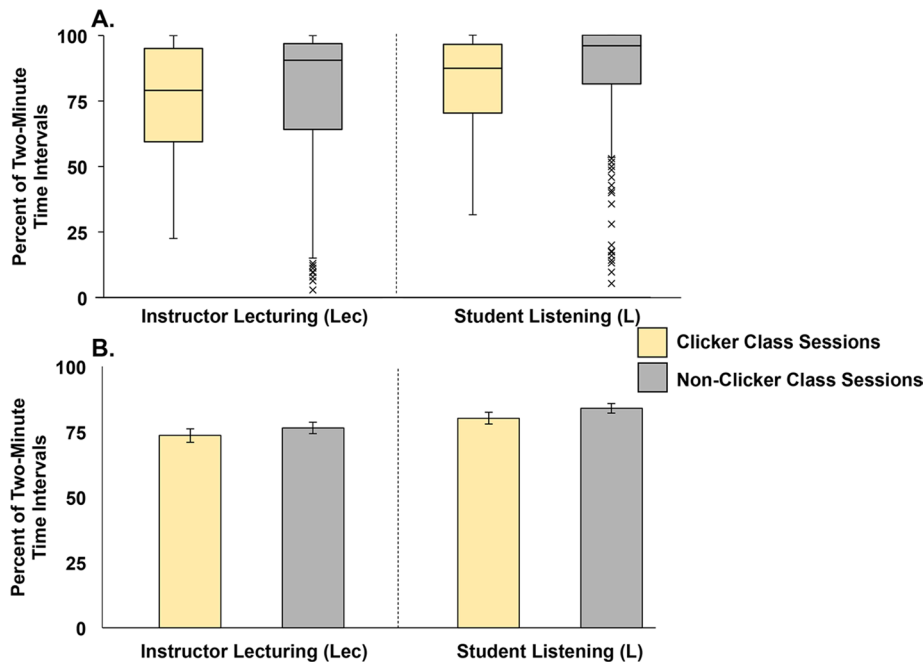
### *Characterizing Instructional Behaviors in Class Sessions Taught with and without Clicker Questions*

To determine whether there are differences in instructional behaviors between classes taught with and without clickers, we first separated out class sessions that included clicker questions. Eighty of the 264 observed class sessions featured at least one clicker question (CQ) instructor code and were thus classified as clicker class sessions (Figure 3). We examined differences in class size and found that the average student enrollment in classes that use clickers (111 students) was significantly higher than that in classes that did not use clickers (72 students) (independent-samples *t* test,  $p < 0.05$ ).

To have a holistic view of the COPUS codes at the same time in both clicker and nonclicker classes, we examined the distribution of collapsed-code percentages and found a range of classroom behaviors (Figure 5). Notably, the non-clicker class sessions range from 0 to 100% Instructor Presenting and Student Receiving, whereas the clicker class sessions have a narrower range from 13 to 87% Instructor Presenting and 19 to 89% Student Receiving.



**Figure 5.** Percentage of collapsed COPUS instructor and student codes for clicker ( $n = 80$ ) and nonclicker ( $n = 184$ ) class sessions. Each horizontal row represents a class session observation. The instructor codes in (A) for clicker class sessions and (B) for nonclicker class sessions are organized by the collapsed code Instructor Presenting. The student codes (C) for clicker class sessions and (D) for nonclicker class sessions are organized by the collapsed code Student Receiving.



**Figure 6.** Comparisons of the COPUS codes instructor lecturing (Lec) and student listening (L) for clicker and nonclicker class sessions. (A) Box-and-whisker plots showing the median and variation for the two class types. The line in the middle of the box represents the median percentage of 2-min time intervals for the class sessions in each group. The top of the box represents the 75th percentile, and the bottom of the box represents the 25th percentile. The space in the box is called the interquartile range (IQR), and the whiskers represent the lowest and highest data points no more than 1.5 times the IQR above and below the box. Data points not included in the range of the whiskers are represented by an "X." (B) Mean percentage of 2-min time intervals with instructor lecturing (Lec) and student listening (L) codes in clicker and nonclicker class sessions. Bars indicate SE.

Because percent-code calculations can be difficult to interpret for individual codes (see *Methods* for further details), we also examined the percentage of 2-min time intervals that included the traditional instructional codes such as instructor lecturing (Lec) and student listening (L). Nonclicker class sessions showed a broader range of percentage of 2-min time intervals that included instructor lecturing (Lec) and student listening (L), and higher median values for these two codes (Figure 6A). However, when comparing means (Figure 6B), there were no statistically significant differences between non-clicker and clicker class sessions (independent-samples *t* test,  $p = 0.4$  for instructor lecturing and  $p = 0.2$  for student listening).

Taken together, these results indicate that University of Maine STEM classes that use clickers (typically characterized by larger enrollments) displayed a narrower range of Instructor Presenting and Student Receiving collapsed-code behaviors compared with nonclicker class sessions. We also observe that clicker and nonclicker class sessions have a similar mean percentage of 2-min time intervals allocated to instructor lecturing (Lec). As a result, students in classes both with and without clickers spend a similar mean percentage of 2-min time intervals listening (L).

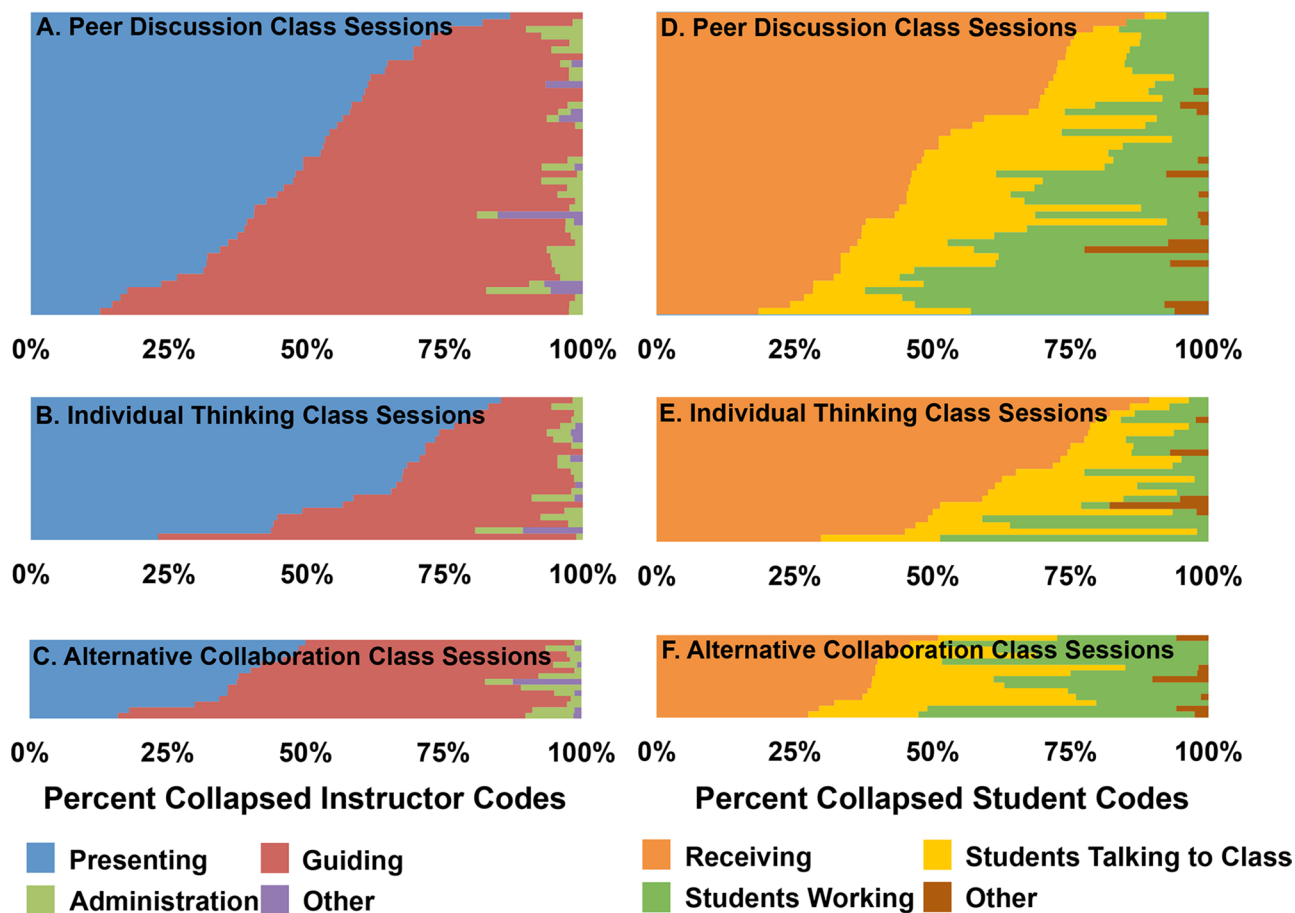
### Documenting Variation in How Clickers Are Implemented at a Campus-wide Level

Because faculty may implement clickers in various ways, possibly in alignment with entirely different pedagogical strategies, we examined classes that used clickers in order to identify some common instructional modes. We first separated out class sessions in which the instructors allowed peer discussion during clicker questions from those that did not (Figure 3, further description in *Methods*), and called this first mode Peer Discussion. The majority of the Peer Discussion class sessions used a combination of both individual and peer discussion during clicker questions. For the class sessions that did not allow peer discussion during clicker questions, we

subdivided the class sessions into Individual Thinking (no peer discussion during the entire class period) and Alternative Collaboration (no peer discussion during the clicker questions but peer discussion during other group activities in the class period). Comparisons of the collapsed instructor and student codes between Peer Discussion, Individual Thinking, and Alternative Collaboration class sessions revealed a range of instructional behaviors (Figure 7), with the Alternative Collaboration class sessions showing the lowest abundance of Instructor Presenting and Student Receiving collapsed codes.

To determine whether the three different modes of clicker use impacted the percentage of time allocated to traditional instructional practices such as instructor lecturing (Lec) and student listening (L), we examined the percentage of 2-min time intervals that included these two codes. Peer Discussion, Individual Thinking, and Alternative Collaboration class sessions are all characterized by a range of percent 2-min time intervals that include instructor lecturing (Lec) and student listening (L), with the Individual Thinking class sessions showing the highest median values (Figure 8A). In addition, Individual Thinking class sessions had a significantly greater mean percentage of 2-min time intervals with instructor lecturing (Lec) and student listening (L) compared with the other two types of class sessions that use clickers (Figure 8B, one-way analysis of variance [ANOVA], Tukey's post hoc test,  $p < 0.05$ , in both cases).

By definition, the Individual Thinking and Alternative Collaboration class sessions had no opportunities for clicker-mediated peer discussion, and the Peer Discussion class sessions contained at least one episode that included peer discussion (Figure 3). Because this definition does not account for possible variation in clicker use within the Peer Discussion class sessions, we also examined all 112 clicker episodes that occurred in the 44 Peer Discussion class sessions. Clicker episodes were classified into the following categories: Individual Thinking Only, Peer Discussion Only, and Individual Thinking and Peer Discussion Combined



**Figure 7.** Percentage of collapsed instructor COPUS codes for (A) Peer Discussion ( $n = 44$ ), (B) Individual Thinking ( $n = 22$ ), and (C) Alternative Collaboration ( $n = 14$ ) class session observations, organized by percent Instructor Presenting. Percentage of collapsed student COPUS codes for (D) Peer Discussion, (E) Individual Thinking, and (F) Alternative Collaboration class session observations organized by percent Student Receiving. Each horizontal bar represents a different class session observation.

(Figure 2). In Peer Discussion class sessions, 75% of the clicker episodes included the opportunity for students to talk to one another (Figure 9), with the most common practice including Individual Thinking and Peer Discussion Combined.

In Spring 2015, the teacher observers provided specific feedback for each class session they observed via an online survey, with a portion of this survey specifically focused on clicker use. In particular, the teachers were asked to provide information about how challenging the clicker questions were for the students based on the voting results (Figure 10). The survey responses suggested that Peer Discussion class sessions more commonly included questions that were challenging for students when compared with Individual Thinking and Alternative Collaboration class sessions.

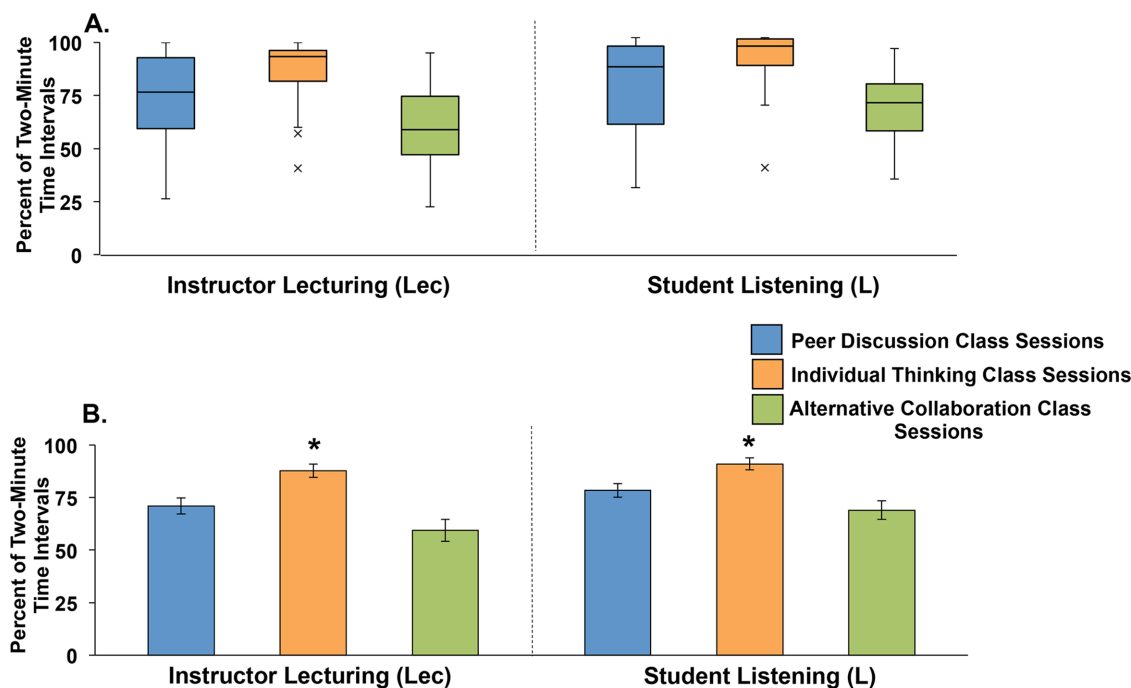
Taken together, these results indicate there are three different predominant modes of clicker use: Peer Discussion, Individual Thinking, and Alternative Collaboration. Among the three different modes of clicker use, the instructors in the Individual Thinking class sessions are spending significantly more time lecturing (Lec) and the students are spending more time listening (L). Alternative Collaboration class sessions tend to include lower abundance of Instructor Presenting and Student Receiving collapsed codes, largely due to the non-

clicker group activities. However, the presence of these activities does not result in significant differences in percentage of 2-min time intervals allocated to instructor lecturing (Lec) and student listening (L) between Peer Discussion and Alternative Collaboration class sessions. In addition to providing an opportunity for peer interaction, clicker questions asked during the Peer Discussion class sessions tended to include questions that provided greater levels of challenge to students.

### *Examining How Individual Instructors Vary Their Implementation of Clicker Questions*

Because instructors can vary their instructional style, we also explored the variation in how individual instructors used clickers. For this analysis, we focused on instructors who were observed teaching with at least two clicker episodes, regardless of the type of episode described in Figure 2. This analysis included 25 instructors from nine different departments. Nineteen of these instructors incorporated at least one opportunity for peer discussion (Figure 11). However, nearly all of the instructors used Individual Thinking Only at some point in their instruction, and this strategy accounted for more than 50% of the episodes for 16 of the 25 instructors.

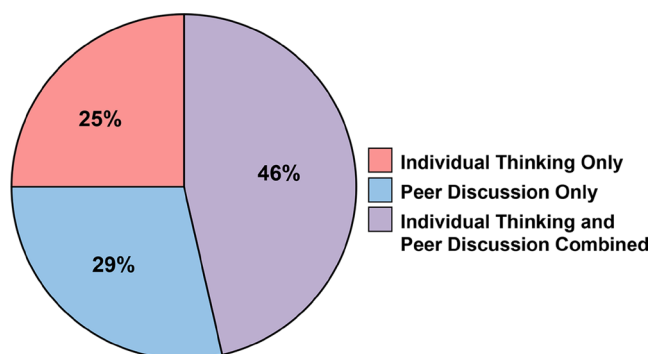




**Figure 8.** Comparisons of the COPUS codes instructor lecturing (Lec) and student listening (L) for Peer Discussion, Individual Thinking, and Alternative Collaboration class sessions. (A) Box-and-whisker plots show the median and variation for the three classroom types. The line in the middle of the box represents the median percentage of 2-min time intervals for the class sessions in each group. The top of the box represents the 75th percentile, and the bottom of the box represents the 25th percentile. The space in the box is called the interquartile range (IQR), and the whiskers represent the lowest and highest data points no more than 1.5 times the IQR above and below the box. Data points not included in the range of the whiskers are represented by an "X." (B) Mean percentage of 2-min time intervals with instructor lecturing (Lec) and student listening (L) code among the three modes of clicker use. Asterisks indicate statistically significant differences, one-way ANOVA, Tukey's post hoc test,  $p < 0.05$ . Bars indicate SE.

## DISCUSSION

Here we discuss the first observation-based, multidisciplinary study of clicker implementation in STEM classes across a single campus. Observations of STEM classrooms revealed that nearly a third of class sessions used clickers (Figure 3). A comparison of class sessions with and without clickers showed that both types of classes had a large range of collapsed-code instructional behaviors (Figure 5) and instructors teaching with clickers allocate a similar percentage of 2-min time intervals to instructor lecturing (Lec) and student listening (L) when compared with the instructors of nonclicker classes



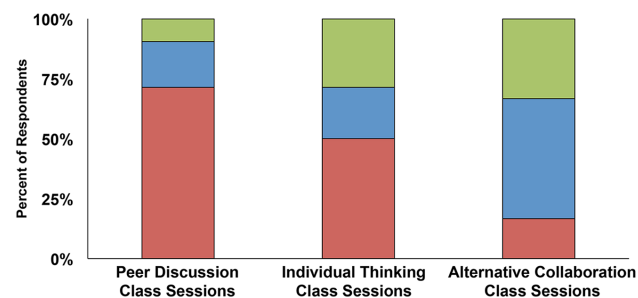
**Figure 9.** Distribution of clicker episodes for Peer Discussion class sessions.

(Figure 6). The results from our study appear to confirm the often articulated concern that adding clickers alone does not guarantee that instructors will spend more time overall on active-engagement, student-centered instruction.

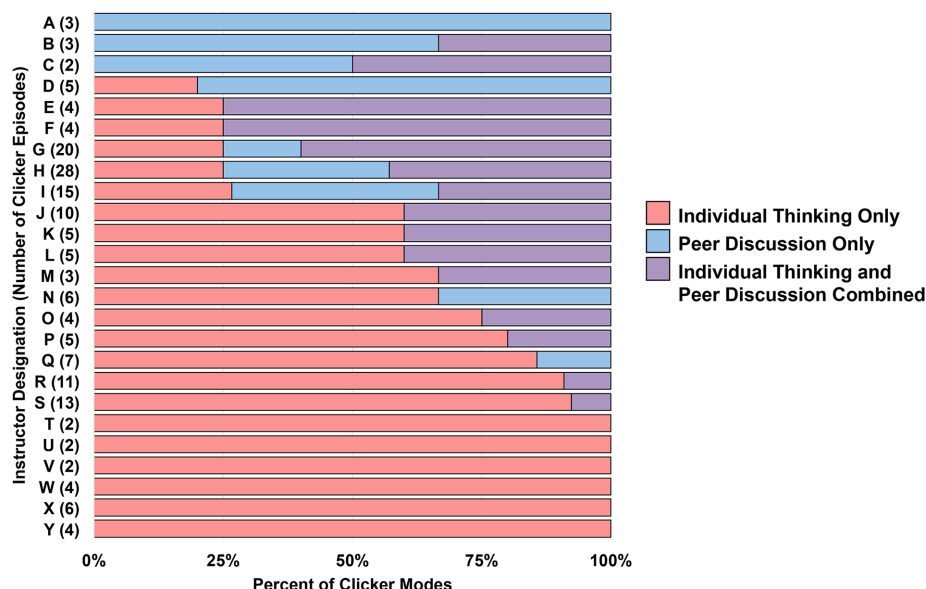
We suspect that part of the reason our data do not show dramatic differences between class sessions that use and do

In general, the clicker questions were:

- Challenging for students: the class vote was often split.
- Easy for students: the majority of the students answered correctly.
- Cannot determine: the instructor did not talk about the class voting results.



**Figure 10.** Observers described how challenging clicker questions were for students based on the voting results shared with the class for the Spring 2015 observations ( $n = 41$  class sessions). The results are shown for each of the three clicker class session modes: Peer Discussion ( $n = 21$  class sessions), Individual Thinking ( $n = 14$  class sessions), and Alternative Collaboration ( $n = 6$  class sessions).



**Figure 11.** Percentage of Individual Thinking Only, Peer Discussion Only, and Individual Thinking and Peer Discussion Combined clicker episodes for each instructor (designated by a letter of the alphabet) with two or more observed clicker episodes. The total number of clicker episodes observed in 2014 and 2015 for each instructor is shown in parentheses. Each horizontal bar represents the percentage of clicker episode modes for a single instructor, organized by percent of Individual Thinking Only episodes.

not use clickers is because we observed three distinct modes of clicker use: Peer Discussion, in which students had at least one opportunity to talk with one another during clicker questions; Individual Thinking, in which no peer discussion was observed; and Alternative Collaboration, in which students had time for discussion, but it was not paired with clicker questions (Figure 3). Our results indicate that for the Peer Discussion class sessions, the majority included clicker questions that combine both individual thinking and group discussion (Figure 3) and that the questions tended to be challenging for students to answer (Figure 10). Furthermore, instructors in the Individual Thinking class sessions spent a significantly larger percentage of 2-min time intervals lecturing (Lec) and the students spent a larger percentage of 2-min time intervals listening (L; Figure 8B). Finally, the Alternative Collaboration class sessions tended to include fewer Instructor Presenting and Student Receiving collapsed codes (Figure 7), but the presence of these activities did not result in significant differences between percentage of time allocated to instructor lecturing (Lec) and student listening (L) between Peer Discussion and Alternative Collaboration class sessions (Figure 8B).

Although studies have shown that student clicker question performance increases during peer discussion when compared with individual thinking (Lasry *et al.*, 2009), little is known about the potential student learning benefits of the Alternative Collaboration style of clicker use. Future work is needed to explore this mode and other types of clicker-supported group work reported in the literature (e.g., Kryjevskaja *et al.*, 2014), especially in cases in which the clicker questions are used to check class understanding after group activities (e.g., Kolber *et al.*, 2014; Smith and Merrill, 2014).

### *Inconsistencies in How Peer Discussion Is Used with Clickers*

Even though our data indicate that there are a variety of ways clicker questions are being used, Individual Thinking

was the only behavior in 28% of the clicker class sessions we observed (Figure 2) and was an instructional strategy used by the majority of instructors (Figure 11). Although these instructors have successfully overcome many obstacles to the implementation of clickers in their classrooms, emphasis on the Individual Thinking strategy may inadvertently limit possible student learning opportunities. For example, if peer discussion is omitted, students may lose the opportunity to build scientific communication skills that are developed by articulating reasoning, evaluating the merits of others' reasoning, and asking peers questions (Turpen and Finkelstein, 2010). In addition, performance increases attributed to peer discussion are lost (Smith *et al.*, 2009; Lasry *et al.*, 2009; Porter *et al.*, 2011; Knight *et al.*, 2015; Barth-Cohen *et al.*, 2016). Furthermore, when peer discussion is omitted, faculty members do not have the opportunity to circulate around the class and listen to student reasoning (Mazur, 1997).

In addition to the lost student and instructor learning opportunities, previous work has shown that students have more negative attitudes about the utility of clickers when they do not discuss the multiple-choice questions (Keller *et al.*, 2007). Student resistance often impacts instructional decisions, and faculty may abandon clicker-supported instruction and other active-learning pedagogies promoted by discipline-based education researchers (Silverthorn, 2006; Henderson and Dancy, 2007). It has also been documented that faculty prioritize personal experience over empirical evidence when making decisions about teaching strategies (Andrews and Lemons, 2015), and the negative experiences associated with nonoptimal clicker implementation may therefore have a long-lasting impact on future instructional decisions.

### *How Can We Encourage Faculty to Include Peer Discussion with Clicker Questions?*

While research has shown there are benefits to allowing students to talk to one another during clicker questions (Smith *et al.*, 2009, 2011; Porter *et al.*, 2011; Knight *et al.*, 2015;

Barth-Cohen *et al.*, 2016), the observed variation in use of the peer discussion portion of clicker implementation is consistent with findings for other research-based pedagogies, which are typically changed and modified during implementation by faculty (Henderson and Dancy, 2007; NRC, 2013). Given that faculty often modify research-based instructional practices, how can we enhance professional development to make sure the peer discussion portion of clicker use is retained during implementation?

One response is to make sure faculty professional development motivates and targets the more nuanced aspects of effective clicker implementation. Considering that 86% of U.S. college faculty recently reported they were familiar with clickers (FTI Consulting, 2015), the majority of professional development audiences likely have some working knowledge of clickers. As such, it is important to move beyond dedicating an entire session to the basics of using a clicker system. Moreover, rather than polling professional development participants about whether or not they have used clickers before, the audience can instead be asked how they use clickers during instruction and can be asked to discuss with their neighbors and report to the group. This approach may reveal innovative ways in which people are using clickers, provides a more detailed picture of participant experience, and can serve as a launching point for motivating the value of peer discussion by drawing upon a combination of findings from research studies (Smith *et al.*, 2009, 2011; Porter *et al.*, 2011; Barth-Cohen *et al.*, 2016) and engaging participants in activities designed to help them identify the features of clicker questions that encourage productive peer discussion.

In addition, recent work has shown that onetime faculty professional development workshops have a limited capacity to create change (Davidovitch and Soen, 2006; Henderson *et al.*, 2011). Instead, faculty need ongoing, in-depth professional development and support (Henderson *et al.*, 2011; PCAST, 2012). For this reason, at the University of Maine, we have started a yearlong faculty professional development program in which faculty meet in rotating pods of three: one individual teaches, one individual observes using the COPUS, and another provides feedback on areas identified in advance by the instructor. Notably, 89% of the faculty members participating in this program have said that encouraging student peer discussion is one of the predominant areas in which they would like assistance, and it will therefore be an ongoing focus of this program.

Finally, there is also a need for clicker question banks that are vetted by the community and include questions that have been shown to encourage productive peer discussion. The work described here indicates that instructors who are using peer discussion in their classes are asking questions that are challenging for students to answer (Figure 10), which are often time-consuming for instructors to write. In addition to presenting clicker questions, a question bank could also include aggregate student voting results, instructor reflections on how to most effectively follow up when student voting results are split among multiple answers, videos of students discussing the clicker questions with one another, and follow-up homework and exam questions that target the concepts from the clicker questions. These supplemental materials, in particular, could 1) foreground the ways in which clicker questions may be used to facilitate student learning, 2) provide some of the scaffolding needed to support effective

implementation, and 3) serve as a flexible resource that faculty may adapt based on the needs of their classrooms.

## CONCLUSION

Our campus-wide, observation-based study of clicker implementation in STEM classrooms revealed that instructors who used clickers demonstrated variation in implementation, with many instructors eliminating peer discussion during some if not all clicker questions. Omitting peer discussion impacts students' ability to articulate their reasoning and to work together to solve problems. In addition, instructors who omit peer discussion lose the chance to listen in on student reasoning and may encounter more student resistance to research-supported instructional techniques. To encourage faculty to include peer discussion, we recommend 1) focusing on peer discussion as an essential component of long-term clicker professional development programs that include multiple opportunities for faculty to learn about using peer discussion with clickers, and 2) establishing clicker question banks that include challenging, higher-order questions for faculty to adapt to their instructional needs.

## ACKNOWLEDGMENTS

The authors appreciate the contributions of the middle and high school teachers who served as observers, including Elizabeth Baker, Toni Barboza, Tristan Bates, Michele Benoit, Stacy Boyle, Virginia Brackett, Elizabeth Connors, Tracy Deschaine, Lauren Driscoll, Andrew Ford, Bill Freudenberger, Kathryn Priest Glidden, Marshall Haas, Kate Hayes, Beth Haynes, Teri Jergenson, Danielle Johnson, Bob Kumpa, Lori LaCombe-Burby, Melissa Lewis, John Mannette, Nichole Martin, Lori Matthews, Nicole Novak, Jennifer Page, Patti Pelletier, Meredith Shelton, Beth Smyth-Handley, Damian Sorenson, Joanna Stevens, Rhonda Stevens, Nancy Stevick, Thomas White, and Michael Witick; and the University of Maine faculty members who welcomed the teachers into their classes for observations. We also thank Ken Akiha, Karen Pelletreau, and Mindi Summers for helpful feedback on this article. This material is based on work supported by the National Science Foundation under grants 1347577 and 0962805.

## REFERENCES

- American Association for the Advancement of Science (2010). *Vision and Change: A Call to Action*, Washington, DC.
- Andrews TC, Lemons PP (2015). It's personal: biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE Life Sci Educ* 14, ar7.
- Barth-Cohen LA, Smith MK, Capps DK, Lewin JD, Shemwell JT, Stetzer MR (2016). What are middle school students talking about during clicker questions? Characterizing small-group conversations mediated by classroom response systems. *J Sci Educ Technol* 25, 50–61.
- Caldwell JE (2007). Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci Educ* 6, 9–20.
- Davidovitch N, Soen D (2006). Using students' assessments to improve instructors' quality of teaching. *J Furth High Educ* 30, 351–376.
- Freeman S, Eddy S, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA* 111, 8410–8415.

- FTI Consulting (2015). U.S. Postsecondary Faculty in 2015: Diversity in People, Goals and Methods, But Focused on Students. Seattle: The Bill & Melinda Gates Foundation. Retrieved from: <http://postsecondary.gatesfoundation.org/wp-content/uploads/2015/02/US-Postsecondary-Faculty-in-2015.pdf>.
- Henderson C, Beach A, Finkelstein N (2011). Facilitating change in undergraduate STEM instructional practices: an analytical review of the literature. *J Res Sci Teach* 48, 952–984.
- Henderson C, Dancy MH (2007). Barriers to the use of research-based instructional strategies: the influence of both individual and situational characteristics. *Phys Rev Spec Top Phys Ed Res* 3, 020102.
- Hora MT, Ferrare JJ (2014). Remeasuring postsecondary teaching: how singular categories of instruction obscure the multiple dimensions of classroom practice. *J Coll Sci Teach* 43, 36–41.
- Hora MT, Oleson A, Ferrare JJ (2013). Teaching Dimensions Observation Protocol (TDOP) User's Manual, Madison: Wisconsin Center for Education Research.
- Keller C, Finkelstein N, Perkins K, Pollock S, Turpen C, Dubson M (2007). Research-based practices for effective clicker use. *AIP Conf Proc* 951, 128–131.
- Knight JK, Wise SB, Rentsch J, Furtak EM (2015). Cues matter: learning assistants influence introductory student interactions during clicker-question discussions. *CBE Life Sci Educ* 14, ar41.
- Kolber BJ, Konsolaki M, Verzi MP, Wagner CR, McCormick JR, Schindler K (2014). Sex-specific differences in meiosis: real-world applications. CourseSource. <http://coursesource.org/courses/sex-specific-differences-in-meiosis-real-world-applications> (accessed 15 December 2015).
- Kryjevskaja M, Boudreaux A, Heins D (2014). Assessing the flexibility of research-based instructional strategies: implementing tutorials in introductory physics in the lecture environment. *Am J Phys* 82, 238–250.
- Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Lane ES, Harris SE (2015). A new tool for measuring student behavioral engagement in large university classes. *J Coll Sci Teach* 44, 83–91.
- Lasry N, Charles E, Whittaker C, Lautman M (2009). When talking is better than staying quiet. *AIP Conf Proc* 1179, 181–184.
- Lund TJ, Pilarz M, Velasco JB, Chakraverty D, Rosploch K, Underlander 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 Sci Educ* 14, ar18.
- Mazur E (1997). Peer Instruction: A User's Manual, Upper Saddle River, NJ: Prentice Hall.
- National Research Council (2013). Adapting to a Changing World: Challenges and Opportunities in Undergraduate Physics Education, Washington, DC: National Academies Press.
- Porter L, Bailey LC, Simon B, Zingaro D (2011). Peer instruction: do students really learn from peer discussion in computing. In: Proceedings of the Seventh International Workshop on Computing Education Research, New York: ACM, 45–52.
- President's Council of Advisors on Science and Technology (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics, Washington, DC: U.S. Government Office of Science and Technology.
- Silverthorn DU (2006). Teaching and learning in the interactive classroom. *Adv Physiol Educ* 30, 135–140.
- Singer SR, Nielsen NR, Schweingruber HA (2012). Discipline-based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering, Washington, DC: National Academies Press.
- Smith MK, Jones FH, Gilbert SL, Wieman CE (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci Educ* 12, 618–627.
- Smith MK, Merrill S (2014). Why do some people inherit a predisposition to cancer? A small group activity on cancer genetics. CourseSource. <http://coursesource.org/courses/why-do-some-people-inherit-a-predisposition-to-cancer-a-small-group-activity-on-cancer#tabs-0-content=0> (accessed 15 December 2015).
- Smith MK, Trujillo C, Su TT (2011). The benefits of using clickers in small-enrollment seminar-style biology courses. *CBE Life Sci Educ* 10, 14–17.
- 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 Sci Educ* 13, 624–635.
- Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT (2009). Why peer discussion improves student performance on in-class concept questions. *Science* 323, 122–124.
- Strukov A (2008). Effects, successes and pitfalls of student (personal) response system (PRS) implementation at the University of Maine, USA: 3-year assessment. In: Proceedings of EdMedia: World Conference on Educational Media and Technology, ed. J Luca and E Weippl, Association for the Advancement of Computing in Education (AACE), 3284.
- Turpen C, Finkelstein ND (2009). Not all interactive engagement is the same: variations in physics professors' implementation of Peer Instruction. *Phys Rev Spec Top Phys Ed Res* 5, 020101.
- Turpen C, Finkelstein ND (2010). The construction of different classroom norms during peer instruction: students perceive differences. *Phys Rev Spec Top Phys Ed Res* 6, 020123.