2022

Revisiting Clickers: In-Class Questions Followed by At-Home Reflections Are Associated with Higher Student Performance on Related Exam Questions

Dana L. Kirkwood-Watts

Emily K. Bremers

Emily Robinson

Kathleen R. Brazeal

Brian Couch

Research Article

JMBE Journal of Microbiology & Biology Education

Revisiting Clickers: In-Class Questions Followed by At-Home Reflections Are Associated with Higher Student Performance on Related Exam Questions

Dana L. Kirkwood-Watts,[a] Emily K. Bremers,[b] Emily A. Robinson,[c]
Kathleen R. Brazeal,[a] and Brian A. Couch[a]
[a]School of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA
[b]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia, USA
[c]Department of Statistics, University of Nebraska, Lincoln, Nebraska, USA

Clicker questions are a commonly used active learning technique that stimulates student interactions to help advance understanding of key concepts. Clicker questions are often administered with an initial vote, peer discussion, and a second vote, followed by broader classroom explanation. While clickers can promote learning, some studies have questioned whether students maintain this performance on later exams, highlighting the need to further understand how student answer patterns relate to their understanding of the material and to identify ways for clickers to benefit a broader range of students. Systematic requizzing of concepts during at-home assignments represents a promising mechanism to improve student learning. Thus, we paired clicker questions with at-home follow-up reflections to help students articulate and synthesize their understandings. This pairing of clickers with homework allowed us to decipher how student answer patterns related to their underlying conceptions and to determine if revisiting concepts provided additional benefits. We found that students answering both clicker votes correctly performed better on isomorphic exam questions and that students who corrected their answers after the first vote did not show better homework or exam performance than students who maintained an incorrect answer across both votes. Furthermore, completing the follow-up homework assignment modestly boosted exam question performance. Our data suggest that longer-term benefits of clickers and associated homework may stem from students having repeated opportunities to retrieve, refine, and reinforce emerging conceptions.

KEYWORDS active learning, clicker questions, clickers, formative assessment, homework, multiple-choice, multiple-true-false, peer instruction, undergraduate

INTRODUCTION

Over the past decades, science, technology, engineering, and math (STEM) classrooms have seen an increased emphasis on active learning (1, 2). When implemented effectively, active learning can improve student engagement, attitudes, and conceptual understanding and support a variety of outcomes, such as improved course performance and decreased achievement gaps (3–6). Many undergraduate instructors have adopted the active learning technique of administering clicker questions with peer instruction (7). A recommended clicker sequence (8) begins with instructors displaying a closed-ended question and students submitting individual answers via an electronic audience response system. Students are then prompted to discuss their answers in small groups and submit another answer in light of their discussions. Following answer submission, the instructor can invite students to share their reasoning and can provide further explanation regarding the answer options.

Clickers started to make their appearance as early as the mid-1990s, gaining popularity by the early 2000s (8–10). Asking clicker questions at strategic points during class helps break up the session, gives students time to process ideas, and highlights key concepts. Importantly, clickers allow the instructor to gauge student understanding and provide an opportunity for students to receive feedback. Clickers have been investigated across a range of course levels, and many reports associate clickers with a variety of benefits (11–14). Students generally express satisfaction with clickers, view them as helpful to their learning, cite specific aspects they value, and recommend their continued use (6, 15–20). The addition of clickers to a lecture, especially

Editor L. Kate Wright, Rochester Institute of Technology
Address correspondence to School of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA. E-mail: bcouch2@unl.edu.
The authors declare no conflict of interest.
Received: 10 March 2022, Accepted: 13 June 2022, Published: 6 July 2022

when coupled with low-risk participation points, increases class attendance and decreases attrition across the semester (7, 21–24). Compared to traditional lecture, course offerings that include clickers can result in greater student learning as measured on concept inventories or final exams (21, 25–28), although it can be difficult to isolate the effect of clickers from other course components and instructional interventions (7).

Student scores commonly increase from a first clicker vote to a second clicker vote that occurs after peer discussion (15, 16, 29, 30). Thus, some researchers have framed the question of clicker impact in terms of whether this improvement stems from peer discussion and associated learning versus students taking additional time to think or obtaining answers from their peers. One study addressed this question by posing an isomorphic question to students immediately after the second vote (31). They found that improved performance on the second vote persisted to the isomorphic question, suggesting that students changing to the correct answer reflected learning from peer discussion. Follow-up work in which students were exposed to different combinations of peer discussion and instructor explanation replicated these findings and showed an additional benefit of instructor explanation (32, 33).

Mixed results have been found, however, with respect to whether clicker-based improvements are retained on longer time scales. In one study, the initial boost seen from first to second vote was not maintained when students were asked similar questions on the subsequent exam (34). In another study conducted by members of our research group, we found that students in higher performance quartiles maintained clicker learning gains, whereas students in lower quartiles did not demonstrate similar benefits on later exams, and this finding held true for questions asked in either a multiple-choice (MC) or multiple-true-false (MTF) format (35). Finally, a study looking at retention after 4 months found nuanced results, namely, that clickers led to better retention of material for students from a nonmajors course but not from a majors course (17).

We sought to develop an instructional intervention that could potentially improve clicker outcomes while simultaneously providing more in-depth information on the understandings that students have after clicker activities. We implemented this intervention in the context of an online assignment based on the role of homework in reinforcing material covered during class (36). We hypothesized that having students revisit targeted clicker questions on homework assignments and explain the reasoning behind the answers would support improved performance on later exam questions addressing the same concepts. Instructional cues and course activities have been shown to affect student behaviors during clicker discussions (37–40), so we reasoned that this later homework activity would encourage students to focus on developing correct reasoning during their discussions and ensure that each student had an opportunity to express their understandings. We also viewed this homework assignment as a way for students to engage in retrieval practice, a process by which later engagement with a concept strengthens learning (41).

Here, we describe results from our approach of giving students clicker questions, followed by a reinforcing homework assignment and later isomorphic exam question. We structured our analyses around four research questions: (i) Does participation in clicker questions and completion of the homework predict performance on isomorphic exam questions? (ii) How do clicker vote patterns relate to student homework explanations? (iii) How do clicker vote patterns relate to performance on isomorphic exam questions? (iv) How do homework explanations relate to performance on isomorphic exam questions?

In each case, we conducted additional exploratory analyses to understand variation across student performance quartiles. By tracking student understanding across this sequence of events, our research aimed to better understand how clickers and homework relate to student learning within the given instructional context.

METHODS

Course context and research design

This study occurred in two equivalent sections of a high-enrollment undergraduate introductory biology course for life sciences majors at a research university ($n = 346$ consenting students; demographics of the class are presented in Table 1; classified as exempt from IRB review, 14314). Both sections were cotaught by two instructors who alternated the weeks in which they took the lead during class. The course consisted of roughly 2 to 4 unique clicker questions per 50-min class session (3 sessions per week), one homework assignment per week, and one exam for each of the four units. Within this broader context, a subset of 16 clicker questions (four per unit) was targeted on the follow-up homework assignment and subsequent exam. Thus, the primary data for this study consist of the pathways each student took from their clicker votes to homework explanations and to subsequent exam answers (Fig. 1).

While the larger pool of clicker questions in the course consisted of MC and MTF questions, the 16 targeted clicker questions were all in the MTF format, thereby providing an efficient way to collect detailed information on student understanding of various conceptions (42–44). Each MTF question stem was followed by four statements with the possibility that one, two, or three of these statements was true for each question (see Fig. S1 in the supplemental material for an example clicker question). Students answered by selecting each statement that they deemed to be true on an electronic iClicker device. Clicker questions were presented to the class for an individual vote (i.e., vote 1), followed by small group discussion with nearby peers and then a second individual vote (i.e., vote 2). After the second vote, the instructor always showed the correct answers, sometimes had select students share their reasoning with the broader class, and always made sure to provide an explanation behind the answer for each statement (regardless of whether the statement was true or false). Clickers comprised 10% of the course grade. For each question, 0.9 points were awarded for participating and 0.1 point was given for answering the four statements correctly.

TABLE 1
Demographics[a] of consenting students (n = 346)

Category	Demographic	n	%
Gender[b]	Female	209	60.4
	Male	137	39.6
Generation status[c]	Continuing generation	260	75.1
	First generation	86	24.9
Race and ethnicity[d]	Non-URM	304	87.9
	URM	42	12.1
Class rank[e]	First yr	213	61.6
	Non-first yr	133	38.4

[a]Demographics were obtained from the institutional data office.
[b]At the time of data collection, the institution did not collect information regarding nonbinary gender identities.
[c]Students were considered continuing-generation if one or both parents had a bachelor's degree. Students were considered first-generation if neither of their parents had a bachelor's degree.
[d]Non-underrepresented (non-URM) included white, Asian, and international students. Underrepresented minority (URM) included black, Hispanic, Hawaiian/Pacific Islander, and American Indian/Alaska Native students.
[e]Class rank was not included in analyses but is provided here for context.

At the end of each week, two clicker questions from the same class day (i.e., targeted clicker questions) were incorporated in the form of a homework question embedded within a larger online assignment. For this homework activity, a targeted clicker question was displayed verbatim along with the correct answers, and students were directed to "provide a separate explanation for why each of the four statements is true or false." Prior to the first assignment, an instructor told students that this activity was designed to help them synthesize their understandings and emphasized how students should deliberately structure

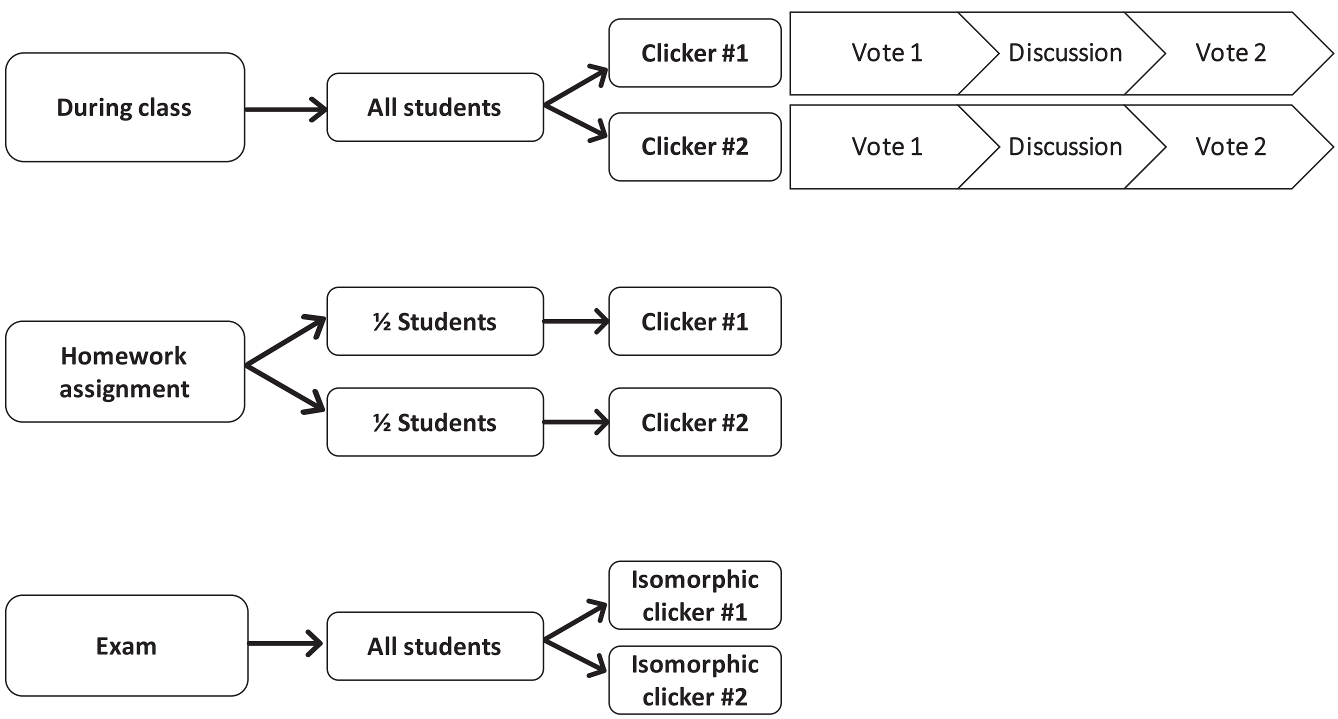

FIG 1. Diagram of the study design, showing pathways for two targeted clicker questions (of 16 total targeted clicker questions). All targeted clicker questions used the multiple-true-false (MTF) format. During class, all students saw the targeted clicker questions. For the first targeted clicker question (clicker 1), students answered individually (vote 1), discussed in small groups, answered individually again (vote 2), and the instructor then went through explanations for the correct answers. This sequence was later repeated for the second targeted clicker question (clicker 2). On the subsequent homework activity occurring at the end of the week, half of the students randomly received one of the clicker questions and the other half randomly received the other clicker question, and students explained why each of the four clicker question statements was true or false. For the exam, all students saw an isomorphic question corresponding to each targeted clicker question.

their explanations to address each MTF statement. While all students received the larger online assignment each week, half of the students randomly received one of the targeted clicker questions and the other half received the other targeted clicker question. Homework was worth 20% of the course grade. The clicker homework question was graded by undergraduate assistants on a scale from 0 to 4 points. These points counted for the assignment grade but were not used for research analyses.

Each targeted clicker question then appeared on the subsequent exam in the form of an isomorphic MTF question, embedded within a larger MTF section. These isomorphic questions presented scenarios and true-false statements that aligned with but differed from the original clicker question, such that they covered similar concepts but required students to have correct understandings to answer the new question (see Fig. S1 in the supplemental material for an example isomorphic exam question). The four unit exams were worth a total of 40% of the course grade, and MTF questions were scored based on how many of the four statements were answered correctly.

Data analysis

While there were 16 targeted clicker questions in this study, we considered each of the four statements per targeted clicker question as a separate item for the purposes of analysis, leading to a total of 64 items. This decision mirrors how MTF questions have been analyzed in previous work and reflects the intention that each item targets a different conceptual aspect of the scenario (45–48).

Members of the research team coded student open-ended explanations for the clicker homework questions, which enabled us to examine how student understanding related to clicker participation and later exam performance. For each MTF statement, we evaluated the corresponding part of the student open-ended explanation by using a binary coding scheme ($n = 9,408$ total coding events). A student's explanation was coded as "(1) demonstrating understanding" if they provided a thorough answer that gave sound reasoning as to "why" a statement was true or false. Explanations were coded as "(0) incomplete/incorrect" if they reiterated the statement without demonstrating an in-depth understanding of the "why" behind the statement or if they provided incorrect reasoning or irrelevant information.

Members of the research team initially worked in pairs to code the open-ended clicker homework explanations. For each MTF statement, we developed a more detailed codebook that provided specific information on the types of student responses reflecting each coding category. For each round, pairs coded 15 explanations independently, calculated their percent agreement, discussed disagreements to consensus, and updated the codebook based on their conversations. Reliability was considered to have been achieved when the two coders reached at least 80% agreement in two consecutive rounds for a given statement, at which point one coder finished coding the remaining explanations.

Statistical analyses were completed in R (version 3.6.3, RStudio version 1.2.5042). Models were run using generalized linear mixed models with a binomial distribution and logit link function in the lme4 package (49). Student was included as a random effect in each model. Demographic variables of gender, race and ethnicity, and first-generation status were incorporated to account for potential confounding effects (50), but these covariates were not a major focus of the current investigation and therefore are not discussed at length here. For certain analyses, students who did not turn in a homework assignment at all were removed in order to specifically compare students who did or did not receive a particular targeted question for reflection.

Model selection was conducted using a backwards stepwise selection process to obtain the model with the lowest Akaike information criterion (AIC) value (51). Starting with a full model, new models were tested by successively omitting interaction terms with the highest nonsignificant *P* values first and then omitting main effects with the highest nonsignificant *P* values. Terms were retained in the model if their removal caused a >2-point increase in AIC relative to the previous model. Additionally, each student was assigned into a single quartile for the semester based on their average performance on the other closed-ended, nontargeted questions across the four unit exams. Following model selection, exploratory analyses were conducted to examine contrasts within student quartile groups. *Post hoc* tests were conducted using lsmeans (52), psych (53), and multcomp (54) packages, and figures were produced using ggplot2 (55).

RESULTS

Research question 1: Does participation in clicker questions and completion of the homework predict performance on isomorphic exam questions?

We sought to determine from an overarching level whether participating in the targeted clicker questions and giving a homework explanation predicted performance on later isomorphic exam questions. Students were considered to have participated in a targeted clicker question if they submitted two votes for that question (irrespective of correctness) and were considered to have completed the associated homework component if they submitted an explanation (irrespective of correctness). At the whole-class level, we found that being present for a targeted clicker question and completing a homework explanation each had significant effects on later exam performance (Fig. 2A and Table 2). The interaction term between these variables was not retained, suggesting that there were no synergistic benefits to completing both a targeted clicker and its associated homework question.

Given the nature of these effects, we separately explored these variables across student quartiles to determine if they had selective benefits for certain students. We found a general upward trend across quartiles when students were absent for the clicker question (Fig. 2B) or did not complete the homework explanation (Fig. 2C). In comparison to those baseline conditions, we observed that being present for a targeted question

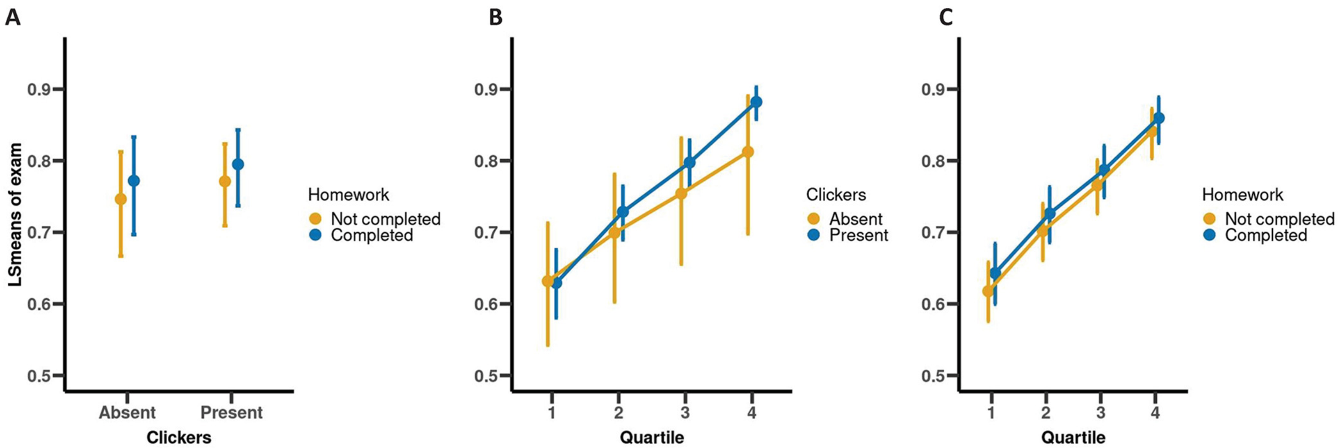

FIG 2. Effect of clicker participation and homework completion on isomorphic exam question performance. Dots represent modeled least-squared means and bars represent 95% confidence intervals. (A) Overall effects of clickers and homework. (B) Clicker effect for each student quartile. (C) Homework effect for each student quartile.

was not associated with improved performance on the later exam question for lower-quartile students but was associated with higher performance for higher-quartile students (Fig. 2B; see also Table S2 in the supplemental material). Conversely, students across different quartiles experienced moderate but even benefits of submitting homework explanations (Fig. 2C; see also Table S3).

Research question 2: How do clicker vote patterns relate to student homework explanations?

The clicker question homework activity presented a unique opportunity to see how clicker vote patterns related to underlying student reasoning, expressed as open-ended explanations regarding the various MTF statements in the clicker question. This model included each clicker vote as a separate binary term (incorrect/correct) as well as an interaction effect between the vote terms. This interaction effect allowed us to identify whether vote 1 (V1) and vote 2 (V2) showed any combined effects on homework explanations. We did not observe significance for the V1*V2 interaction or the clicker vote main effects in the best-fit model, suggesting an equivalence across the various clicker patterns (Table 3). However, we noted that the interaction term was retained and approached significance in the final model ($P = 0.055$), which led us to visualize how each of the different clicker patterns related to homework explanations (Fig. 3). We saw a trend where students who answered both votes correctly were more likely than any of the other clicker vote patterns to provide a homework explanation that demonstrated understanding, but we did not conduct *post hoc* significance testing because the interaction term was not previously significant.

Although we did not detect a formal effect of clicker vote patterns on homework explanations, we wanted to determine whether clicker participation was associated with any benefit on later homework explanations compared to being absent. Since they were statistically equivalent, we collapsed all four clicker vote patterns into a single group, representing when students were present for the targeted clicker question. We found that being present for the clicker question predicted a higher likelihood of providing a homework explanation that demonstrated understanding (Fig. 4A; see also Table S5 in the supplemental material). When we explored this result by

TABLE 2
Effect of clicker participation and homework completion on isomorphic exam question performance[a]

Factor	Estimate	SE	*Z* value	*P* value
(Intercept)	1.064	0.142	7.507	**<0.001**
Clicker participation	0.392	0.085	4.636	**<0.001**
Homework completed	0.140	0.038	3.732	**<0.001**
Generation status (first generation)	−0.002	0.168	−0.012	0.990
Race and ethnicity (URM)	0.032	0.212	0.151	0.880
Clicker participation*generation status	−0.221	0.157	−1.410	0.159
Clicker participation*race and ethnicity	−0.292	0.193	−1.513	0.130

[a]Logit (exam score) was calculated as follows: clicker participation + homework completion + generation status + race and ethnicity + clicker participation*generation status + clicker participation*race and ethnicity (see Table S1 in the supplemental material for information on model selection). *P* values in boldface indicate significance ($P < 0.05$). SE, standard error.

TABLE 3
Effect of clicker vote patterns on homework explanations[a]

Factor	Estimate	SE	*Z* value	*P* value
(Intercept)	0.385	0.103	3.746	**<0.001**
Vote 1 correct	−0.176	0.145	−1.214	0.225
Vote 2 correct	−0.073	0.104	−0.696	0.487
Gender (male)	−0.389	0.152	−2.555	**0.011**
Generation status (first generation)	−0.244	0.110	−2.229	**0.026**
Vote 1*Vote 2	0.294	0.153	1.918	0.055
Vote 1*gender	0.228	0.120	1.901	0.057
Vote 2*gender	0.234	0.152	1.545	0.122

[a]Logit (homework explanation) was calculated as follows: V1 + V2 + gender + generation status + V1*V2 + V1*gender + V2*gender (see Table S4 in the supplemental material for information on model selection). *P* values in boldface indicate significance ($P < 0.05$).

quartile, we saw that the lowest quartile being present in class did not significantly improve their later homework explanations (Fig. 4B; see also Table S5). Conversely, the top three quartiles (Q2, Q3, and Q4) appeared to experience a larger benefit of being present, but this only reached statistical significance for Q2 and Q4.

Research question 3: How do clicker vote patterns relate to performance on isomorphic exam questions?

We next investigated how clicker vote patterns related to student performance on isomorphic exam questions. This model predicted student performance on isomorphic exam questions as the outcome variable and again included each clicker vote along with an interaction term. In this case, we found that the V1*V2 interaction term had a significant effect (Table 4), so we visualized and conducted *post hoc* analysis of the various pathways that students could take through the clicker question sequence (Fig. 5; see also Table S7 in the supplemental material). We found that answering both votes correctly (i.e., correct-correct) was associated with a student scoring significantly higher on the corresponding exam question compared to all other vote patterns (i.e., incorrect-correct, incorrect-incorrect, correct-incorrect), while none of these other patterns predicted significantly higher exam performance relative to each other.

We then analyzed how these clicker patterns compared when a student was absent for a targeted clicker question. For

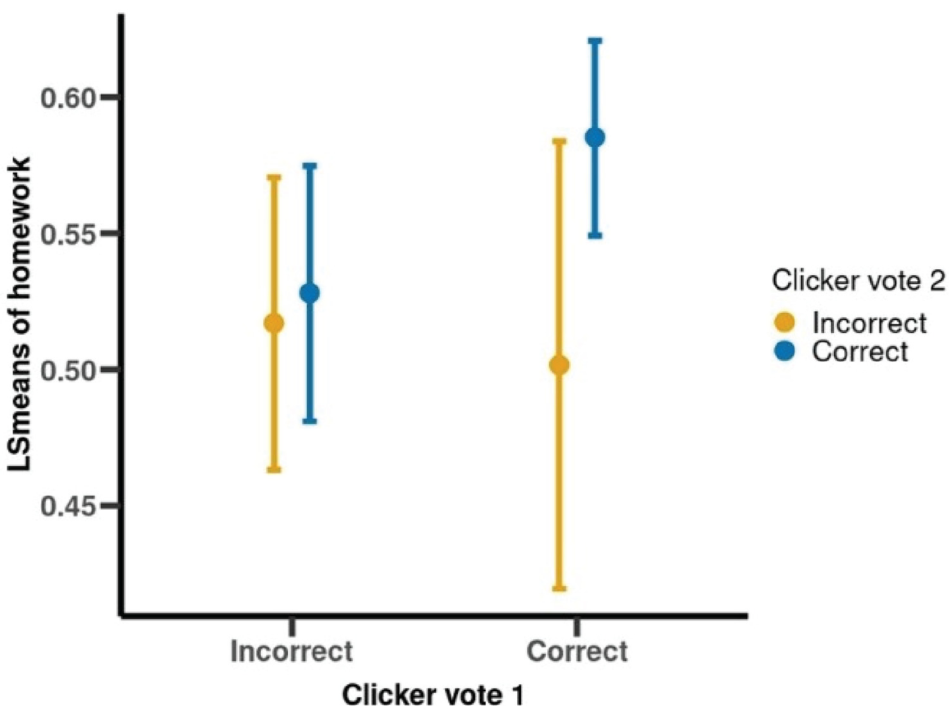

FIG 3. Effect of clicker vote patterns on homework explanations. Dots represent modeled least-squared means and bars represent 95% confidence intervals.

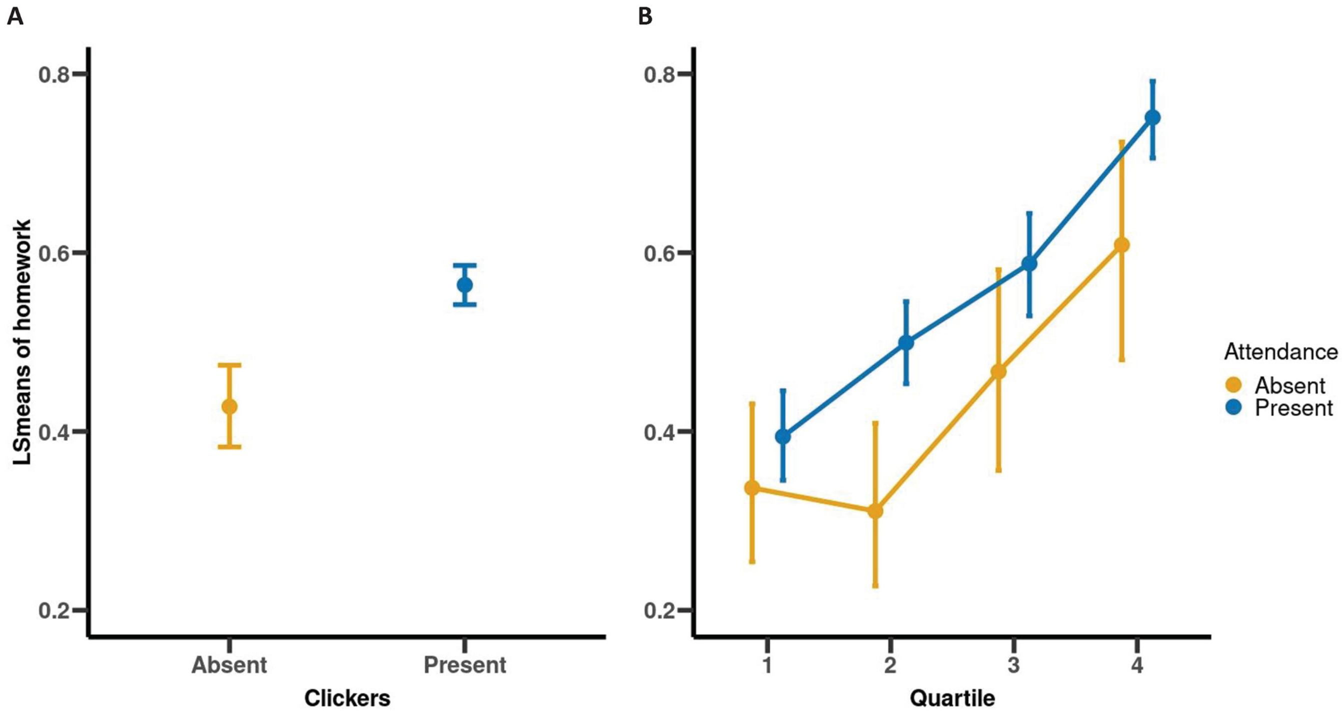

FIG 4. Effect of clicker participation on homework explanations overall (A) and for each student quartile (B). Dots represent modeled least-squared means and bars represent 95% confidence intervals.

this analysis, we collapsed the three vote patterns that did not significantly differ from each other (i.e., incorrect-correct, incorrect-incorrect, correct-incorrect). We discovered that students answering both votes correctly scored significantly higher than either of the other groups and that students with any other clicker pattern also performed significantly higher on the exam questions than students who were absent (Fig. 6A; see also Table S8).

We saw a more nuanced trend emerge when we explored student quartiles (Fig. 6B; see also Table S9). Compared to being absent, the lowest quartile saw no significant benefits of either answering correctly on both votes or having any other clicker pattern. Conversely, the top three quartiles saw an apparent benefit of answering both votes correctly, reaching statistical significance for Q2 and Q4. For the three top quartiles, having any other clicker pattern was not significantly better than being absent.

Research question 4: How do homework explanations relate to performance on isomorphic exam questions?

While the clicker pattern analysis provided insight into how student experiences during class related to their later understandings, we also wanted to see how student homework

TABLE 4
Effect of clicker vote patterns on isomorphic exam question performance[a]

Factor	Estimate	SE	*Z* value	*P* value
(Intercept)	1.265	0.091	13.950	**<0.001**
Vote 1 correct	−0.227	0.104	−2.175	**0.030**
Vote 2 correct	−0.086	0.071	−1.207	0.228
Homework completed	−0.062	0.031	−1.974	**0.048**
Gender (male)	−0.048	0.094	−0.511	0.610
Generation status (first generation)	−0.049	0.126	−0.387	0.699
Race and ethnicity (URM)	−0.231	0.118	−1.959	0.050
Vote 1*Vote 2	0.535	0.110	4.849	**<0.001**
Vote 1*gender	0.166	0.081	2.043	**0.041**
Vote 2*generation status	−0.158	0.110	−1.443	0.149

[a]Logit (exam score) was calculated as follows: V1 + V2 + homework completion + gender + generation status + V1*V2 + V1*gender + V2*generation status (see Table S6 in the supplemental material for information on model selection). *P* values in boldface indicate significance ($P < 0.05$).

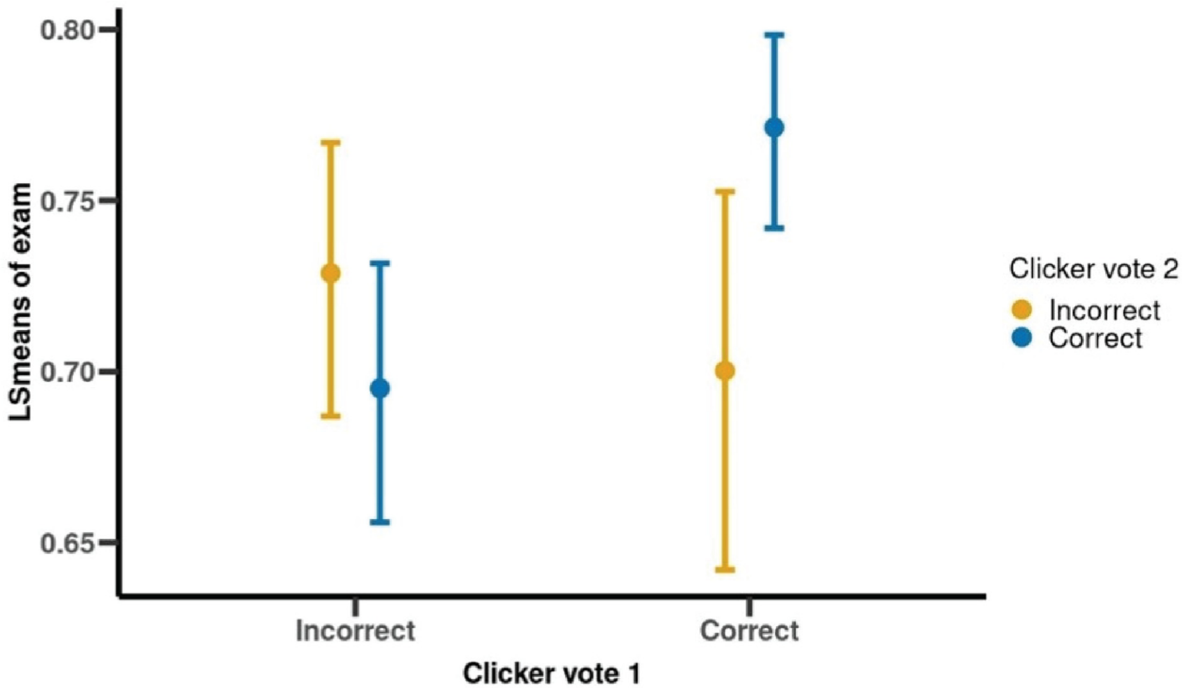

FIG 5. Effect of clicker vote patterns on isomorphic exam question performance. Dots represent modeled least-squared means and bars represent 95% confidence intervals.

explanations related to their performance on isomorphic exam questions. Students who gave a homework explanation that demonstrated understanding were more likely than students who gave an incomplete or incorrect explanation to answer the isomorphic exam question correctly (Fig. 7A and Table 5). Furthermore, this relationship was maintained across all quartiles, suggesting that a student's homework explanation generally predicted how well they performed on the related exam question (Fig. 7B; see also Table S11 in the supplemental material).

DISCUSSION

This study investigated various associations between student participation in clicker questions, their opportunity and ability to explain concepts on related homework activities, and their subsequent performance on isomorphic exam questions. Overall, we detected significant benefits of participating in clicker questions and completing the homework activity (Fig. 1A and Table 2), suggesting that each of

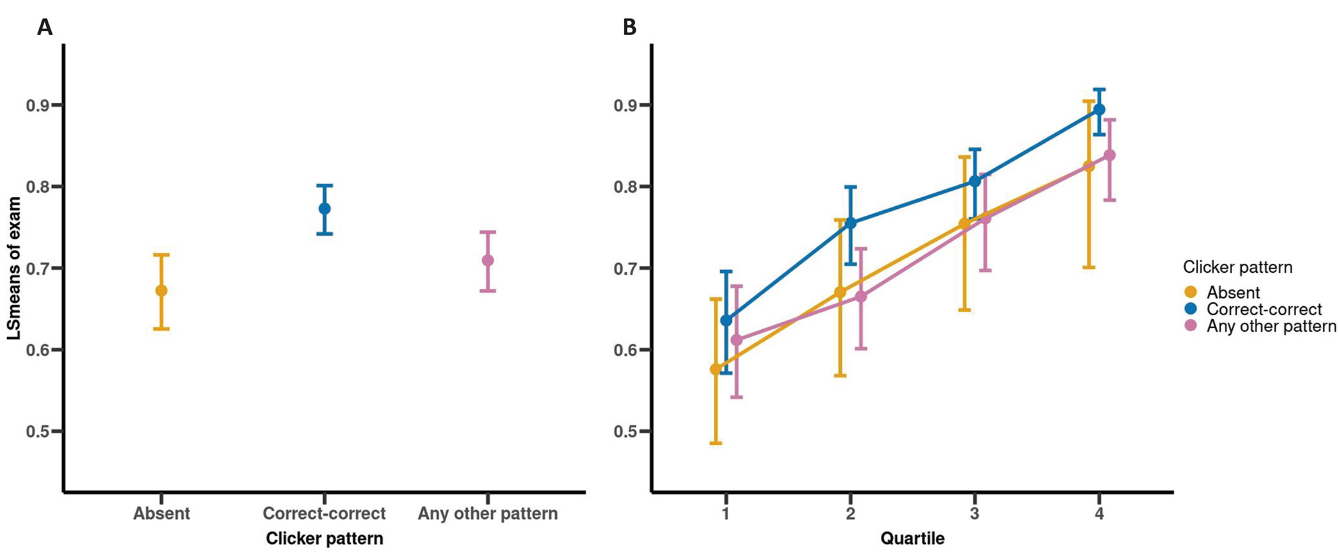

FIG 6. Effect of clicker vote patterns on isomorphic exam question performance compared to being absent overall (A) and for each student quartile (B). Dots represent modeled least-squared means and bars represent 95% confidence intervals. Absent, did not submit clicker votes; correct-correct, answer both votes correctly; any other pattern, submitted incorrect-correct, incorrect-incorrect, or correct-incorrect votes.

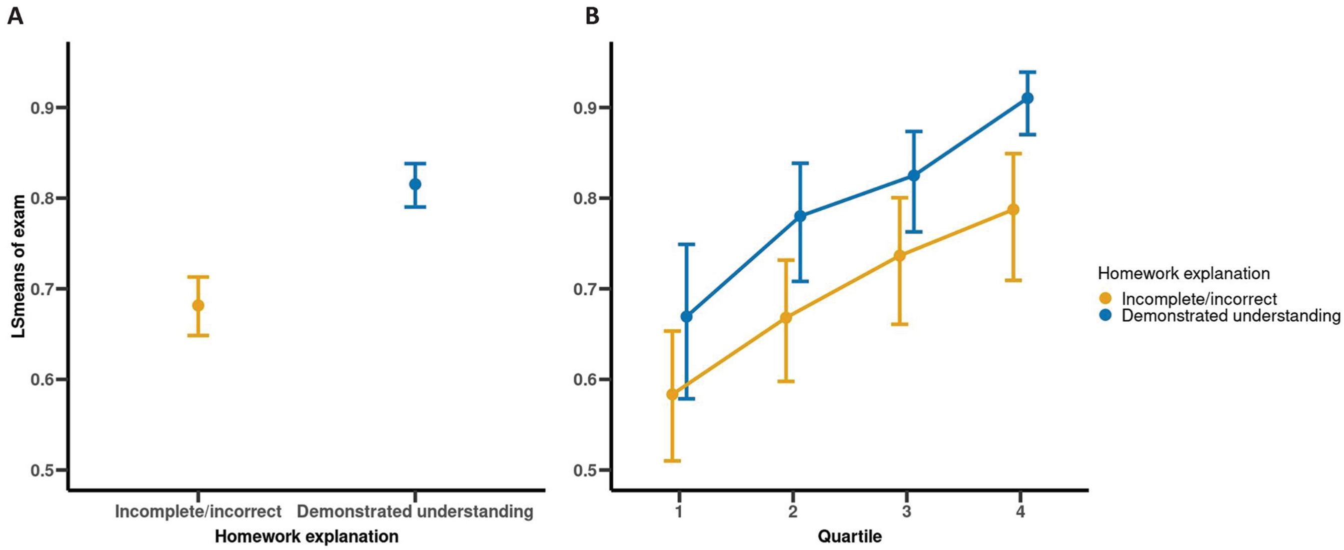

FIG 7. Effect of homework explanation on isomorphic exam question performance overall (A) and for each student quartile (B). Dots represent modeled least-squared means and bars represent 95% confidence intervals.

these instructional approaches had a net positive influence on student outcomes. This data set also provided a means by which to better understand the different ways that students progressed through these different stages.

We analyzed how clicker vote patterns related to homework explanations and isomorphic exam questions, two delayed measures of the learning that students might have achieved as a result of the clicker process. We found that being present for a clicker question generally predicted whether a student would demonstrate understanding on their later homework explanation (Fig. 4A; see also Table S5 in the supplemental material). Our statistical analyses do not allow us to attribute this effect to any particular vote pattern (Table 3), but we noted that students answering both clicker votes correctly had a tendency to perform better on the homework (Fig. 3). Given that all students who attended class heard an explanation for each MTF statement and could take notes that they might then use for the homework question, we can see how being present could have given a general boost to students and how the shared experience of hearing an explanation might have made students more similar to each other in terms of how they described the clicker answers. For isomorphic exam questions, we again saw a general benefit of being present for a clicker question, but the improved performance occurred predominantly when students answered both clicker votes correctly (i.e., correct-correct) (Fig. 5; see also Table S7). Interestingly, in both cases, students who changed from an incorrect to a correct clicker vote did not perform any better than students who were incorrect on both votes, suggesting that the improved performance they experienced as a result of peer discussion could not be recapitulated at later time points.

Finally, we explored how the effects varied by student quartile. Unfortunately, the lowest-quartile (Q1) students saw little benefit from participating in clicker questions compared to being absent. Students in the higher quartiles saw more positive outcomes, showing various effects of participating in clicker questions on subsequent homework explanations and isomorphic exam questions, culminating with students in the top quartile (Q4) demonstrating the greatest overall benefits of clicker participation on exam performance (Fig. 2B; see

TABLE 5
Effect of homework explanation on isomorphic exam question performance[a]

Factor	Estimate	SE	*Z* value	*P* value
(Intercept)	0.728	0.103	7.070	<0.001
Clicker participated	0.266	0.091	2.915	0.004
Homework explanation (demonstrated understanding)	0.621	0.068	9.158	<0.001
Gender (male)	−0.060	0.092	−0.654	0.513
Generation status (first generation)	−0.140	0.087	−1.615	0.106
Homework explanation*gender	0.206	0.110	1.867	0.062

[a]Logit (exam score) was calculated as follows: clicker participation + homework explanation + gender + generation status + homework explanation*gender (see Table S10 in the supplemental material for information on model selection). *P* values in boldface indicate significance ($P < 0.05$).

also Table S2). Conversely, the positive experience of providing a homework explanation had a more even effect (Fig. 2C; see also Table S3), and students who were able to provide a correct homework explanation were more likely to perform well on the related exam question across quartiles (Fig. 7; see also Tables S10 and S11 in the supplemental material).

Altogether, our findings suggest a potential role of clickers and associated homework explanations in reinforcing nascent conceptions. Indeed, answering both clicker votes correctly and providing a homework explanation that demonstrated understanding each predicted a higher likelihood that a student would perform better on the later exam question. While this could simply represent a selection effect for concepts that students already knew, we noted that the effect required students to participate in clicker questions (i.e., it was greater than the knowledge they had when absent, even for higher-performing students) and that the effect specifically occurred for the correct-correct pattern (i.e., just being correct on the first or second vote did not have the same effect). This result resonates with a prior study that examined student reasoning before and after peer discussion and found that explanation quality improved for students who correctly answered both clicker votes, suggesting that peer discussion specifically enabled these students to enhance their understandings (56). With respect to homework, students in our study were randomly assigned to address only one of the two targeted clicker questions that occurred in a given week, so the homework explanation effect can be linked more directly to the act of seeing and explaining a particular question. Additional research would be needed to distinguish potential effects due to cuing students toward test content, spending more time with certain concepts, and writing explanations to targeted questions.

Our finding that the highest-quartile (Q4) students experienced the most consistent benefits of clicker participation resonates with our previous clicker research (35) and other studies finding that active learning differentially benefits top students. This phenomenon has been observed in contexts utilizing small group work, such as class periods replaced with small group exercises (57) or courses taught using the SCALE-UP approach (58). Higher-performing students have also been seen to benefit more from retrieving versus copying, suggesting that the nature of the task can influence how different groups learn (41). We speculate that higher-performing students in our study may have driven their peer discussion groups, and this experience enabled them to articulate their reasoning and refine their understandings during the subsequent instructor explanation. In turn, lower-performing students may have had less opportunity to explain their ideas during peer discussion, thus limiting their ability to retrieve, formulate, and advance their conceptual understanding.

One limitation of our study was that we were unable to distinguish clicker participation from other class activities occurring that day, such as lecturing on the given topic. Previous research has attempted to disentangle these effects by comparing targeted exam question performance to alternative course offerings that did not include the corresponding clicker question or to various types of nontargeted exam questions and has generally attributed the resulting benefits to clicker questions (17, 24, 59–61). Our research here focused more directly on understanding clicker vote patterns and differential benefits of the various course structures.

This research reflects the challenges associated with implementing and studying activities embedded within a complex course environment, and we recognize how the instructional context shaped student engagement in ways that might differ from other courses (38, 40, 62). Our statistical models also revealed effects of gender and generation status, and while we did not examine these effects in detail, these findings further underscore the complex social dimensions underlying course activities and performance. Overall, our results provide support for the use of clicker questions and associated homework to guide and reinforce emerging understandings. However, additional work is needed to identify clicker approaches that provide benefits for lower-performing students and in cases where students switch from incorrect to correct votes. This can potentially be achieved through more deliberate attention to the composition, norms, or procedures of peer discussion groups. For example, previous research has suggested that lower-performing students may benefit from participating in more homogeneous groups because this allows them to work through activities without the influence of a higher-performing peer (63). Students engage in a variety of behaviors during peer discussions (64), and so modeling and monitoring discussion behaviors more explicitly might also help students engage with incorrect ideas and develop rationales for correct answers.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

This work was funded by grants from the National Science Foundation (NSF; DUE-1610621 and DUE-2044243). We received technical support from Jessica Mercer, who was paid using funds from the associated NSF grant (DUE-1610621). No other individuals provided writing or other assistance, and there were no potential conflicts for study participants. The study sponsors had no role in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; or in the decision to submit the report for publication.

REFERENCES

1. American Association for the Advancement of Science. 2011. Vision and change in undergraduate biology education: a call to action. AAAS, Washington, DC.

2. Ambrose SA, Bridges MW, DiPietro M, Lovett MC, Norman MK, Mayer RE. 2010. How learning works: seven research-based principles for smart teaching. Jossey-Bass, San Francisco, CA.
3. Prince M. 2004. Does active learning work? A review of the research. J Eng Educ 93:223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x.
4. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 111:8410–8415. https://doi.org/10.1073/pnas.1319030111.
5. Theobald EJ, Hill MJ, Tran E, Agrawal S, Arroyo EN, Behling S, Chambwe N, Cintrón DL, Cooper JD, Dunster G, Grummer JA, Hennessey K, Hsiao J, Iranon N, Jones L, Jordt H, Keller M, Lacey ME, Littlefield CE, Lowe A, Newman S, Okolo V, Olroyd S, Peecook BR, Pickett SB, Slager DL, Caviedes-Solis IW, Stanchak KE, Sundaravardan V, Valdebenito C, Williams CR, Zinsli K, Freeman S. 2020. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proc Natl Acad Sci U S A 117:6476–6483. https://doi.org/10.1073/pnas.1916903117.
6. Armbruster P, Patel M, Johnson E, Weiss M. 2009. Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. CBE Life Sci Educ 8:203–213. https://doi.org/10.1187/cbe.09-03-0025.
7. Caldwell JE. 2007. Clickers in the large classroom: current research and best-practice tips. CBE Life Sci Educ 6:9–20. https://doi.org/10.1187/cbe.06-12-0205.
8. Mazur E. 1996. Peer instruction: a user's manual. Addison-Wesley, Upper Saddle River, NJ.
9. Dufresne RJ, Gerace WJ, Leonard WJ, Mestre JP, Wenk L. 1996. Classtalk: a classroom communication system for active learning. J Comput High Educ 7:3–47. https://doi.org/10.1007/BF02948592.
10. Fagen AP, Crouch CH, Mazur E. 2002. Peer instruction: results from a range of classrooms. Phys Teach 40:206–209. https://doi.org/10.1119/1.1474140.
11. Vickrey T, Rosploch K, Rahmanian R, Pilarz M, Stains M. 2015. Research-based implementation of Peer Instruction: a literature review. CBE Life Sci Educ 14:es3. https://doi.org/10.1187/cbe.14-11-0198.
12. Hunsu NJ, Adesope O, Bayly DJ. 2016. A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. Comput Educ 94:102–119. https://doi.org/10.1016/j.compedu.2015.11.013.
13. Kay RH, LeSage A. 2009. Examining the benefits and challenges of using audience response systems: a review of the literature. Comput Educ 53:819–827. https://doi.org/10.1016/j.compedu.2009.05.001.
14. Chien Y-T, Chang Y-H, Chang C-Y. 2016. Do we click in the right way? A meta-analytic review of clicker-integrated instruction. Educ Res Rev 17:1–18. https://doi.org/10.1016/j.edurev.2015.10.003.
15. Cortright RN, Collins HL, DiCarlo SE. 2005. Peer instruction enhanced meaningful learning: ability to solve novel problems. Adv Physiol Educ 29:107–111. https://doi.org/10.1152/advan.00060.2004.
16. Giuliodori MJ, Lujan HL, DiCarlo SE. 2006. Peer instruction enhanced student performance on qualitative problem-solving questions. Adv Physiol Educ 30:168–173. https://doi.org/10.1152/advan.00013.2006.
17. Crossgrove K, Curran KL. 2008. Using clickers in nonmajors- and majors-level biology courses: student opinion, learning, and long-term retention of course material. CBE Life Sci Educ 7:146–154. https://doi.org/10.1187/cbe.07-08-0060.
18. Brazeal KR, Couch BA. 2017. Student buy-in toward formative assessments: the influence of student factors and importance for course success. J Microbiol Biol Educ 18:18.1.6. https://doi.org/10.1128/jmbe.v18i1.1235.
19. Brazeal KR, Brown TL, Couch BA. 2016. Characterizing student perceptions of and buy-in toward common formative assessment techniques. CBE Life Sci Educ 15:ar73. https://doi.org/10.1187/cbe.16-03-0133.
20. Buil I, Catalán S, Martínez E. 2016. Do clickers enhance learning? A control-value theory approach. Comput Educ 103:170–182. https://doi.org/10.1016/j.compedu.2016.10.009.
21. Knight JK, Wood WB. 2005. Teaching more by lecturing less. Cell Biol Educ 4:298–310. https://doi.org/10.1187/05-06-0082.
22. Burnstein RA, Lederman LM. 2001. Using wireless keypads in lecture classes. Phys Teach 39:8–11. https://doi.org/10.1119/1.1343420.
23. Lasry N, Mazur E, Watkins J. 2008. Peer instruction: from Harvard to the two-year college. Am J Phys 76:1066–1069. https://doi.org/10.1119/1.2978182.
24. Shapiro A. 2009. An empirical study of personal response technology for improving attendance and learning in a large class. J Scholarsh Teach Learn 9:13–26.
25. Crouch CH, Mazur E. 2001. Peer instruction: ten years of experience and results. Am J Phys 69:970–977. https://doi.org/10.1119/1.1374249.
26. McConnell DA, Steer DN, Owens KD, Knott JR, Van Horn S, Borowski W, Dick J, Foos A, Malone M, McGrew H, Greer L, Heaney PJ. 2006. Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. J Geosci Educ 54:61–68. https://doi.org/10.5408/1089-9995-54.1.61.
27. Mora G. 2010. Peer instruction and lecture tutorials equally improve student learning in introductory geology classes. J Geosci Educ 58:286–296. https://doi.org/10.5408/1.3559693.
28. Balta N, Michinov N, Balyimez S, Ayaz MF. 2017. A meta-analysis of the effect of peer instruction on learning gain: identification of informational and cultural moderators. Int J Educ Res 86:66–77. https://doi.org/10.1016/j.ijer.2017.08.009.
29. Lasry N, Charles E, Whittaker C. 2016. Effective variations of peer instruction: the effects of peer discussions, committing to an answer, and reaching a consensus. Am J Phys 84:639–645. https://doi.org/10.1119/1.4955150.
30. Rao SP, DiCarlo SE. 2000. Peer instruction improves performance on quizzes. Adv Physiol Educ 24:51–55. https://doi.org/10.1152/advances.2000.24.1.51.
31. 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. https://doi.org/10.1126/science.1165919.
32. Smith MK, Wood WB, Krauter K, Knight JK. 2011. Combining peer discussion with instructor explanation increases student learning from in-class concept questions. CBE Life Sci Educ 10:55–63. https://doi.org/10.1187/cbe.10-08-0101.

33. Zingaro D, Porter L. 2014. Peer instruction in computing: the value of instructor intervention. Comput Educ 71:87–96. https://doi.org/10.1016/j.compedu.2013.09.015.
34. Watkins EP, Sabella MS. 2008. Examining the effectiveness of clickers on promoting learning by tracking the evolution of student responses. AIP Conf Proc 1064:223–226. https://doi.org/10.1063/1.3021260.
35. Hubbard JK, Couch BA. 2018. The positive effect of in-class clicker questions on later exams depends on initial student performance level but not question format. Comput Educ 120:1–12. https://doi.org/10.1016/j.compedu.2018.01.008.
36. Strandberg M. 2013. Homework: is there a connection with classroom assessment? A review from Sweden. Educ Res 55:325–346. https://doi.org/10.1080/00131881.2013.844936.
37. Knight JK, Wise SB, Southard KM. 2013. Understanding clicker discussions: student reasoning and the impact of instructional cues. CBE Life Sci Educ 12:645–654. https://doi.org/10.1187/cbe.13-05-0090.
38. Turpen C, Finkelstein ND. 2010. The construction of different classroom norms during peer instruction: students perceive differences. Phys Rev Spec Top Phys Educ Res 6:e020123. https://doi.org/10.1103/PhyRevSTPER.6.020123.
39. Lucas A. 2009. Using peer instruction and I-clickers to enhance student participation in calculus. PRIMUS 19:219–231. https://doi.org/10.1080/10511970701643970.
40. 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 Educ Res 5: e020101. https://doi.org/10.1103/PhysRevSTPER.5.020101.
41. Carpenter SK, Lund TJS, Coffman CR, Armstrong PI, Lamm MH, Reason RD. 2016. A classroom study on the relationship between student achievement and retrieval-enhanced learning. Educ Psychol Rev 28:353–375. https://doi.org/10.1007/s10648-015-9311-9.
42. Brassil CE, Couch BA. 2019. Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian analysis. Int J STEM Educ 6:16. https://doi.org/10.1186/s40594-019-0169-0.
43. Couch BA, Hubbard JK, Brassil CE. 2018. Multiple–true–false questions reveal the limits of the multiple–choice format for detecting students with incomplete understandings. Bioscience 68:455–463. https://doi.org/10.1093/biosci/biy037.
44. Frisbie DA. 2005. The multiple true-false item format: a status review. Educ Meas Issues Pract 11:21–26. https://doi.org/10.1111/j.1745-3992.1992.tb00259.x.
45. Couch BA, Wright CD, Freeman S, Knight JK, Semsar K, Smith MK, Summers MM, Zheng Y, Crowe AJ, Brownell SE. 2019. GenBio-MAPS: a programmatic assessment to measure student understanding of Vision and Change core concepts across general biology programs. CBE Life Sci Educ 18:ar1. https://doi.org/10.1187/cbe.18-07-0117.
46. Semsar K, Brownell S, Couch BA, Crowe AJ, Smith MK, Summers MM, Wright CD, Knight JK. 2019. Phys-MAPS: a programmatic physiology assessment for introductory and advanced undergraduates. Adv Physiol Educ 43:15–27. https://doi.org/10.1152/advan.00128.2018.
47. Summers MM, Couch BA, Knight JK, Brownell SE, Crowe AJ, Semsar K, Wright CD, Smith MK. 2018. EcoEvo-MAPS: an ecology and evolution assessment for introductory through advanced undergraduates. CBE Life Sci Educ 17:ar18. https://doi.org/10.1187/cbe.17-02-0037.
48. Couch BA, Wood WB, Knight JK. 2015. The Molecular Biology Capstone assessment: a concept assessment for upper-division molecular biology students. CBE Life Sci Educ 14:ar10. https://doi.org/10.1187/cbe.14-04-0071.
49. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01.
50. Theobald R, Freeman S. 2014. Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. CBE Life Sci Educ 13:41–48. https://doi.org/10.1187/cbe-13-07-0136.
51. Kolpikova EP, Chen DC, Doherty JH. 2019. Does the format of preclass reading quizzes matter? An evaluation of traditional and gamified, adaptive preclass reading quizzes. CBE Life Sci Educ 18:ar52. https://doi.org/10.1187/cbe.19-05-0098.
52. Lenth RV. 2016. Least-squares means: the R package lsmeans. J Stat Softw 69:1–33. https://doi.org/10.18637/jss.v069.i01.
53. Revelle W. 2021. psychTools: tools to accompany the "psych" package for psychological research (2.1.3). https://CRAN.R-project.org/package=psychTools. Retrieved 3 May 2021.
54. Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. Biom J 50:346–363. https://doi.org/10.1002/bimj.200810425.
55. Gómez-Rubio V. 2017. Book review. Ggplot2: elegant graphics for data analysis (2nd edition). J Stat Softw 77:1–3. https://doi.org/10.18637/jss.v077.b02.
56. Brooks BJ, Koretsky MD. 2011. The influence of group discussion on students' responses and confidence during peer instruction. J Chem Educ 88:1477–1484. https://doi.org/10.1021/ed101066x.
57. Marbach-Ad G, Rietschel CH, Saluja N, Carleton KL, Haag ES. 2016. The use of group activities in introductory biology supports learning gains and uniquely benefits high-achieving students. J Microbiol Biol Educ 17:360–369. https://doi.org/10.1128/jmbe.v17i3.1071.
58. Beichner RJ, Saul JM. 2003. Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) Project, p 1–17. *In* Proceedings of the International School of Physics Enrico Fermi, Italian Physical Society, Bologna, Italy.
59. Shapiro AM, Sims-Knight J, O'Rielly GV, Capaldo P, Pedlow T, Gordon L, Monteiro K. 2017. Clickers can promote fact retention but impede conceptual understanding: the effect of the interaction between clicker use and pedagogy on learning. Comput Educ 111:44–59. https://doi.org/10.1016/j.compedu.2017.03.017.
60. Levesque AA. 2011. Using clickers to facilitate development of problem-solving skills. CBE Life Sci Educ 10:406–417. https://doi.org/10.1187/cbe.11-03-0024.
61. Pan SC, Cooke J, Little JL, McDaniel MA, Foster ER, Connor LT, Rickard TC. 2019. Online and clicker quizzing on jargon terms enhances definition-focused but not conceptually focused biology exam performance. CBE Life Sci Educ 18:ar54. https://doi.org/10.1187/cbe.18-12-0248.
62. Brazeal KR, Brown TL, Couch BA. 2021. Connecting activity

implementation characteristics to student buy-in toward and utilization of formative assessments within undergraduate biology courses. J STEM Educ Res 4:329–362. https://doi.org/10.1007/s41979-021-00054-2.

63. Jensen JL, Lawson A. 2011. Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. CBE Life Sci Educ 10:64–73. https://doi.org/10.1187/cbe.10-07-0089.

64. James MC, Willoughby S. 2011. Listening to student conversations during clicker questions: what you have not heard might surprise you! Am J Phys 79:123–132. https://doi.org/10.1119/1.3488097.