Response switching and self-efficacy in Peer Instruction classrooms

Kelly Miller,[1,*] Julie Schell,[1] Andrew Ho,[2] Brian Lukoff,[1] and Eric Mazur[1]

[1]*Department of Physics and Division of Engineering and Applied Sciences,*
Harvard University, Cambridge, Massachusetts 02138, USA
[2]*Graduate School of Education, Harvard University, 6 Appian Way, Cambridge,*
Massachusetts 02138, USA

Peer Instruction, a well-known student-centered teaching method, engages students during class through structured, frequent questioning and is often facilitated by classroom response systems. The central feature of any Peer Instruction class is a conceptual question designed to help resolve student misconceptions about subject matter. We provide students two opportunities to answer each question—once after a round of individual reflection and then again after a discussion round with a peer. The second round provides students the choice to "switch" their original response to a different answer. The percentage of right answers typically increases after peer discussion: most students who answer incorrectly in the individual round switch to the correct answer after the peer discussion. However, for any given question there are also students who switch their initially right answer to a wrong answer and students who switch their initially wrong answer to a different wrong answer. In this study, we analyze response switching over one semester of an introductory electricity and magnetism course taught using Peer Instruction at Harvard University. Two key features emerge from our analysis: First, response switching correlates with academic self-efficacy. Students with low self-efficacy switch their responses more than students with high self-efficacy. Second, switching also correlates with the difficulty of the question; students switch to incorrect responses more often when the question is difficult. These findings indicate that instructors may need to provide greater support for difficult questions, such as supplying cues during lectures, increasing times for discussions, or ensuring effective pairing (such as having a student with one right answer in the pair). Additionally, the connection between response switching and self-efficacy motivates interventions to increase student self-efficacy at the beginning of the semester by helping students develop early mastery or to reduce stressful experiences (i.e., high-stakes testing) early in the semester, in the hope that this will improve student learning in Peer Instruction classrooms.

I. INTRODUCTION

Peer Instruction, a student-centered teaching methodology developed in the 1990s [1], engages students during class through a sequence of questioning and discussion. Questions are generally conceptual in nature and probe students' abilities to apply their understanding to solve conceptual problems. These questions, called "ConcepTests," are the centerpiece of Peer Instruction. Students first respond to a ConcepTest individually (first round) and then respond again to the same question after discussing with a peer (second round). The students' responses are typically recorded via classroom response systems, which allow the instructor to track the class-wide percentage of right answers between the two rounds of questioning. This percentage almost always increases for a given question after peer discussion (that is, between the first and second rounds of questioning). However, for any given question, there are also both students who switch their initially right answer to a wrong answer and students who switch their initially wrong answer to a different wrong answer. Understanding the difference between the different types of response switching helps provide insight into student cognition in Peer Instruction environments. We pose three research questions. First, how often does response switching occur and in which direction? Second, what is the relationship between response switching and predetermined student characteristics, specifically gender, precourse physics knowledge, and precourse self-efficacy? Third, what is the relationship between response switching and ConcepTest difficulty?

While the relationship between precourse knowledge and student achievement is well studied, there is increasing interest in the relationship between noncognitive dimensions, such as self-efficacy and academic performance. Self-efficacy refers to an individual's belief that one can successfully complete a task [2]. Self-efficacy is a strong predictor for performance in science courses [3–7] as well

---

*kmiller@harvard.edu

as science career choices [4,5,8,9]. Self-efficacy also influences a number of factors that are relevant to learning in a Peer Instruction environment, such as perseverance and self-regulated learning [10]. Students with higher self-efficacy are more persistent, harder working, participate more readily, and experience fewer negative emotions in the face of difficulty than students with lower self-efficacy [11]. In their study on the effects of self-efficacy on student behavior during conceptual learning, Bouffard-Bouchard *et al.* [12] found that "efficacious students were better at monitoring their working time, more persistent, less likely to reject correct hypotheses prematurely, and better at solving conceptual problems than inefficacious students of equal ability" (Zimmerman [13]). In this paper, we show that response switching is related to students' precourse self-efficacy. Given this relationship between student response switching and self-efficacy, it is possible that improving a student's self-efficacy would enhance the effectiveness of Peer Instruction for that student. Recent work has identified a technique for identifying events in small group settings that impact self-efficacy in physics learning [14]. These techniques have exciting implications for improving students' self-efficacy in Peer Instruction environments.

II. METHODS

We gathered ConcepTest (CT) response data over an entire semester in an introductory, calculus-based electricity and magnetism class taught using Peer Instruction at Harvard University, by one of the authors (E. M.), a professor with over twenty years of experience teaching with this method. CTs are short conceptual questions that focus on a single topic [1]. The class had 91 students (50 male students and 41 female students), the majority of whom were engineers or premedical students. The class met twice a week for 90 min and during each class, somewhere between five and nine CTs were posed. In total, 83 CTs were posed over the course of the entire semester. Students answered each CT in two rounds of questioning by entering their responses via Learning Catalytics, an online classroom response system. Students did not receive credit based on the correctness of their answer; rather, credit was only awarded for participation.

We divided each pair of CT responses given in the two rounds of Peer Instruction into one of five categories: (1) right-right (RR), the question is answered correctly during both rounds, (2) wrong-to-right (WR), the question is answered incorrectly in round 1 and correctly in round 2, (3) right-to-wrong (RW), the question is answered correctly in round 1 and incorrectly in round 2 (4) wrong-wrong same (WW-S), the question is answered with the same incorrect response in both rounds, and (5) wrong-wrong different (WW-D), the question is answered with a different wrong response in both rounds. Only questions answered in

both rounds were included in the analysis. If a student responded to a question in round 1 but not in round 2 (or vice versa), we did not include those responses in our analysis.

We applied a two-parameter item response model [15] to the first round of responses across all items to estimate the difficulty of each item (b parameter). The item response model helps to scale the difficulty of each question to support the generalization of the item difficulties to other populations of students.

The Conceptual Survey of Electricity and Magnetism (CSEM) [16] and a self-efficacy survey (the Peer Instruction Self-Efficacy Instrument) [17] were each administered twice, once as pretests after the second class in the semester, and again as post-tests at the end of the semester (after the final exam but before students received their final grade). The Peer Instruction Self-Efficacy Instrument (PISE) [17] was developed by two of the authors and is based on the Sources of Self-Efficacy in Science Courses (SOSESC) survey [18] and Bandura [11]. This survey is comprised of 21 items scored on a five-point Likert scale. Students are asked if they "strongly agree," "agree," are "neutral," "disagree," or "strongly disagree" with statements about whether they think they will be successful in a number of physics-related tasks (e.g., solving difficult physics problems or communicating physics successfully to a peer). The complete survey is in Appendix B. All items on the PISE have point biserial coefficients [19] greater than 0.2 except for item 13, which was dropped from analysis. The reliability estimated by Cronbachs alpha [20] is 0.88 for the pretest and 0.85 for the post-test.

III. RESULTS

A. Descriptive statistics on switching

Figure 1 shows the extent to which students switch their CT responses between the first and second round of questioning. Figure 1(a) shows the fraction of CTs that are switched (in any direction) out of the total number of CTs that each student answers. Figures 1(b)–1(d) show the fraction of CTs that are switched in a specific direction out of the total number of CTs that are switched [(b): wrong to right; (c): wrong to different wrong; (d): right to wrong]. On average, students switch 44% of the CTs they answer over the course of the semester. Of the switched responses, 73% are from wrong to right, 17% from wrong to a different wrong, and 10% from right to wrong.

B. Normalizing switching

When response switching is measured as a fraction of all responses that are switched as in Fig. 1, the response switching is confounded with the frequency of right (or wrong) answers in round 1. Normalizing the variables with respect to the response in round 1 provides us with an
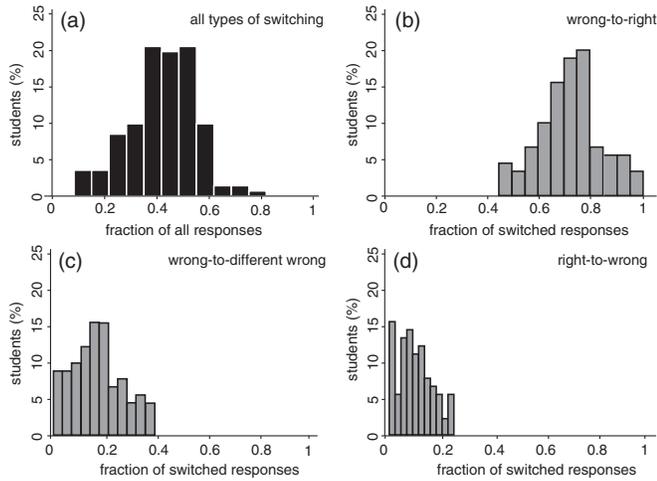
FIG. 1.   (a) Fraction of all responses that are switched (top left). (b) Fraction of all switched responses that are switched from wrong to right (top right). (c) Fraction of all switched responses that are switched from wrong to a different wrong (bottom left), and (d) fraction of all switched responses that are switched from right to a wrong (bottom right).

adjusted measure of the response switching, independent of how many times a student was right (or wrong) in round 1. Figure 2 illustrates the need for this normalization. In Fig. 2, each data point represents the relationship between the fraction of answered items that were switched from wrong to right and the number of items a student answered incorrectly in round 1. When the WR transition is not normalized with respect to the round 1 response, the number of wrong round 1 responses confounds the number
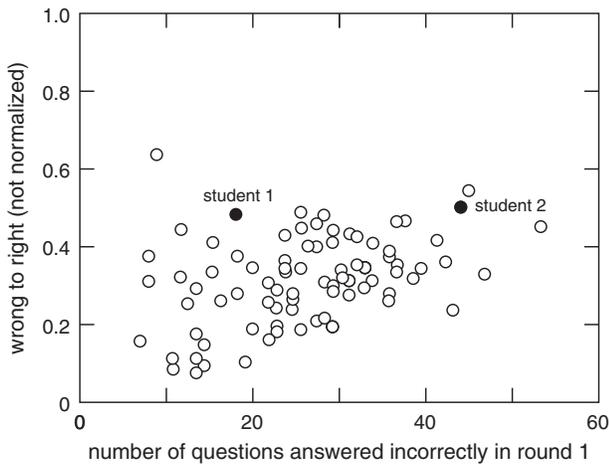
FIG. 2.   Fraction of all CT responses that each student switches from wrong to right plotted as a function of the number of questions the student answers incorrectly in round 1. Students 1 and 2 both switch approximately 50% of the questions they answer from wrong to right. These two students are not directly comparable, however, because student 1 has far fewer incorrect answers in round 1 than student 2. Normalizing with respect to the number of incorrect answers in round 1 allows us to compare these two students.

of wrong to right switches. To illustrate this point, consider the two students highlighted in Fig. 1, both of whom switch approximately 50% of their responses from wrong to right. Student 1 answered less than 20 items incorrectly to begin with and therefore switched to the right answer fewer than 10 times. Student 2, on the other hand, had more than forty wrong responses in round 1 and therefore switched from wrong to right more than 20 times. Normalizing adjusts the response transition (WR) to account for the frequency of initial wrong answers.

To calculate each of the normalized transition variables we first sum the number of times each student's response falls into each of the five transition categories. Then, we normalize each of these sums to express them as a percentage of the times that the first response was wrong or right. The normalized version of the RR transition is computed by dividing the total number of questions where a student was right in both rounds by the number of right answers they provided in round 1. The WR transition is computed by dividing the total number of questions where students switch their answer from wrong to right by the number of wrong answers they provided in round 1. The RW transition was computed by dividing the total number of questions where a student switched their answer from right to wrong by the number the right answers they provided in round 1. A summary of how these normalized transition variables are calculated is provided in Appendix A.

C. Switching and self-efficacy

We find that response switching is a function of students' precourse self-efficacy. Students with low self-efficacy are both more likely to switch their responses and more likely to switch in a "negative" direction (from right to wrong and from wrong to a different wrong) than students with high self-efficacy. Students with high self-efficacy are much more likely to switch from wrong to right than students with low self-efficacy. Figure 3 shows the average normalized percentage of switched responses for students with low and high self-efficacy.

Table I displays the standardized regression parameters and significance metrics for two models, each of which predicts the proportion of responses switched from right to wrong (RW), wrong to right (WR), and wrong to a different wrong (WW-D), normalized with respect to the first response. In each set of two models, Model 1 controls for precourse student self-efficacy only whereas Model 2 controls for both precourse student self-efficacy and CSEM scores. Students with high precourse self-efficacy switch from right to wrong and from wrong to a different wrong significantly less often, and switch from wrong to right significantly more often than students with low self-efficacy ($p < 0.001$). This is true even when incoming physics knowledge is controlled for, indicating that the self-efficacy measurement is not simply a proxy for incoming
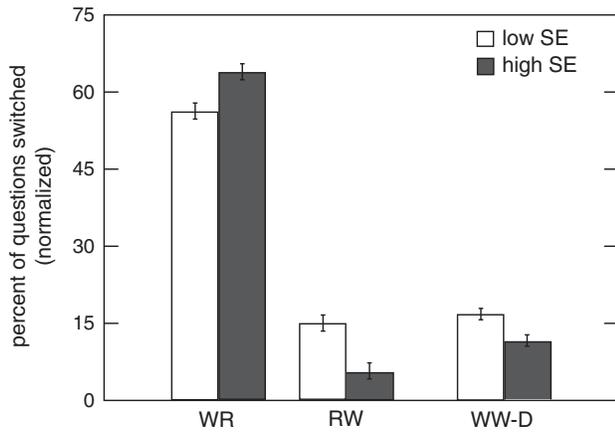
FIG. 3.   Average percent of questions switched in each of the three directions (wrong to right, right to wrong, wrong to different wrong) normalized with respect to the first response, for students with high and low self-efficacy

FIG. 4.   Average right-to-wrong switching (normalized) for students with different levels of agreement with the statements "I usually don't worry about my ability to solve physics problems" (left) and "I can communicate science effectively" (right). No students strongly disagreed with the statement "I can communicate science effectively" and therefore this column is missing from the figure on the right.

physics knowledge. The magnitude of the standardized coefficients represents their relative predictive power in each of the models; a comparison of the self-efficacy coefficient to the CSEM coefficient in Model 2 indicates that self-efficacy is more predictive of response switching than incoming physics knowledge.

Figure 3 shows the predicted proportion of questions switched in each of the three directions, based on model 2, for students with high and low self-efficacy. Students with self-efficacy scores at least 1 standard deviation greater than the mean were classified as those with high self-efficacy and compared to students with low self-efficacy (scores less than one standard deviation less than the mean). Students with high self-efficacy switch their responses from wrong to right more than students with low self-efficacy ($p < 0.05$). Students with low self-efficacy switch their responses from right to wrong ($p < 0.005$) and from wrong to a different wrong ($p < 0.05$) more than students with high self-efficacy. There are no statistically significant correlations between students' self-efficacy and their

response patterns that do not involve switching (RR, WW-S).

Students' responses to two individual items on the Peer Instruction Self-Efficacy Survey correlate strongly with switching from right to wrong. Figure 4 shows average right to wrong response switching (normalized) for students with different levels of agreement or disagreement with the statements "I usually don't worry about my ability to solve physics problems" (item 10) and "I can communicate science effectively" (item 20). A one-way analysis of variance indicates that students who strongly disagree with item 10 switch from right to wrong significantly more than students who agree (or disagree less strongly) with that item ($p < 0.001$). Similarly, students who disagree with item 20 switch from right to wrong significantly more than students who agree with that item ($p < 0.05$). These relationships are significant even after controlling for students' incoming CSEM scores. Therefore, independent of their actual physics ability, students with a low assessment of their problem solving and communicating science abilities are significantly more likely to switch their responses from right to wrong than students with a high assessment of those abilities.

TABLE I.   Standardized coefficients for linear regression models predicting proportion of responses switched from right to wrong (RW), wrong to right (WR), and wrong to a different wrong (WW-D), all normalized with respect to the first response. Model 1 controls for precourse student self-efficacy only, whereas Model 2 controls for both precourse student self-efficacy and CSEM scores.

	RW (normalized)		WR (normalized)		WW-D (normalized)	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Self-efficacy	-0.32^{b}	-0.31^{b}	0.20	0.26^{a}	-0.32^{b}	-0.25^{a}
CSEM		-0.22^{a}		-0.16		0.005
R^2	0.10	0.23	0.04	0.07	0.11	0.07
RMSE	0.96	0.81	0.99	0.94	0.95	0.90

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

D. Switching and gender

Table II shows that a gender difference exists in the fraction of responses that are switched. Female students switch 15% more overall ($p < 0.05$) and 45% more from right to wrong ($p < 0.05$) than male students—but this gender difference disappears when controlling for self-efficacy. The reason this difference disappears is female students have lower precourse self-efficacy than male students. Female students score 10% lower than male students on the precourse Peer Instruction Self-Efficacy Instrument [17]. We also find that female students switch less from wrong to right and more from wrong to a different wrong than male students, although neither of these differences is statistically significant. Table II displays the standardized coefficients for three different linear regression models predicting the ratio of overall switching, right to wrong switching, wrong to right switching, and wrong to different wrong switching. Model 1 controls for gender only, model 2 controls for gender and precourse self-efficacy, and model 3 controls for gender, self-efficacy, and precourse CSEM scores. Gender alone is a significant predictor for overall switching and right to wrong switching ($p < 0.05$) but when self-efficacy is added to the model, it is no longer a significant predictor in either case.

E. Switching and CT difficulty

To investigate the relationship between ConcepTest difficulty and student switching, we shift our focus to item level switching by looking at what percentage of students switch for each individual ConcepTest. Figure 5 shows the percent of students who switch (in any direction) as a function of CT difficulty. Each point on Fig. 5 represents an individual CT.

The difficulty of each CT (b-parameter) is estimated using a 2PL item response theory model [15] and this is plotted against the fraction of students that switched their response from the first to the second round. As the difficulty of the item increases, so too does the percentage
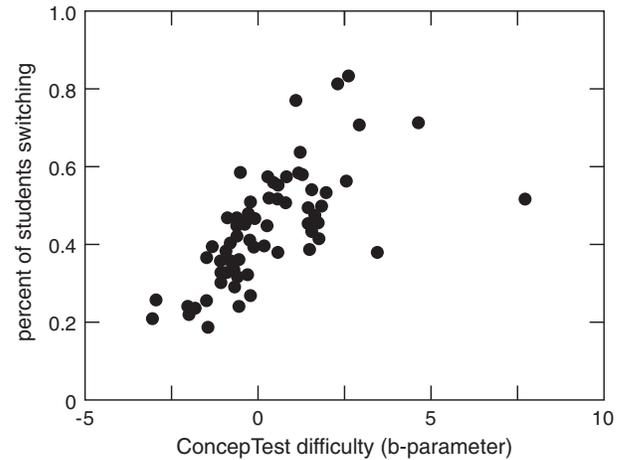
FIG. 5. Fraction of students who switch (in any of the three directions) plotted for each individual ConcepTest as a function of CT difficulty (obtained from a 2PL item response theory model).

of students who switch their response. With increasing item difficulty, students are more likely to switch from right to wrong ($p < 0.001$) or from wrong to a different wrong ($p < 0.05$) and less likely to switch from wrong to right ($p < 0.05$). The correlation between switching and CT difficulty is 0.54 ($p < 0.001$). Table III shows the correlations between CT difficulty (b parameter) and the fraction of students who, for each question (1) switch from right to wrong divided by the number of students who answered correctly in round 1, (2) switch from wrong to right divided by the number of students who answered incorrectly in round 1, (3) switch from wrong to a different wrong divided by the number of students who answered incorrectly in round 1, (4) have the same wrong answer in both rounds divided by the number of students who answered incorrectly in round 1, and (5) have the right answer in both rounds divided by the number of students who answered correctly in round 1. Table III shows that response switching is related to the difficulty of the item.

TABLE II. Standardized coefficients for linear regression models predicting proportion of responses: switched out of all responses (switched), switched from right to wrong out of all correct responses in the first round (RW normalized), switched from wrong to right out of all incorrect responses in the first round (WR normalized), and switched from wrong to a different wrong out of all incorrect responses in the first round (WW-D normalized). Model 1 controls for gender, model 2 controls for gender and precourse student self-efficacy, and model 3 controls for gender and both pre-course student self-efficacy and CSEM scores.

	Switched			RW (normalized)			WR (normalized)			WW-D (normalized)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Gender	0.47[a]	0.29	0.21	0.43[a]	0.27	0.21	-0.10	0.03	0.22	0.27	0.09	0.02
Self-efficacy		−0.29[b]	−0.18		−0.28	−0.28[a]		0.20	0.29[a]		−0.31[b]	−0.24[a]
CSEM			−0.46[c]			−0.19			−0.13			0.007
R^2	0.05	0.13	0.40	0.04	0.12	0.24	0.003	0.04	0.08	0.02	0.11	0.07
RMSE	0.98	0.95	0.76	0.08	0.96	0.81	1.01	1.0	0.94	0.99	0.94	0.90

[a]$p < 0.05$.
[b]$p < 0.01$.
[c]$p < 0.001$.

TABLE III. Correlation between the difficulty of each item (as estimated using a 2PL IRT model) and the fraction of students who, for that item, have a different wrong answer in each round (WW-D), switch from the right answer to a wrong answer (RW), switch from a wrong answer to the right answer (WR), have the same wrong answer in both rounds, and have the right answer in both rounds.

	RW	WR	WW-D	WW-S	RR
Item difficulty	0.38^b	-0.24^a	0.25^a	0.10	-0.42^b

[a]$p < 0.05$.
[b]$p < 0.001$.

IV. DISCUSSION

A. Class-wide response switching

The class-wide percentage of correct answers in each round of questioning provides real-time feedback of student understanding and is used, by the instructor, to guide the class during Peer Instruction. Ideally, students who initially respond incorrectly switch to the right answer after the discussion with their peers. Similarly, in the ideal situation students who initially respond correctly do not switch to a wrong answer after the discussion. We find that, on average, students switch on 44% of their responses and that the vast majority of these switches (73%) is from wrong to right. Although this percentage is high, there is still a significant proportion of CT switching (27%) in directions that are negatively associated with student learning (right to wrong and wrong to different wrong).

B. Switching and self-efficacy

Students with low self-efficacy have been shown to be less persistent and experience more negative emotions in the face of difficulty than students with high self-efficacy [11]. Bouchard [12] found that students with low self-efficacy are more likely to reject correct hypotheses prematurely and struggle with conceptual problem solving than students of the same ability with high self-efficacy. Our results show that students who have low confidence in their ability to solve problems and communicate science score are also significantly more likely to switch the right answer to a wrong one. This behavior is, at best, frustrating for students, and, at worst, contributing to an even further decrease in academic self-efficacy. Recent work has shown interactions in small group settings can impact self-efficacy in physics learning [14]. This work has involved the development of a framework to identify specific events (referred to as self-efficacy opportunities) that take place during small group problem solving sessions. These self-efficacy opportunities have been shown to directly influence students confidence in their ability to solve physics problems [14]. This framework suggests a methodology for studying how self-efficacy develops during peer interactions. Interventions to increase student self-efficacy at the beginning of the semester could improve student learning in interactive environments such as Peer Instruction classrooms.

C. Switching and CT difficulty

It is important for instructors to understand that they have some measure of control over the switching that occurs in their classrooms via the difficulty of the ConcepTests. Instructors should consider scaffolding more difficult CTs by building up to them with a series of less difficult questions. Cognitive science researchers recommend scaffolding as an instructional strategy to support students with the difficult task of transferring learning [21]. Research has shown that prefacing more difficult, synthesis problems with a sequence of related, but more basic conceptual questions, helps students answer the more difficult problems [22]. Presenting easier, warm-up questions before a difficult question could help students break difficult concepts up into smaller, more cognitively manageable chunks. Because ConcepTests often require students to apply conceptual understanding in new contexts, or transfer their learning, it is possible that scaffolding difficult ConcepTests may assist with positive switching transitions. A future study of CT response patterns to a series of scaffolded questions would prove interesting in providing further insight into the relationship between switching and CT difficulty.

V. CONCLUSION

Two results from our analysis of student ConcepTest switching behavior in a Peer Instruction environment have direct implications for classroom practice. The first is that CT switching behavior is a function of students precourse self-efficacy. Students with low precourse self-efficacy are more likely to switch to a wrong response and less likely to switch to the right response. We observe a similar switching behavior as the difficulty of the item increases. Understanding that students switch to the wrong response more often with difficult questions is informative because it indicates that instructors may need to provide better scaffolding for those questions. The strong connection between CT switching and self-efficacy suggests that interventions to increase student self-efficacy at the beginning of the semester might improve students experiences during Peer Instruction and help students take better advantage of this teaching strategy. Such interventions could include helping students build a sense of mastery, providing modeling experiences, social persuasion about students capabilities to succeed, and reducing stressful or anxiety provoking situations for students, such as high-stakes testing early in the semester.

out the study, and K. M. analyzed the results. J. S. and B. L. designed and implemented the self-efficacy survey used in this study (Peer Instruction Self-Efficacy Instrument). E. M. and A. H. supervised the research and the development of the manuscript. K. M. wrote the first draft of the manuscript; all authors subsequently took part in the revision process and approved the final copy of the manuscript.

APPENDIX A

$$\text{Right to Right (norm)} = \frac{\text{Right-Right}}{\text{Right-Round1}}$$

$$\text{Right to Wrong (norm)} = \frac{\text{Right-Wrong}}{\text{Right-Round1}}$$

$$\text{Wrong to Right (norm)} = \frac{\text{Wrong-Right}}{\text{Wrong-Round1}}$$

$$\text{Wrong to Same Wrong (norm)} = \frac{\text{Wrong-Wrong(same)}}{\text{Wrong-Round1}}$$

$$\text{Wrong to Diff Wrong (norm)} = \frac{\text{Wrong-Wrong(diff)}}{\text{Wrong-Round1}}$$

APPENDIX B

1. Peer Instruction self-efficacy instrument

Rank your level of agreement with each of the following statements using the following scale:

(1) strongly disagree
(2) disagree
(3) neutral
(4) agree
(5) strongly agree

1) I enjoy learning about science.
2) I enjoy learning about physics.
3) I often do well in science courses.
4) I often do well in non-science courses.
5) I identify with students who do well on exams and quizzes in science courses.
6) I expect to receive an A- or higher in this course.
7) I am confident I can do the work required for this course.
8) Doing laboratory experiments and write-ups comes easy to me.
9) I am often able to help my classmates with physics in the laboratory or in section.
10) I usually don't worry about my ability to solve physics problems.
11) When I come across a tough physics problem, I work at it until I solve it.
12) I get a sinking feeling when I think of trying to tackle difficult physics problems.
13) I like hearing about questions that other students have about the reading.
14) I am usually confident of my answers to the Learning Catalytics questions before I talk to a neighbor.
15) I am usually confident that I can convince my neighbor of my answer to Learning Catalytics questions.
16) I know how to explain my answers to Learning Catalytics questions in a way that helps others understand my answer.
17) My peers know how to explain their answers to Learning Catalytics questions in a way that helps me understand their answer.
18) Listening to my neighbors talk about their answers increases my confidence when responding to the same Learning Catalytics question a second time.
19) Practicing answering Learning Catalytics questions in class makes it easier for me to do physics problems at home.
20) I can communicate science effectively.
21) I can communicate physics effectively.

[1] E. Mazur, *Peer Instruction A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).

[2] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. **84**, 191 (1977).

[3] S. Andrew, Self-efficacy as a predictor of academic performance in science, J. Adv. Nurs. **27**, 596 (1998).

[4] R. W. Lent, S. D. Brown, and K. C. Larkin, Self-efficacy in the prediction of academic performance and perceived career options, J. Counsel. Psychol. **33**, 265 (1986).

[5] R. W. Lent, S. D. Brown, and K. C. Larkin, Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking, J. Counsel. Psychol. **34**, 293 (1987).

[6] K. D. Multon, S. D. Brown, and R. W. Lent, Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation, J. Counsel. Psychol. **38**, 30 (1991).

[7] J. Pietsch, R. Walker, and E. Chapman, The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school, J. Educ. Psychol. **95**, 589 (2003).

[8] J. Dalgety and R. K. Coll, Exploring first-year science students' chemistry self-efficacy, Int. J. Sci. Math. Educ. **4**, 97 (2006).

[9] P. Luzzo, Darrell Anthony Hasper, K. A. Albert, M. A. Bibby, J. Martinelli, and A. Edward, Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students, J. Counsel. Psychol. **46**, 233 (1999).

[10] G. Trujillo and K. D. Tanner, Considering the role of affect in learning: monitoring students' self-efficacy, sense of belonging, and science identity, CBE Life Sci. Educ. **13**, 6 (2014).

[11] A. Bandura, *Self-Efficacy: The Exercise of Control* (Freeman, New York, 1997).

[12] T. Bouffard-Bouchard, S. Parent, and S. Larivee, Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students, Int. J. Behav. Dev. **14**, 153 (1991).

[13] B. Zimmerman, Self-efficacy: An essential motive to learn, Contemp. Educ. Psychol. **25**, 82 (2000).

[14] V. Sawtelle, E. Brewe, R. M. Goertzen, and L. H. Kramer, Identifying events that impact self-efficacy in physics learning, Phys. Rev. ST Phys. Educ. Res. **8**, 020111 (2012).

[15] D. Thissen and S. Lynne, A taxonomy of item response models, Psychometrika **51**, 567 (1986).

[16] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).

[17] J. Schell and B. Lukoff, Peer instruction self-efficacy instrument [Developed at Harvard University]. Unpublished Instrument (2010).

[18] H. Fencl and K. Scheel, Pedagogical approaches, contextual variables, and the development of student self-efficacy in undergraduate physics courses, AIP Conf. Proc. **720**, 173 (2004).

[19] P. Kline, *A Handbook of Test Construction: Introduction to Psychometric Design* (Methuen, London, New York, 1986).

[20] L. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika **16**, 297 (1951).

[21] S. A. Ambrose, M. W. Bridges, M. DiPietro, M. C. Lovett, and M. K. Norman, *How Learning Works: Seven Research-Based Principles for Smart Teaching* (John Wiley & Sons, New York, 2010).

[22] L. Ding, N. Reay, A. Lee, and L. Bao, Exploring the role of conceptual scaffolding in solving synthesis problems, Phys. Rev. ST Phys. Educ. Res. **7**, 020109 (2011).