

## Essay

# Research-Based Implementation of Peer Instruction: A Literature Review

Trisha Vickrey, Kaitlyn Rosploch, Reihaneh Rahmanian, Matthew Pilarz,  
and Marilyne Stains

Department of Chemistry, University of Nebraska–Lincoln, Lincoln, NE 68588

Submitted November 6, 2014; Revised December 7, 2014; Accepted December 10, 2014

Monitoring Editor: Daron Barnard

Current instructional reforms in undergraduate science, technology, engineering, and mathematics (STEM) courses have focused on enhancing adoption of evidence-based instructional practices among STEM faculty members. These practices have been empirically demonstrated to enhance student learning and attitudes. However, research indicates that instructors often adapt rather than adopt practices, unknowingly compromising their effectiveness. Thus, there is a need to raise awareness of the research-based implementation of these practices, develop fidelity of implementation protocols to understand adaptations being made, and ultimately characterize the true impact of reform efforts based on these practices. Peer instruction (PI) is an example of an evidence-based instructional practice that consists of asking students conceptual questions during class time and collecting their answers via clickers or response cards. Extensive research has been conducted by physics and biology education researchers to evaluate the effectiveness of this practice and to better understand the intricacies of its implementation. PI has also been investigated in other disciplines, such as chemistry and computer science. This article reviews and summarizes these various bodies of research and provides instructors and researchers with a research-based model for the effective implementation of PI. Limitations of current studies and recommendations for future empirical inquiries are also provided.

## INTRODUCTION AND BACKGROUND

Discipline-based education researchers have responded to calls (President's Council of Advisors on Science and Technology [PCAST], 2010, 2012) for instructional reforms at the postsecondary level by developing and testing new instructional pedagogies grounded in research on the science of learning (Handelsman *et al.*, 2004; National Research Council [NRC], 2011, 2012). These research-based pedagogies significantly increase both student learning and attitudes

toward science (NRC, 2011, 2012). Peer instruction (PI), which was first introduced by Eric Mazur in 1991 (Mazur, 1997), is an example of a research-based pedagogy. In PI, traditional lecture is intermixed with conceptual questions targeting student misconceptions. Following a mini-lecture, students are asked to answer a conceptual question individually and vote using either a flash card or a personal response system commonly called a "clicker." If a majority of students respond incorrectly, the instructor then asks students to convince their neighbors that they have the right answer. Following peer discussion, students are asked to vote again. Finally, the instructor explains the correct and incorrect answers (Mazur, 1997; Crouch and Mazur, 2001). It is important to note that, although PI is commonly associated with clickers and there have been helpful reviews on best practices for clicker use (Caldwell, 2007; MacArthur *et al.*, 2011), this article is focused on PI, a specific, evidence-based pedagogy that can be effectively implemented with or without clickers.

PI has been primarily disseminated and adopted by physics instructors (Henderson *et al.*, 2012) but has also been

CBE Life Sci Educ March 2, 2015 14:es3

DOI:10.1187/cbe.14-11-0198

Address correspondence to: Marilyne Stains (mstains2@unl.edu).

© 2015 T. Vickrey *et al.* CBE—Life Sciences Education © 2015 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.

widely adopted by faculty members in the biological sciences and other science, technology, engineering, and math (STEM) fields (Borrego *et al.*, 2011). However, a recent study indicates that instructors adapt the PI model when implementing it in their classrooms, often eliminating either the individual voting or peer discussion steps, which are critical to the effectiveness of the pedagogy (Turpen and Finkelstein, 2009). Modification of evidence-based instructional practices has been associated with reduced learning gains in other studies (Andrews *et al.*, 2011; Henderson *et al.*, 2012; Chase *et al.*, 2013). While sometimes necessary, modifications are often made without fully realizing how they will impact effectiveness. The lack of knowledge of adequate ways to adapt a practice is due to two major reasons: 1) for most evidence-based instructional practices, few empirical studies have been conducted to identify the critical elements of the practice that make them effective; 2) for the few practices for which this type of research exists, the studies have been reported in various fields and journals, making it difficult for instructors and researchers alike to have a comprehensive, research-based description of the most effective implementation. PI falls into the latter group. There have been numerous studies exploring various aspects of the PI model, but these studies have been disseminated in a variety of ways. This article is intended to provide a comprehensive review of these studies. In particular, we summarize studies demonstrating the effectiveness of PI, describing the stakeholders' views on PI, and identifying the critical aspects of PI implementation. We foresee that this article will be used by researchers to design instruments that measure fidelity of implementation of PI, professionals involved in professional development to provide them with resources for their sessions, and instructors from multiple STEM disciplines interested in implementing this practice.

## METHOD

The search of articles for this literature review was constrained by the following parameters: 1) studies had to be conducted at the college level and in STEM courses; 2) studies had to report results that could only be attributed to the implementation of PI; 3) studies in which PI was implemented as part of a set of several evidence-based instructional practices and that only provided results for the set of practices were not included; and 4) implementation of PI had to follow the steps described by Mazur (1997), which have been associated with measurable learning gains (see *Impact of PI on Students*). Each step is discussed in more detail in the *Evidence-Based Implementation of PI* section. PI was used in combination with the following keywords in the ERIC, Web of Science, and Google Scholar databases: learning gains, retention, flash cards, clickers, personal response system, problem solving, concept inventory, concept test, voting, histogram, and peer discussion. The studies that met the criteria for this literature review are included in Supplemental Table 1.

## IMPACT OF PI ON STUDENTS

Studies have measured the impact of PI on learning gains, problem-solving skills, and student retention.

## Are There Measurable Learning Gains with the Use of PI?

The impact of PI on student learning has been most commonly measured in physics through the calculation of normalized learning gains. Normalized learning gains were first introduced by Hake (1998) in a widely cited study demonstrating the positive impact of active-learning instruction in comparison with traditional lecture. Normalized learning gains are calculated when a conceptual test, typically a concept inventory (Richardson, 2005), is implemented both at the beginning and end of a semester/unit/chapter. The actual gain in a student's score is divided by the maximal possible gain,  $((\text{posttest} - \text{pretest}) / (100 - \text{pretest}) \times 100)$ , which allows a valid comparison of gains between students with different pretest scores. In a longitudinal study, Crouch and Mazur (2001) explored the impact of PI compared with traditional lecture on student learning in algebra- and calculus-based introductory physics courses at Harvard University. At the beginning and end of a semester, they administered a conceptual test, the Force Concept Inventory (FCI; Hestenes *et al.*, 1992), to measure changes in normalized learning gains as they implemented either the PI pedagogy alone or a combination of PI and just-in-time teaching (Novak, 1999; Simkins and Maler, 2009) pedagogies. During the 10 yr of data collection, Crouch and Mazur (2001) observed normalized learning gains that were regularly twice as large as those observed with traditional lecture, even when implementing PI alone.

To further validate the positive impact of PI on student learning, these authors collected survey data from other current and past implementers of PI who had administered the FCI (Fagen *et al.*, 2002). The survey was posted on the Project Galileo website and directly emailed to more than 2700 instructors. The authors identified 384 instructors who were current or former PI users. They were able to obtain matched pre-post FCI data from 108 of these instructors representing 11 different institutions, including 2-yr, 4-yr, and research-intensive institutions, and 30 different courses. In 90% of these courses, they found medium normalized learning gains (medium  $g$  ranges from 0.30 to 0.70) with only three courses falling below that range. According to Hake's (1998) study, medium normalized learning gains are typically not achieved in traditionally taught courses. Another study by Lasry *et al.* (2008) compared the impact of the first implementation of PI in physics courses at Harvard University with that of implementation at a 2-yr college at which student's preinstructional background in physics is lower. Their quasi-experimental study demonstrated that students in these two settings achieved similar normalized learning gains ( $g = 0.50$  at Harvard University and  $g = 0.49$  at the 2-yr college).

The impact of PI on learning has been studied in disciplines other than physics as it has gained popularity. In the geosciences, McConnell *et al.* (2006) determined that the average difference between post- and pretest scores on the Geosciences Concepts Inventory (GCI; Libarkin and Anderson, 2005) was greater with PI pedagogy, and Mora (2010) reported greater normalized learning gains on the GCI compared with traditional lecture. Moreover, students in an introductory computer science course implementing PI scored half a letter grade higher on a final examination compared with peers in a lecture-based course covering the same topics

(Simon *et al.*, 2013b). Similar improvements on final examinations have been observed for students in two separate studies of calculus courses implementing PI (Miller *et al.*, 2006; Pilzer, 2001). However, a study in computer science reported less remarkable improvements on final examination scores in a course comparing PI and pure lecture (Zingaro, 2014). In this study, PI was associated with a 4.4% average increase on the final examination, but this finding was not significant ( $p = 0.10$ ). While there have been additional PI studies in other disciplines, they have focused on the benefits of specific steps of the PI sequence on student learning. Those will be presented later in the paper.

Extensive research thus indicates that PI is effective in promoting students' conceptual understanding in a variety of STEM disciplines and courses across various institutions.

### ***Does PI Improve Problem-Solving Skills?***

Studies have also focused on characterizing the impact of PI on students' problem-solving skills. In a study by Cortright *et al.* (2005), PI was introduced in an exercise physiology course. Students were randomly assigned into a PI group or a non-PI group in which students were presented with the in-class concept test but were instructed to answer the questions individually rather than discussing them with their peers. Students in the PI group improved significantly ( $p = 0.02$ ) in their ability to answer questions designed to measure mastery of the material. Importantly, the PI group's ability to solve novel problems (i.e., transfer knowledge) was significantly greater compared with that of the non-PI group (Cortright *et al.*, 2005).

In another study, PI was introduced in a veterinary physiology course (Giuliodori *et al.*, 2006). Giuliodori and colleagues compared student responses before and after peer discussion to determine whether or not PI improved students' ability to solve problems requiring qualitative predictions (increase/decrease/no change) about perturbations to physiological response systems (i.e., integration of multiple concepts and transfer ability). The number of students correctly answering questions improved significantly after peer discussion (Giuliodori *et al.*, 2006). Moreover, in a comparison of ability to transfer knowledge, students in an entomology course for nonmajors using PI scored significantly higher ( $p < 0.05$ ) on a near-transfer task (e.g., application of prior learning to a slightly different situation) compared with students in a non-PI group (Jones *et al.*, 2012). These studies suggest that PI improves students' ability to apply material to novel problems.

In addition to improvements on multiple-choice and qualitative questions, PI has been associated with learning gains on quantitative questions (Crouch and Mazur, 2001). The previously mentioned longitudinal study conducted at Harvard University in physics courses compared quantitative problem solving in courses with and without PI. A final examination consisting entirely of quantitative problems was administered after the first year of instruction with PI. The mean score on the exam was statistically significantly higher in the course with PI compared with traditional lecture. Thus, PI pedagogy can enhance both qualitative and quantitative problem-solving skills.

### ***How Does PI Affect Attrition Rates?***

Students' persistence in STEM fields is a critical concern at the forefront of federal and national initiatives (National

Science Foundation, 2010; PCAST, 2012). Several studies have examined student retention rate in courses using PI. In the instructor survey study conducted by Mazur and colleagues, instructors implementing PI reported lower student attrition rates compared with those using traditional lecture (Crouch and Watkins, 2007). In the study comparing the implementation of PI in physics courses at Harvard University and at a 2-yr college (Lasry *et al.*, 2008), the dropout rate (difference between the number of students enrolled and the number of students taking the final exam) decreased by 15.5% between the traditional lecture (20.5%) and PI (5%) sections at the 2-yr college. Similarly, the implementation of PI at Harvard University reduced the dropout rate by more than half to a rate consistently  $< 5\%$  (Lasry *et al.*, 2008). Increased retention and lower failure rates in courses with PI have also been reported from a retrospective study of more than 10,000 students in lower- and upper-division computer science courses (Porter *et al.*, 2013).

## **STAKEHOLDERS' VIEWS ON PI**

### ***What Do Students Think about PI?***

Students' resistance to instructional practices that differ from their expectation (i.e., traditional lecture) has been reported as an important barrier to instructors' continued implementation of evidence-based instructional practices (Felder and Brent, 1996). Thus, positive student reception of new instructional practices is important. Several studies have investigated this particular aspect of PI. For example, the longitudinal study conducted by Mazur and colleagues found no difference in students' course evaluations before and after implementation of PI (Crouch and Mazur, 2001). On the other hand, the instructor survey study conducted by Mazur and colleagues found that out of the 384 instructors, 70% reported obtaining higher course evaluations from students in PI classes compared with course evaluations for traditional courses. Despite these overall positive results, 17% of instructors reported a mixed response from students, while 5% reported a negative response. Additionally, a small percent (4%) of instructors who reported that their students had a positive response to PI indicated that the response was initially negative (Fagen, 2003).

In another study, student opinion of PI was compared between majors in a genetics course and nonmajors in an introductory biology course (Crossgrove and Curran, 2008). Each group answered a student opinion survey containing 11 questions. The average Likert scores for all but two questions were not significantly different between the groups. In particular, nonmajors thought that PI improved their exam performance, whereas majors thought this to a lesser extent ( $p < 0.001$ ). The authors also found that the nonmajors were more inclined to encourage the instructors to continue using PI, whereas majors were more ambivalent ( $p < 0.05$ ). Student feedback regarding the continued use of PI in their own and other's courses has also been explored. For example, in introductory computer science (Simon *et al.*, 2010), exercise science (Cortright *et al.*, 2005), preparatory engineering (Nielsen *et al.*, 2013), engineering mechanics (Boyle and Nicol, 2003), and veterinary physiology courses (Giuliodori *et al.*, 2006), students generally recommend that PI be used in other and/or future courses.

Interestingly, researchers have identified specific aspects of PI that students appreciate. For example, students report that they value the immediate feedback PI provides (Cortright *et al.*, 2005; Giuliadori *et al.*, 2006; Crossgrove and Curran, 2008; Simon *et al.*, 2013a). Moreover, in an analysis of 84 open-ended surveys from students enrolled in an introductory computer science course implementing PI pedagogy, Simon *et al.* (2013a) found that students felt that PI improved their relationship with their instructors, a finding also observed among students in an exercise science course (Cortright *et al.*, 2005).

Most importantly, students overwhelmingly report that PI helps them learn course material (Cortright *et al.*, 2005; Ghosh and Renna, 2006; Giuliadori *et al.*, 2006; Porter *et al.*, 2011b; Nielsen *et al.*, 2013; Simon *et al.*, 2010, 2013a). Indeed, PI has been shown to significantly impact student self-confidence (Gok, 2012; Zingaro, 2014). For example, students in two sections of an introductory computer science course, one with PI and one with traditional lecture, were asked to rate their self-confidence on a variety of programming tasks (Ramalingam and Wiedenbeck, 1999) at the beginning and end of the semester (Zingaro, 2014). Students enrolled in the PI course had statistically significant gains ( $p = 0.015$ ) in self-efficacy compared with those enrolled in traditional lecture courses (Zingaro, 2014). In another study, the self-efficacy of students enrolled in algebra-based physics courses with or without PI was compared (Gok, 2012). Similarly, the self-efficacy of the students increased significantly with PI ( $p = 0.041$ ) compared with traditional lecture.

Overall, students have neutral to positive views on PI and seem to recognize its value over traditional teaching.

### **What Do Instructors Think about PI?**

In addition to getting students' opinions, it is also important to gain insight into instructors' experiences with PI implementation. The instructor survey study conducted by Mazur and colleagues found that 90% of these instructors reported having a positive experience, 79% indicated that they would continue implementing PI, and another 8% reported they would probably use PI again (Fagen *et al.*, 2002). This positive response was echoed in a study conducted by Porter *et al.* (2011a), who observed that PI was beneficial because it "enables instructors to dynamically adapt class to address student misunderstandings, engages students in exploration and analysis of deep course concepts, and explores arguments through team discussions to build effective, appropriate communication skills" (p. 142). Although a large number of faculty members have reported using PI (Fagen *et al.*, 2002), most research has focused on student perception and learning rather than faculty experience. More information about instructors' perception of this pedagogy is needed to help inform the successful implementation of PI.

### **ACADEMIC SETTINGS IN WHICH PI HAS BEEN IMPLEMENTED**

PI has been implemented in a variety of academic settings (see Supplemental Table 1 for a complete list of studies). For example, out of the 384 PI users surveyed by Mazur and colleagues, 67% taught at universities, 19% taught at 4-yr

colleges, 5% taught high school, and smaller percentages taught at other institutions, such as 2-yr and community colleges (Fagen *et al.*, 2002). PI has also been implemented in different subject areas, course levels, and class sizes. In particular, research has been conducted in astronomy, biological sciences, calculus, chemistry, computer science, geosciences, economics, educational psychology, engineering, entomology, medical/veterinary courses, philosophy, and physics courses (see Supplemental Table 1). Of the research articles cited in this paper, 84% used PI in lower-level courses, 12% used it in upper-level undergraduate courses, and 4% used it in medical/veterinary school courses. Furthermore, 25% used PI in small classes (< 50 students), while 75% used it in large classes.

### **EVIDENCE-BASED IMPLEMENTATION OF PI**

There are several great resources describing best practices for implementing PI (Mazur, 1997; Crouch and Watkins, 2007). However, these guidelines do not necessarily provide empirical data to support best practices. In the next sections, we describe the results of empirical studies that have tested many of these guidelines. The discussion of guidelines will follow the order of the model for PI presented and researched by Mazur (1997):

1. Question posed
2. Students given time to think
3. Students record individual answers
4. Students convince their neighbors (peer discussion)
5. Students record revised answers
6. Feedback to teacher: tally of answers
7. Instructor's explanation of correct answer

### **Why Does the Type of Question Posed Matter? (Step 1)**

PI is intended to address misconceptions in a specific content area and foster conceptual understanding. To achieve this intended outcome, the type of question asked during each PI event should have an explicit pedagogic purpose (Beatty *et al.*, 2006); however, both the difficulty level of the question and the extent to which instructors ask conceptual questions varies. In an ethnographic study exploring the ways six physics instructors implemented PI, Turpen and Finkelstein (2009) found that the content-related questions instructors asked could be classified as conceptual, recall, or algorithmic. The percent of conceptual questions ranged from 64 to 85%, recall ranged from 4 to 24%, and algorithmic from 0 to 11%. Although instructors primarily asked conceptual questions, recall questions appeared common. Asking recall questions may be appealing to instructors; however, research suggests that asking higher-order questions yields better student results.

In a study in a large medical physiology course for first-year students, Rao and DiCarlo (2000) compared question type with the percentage of correct answers on the individual vote versus after peer discussion. Questions were classified as either testing recall, intellectual (i.e., comprehension, application, and analytical ability), or synthesis and evaluation

skills (Rao and DiCarlo, 2000). The researchers found a statistically significant difference in the proportion of students answering correctly after peer discussion for all three types of questions. However, on questions testing synthesis and evaluation skills, students exhibited the greatest improvement: from  $73.1 \pm 11.6\%$  to  $99.8 \pm 0.24\%$  (this is compared with  $94.3 \pm 1.8\%$  to  $99.4 \pm 0.4\%$  for recall questions and  $82.5 \pm 6.0\%$  to  $99.1 \pm 0.9\%$  for questions testing intellectual abilities). These results imply that students benefited most from peer discussion on difficult questions. Albeit, the percent of students arriving at the correct answer for each type of question before peer discussion was already large, calling into question the actual difficulty level of the questions asked.

A study by Smith *et al.* (2009) also suggests that students improve the most when instructors ask difficult questions. In a genetics course for biology majors, the percent of correct student responses before and after peer discussion was compared for questions classified as having a low, medium, or high level of difficulty. Question difficulty was based on the number of students who obtained a correct answer on the individual vote. Student learning gains were the greatest when students were asked difficult rather than easy questions (Smith *et al.*, 2009). Smith's study was replicated by Porter *et al.* (2011b) in two computer science courses, and the same general trend was observed. In another study by Knight *et al.* (2013), students' discussion in an upper-division developmental biology course was compared with question difficulty, which was determined by analyzing each question using Bloom's taxonomy. When instructors asked higher-order questions, student discussions became more sophisticated, which was associated with larger learning gains.

The benefits of PI are accentuated when higher-level intellectual questions are used in favor of basic recall questions. These results support using PI for more challenging questions. Several concept inventories have been developed in biology that could be used as resources for PI questions (Williams, 2014).

### ***Does Individual Thinking Matter? (Steps 2 and 3)***

The second and third steps in Mazur's PI model consist of having students think for themselves and commit to their answers by voting. Uncovering the relationship between individual student voting and student perception of or learning during PI is important, because professors report skipping these steps and going directly into peer discussion after posing a question (Turpen and Finkelstein, 2009). Several studies have explored whether or not the time for students to think and answer questions individually is necessary. For example, Nicol and Boyle (2003) utilized a mixed-method design to compare engineering students' perceptions of PI when the full model was used (steps 1–6) versus a modified method consisting of peer discussion followed by class-wide discussion (with steps 2 and 3 skipped). They found that, while students thought that both methods enhanced their understanding of the concepts, 82% of the students indicated they prefer answering the question individually before discussing it with their neighbors. Of the students who explained their preference, 80% made comments suggesting that the individual response time forced them to think about and identify an answer to the question; they felt that this led them to be more active and engaged during the

peer discussion. Moreover, 90% of students agreed that "a group discussion after an individual response leads to deeper thinking about the topic." In contrast, students indicated that starting with the peer discussion often led them to be more passive, accepting answers from more confident students without thinking critically.

Nielsen *et al.* (2014) further explored this issue with a mixed-method project that used surveys, interviews, and classroom observations. They implemented PI (steps 1–6) and modified PI without steps 2 and 3 in an introductory physics course for engineers (four sections;  $n = 279$  total). PI and modified PI were implemented in all sections alternatively. The results of focus group interviews ( $n = 16$  students) and surveys ( $n = 109$ ) were similar to the study from Nicol and Boyle (2003): the majority of students felt that the individual time was necessary to help them form their opinion without being influenced by others (Nielsen *et al.*, 2014).

It is important not only to evaluate students' perceptions of individual voting, but also to quantitatively and qualitatively analyze students' discussions with PI compared with modified PI. Nielsen *et al.* (2014) found a statistically significant increase in the amount of time spent talking with the PI model compared with the modified PI model. They attributed this increase to time spent on argumentation (i.e., presentation of an idea or explanation; 90% increase in argumentation time medians), indicating an improvement in discussion quality.

This series of studies thus indicates that students' individual commitments to an answer before peer discussion improves students' learning experiences and that steps 2 and 3 should not be skipped during implementation. It is important to note, however, that the studies discussed in this section do not report learning gains for individual thinking and that, while students' perception and ability to construct arguments is important from a pedagogical standpoint, more research is needed to confirm the importance of individual voting.

### ***Are Clickers Necessary? (Step 3)***

PI is often associated with classroom response systems or clickers. However, not everyone using clickers is conducting PI. Likewise, PI can be conducted without the use of clickers. Several studies have investigated the effect of PI when other voting methods, such as flash cards or ABCD cards have been used. For example, Lasry (2008) compared PI using either flash cards or clickers in two different sections of an algebra-based mechanics course taught by the same instructor. Student learning was measured by comparing learning gains on the FCI and an examination between the two sections. While both the flash cards and clicker groups improved on the FCI, no statistical differences in learning gains were observed between the two groups on either the FCI or the examination. Flash cards have proven effective in other studies. For example, Cortright *et al.* (2005) studied a physiology course using flash cards and found that students' problem-solving skills improved during PI.

Another study by Brady *et al.* (2013) used a quasi-experimental design to investigate differences in students' metacognitive processes and performance outcomes in sections of a psychology course that implemented PI with clickers versus other sections of the same course that implemented

PI with paddles (i.e., a low-technology flash card system provided by the primary researcher). Results demonstrate that the use of paddles increased metacognition skills, while the use of clickers resulted in higher performance outcomes. This study indicates that there may be a metacognitive advantage to using non clicker response systems for PI, but additional studies are needed to confirm this finding.

Overall, these studies indicate that PI can be effectively implemented with clickers or with low-tech voting tools.

### ***Does Showing the Distribution of Answers after the First Vote Matter? (Step 3)***

When personal response devices such as clickers are used, the instructor has the option to share with her/his class the distribution of students' answers following the first vote. Several studies have examined the impact of this option on students' behavior during peer discussion.

Perez *et al.* (2010) investigated whether showing the distribution of answers (as a bar graph) following the first vote (step 2) biased students' second vote (step 4). They implemented a crossover research design in a freshman-level biology majors course (eight sections, 629 students participated in the study): in one treatment, students were shown the bar graph before peer discussion, and in the other treatment, they were not. Thus, each treatment group saw the bar graph after the first vote 50% of the time. Instructors in each section used the exact same set of questions. The responses of students seeing the bar graph before peer discussion were compared with those who did not. They found that when students saw the bar graph after the first vote, they were 30% more likely to change their answer to the most common one. This bias was more pronounced on difficult questions, and it appeared to account for 5% of the learning gains observed between the first and second vote. This study suggests that when instructors display the bar graph, students may think that the most common answer is correct rather than constructing a correct response through discussion with their peers. Indeed, data from a qualitative study on the impact of showing the bar graph after the first vote support these findings (Nielsen *et al.*, 2012). Group interviews revealed that students perceive the most common answer to be the most correct, and students are less willing to defend an answer if it is not the most common one.

Student bias toward the most common answer when the distribution of answers is shown was not observed, however, in a similar study in chemistry (Brooks and Koretsky, 2011). Two cohorts of students in a thermodynamics course ( $n = 128$  students) were compared: one saw the distribution after the first vote, while the other one did not. They found that both groups of students had a similar tendency to select the consensus answer regardless of seeing the distribution. Moreover, they found no difference in the quality of the explanations students wrote to justify their answers. They did, however, see a difference in students' confidence in their answers: students who saw the bar graph after the first vote were statistically more confident when their answer matched the consensus answer, even if the consensus answer was incorrect (Brooks and Koretsky, 2011).

More research is needed to fully understand the effect of displaying the bar graph after the first vote. Based on the results of the few studies investigating this issue, it seems

that it may be most effective to show the difference in the distribution of answers between the first and second vote after peer discussion. This approach would limit the bias toward the consensus answer observed in some studies (Perez *et al.*, 2010; Nielsen *et al.*, 2012), while not only enhancing the confidence of students who had the correct answer in the first vote but also maintaining the integrity of student discussion.

### ***When Is It Appropriate to Engage Students in Peer Discussion? (Step 4)***

The analysis from the *Why Does the Type of Question Posed Matter?* section suggested that the benefits of student discussion on learning vary based on the proportion of correct responses on the initial vote; indeed, limited learning gains between the individual vote and the revote were observed on easy questions (Rao and DiCarlo, 2000; Smith *et al.*, 2009; Porter *et al.*, 2011b; Knight *et al.*, 2013). In their longitudinal analysis, Crouch and Mazur (2001) found that the largest improvement in moving toward the correct answer on a revote occurred when the initial individual answer was correct for ~50% of the class. Nevertheless, there were still substantial learning gains when the initial percent of correct responses was between 35 and 70%. This empirical study indicates that students should be engaged in peer discussion when the percent of correct answers on the individual vote falls between 35 and 70%. Below 35%, the concept may need to be further described or a salient hint provided. Subsequent studies suggest that students may still benefit from talking to one another, even when only a small proportion (< 35%) of the class obtained the correct response (Simon *et al.*, 2010; Smith *et al.*, 2009). Above 70%, the instructor should skip to the explanation of the answer.

More research is needed to optimize guidelines for step 4. Regardless of the proportion of correct answers on the initial vote, students seem to benefit from peer discussion.

### ***Does Peer Discussion Matter? (Steps 4 and 5)***

Peer discussion is the most recognizable feature of the PI model, and much of this review has been devoted to reporting on the learning gains observed after students' discussions. As such, it is important to understand the role of small group discussion in PI as well as to determine whether or not the observed improvement in student response is more than students with incorrect answers simply copying those who are correct. Smith *et al.* (2009) investigated this issue in a one-semester undergraduate genetics class ( $n = 350$ ). During the semester, students were asked 16 sets of paired questions testing the same concepts but with different cover stories. The first question, Q1, was given, and students voted individually, discussed the question with their peers, and voted again (Q1<sub>ad</sub>). Then, students were given a second question, Q2, testing the same concept as the first. The proportion of correct answers for each question was compared. They found that the proportion of correct answers for Q2 was significantly greater than for Q1 and Q1<sub>ad</sub>, and out of the students who initially answered Q1 incorrectly but Q1<sub>ad</sub> correctly, 77% went on to answer Q2 correctly. Thus, when students do not initially understand a concept, they can discuss the material with a peer and then apply this information to answer a similar question correctly. Interestingly, a statistical analysis

of students who answered Q1 incorrectly but Q2 correctly suggests that some students did not belong to a discussion group with a student who knew the correct answer. These students were presumably able to arrive at the correct answer through peer discussion. Porter *et al.* (2011b) replicated the previous study in two different upper-division computer science classes ( $n = 96$  total). Similarly, they found that more students answered Q2 correctly compared with Q1 and that peer discussion improved learning gains for 13–20% of the students. Peer discussion has also been shown to improve the proportion of correct responses on the revote in general chemistry courses (Bruck and Towns, 2009).

Although the proportion of correct answers increases after peer discussion, alternative hypotheses, such as the extra time allowed for individual reflection or to process information, could also explain these learning gains. Lasry *et al.* (2009) designed a crossover study in three algebra-based introductory physics courses ( $n = 88$ ) to test whether peer discussion or other metacognitive processes, such as reflection or time on task, explained the learning gains associated with PI. Students voted on a question individually and then were assigned one of three tasks: peer discussion, silent reflection on answers, or distraction by cartoon. All groups were asked to vote again. The learning gains were highest when students engaged in peer discussion. These results suggest that the improvement observed after peer discussion is not due to another metacognitive process.

To assess the impact of student discussions, Brooks and Koretsky (2011) examined the relationship between student reasoning before and after peer discussion. Students recorded an explanation for their responses to concept tests before and after discussing them with their peers. Each explanation was then analyzed for both depth and accuracy. The quality of explanations from students who had responded correctly on both the initial vote and the revote improved following peer discussion. Although these students had the correct answer initially, they gained a more in-depth understanding of the concepts after peer discussion. Even though student explanations improve after peer discussion, the actual quality of the discussion appears variable. James and Willoughby (2011) recorded 361 peer discussions from four different sections of an introductory-level astronomy course. When they compared student responses with the recorded conversations, they found that 26% of the time student responses implied understanding, while the quality of conversations suggested otherwise. Furthermore, in 62% of the recorded conversations, student discussions included incorrect ideas or ideas that were unanticipated.

Taken together, these studies suggest that, although peer discussion positively impacts student learning, the improvements observed between the first and second vote may overestimate student understanding.

### **How Much Time Should Be Given to Students to Enter Their Votes? (Steps 5 and 6)**

Once a question is asked, instructors must allow students enough time to respond thoughtfully while still maximizing class time. Faculty members have reported variations in the voting time they allowed students during PI (Turpen and Finkelstein, 2009). For example, during a semester-long observation of six physics instructors, the average voting time

given to students varied from  $100 \pm 5$  s to  $153 \pm 10$  s. Out of these six instructors, two had large SDs of more than 100 s. Not only were there variations in response times given between classes, but also within classes (Turpen and Finkelstein, 2009). Thus, establishing guidelines for the optimal amount of time that instructors should allow is important. Unfortunately, we were only able to identify one study on this topic for this review. Miller *et al.* (2014) examined the difference in response times between correct and incorrect answers in two physics courses with PI: one at Harvard University and one at Queen's University (Kingston, Ontario, Canada). In both classes, the proportion of correct to incorrect answers decreased when ~80% of the students had responded, suggesting that incorrect answers take more time than correct ones. Additionally, these researchers examined student response times for questions classified as either easy or difficult. For easy questions, students who answered incorrectly took significantly ( $p < 0.001$ ) more time than those who answered correctly, while for difficult questions, students took more time to answer regardless of correct or incorrect responses. Although more research is needed to confirm Miller and co-workers' findings, these data suggest that instructors should alert students that they will terminate polls (i.e., issue a final countdown) after ~80% of students have voted, particularly when posing less difficult questions.

### **How Does the Role of the Instructor Affect PI? (Step 6)**

Although PI's critical feature is peer discussion, the instructor's explanation of concept tests also influences the effectiveness of PI. Several studies have investigated the impact of instructors' explanations at the end of the PI cycle on student learning (Smith *et al.*, 2011; Zingaro and Porter, 2014a,b). Smith *et al.* (2011) used pairs of isomorphic questions (i.e., two different questions assessing the same concept) to compare the impact of three different instructor interventions in two genetics courses: one for biology majors ( $n = 150$  students) and one for nonmajors ( $n = 62$  students). The three experimental conditions were as follows:

1. Peer discussion only: students answer the first question (Q1) according to the PI model. After the revote, the instructor provides the correct answer without explanation, and students answer the isomorphic question (Q2).
2. Instructor explanation only: students answer Q1 individually. The instructor explains the answer to the class, and then students answer Q2.
3. Peer discussion and instructor explanation: students answer Q1 according to the PI model. After the revote, the instructor explains the answer, and students answer Q2.

Significantly larger learning gains ( $p < 0.05$ ), as measured by the normalized change in scores between the two questions, were observed in the third intervention, which combined PI with instructor explanations (Smith *et al.*, 2011). Moreover, these learning gains were observed across both courses and for students at all levels of ability (low, medium, and high performing as determined by the mean scores on the first question).

An analogous experiment by Zingaro and Porter (2014a) in an introductory computer science class ( $n = 126$ ) yielded

similar results. Students experienced larger learning gains with PI through the combination of peer discussion and instructor explanation compared with student discussion alone (81 vs. 69% correct on Q2, respectively). In addition, instructor explanation resulted in the largest gains in learning when the question was more difficult. In a subsequent study, Zingaro *et al.* (2014b) also found that the combination of peer discussion and instructor explanation compared with peer discussion alone was positively correlated to performance on the final exam.

Instructor behaviors, such as cuing discussion, also appear to impact PI implementation. For instance, Knight *et al.* (2013) measured the impact of instructional cues on student discussion during PI in an upper-division developmental biology course. In this study, the instructor either framed peer discussion as “answer-centered” (i.e., asking students to discuss answers) or “reasoning-centered” (i.e., asking students to discuss the reasons behind their answers). The resulting student discussions were then analyzed according to a scoring system measuring the quality of student reasoning. The quality of reasoning was significantly ( $p < 0.01$ ) higher in the reasoning-cued condition compared with the answer-cued one (Knight *et al.*, 2013). Other studies have indicated that not only do students perform better on concept tests when given specific guidelines for peer discussion (Lucas, 2009) but also that students place more value on articulating their responses when instructors emphasize “sense making” over “answer making” (Turpen and Finkelstein, 2010).

Together, these data suggest that it is important for instructors to discuss answers to concept tests with students and to communicate expectations for peer discussion clearly with a focus on sense making.

### Does Grading Matter?

When using a personal response system during PI, instructors have the option of awarding points for student responses. Awarding points for correct answers (high-stakes grading) versus participation (low-stakes grading) has been shown to impact the dynamics of peer discussion (James, 2006; James *et al.*, 2008; Turpen and Finkelstein, 2010). James (2006) compared student discussion practices in two introductory astronomy classes taught by two different instructors: one a standard class for nonmajors, the other multidisciplinary and focused on space travel. One instructor adopted high-stakes grading, wherein student responses accounted for 12.5% of their total grade. In this class, students were awarded full credit for a correct response but one-third credit for an incorrect one. The second instructor adopted low-stakes grading, wherein student responses accounted for 20% of their total grade, but students were awarded full credit for both correct and incorrect answers. During the semester, 12–14 pairs of student discussions in each class were recorded on three separate occasions. For each conversation pair, James analyzed the conversation bias (i.e., the extent to which one student compared with the other dominated a conversation). The conversation bias among partners was significantly higher ( $p = 0.008$ ) in the classroom with high-stakes grading incentives. Conversation bias was correlated to course grade in the high-stakes classroom but not in the low-stakes one.

In an extension of this research, James *et al.* (2008) conducted a study examining student discourse in three

introductory astronomy classes taught by two different instructors over two semesters. Instructor A taught two semesters, adopting high-stakes grading practices during the first semester and low-stakes grading practices the following semester. Instructor A implemented PI identically each semester and used the same concept test questions. Instructor B adopted a low-stakes grading approach during the first semester but was not observed the following semester. The conversation bias of student pairs was analyzed using two different techniques: one that categorized student ideas (idea count) and another that accounted for the amount of time a student spent talking (word count). When instructor A switched from a high- to low-stakes approach, conversation bias decreased significantly in both the idea count ( $p = 0.025$ ) and the word count ( $p = 0.044$ ) analyses. There was no significant difference in conversation bias between student pairs in either of the low-stakes classrooms taught by the two different instructors. James’ research suggests that when grading incentives favor correct answers, the student with the most knowledge dominates the discussion. Unfortunately, the experimental unit in James’ studies was the students rather than the classroom, which may underestimate the variability of the students, likely confounding these results. However, research by Turpen and Finkelstein (2010) also supports low-stakes grading approaches. For example, they found that high-stakes grading incentives were associated with reduced peer collaboration on questions.

Overall, peer discussion appears to benefit students the most when instructors award participation points for answering questions during peer discussion rather than awarding points for answering questions correctly. It is important to note, however, that although conversation bias improved, more research is needed to link this improvement directly to learning gains.

### LIMITATIONS OF CURRENT RESEARCH ON PI

Although PI is one of the most researched evidence-based instructional practices, this research has several shortcomings.

First, more research is needed to resolve some of the uncertainties in implementation addressed in this review. For example, the impact of high-stakes grading on peer discussion and the consequences of displaying the voting results before peer discussion is unclear. In addition, there is uncertainty regarding the optimal proportion of correct answers needed on the initial vote to engage students in peer discussion.

Second, little is understood about the relationship between PI and individual student characteristics or students’ prior knowledge. Although Mazur and colleagues (Lorenzo *et al.*, 2006) reported a reduced gender gap in classes with PI compared with those without, it is unclear whether similar gains occur in other disciplines and whether PI benefits some students more than others. Some research suggests that there is, indeed, a relationship between student characteristics and performance on PI concept tests. For example, Steer *et al.* (2009) evaluated responses on concept tests from 4700 students enrolled in five different earth science classes at a community college. Female students from underrepresented groups were significantly more likely to change from an incorrect response before peer discussion to a correct one afterward. These students were less likely to have a correct



answer during the individual voting period; therefore, they were able to make larger gains. Similar improvements by underrepresented students have been reported in calculus courses (Miller *et al.*, 2006). Although these results are encouraging, analyzing raw changes in scores does not account for the fact that students with low scores are able to gain more than students who score higher initially. Other studies have shown that gender differences become insignificant once prior knowledge is controlled for (Miller *et al.*, 2014). For example, males and females exhibited significantly different response times on concept tests during both the initial voting period and revote. However, once students' pre-course knowledge and self-efficacy were account for, these differences became insignificant.

Learning gains in PI courses have been analyzed alongside prior achievement in other studies. For example, in a study of 1236 earth science students, prior student achievement (measured by ACT score) predicted the number of correct responses to concept tests during PI (Gray *et al.*, 2011). Unfortunately, overall learning gains were not reported in this study. In another study in computer science courses with and without PI, students' high school background was compared with scores on a final examination (Simon *et al.*, 2013b). PI was most effective among students indicating that the majority of their high school attended college. For students who indicated that the majority of their high school did not attend college, there was a slight negative, but insignificant, effect of PI versus traditional lecture.

These results imply there is a relationship between the learning gains observed with PI and individual student characteristics such as race, gender, and prior achievement. Thus, the true impact of PI cannot be realized without controlling for student characteristics (Theobald and Freeman, 2014). Researchers must account for the variation between students in classes and between instructors. For example, individual students enrolled in a STEM course for nonmajors in which the instructor has implemented PI should not be directly

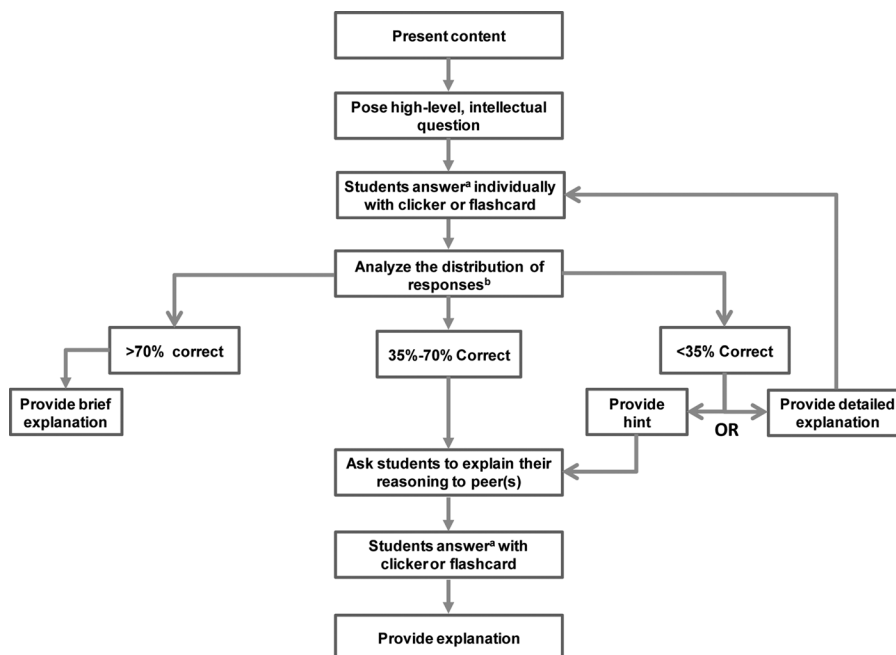
compared with individual students in another STEM course for majors in which the instructor has not implemented PI. Similarly, differences in instructors' characteristics, such as fidelity of implementation of PI, demographics, and teaching experience/training must be considered during analysis.

Methodological shortcomings also include inappropriate selection of unit of analysis, which increases the likelihood of type I errors. In particular, this literature commonly reports individuals rather than the classroom in which PI was implemented as the unit of analysis. Individual students within a class are more likely to have similar characteristics beyond those that can be accounted for, and observations from these students are therefore not independent. Choosing the classroom as the unit of analysis ensures independence. Unfortunately, if the unit of analysis is at the aggregate level, studies may be underpowered due to smaller sample size. Thus, it may be useful for researchers to increase the number of classrooms for comparison in their studies or replicate prior studies in order to facilitate meta-analyses.

Finally, the majority of the studies surveyed in this review has utilized classroom response systems, which have grown increasingly more sophisticated. For example, bring-your-own-device technologies are now available and allow instructors to ask open-ended questions. While these new technologies allow instructors to ask higher-level questions, they may unintentionally increase student distraction (Duncan *et al.*, 2012). As institutions and individuals using PI adopt these new technologies, the impact of bring-your-own-device tools on student learning will need to be determined.

## CONCLUSIONS

In this review, we provide an analysis of student outcomes associated with PI and features critical to successful implementation. In comparison with traditional lecture, this pedagogy overwhelmingly improves students' ability to



**Figure 1.** Research-based implementation of PI. <sup>a</sup>See section *How Much Time Should Be Given to Students to Enter Their Vote?* <sup>b</sup>See section *Does Showing the Distribution of Answers after the First Vote Matter?*

solve conceptual and quantitative problems and to apply knowledge to novel problems. Students value PI as a useful learning tool and are more likely to persist in courses utilizing it. Likewise, instructors value the improved student engagement and learning observed with PI.

From our analysis of the research, we propose a revised, evidence-based model of the steps of PI (Figure 1). This model could guide practitioners in an effective implementation of PI that would lead to the most positive student outcomes. It could also inform researchers in the design of protocols measuring instructors' fidelity in PI implementation.

## ACKNOWLEDGMENTS

We acknowledge the financial support from the Department of Chemistry and the University of Nebraska–Lincoln. T.V. was funded by the National Science Foundation through grants 1256003 and 1347814.

## REFERENCES

- Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE Life Sci Educ* 10, 394–405.
- Beatty ID, Gerace WJ, Leonard WJ, Dufresne RJ (2006). Designing effective questions for classroom response system teaching. *Am J Phys* 74, 31.
- Borrego M, Cutler S, Froyd J, Prince M, Henderson C (2011). Faculty use of research based instructional strategies. In: Proceedings of the Australasian Association for Engineering Education Conference 2011: Developing Engineers for Social Justice: Community Involvement, Ethics & Sustainability 5–7 December 2011, Fremantle, Western Australia. Barton, ACT: Engineers Australia, 448–453.
- Boyle JT, Nicol DJ (2003). Using classroom communication systems to support interaction and discussion in large class settings. *Assoc Learn Technol J* 11, 343–57.
- Brady M, Seli H, Rosenthal J (2013). “Clickers” and metacognition: a quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Comput Educ* 65, 56–63.
- Brooks BB, Koretsky MDM (2011). The influence of group discussion on students' responses and confidence during peer instruction. *J Chem Educ* 88, 1477–1484.
- Bruck AD, Towns MH (2009). Analysis of classroom response system questions via four lenses in a general chemistry course. *Chem Educ Res Pract* 10, 291–295.
- Caldwell JE (2007). Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci Educ* 6, 9–20.
- Chase A, Pakhira D, Stains M (2013). Implementing process-oriented, guided-inquiry learning for the first time: adaptations and short-term impacts on students' attitude and performance. *J Chem Educ* 90, 409–416.
- Cortright RN, Collins HL, DiCarlo SE (2005). Peer instruction enhanced meaningful learning: ability to solve novel problems. *Adv Physiol Educ* 29, 107–111.
- 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.
- Crouch CH, Mazur E (2001). Peer instruction: ten years of experience and results. *Am J Phys* 69, 970.
- Crouch C, Watkins J (2007). Peer instruction: engaging students one-on-one, all at once. In: *Reviews of Research-Based Reform Curricula in Introductory Physics*, ed. EF Redish and P Cooney, College Park, MD: American Association of Physics Teachers, 1–55.
- Duncan DK, Hoekstra AR, Wilcox BR (2012). Digital devices, distraction, and student performance: does in-class cell phone use reduce learning? *Astron Educ Rev* 11, 1.
- Fagen AP (2003). Assessing and Enhancing the Introductory Science Course in Physics and Biology: Peer Instruction, Classroom Demonstrations, and Genetics Vocabulary. [http://mazur.harvard.edu/sentFiles/Mazurpubs\\_537.pdf](http://mazur.harvard.edu/sentFiles/Mazurpubs_537.pdf) (accessed 5 October 2014).
- Fagen AP, Crouch C, Mazur E (2002). Peer instruction: results from a range of classrooms. *Phys Teach* 40, 206.
- Felder RM, Brent R (1996). Navigating the bumpy road to student-centered instruction. *Coll Teach* 44, 43–47.
- Ghosh S, Renna F (2006). Technology in Support of Good Pedagogy: Electronic Response Systems and Peer Instruction in an Economics Classroom, 22. <http://ssrn.com/abstract=888544> or <http://dx.doi.org/10.2139/ssrn.888544>.
- Giuliodori MJ, Lujan HL, DiCarlo SE (2006). Peer instruction enhanced student performance on qualitative problem-solving questions. *Adv Physiol Educ* 30, 168–173.
- Gok T (2012). The effects of peer instruction on student's conceptual learning and motivation. *Asia-Pacific Forum Sci Learn Teach* 13, article 10.
- Gray K, Owens K, Liang X, Steer D (2011). Assessing multimedia influences on student responses using a personal response system. *J Sci Educ Technol* 21, 392–402.
- Hake R (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66, 64–74.
- Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM, *et al.* (2004). Scientific teaching. *Science* 304, 521–522.
- Henderson C, Dancy M, Niewiadomska-Bugaj M (2012). Use of research-based instructional strategies in introductory physics: where do faculty leave the innovation-decision process? *Phys Rev Spec Top Phys Educ Res* 8, 020104.
- Hestenes D, Wells M, Swackhamer G (1992). Force Concept Inventory. *Phys Teach* 30, 141–158.
- James MC (2006). The effect of grading incentive on student discourse in peer instruction. *Am J Phys* 74, 689.
- James MC, Barbieri F, Garcia P (2008). What are they talking about? Lessons learned from a study of peer instruction. *Astron Educ Rev* 7, 37–43.
- 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.
- Jones ME, Antonenko PD, Greenwood CM (2012). The impact of collaborative and individualized student response system strategies on learner motivation, metacognition, and knowledge transfer. *J Comput Assist Learn* 28, 477–487.
- 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.
- Lasry N (2008). Clickers or flashcards: is there really a difference? *Phys Teach* 46, 242.
- Lasry N, Charles E, Whittaker C, Lautman M (2009). When talking is better than staying quiet. In: *Physics Education Research Conference*, ed. M Sabella, C Henderson, and C Singh, Melville, NY: American Institute of Physics, 181–184.
- Lasry N, Mazur E, Watkins J (2008). Peer instruction: from Harvard to the two-year college. *Am J Phys* 76, 1066.
- Libarkin JC, Anderson SW (2005). Assessment of learning in entry-level geoscience courses: results from the Geoscience Concept Inventory. *J Geosci Educ* 53, 394–401.
- Lorenzo M, Crouch CH, Mazur E (2006). Reducing the gender gap in the physics classroom. *Am J Phys* 74, 118.

- Lucas A (2009). Using peer instruction and i-clickers to enhance student participation in calculus. *Primus* 19, 219–231.
- MacArthur J, Jones L, Suits J (2011). Faculty viewpoints on teaching large-enrollment science courses with clickers. *J Comput Math Sci Teach* 30, 251–270.
- Mazur E (1997). *Peer Instruction: A User's Manual*, Upper Saddle River, NJ: Prentice Hall.
- McConnell DA, Steer DN, Owens KD, Knott JR, Dick J, Heaney PJ (2006). Using ConcepTests to assess and improve student conceptual understanding in introductory geoscience courses. *J Geosci Educ* 54, 61–68.
- Miller K, Lasry N, Lukoff B, Schell J, Mazur E (2014). Conceptual question response times in peer instruction classrooms. *Phys Rev Spec Top Phys Educ Res* 10, 020113.
- Miller RLR, Santana-Vega E, Terrell MS (2006). Can good questions and peer discussion improve calculus instruction? *Probl Resour Issues Math Undergrad Stud* 16, 193–203.
- Mora G (2010). Peer instruction and lecture tutorials equally improve student learning in introductory geology classes. *J Geosci Educ* 58, 286.
- National Research Council (NRC) (2011). *Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education: Summary of Two Workshops*, Washington, DC: National Academies Press.
- NRC (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, Washington, DC: National Academies Press.
- National Science Foundation (2010). *Preparing the Next Generation of STEM Innovators: Identifying and Developing Our Nation's Human Capital*, Arlington, VA.
- Nicol DJ, Boyle JT (2003). Peer instruction versus class-wide discussion in large classes: a comparison of two interaction methods in the wired classroom. *Stud High Educ* 28, 458–473.
- Nielsen KL, Hansen G, Stav JB (2013). Teaching with student response systems (SRS): teacher-centric aspects that can negatively affect students' experience of using SRS. *Res Learn Technol* 21, 18989.
- Nielsen KL, Hansen G, Stav JB (2014). How the initial thinking period affects student argumentation during peer instruction: students' experiences versus observations. *Stud High Educ* 3, 1–15.
- Nielsen KL, Hansen-Nygård G, Stav JB (2012). Investigating peer instruction: how the initial voting session affects students' experiences of group discussion. *ISRN Educ* 2012, article 290157.
- Novak GM (1999). *Just-in-Time Teaching: Blending Active Learning with Web Technology*, Prentice Hall Series in Educational Innovation, Upper Saddle River, NJ: Prentice Hall, 188.
- Perez KE, Strauss EA, Downey N, Galbraith A, Jeanne R, Cooper S, Madison W (2010). Does displaying the class results affect student discussion during peer instruction? *CBE Life Sci Educ* 9, 133–140.
- Pilzer S (2001). Peer Instruction in physics and mathematics. *Probl Resour Issues Math Undergrad Stud* 11, 185–192.
- Porter L, Bailey-Lee C, Simon B (2013). Halving fail rates using peer instruction: a study of four computer science courses. In: *SIGCSE '13: Proceedings of the 44th ACM Technical Symposium on Computer Science Education*, New York: ACM Press, 177–182.
- Porter L, Bailey-Lee C, Simon B, Cutts Q, Zingaro D (2011a). Experience report: a multi-classroom report on the value of peer instruction. In: *Proceedings of the 16th Annual Joint Conference on Innovation and Technology in Computer Science Education*, New York: ACM Press, 138–142.
- Porter L, Bailey-Lee C, Simon B, Zingaro D (2011b). Peer instruction: do students really learn from peer discussion in computing? In: *ICER '11: Proceedings of the Seventh International Workshop on Computing Education Research*, New York: ACM Press.
- President's Council of Advisors on Science and Technology (2010). *Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future*, Washington, DC: Executive Office of the President, President's Council of Advisors on Science and Technology.
- 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.
- Ramalingam V, Wiedenbeck S (1999). Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. *J Educ Comput Res* 19, 367–381.
- Rao SP, DiCarlo SE (2000). Peer instruction improves performance on quizzes. *Adv Physiol Educ* 24, 51–55.
- Richardson J (2005). Concept inventories: tools for uncovering STEM students' misconceptions. In: *Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics (STEM) Education*, Washington, DC: American Association for the Advancement of Science, 19–25.
- Simkins S, Maler M (2009). *Just in Time Teaching: Across the Disciplines, and across the Academy*, Sterling, VA: Stylus, 224.
- Simon B, Esper S, Porter L, Cutts Q (2013a). Student experience in a student-centered peer instruction classroom. *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research—ICER '13*, New York: ACM Press, 129.
- Simon B, Kohanfars M, Lee J (2010). Experience report: peer instruction in introductory computing. *SIGCSE '10: Proceedings of the 41st ACM Technical Symposium on Computer Science Education*, New York: ACM Press, 341–345.
- Simon B, Parris J, Spacco J (2013b). How we teach impacts student learning: peer instruction vs. lecture in CS0. *SIGCSE '13: Proceedings of the 44th ACM Technical Symposium on Computer Science Education*, New York: ACM Press, 41.
- 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.
- 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.
- Steer D, McConnell D, Gray K, Kortz K, Liang X (2009). Analysis of student responses to peer-instruction conceptual questions answered using an electronic response system: trends by gender and ethnicity. *Sci and Educ* 18, 30–38.
- 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.
- Turpen C, Finkelstein N (2009). Not all interactive engagement is the same: variations in physics professors' implementation of peer instruction. *Phys Rev Spec Top Phys Educ Res* 5, 20101.
- 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, 020123.
- Williams K (2014). *Concept Inventories/Conceptual Assessments in Biology* San Diego State University Center for Teaching and Learning. <http://go.sdsu.edu/dus/ctl/cabs.aspx> (accessed 6 November 2014).
- Zingaro D (2014). Peer instruction contributes to self-efficacy in CS1. *SIGCSE '14: Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, New York: ACM Press, 373–378.
- Zingaro D, Porter L (2014a). Peer instruction in computing: the value of instructor intervention. *Comput Educ* 71, 87–96.
- Zingaro D, Porter L (2014b). Peer instruction: a link to the exam. *Proceedings of the 19th Annual Conference on Innovation and Technology in Computer Science Education*, New York: ACM Press, 255–260.