Peer Instruction in Computer Science at Small Liberal Arts Colleges

Leo Porter[1], Saturnino Garcia[2], John Glick[2], Andrew Matusiewicz[1] and Cynthia Taylor[3]

[1]Department of Mathematics and Computer Science, Skidmore College

[2]Department of Mathematics and Computer Science, University of San Diego

[3]Department of Computer Science, Oberlin College

ABSTRACT

Peer Instruction (PI) has been shown to be successful at improving pass-rates and improving retention of majors in large classes at large research-intensive institutions. At these institutions, students have been shown to learn from peer discussion in PI and both students and faculty have reported that they value PI in their classrooms. However, little is known about the effectiveness of PI in small classrooms at teaching-focused liberal arts colleges. This study evaluates results from seven lower-division classes and four upper-division classes taught at three different liberal arts institutions using PI. In these classes, PI experienced similar success as that reported at large-research intensive universities, both in terms of student learning from peer discussion and from student attitudinal surveys. Most notably, of 137 surveyed students, 91% recommend more faculty use PI in their classes.

Categories and Subject Descriptors

K.3.2 [**Computer Science Education**]: Computer and Information Science Education

Keywords

Peer Instruction, Active Learning, Liberal Arts Colleges

1. INTRODUCTION

Peer Instruction (PI) is a pedagogical practice designed to support student engagement in lectures and improve learning outcomes. PI centers on multiple-choice questions that students answer individually before discussing in small groups and answering again. This group vote is then followed by an instructor-led, class-wide discussion. PI has been shown to significantly improve student learning on concept inventories in physics [1, 6, 7]. Recently, these positive results have led to the adoption of PI in computer science at large, research-intensive institutions [15, 18].

At these institutions, PI in computing has gained significant momentum with increasingly positive supporting evidence. At these large institutions, PI has been shown to be highly valued by students, with a vast majority desiring other instructors adopt this practice in their classrooms [15, 11]. Instructors, as well as students, have been shown to value PI [11]. Students have been shown to learn from the peer discussion portion of the PI process by applying that new learning to isomorphic questions [12]. Most recently, PI has been shown to reduce fail rates [10], improve final exam scores [17], to be a valued component of CS-Principles [4], and contribute (with other best practices) to improved retention in CS1 [13].

Despite PI's success at large institutions, little is known of the effectiveness of PI in computer science in small, liberal arts colleges. Intuitively, small classes at liberal arts colleges are believed to offer higher levels of engagement between students and faculty. Instructors at these institutions may be more focused on teaching than their research-intensive counterparts, and thus be more apt to employ effective pedagogical practices. In fact, instructors at liberal arts institutions have espoused skepticism toward PI for exactly these reasons, claiming it is a technique whose effectiveness is limited to large institutions, since such engagement already occurs in their classrooms. Are they correct? Would the perceived generally higher levels of existing student engagement in the liberal arts classroom decrease the relative value of PI? In turn, would fewer students desire other faculty adopt PI at liberal arts colleges than at larger institutions?

To address skepticism of PI at small, liberal arts institutions, a group of five instructors adopted PI in their classes at three liberal arts institutions. These instructors included both junior and senior instructors, and both those experienced in developing PI materials and those adopting materials from others. The instructors adopted PI in eleven different classes—spanning four separate courses—to measure the impact on 201 enrolled students.

After analyzing these courses we found that student performance on in-class questions was comparable to that reported at large institutions as were the overwhelmingly positive student attitudes toward PI. These results strongly support more widespread adoption of PI in computer science classes at liberal arts institutions.

The contributions of this work include:

ITiCSE'13, July 1–3, 2013, Canterbury, England, UK.

Table 1: Per-Course Statistics. Note that L-E and L-G represent two separate but identical sections of the same course that have been merged for analysis.

	L-A	L-B	L-C	L-D	L-E	L-F	L-G	U-A	U-B	U-C	U-D
Course	CS1	CS1	CS1	CS1	CS1	CS1	CS1-E	Arch	Org	OS	Theory
Term	FA11	SP12	FA12	FA12	FA12	FA12	FA12	FA11	SP12	FA12	SP12
Enrolled Students	19	17	19	18	19,13	17	19,25	4	10	9	13
(S)enior/(J)unior Faculty	J	J	J	J	S	J	J	J	J	J	S
(D)eveloper or (A)dopter	A	A	A	A	A	A	A	D	D	D	A
Clicker Data			✓	✓	✓				✓	✓	✓
Attitudinal Survey Data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

1. We analyze results from three lower-division and three upper-division PI classes at liberal arts colleges to determine *Normalized Gain (NG)*, the improvement from individual vote to group vote. Averaged by class, NG is 42% for lower-division classes and 53% for upper-division classes. These results compare favorably to the results ranging from 29% to 41% in lower-division courses at large institutions previously reported in the literature [15, 18].

2. We report student attitudinal survey results on the value of PI in the liberal arts classroom. The vast majority (91%) of responding students recommend other instructors use PI in their classrooms, an important consideration for teaching-oriented, liberal arts institutions.

3. We discuss some potential limitations to these results as well as provide recommendations for PI adopters specific to small, liberal arts classrooms.

2. BACKGROUND

PI was established as a pedagogical practice initially in physics to improve student understanding of course concepts [1]. The practice consists of pre-class preparation, often in the form of online reading quizzes [2]. In class, students are posed questions which they think about and answer individually (solo vote), often using a classroom response system. The students then discuss the topic with a small group of classmates, and then answer the question again (group vote). The instructor then leads a class-wide discussion, and based on the quality of student responses, dynamically adjusts the next topic. For example, if most of the class answers correctly, the instructor may move forward. If most of the class answers incorrectly, the instructor may step back and ask more questions or hold a mini-lecture on the topic.

PI first appeared in the computer science education research literature in 2010 [15, 18]. Absent access to established concept inventories, student success was measured in terms of student attitudes toward PI and in terms of normalized gain (a measure of the improvement between individual and solo vote carefully defined in Section 3) [15, 18]. Normalized gain in the study of CS1 was 41% [15] and in remedial CS1 (for students repeating the course) was 29% [18]. We will compare our classes against these results to determine if similar gains are experienced in liberal arts classes.

The improvement from solo to group vote was subject to criticism, as students may just be copying the right answer from perceived stronger students. In biology, an isomorphic (same concept) second question was asked immediately after the group vote to test if students could apply their new learning to a new question. In biology, the result was that a majority of the learning that occurred during peer discussion was retained on the isomorphic question [16]. This important study was replicated in computer science in upper-division computer architecture and theory of computation classes with similar findings—that students can apply what they learned from peer discussion on another, same-concept, question [12]. These studies have implications for this work as they suggest that the majority of the gains we report here from group discussion reflect student learning.

Based on this initial research, a wave of recent publications on PI in computer science demonstrates that reading quiz performance correlates with course performance [19] and both students and instructors value PI [11]. Perhaps more significantly, PI has been shown to more than halve the fail rates in courses at a large institution [10] and that PI, combined with other CS1 best practices, can improve major retention by nearly a third [13]. Other work has aimed to identify why PI confers these advantages, including the suggestion that PI enables a form of cognitive apprenticeship for students [3].

All of these studies on PI in computer science have been conducted in large classes at large research-intensive universities. Although clickers alone have been shown to be valued by students in a CS0/Information Technology class at a liberal arts institution [8], this is the first investigation, to our knowledge, of PI in computer science classes at liberal arts institutions with small classes.

3. METHODOLOGY

All involved courses adopted the PI methodology and were taught at small, selective, private, liberal-arts institutions. Table 1 provides details regarding: whether the course was (L)ower- or (U)pper-division; the number of students enrolled in the class; the number of students who filled out the survey; the (J)unior/(S)enior status of the faculty instructor; whether the instructor (D)eveloped or (A)dopted the PI materials used for the course; and the data available for the study (clicker results / attitudinal surveys).

All courses adopted PI using clickers. In all classes, students were graded on participation in clicker questions and were not graded for answering questions correctly. In all except course L-D, students always discussed each question. In L-D, group discussion was omitted at times per instructor discretion.

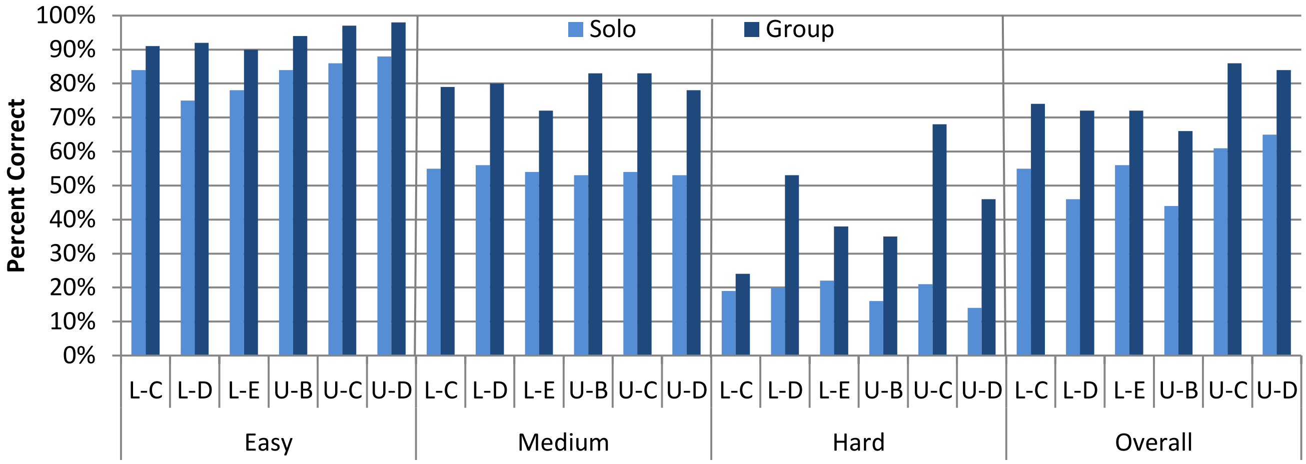

Figure 1: Organized by question difficulty, average per-question correctness on solo and group responses per course.

All CS1 courses were taught using variations of materials from the first PI CS1 study [15]. This course includes media computation [5] and pair programming [9]. All CS1 courses used Java with the exception of L-G, which used C++. In addition to CS1, courses included: a CS1-style class for engineers (L-G), Computer Architecture (U-A), Computer Organization (U-B), Operating Systems (U-C), and Automata / Theory of Computation (U-D).

3.1 Question Analysis

Prior work in other sciences has used established concept inventories to measure student learning [1]. The general absence of similar concept inventories in computer science has caused prior work to measure learning gains based on the improvement from the solo to group vote. Such a measurement has limitations in that it does not measure long-term retention nor does it measure the impact of instructor-led class-wide discussion. Despite these limitations, these learning gains (improvement from solo to group) have been shown to persist on subsequent isomorphic (same topic) questions in biology [16] and, recently, computer science [12].

Student responses on solo and group votes were recorded using i-clicker software in three lower-division and three upper-division courses, as indicated by Table 1. We categorize questions as being either easy, medium, or hard based on the percentage of who answered correctly during the solo vote; less than 35% correct is considered hard, 35%-74% correct is medium, and 75% correct and above is easy. This categorization follows accepted conventions [1, 16].

Improvement between solo and group vote is measured as Normalized Gain (NG). NG is calculated using the following equation [1]:

$$NG = \frac{group - solo}{(1.0 - solo)} \tag{1}$$

NG scales the improvement from solo to group vote by the fraction of students who initially responded incorrectly. This scaling allows for comparisons between the improvement of both easy and hard questions. However, it may penalize hard questions, as hard questions will require more dramatic changes in student correctness to achieve the same NG as easy questions.

3.2 Attitudinal Responses

Students were given voluntary end of term surveys to complete regarding the course based on the surveys given in prior work [11, 15]. Student responses were measured on a 6-point forced-choice Likert scale (no neutral). This means that on the survey, students were presented with statements with which they could choose to disagree very strongly, disagree strongly, disagree, agree, agree strongly, or agree very strongly.

4. RESULTS

In this section, we present average improvement between individual and group votes on clicker questions, in order to demonstrate that students are learning from discussion with their peers. We also provide the quantitative results of our attitudinal study, as well as select student quotes, showing that the students we surveyed are strongly in favor of instructors adopting peer instruction.

4.1 Question Analysis

Figure 1 shows the per-course percentage of students that answered correctly on individual and group responses, separated by question difficulty. All courses experience similar results for medium difficulty questions, which is the difficulty level recommended by prior work [1].

The most notable difference between classes appears between the improvements in group correctness for hard difficulty questions. Students in L-C showed little improvement from individual to group voting (<5%) while students in U-C showed an improvement of nearly 30%. In post-class discussions with the instructors, We identified a possible cause for this large difference after discussions with the instructors of these two courses. In course L-C the instructor rarely intervened in group discussions, instead only observing student interactions. In course U-C, the instructor was involved in nearly every discussion, frequently answering student questions. The U-C instructor's involvement in these conversations likely improved their performance above that which

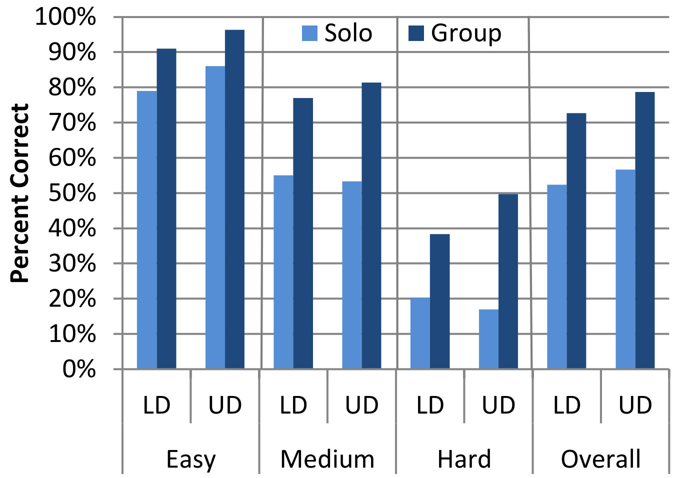

Figure 2: Grouped by question difficulty, average per-question correctness on solo and group responses averaged across lower-division (LD) and upper-division (UD) classes.

Table 2: Normalized Gain (NG) per course.

Course	Easy	Medium	Hard	Overall
L-C	45%	54%	7%	43%
L-D	70%	54%	41%	48%
L-E	54%	39%	20%	36%
U-B	60%	63%	22%	40%
U-C	79%	65%	60%	64%
U-D	84%	52%	37%	54%
Avg.	65%	55%	31%	48%

they might have achieved on their own. This involvement can still be viewed as being within the intent of the PI process. We will further address this issue in our recommendations for potential adopters in Section 5.

4.1.1 Course Level (Upper/Lower Division)

Student correctness on individual and group votes appears in Figure 2. In this figure, student correctness was averaged by lower-division and upper-division courses to better explore differences between lower-division and upper-division courses. Both lower- and upper-division courses experience similar improvements from individual to group vote, meaning both lower-division and upper-division students benefit from group discussion.

One subtle difference between lower-division and upper-division classes is that the raw improvement between students is higher for medium and for hard in the upper-division courses than in the lower-division courses. Although this improvement may stem from general differences in question content, instructor involvement, and students, we suspect upper-division students may possess higher levels of academic maturity and therefore may be better equipped to reason through difficult concepts.

4.2 Normalized Gain

Table 2 provides the normalized gains per course. Prior work at large, research-intensive universities has reported normalized gain between 29% and 41%. All our courses fall within or above that range. By this measure, PI is similarly effective in computer science both in large and small classes.

4.3 Attitudinal Survey

Table 3 provides results from a student attitudinal survey administered at the end of the term. The questions were asked on a 6-point forced-choice Likert scale (no neutral). We summarize the various levels of agreement (agree, strong agree, and very strong agree) here as just agree. We report results from the following questions:

- **Solo Voting:** "Thinking about clicker questions on my own, before discussing with people around me, helped me learn the course material."
- **Group Discussion:** "Discussing course topics with my seatmates in class helped me better understand the course material."
- **Immediate Feedback:** "The immediate feedback from clickers helped me focus on weaknesses in my understanding of the course material."
- **Clickers with Discussion:** "Clickers with discussion is valuable for my learning."
- **Recommend Approach:** "I recommend that other instructors use our approach (reading quizzes, clickers, in-class discussion) in their courses."

One interesting result from Table 3 is that course U-B experiences only 72% of the students recommending other faculty use this approach. This result lags behind the average of 91% for the other courses. In this course a single student publicly voiced displeasure with the PI methodology multiple times. Despite this dissenting student's openly-stated opinions, the fact that five of the seven students (72%) still recommended PI is a strong testament to the students' affinity for PI.

Despite these anomalies, the overall student responses were quite positive. Averaged by class, we find that students value thinking about the question on their own (98%), value discussion with classmates (97%), value feedback on their understanding of course material (93%), and find it valuable for their learning (95%). Moreover, an overwhelming majority of students (91%) recommend other instructors adopt this technique.

4.4 Student Remarks

Student remarks from the survey may provide a more complete picture of just why students favored PI courses. Students described PI classes as being more active/engaging than regular lecture, and described discussion as being helpful even when they knew the answer.

- "I cannot overstate how essential clickers are to staying awake in class."
- "Helping to explain the material to the people around me was immensely useful in understanding the material myself."

Overall, they describe peer instruction as more focused on understanding of the material than more traditional lecture.

Table 3: For each course, the percentage of students who agree with the statement based on an attitudinal results survey. The full text of the questions can be found in Section 4.3.

Course	L-A	L-B	L-C	L-D	L-E	L-F	L-G	U-A	U-B	U-C	
Survey Participants	15	17	19	13	18	9	27	4	7	8	Avg
Solo voting helps me learn	100%	100%	100%	100%	100%	100%	96%	100%	86%	100%	98%
Group discussion helped me understand	100%	100%	90%	100%	94%	100%	100%	100%	86%	100%	97%
Immediate feedback helped me identify my weaknesses	100%	94%	100%	85%	100%	100%	89%	100%	72%	88%	93%
Clickers with discussion helps me learn	93%	94%	100%	100%	94%	89%	93%	100%	86%	100%	95%
Recommend Approach	100%	94%	95%	92%	100%	78%	81%	100%	72%	100%	91%

- "I come away from [PI] lectures awake, energized, and feeling awesome about knowing the material. That's 'cause my job is to understand and pay attention, not just pay attention and hope for the best."

Students also favored the focus on doing the reading ahead of time, allowing for class time spent on deeper understanding of the material:

- "In other courses, I feel like more like an 'information sponge'. Material is thrown at me, and I have to soak up and take notes on as much as I can and hope that I'm taking notes on the right thing for the next exam. In this class, it felt like I did my initial learning with the readings, and then I took that learning into class and used class to solidify what I knew. I felt much better prepared for quizzes and exams in this class than i normally do."

These quotes support the quantitative survey results indicating that students value in-class discussion with their peers, and feel they are learning from it.

5. DISCUSSION

We generally shared the intuition of our colleagues at small liberal arts institutions: that the high-levels of student satisfaction with standard lecture classes at our institutions would cause students to be less receptive to new pedagogical practices. Thus, we were surprised that students so overwhelmingly valued PI in our courses (with some exceptions). In this section we discuss possible limitations of this study and offer liberal-arts institution specific recommendations.

5.1 Limitations

Although the student feedback is strongly in favor of peer instruction, we do not know if this could be explained by the quality of the instructors in this study. It is possible that the instructors willing to try PI in their classes at liberal arts institutions are "better" than a standard instructor. We believe this is likely not the case as only one of the instructors is a senior instructor and the other four were all junior faculty (two tenure-track faculty in their first two years, two visiting or adjunct faculty in their first year). However, if these instructors are better perceived by students, the replicability of these results may be jeopardized.

The student attitudinal survey was entirely voluntary and had good participation (72% of the students responded). However, 28% of students' opinions were not directly measured. Also, one course was not surveyed at all (U-D). Hence, our attitudinal results only measure those students who responded, and we cannot be sure of the opinions of those students who did not respond.

Although the intent of measuring the improvement from solo to group vote was primarily to compare against the success of PI at larger institutions, the studies on isomorphic questions [12, 16] may allow us to conclude that these improvements from peer discussion reflect student learning. However, the studies using isomorphic questions were performed with instructors not speaking with student groups during peer discussion and were performed at large institutions. As these facets do not apply to our study, this limits the extent to which we can conclude student learning is represented from the solo to group vote improvement. Validating student learning in a similar manner at a small institution could be the topic of future work.

5.2 Recommendations

A number of resources and best practices for PI adopters are available online [14]. In addition to those best practices, we recommend two additional practices.

First, the authors observed a student perception, among those students not recommending PI, that the clicker questions were a form of assessment. In turn, these students felt that incorrect answers to the in-class questions were failures on their part. This misperception arose despite all classes grading clicker questions on participation, not correctness. The authors found from subsequent discussions with students that once this misperception appears, it can be hard to overcome.

To address this misperception, we recommend emphasizing that questions in class are for learning, not assessment. Discuss that you will ask questions which most students will not get correct to help motivate lines of inquiry. Point out that if every student were correct on the individual vote, class would be just a validation of what they already know, instead of an opportunity for learning. However, students should be concerned about incorrect responses if they are consistently incorrect when most of the class answers correctly as it may reflect that they are falling behind.

Second, some of our instructors valued being more involved with students during the peer-discussion portion of the questions than is possible in a large class (just by the sheer number of groups). With three to seven groups in a small class, an instructor can try to engage with a larger fraction of those groups during the discussion. The primary value of this is not to point students towards the correct answer, but to help guide the discussion. By doing so, the instructor can ask the questions they hope the students are asking each other. For example, ask why a distractor answer is incorrect, and in what cases it might be correct. Similarly, ask if they could solve variants of the question. Guiding their thinking about the questions and their approach to discussion is highly valuable, especially early in the course.

6. CONCLUSION

Studies of Peer Instruction (PI) in computer science in large classes at large institutions have been positive, reporting high student satisfaction, improved exam scores, and higher passing rates. To what degree can PI be successful at institutions with small classrooms and teaching-focused faculty? In this study, five instructors used PI in seven lower-division classes and four upper-division classes at three different liberal arts institutions with small classes (i.e. those with 25 or fewer enrolled students). Impacting 201 enrolled students, we find that student learning from peer discussion in these liberal arts institutions is comparable to that found at larger institutions. Moreover, student opinions of PI remain overwhelmingly positive, with an average of 91% of students recommending that more faculty use PI in their classes.

7. ACKNOWLEDGEMENTS

Thank you to the anonymous reviewers for their contributions to this work. This work was supported in part by NSF grant 1140731.

8. REFERENCES

[1] C. H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69:970–977, 2001.

[2] C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur. Peer instruction: Engaging students one-on-one, all at once. In E. F. Redish and P. J. Cooney, editors, *Research-Based Reform of University Physics*. American Association of Physics Teachers, 2007.

[3] Q. Cutts, S. Esper, M. Fecho, S. R. Foster, and B. Simon. The abstraction transition taxonomy: Developing desired learning outcomes through the lens of situated cognition. In *Proc. of ICER '12*, New York, NY, USA, 2012. ACM.

[4] Q. Cutts, S. Esper, and B. Simon. Computing as the 4th r: a general education approach to computing education. In *Proc. of ICER '11*, pages 133–138. ACM, 2011.

[5] A. Forte and M. Guzdial. Motivation and nonmajors in computer science: identifying discrete audiences for introductory courses. *Education, IEEE Transactions on*, 48(2):248–253, 2005.

[6] R. Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66:64, 1998.

[7] N. Lasry, E. Mazur, and J. Watkins. Peer instruction: From harvard to the two-year college. *American Journal of Physics*, 76:1066–1069, 2008.

[8] M. Martyn. Clickers in the classroom: An active learning approach. *EDUCAUSE Quarterly*, 30, No. 2:71–74, 2007.

[9] C. McDowell, L. Werner, H. Bullock, and J. Fernald. Pair programming improves student retention, confidence, and program quality. *Communications of the ACM*, 49(8):90–95, 2006.

[10] L. Porter, C. Bailey-Lee, and B. Simon. Halving fail rates using peer instruction: A study of four computer science courses. In *Proc. of the 44th SIGCSE*, 2013.

[11] L. Porter, C. Bailey Lee, B. Simon, Q. Cutts, and D. Zingaro. Experience report: a multi-classroom report on the value of peer instruction. In *Proc. of ITiCSE '11*, 2011.

[12] L. Porter, C. Bailey-Lee, B. Simon, and D. Zingaro. Peer instruction: Do students really learn from peer discussion in computing? In *Proc. of ICER '11*, New York, NY, USA, 2011. ACM.

[13] L. Porter and B. Simon. Retaining nearly one-third more majors with a trio of instructional best practices in cs1. In *Proc. of the 44th SIGCSE*, 2013.

[14] B. Simon. Peer instruction for computer science. `http://peerinstruction4cs.org`, 2012. [Online; accessed 1-Jan.-2013].

[15] B. Simon, M. Kohanfars, J. Lee, K. Tamayo, and Q. Cutts. Experience report: Peer instruction in introductory computing. In *Proc. of the 41st SIGCSE*, pages 341–345, New York, NY, USA, 2010. ACM.

[16] M. K. Smith, W. B. Wood, W. K. Adams, C. Wieman, J. K. Knight, N. Guild, and T. T. Su. Why peer discussion improves student performance on in-class concept questions. *Science*, 323:122–124, 2009.

[17] J. Spacco, J. Parris, and B. Simon. How we teach impacts learning: peer instruction vs. lecture in cs0. In *Proc. of the 44th SIGCSE*, 2013.

[18] D. Zingaro. Experience report: Peer instruction in remedial computer science. In *In Proc. of Ed-Media 2010*. AACE, 2010.

[19] D. Zingaro, C. Lee, and L. Porter. Peer instruction in computing: the role of reading quizzes. In *Proc. of the 44th SIGCSE*, 2013.