A Comparison of Peer Instruction and Collaborative Problem Solving in a Computer Architecture Course

Geoffrey L Herman
University of Illinois at Urbana-Champaign
Urbana, Illinois
glherman@illinois.edu

Sushmita Azad
University of Illinois at Urbana-Champaign
Urbana, Illinois
sazad2@illinois.edu

ABSTRACT

While substantial research has demonstrated that active learning pedagogies are better for learning than passive lectures, we need to understand the trade-offs between different active learning pedagogies. Computer Architecture at Midwestern University has historically been taught using active lectures, introducing content with a few clicker questions. In Fall 2018 (N = 363), short video lectures were made available to students as a supplemental resource. In Spring 2019, the instructor flipped the course, requiring students to watch the video lectures and complete an assignment before attending class. Two versions of the course were taught concurrently, using the same homework assignments, machine problems, and examinations but with different in-class pedagogies. Version SP19PI (N = 179) was taught using peer instruction. Version SP19CP (N = 73) was taught using collaborative problem solving, organizing students into teams of 3 to work on problems. Students completed surveys that measured their perceptions of time spent on the course, course difficulty, perceptions of stress, and sense of belonging. We compare students' performance on midterm exams and their non-cognitive outcomes to examine the relative effects of these different active learning pedagogies. We find that both flipped offerings (peer instruction and collaborative problem solving) benefited students beyond active lectures. Peer instruction (SP19PI) made learning more efficient. Collaborative instruction (SP19CP) provided greater social support for learning and eliminated gender grade disparities.

CCS CONCEPTS

• **Applied computing → Education**; **Collaborative learning**; *Computer-assisted instruction*; E-learning;

KEYWORDS

flipped classroom; peer instruction; collaborative problem solving; active learning; motivation

ACM Reference Format:
Geoffrey L Herman and Sushmita Azad. 2020. A Comparison of Peer Instruction and Collaborative Problem Solving in a Computer Architecture Course. In *The 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20), March 11–14, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3328778.3366819

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGCSE '20, March 11–14, 2020, Portland, OR, USA
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6793-6/20/03.
https://doi.org/10.1145/3328778.3366819

1 INTRODUCTION

Research in STEM and computing has demonstrated that active learning pedagogies are better for students' learning when compared with traditional passive lecturing [2, 8]. Active learning pedagogies are broadly defined as teaching methods that require students to actively construct their own understanding of presented content rather than passively receiving information [8]. In their meta-analysis on active learning, Freeman et al. argue that the positive effects of active learning are so strong and robust across contexts, that we need to stop using traditional lecturing as our baseline measurement and instead should examine the relative merits of different active learning pedagogies [8]. We present a study to compare the relative effects of a lecture with elements of active learning with two evidence-based, active learning pedagogies — peer instruction [18] and collaborative problem solving [10, 11] — on student outcomes in a computer architecture course.

(1) What are the relative effects of peer instruction and collaborative problem solving on students' learning?
(2) What are the relative effects of peer instruction and collaborative problem solving on students' affective or non-cognitive outcomes?

1.1 Context of the research study

At the University of Illinois at Urbana-Champaign, Computer Architecture is a large enrollment (300-400 students per term), required course for Computer Science majors. Prior to Spring 2019, the course was taught with lectures three days per week and a discussion/laboratory section once per week. Lectures introduced students to the course content. Lectures were interspersed with about three, multiple-choice questions (MCQs) to keep students engaged and to address common misconceptions. The discussion/laboratory discussion sections were taught by graduate teaching assistants and undergraduate course assistants. These sections guided students through problem solving exercises to prepare them for each week's machine problem assignment (i.e., long homework coding problems). Thus, the course could be described as using elements of active learning throughout, but it was not committed to any particular evidence-based pedagogical method.

Using video lectures that were first made available to students during Fall 2018, the instructor flipped the classroom during Spring 2019: Students were introduced to course content outside class with the video lectures so that class time could be spent engaging students in more active learning. Students completed a short, online homework assignment that included open-ended conceptual questions before coming to class. The instructor used students' answers to these questions to guide the active learning pedagogies.

To maximize the value of the flipped course design, we compare the potential benefits of two different evidence-based instructional practices: *Peer Instruction* [18] and *Context-Rich Collaborative Problem Solving* [10, 11]. We concurrently taught two variants of the course: SP19PI and SP19CP. SP19PI was taught using peer instruction, which is a widely used active learning pedagogy in computing [21, 26]. SP19CP was taught using context-rich, collaborative problem solving [10, 11]. Teams of students were given a worksheet with complex, open-ended problems and encouraged to work together as the instructor floated between teams.

2 LITERATURE REVIEW

Constructivist Theory argues that knowledge cannot merely be transmitted from an instructor to a student [6, 20]. Rather, a student must be actively involved in the construction and reconstruction of their own knowledge throughout the learning process. Active learning pedagogies are generally based upon Constructivist Theory, seeking to create opportunities and environments in which students are expected to be active participants in learning rather than passive recipients of knowledge [8].

2.1 Peer Instruction

Peer Instruction was popularized by Eric Mazur [18]. In peer instruction, the instructor typically guides students through a series of MCQs that aim to elicit common misconceptions about a topic. Peer instruction frequently iterates through the following steps [4].

(1) Instructor introduces or reviews a topic.
(2) Instructor poses a multiple-choice question about the topic.
(3) Students attempt to answer the question by themselves, committing to an answer through some response system such as a clicker or response card.
(4) Instructor shares student responses and students are encouraged to discuss their answers with peers before submitting their answer a second time.
(5) The instructor debriefs with students based on their responses before posing additional questions or progressing to the next topic.

Peer instruction is an effective teaching method in many disciplines, including a variety of computing courses [21, 26].

To implement peer instruction well, instructors must carefully craft MCQs that are both approachable yet challenging [5]. If these questions are too easy, students can easily become bored and disengage from learning activities. If questions are too hard, students can become overwhelmed and discouraged. Additionally, like lecturing, peer instruction constrains all students to progress at the same pace, regardless of their mastery of the material.

2.2 Context-Rich Collaborative Problem Solving

Context-Rich Collaborative Problem Solving was developed by Heller [10, 11]. This pedagogy engages students through two primary mechanisms: context and collaboration. In STEM, instructors frequently give students context-less problems that focus on drilling a single skill or concept. Drilling problems are valuable for developing some core competencies but are limited at helping students identify when or how to apply these competencies in disciplinary contexts. By engaging students in rich disciplinary contexts, students have an opportunity to learn how and when to apply their competencies and are given what are hopefully more motivational problems that help them understand why these problems are important to solve. Because this additional context makes problems more complex, students are organized into teams so that they can mutually support each other through these challenging problems.

The instructor may provide some initial instruction to help students get started, but student teams primarily work at their own pace [11]. The instructor rotates between groups, seeking not to provide students with answers but to ask students probing questions to help them discover answers for themselves and make continuous progress throughout the class period. The instructor may interrupt group work to provide some instruction for all groups if she perceives that many groups are struggling with the same difficulties. This type of open-ended problem solving can be motivating for students as they can progress at their own speed and experience more control over their learning [19]. Providing students with the opportunity to collaborate can also potentially help students develop friendships and see more of their peers as resources [14].

Creating high quality collaborative exercises can be more difficult than creating peer instruction questions. If problems are too simple, students have no reason to collaborate. If problems are too difficult, students can get even more stuck than with peer instruction [10]. Additionally, students may be frustrated by an instructor who primarily asks them questions rather than providing them answers as they may expect should be the instructors' role [7].

2.3 Flipped Classrooms

The flipped classroom (or inverted classroom) is an instructional strategy that is becoming increasingly used to support the use of active learning in the classroom [3, 15, 22]. In a flipped classroom, students are responsible for familiarizing themselves with the content before coming to class [15]. The instructor may assign readings or videos to watch before class, so that class time can be used for more active learning. It is commonly recommended that students complete some type of pre-class quiz to encourage students to complete pre-class activities or help the instructor tailor the in-class activities to students' levels of mastery [15].

Flipped classrooms have been shown to potentially have many benefits such as increased flexibility for students to familiarize themselves with course content at their own pace, re-reading texts or re-watching videos [15]. Since instructors do not need to focus on "covering" the content, they are more free to focus on interacting with students and reviewing particularly challenging content. These affordances can improve student engagement and learning.

3 EXPERIMENTAL DESIGN: COMPONENTS OF COMPUTER ARCHITECTURE

We performed a quasi-experimental study. In this section, we describe the components of the course. Table 1 highlights which components we attempted to control between offerings and which were varied. The same faculty instructor was the only instructor for all offerings. Because our goal in this study was to focus on the effects of the pedagogies used in the classroom, we endeavored to keep as many assignments and learning resources the same or parallel across offerings. For example, when teaching students about finite state machines, all students were given a guiding example of a simple candy dispenser. In peer instruction, students were given MCQs at critical points such as deciding how many flip-flops would be needed to store the state. In collaborative problem-solving, students were simply given the specification for the candy dispenser and were given hints to think about how many flip-flops would be needed in the final design of the sequential logic circuit.

This constraint meant that we did not make some changes that may have been optimal for each particular pedagogy. This decision minimizes confounders deriving from the creation of new learning materials or learning activities outside the classroom. However, this design means that any differences may be more conservative than if tailored resources and policies were created for each pedagogy.

Class sizes were constrained by facilities. FA18 was taught in two sections in a lecture hall with 200 seats. SP19PI was taught in the same lecture hall. SP19CP was taught in the largest active learning classroom available on campus with 90 seats. All teaching resources are available upon request by contacting the lead author.

3.1 Attendance and Weekly Quizzes

In all three offerings, students were awarded 3% course credit for attendance during the primary pedagogy. Attendance points were awarded with clicker questions. An additional 2% course credit was awarded for attendance through a short (one problem, < 10 minutes) weekly quizzes. Quizzes were given during once-a-week discussion section (FA18 and SP19PI) or lecture (SP19CP) and were meant to be easy (90% average).

3.2 Video Lectures

Video lectures were filmed during Fall 2018 using screen captured powerpoint slides with instructor voice-over. Videos are hosted on a YouTube-like platform [13] with links embedded in students' homework assignments. Videos were less than ten minutes each. About 15 minutes of video lecture content was required before each class meeting. Videos were filmed using principles from Multimedia Learning Theory [16, 17]. Slides had minimal, if any text, so that voice narration and visual space did not compete for students' working memory. Video lectures using these principles have been shown to improve students' preparation for flipped classrooms [24].

3.3 Homework and Machine Problems

Online homework assignments were offered through a proprietary learning management system (PrairieLearn [1]). This system encourages the creation of question generators that create randomized instances of a question within a certain parameter space. For example, a question about how many bits are needed to encode N pieces of information might randomize N to be between 33 and 255 to help students identify the $\lceil log_2 N \rceil$ solution for this class of problems. The randomizing lets students attempt as many variants of each problem as they wish.

Each week, two hours of class time were devoted to helping students prepare to complete coding-intensive assignments called machine problems. Students were encouraged to work in teams to complete the assignments. Machine problems were completed in a variety of languages such as C, Verilog, and MIPS assembly.

3.4 Exams and Second-chance exams

Starting in the fourth week, students took an exam every other week. They also had the option to re-take a different version of each exam in the immediately following week for full grade replacement if they wanted to improve their score on the exam [12]. For example, week 4 students took a required exam 1, week 5 students could elect to take a new, different variant of exam 1, week 6 students took a required exam 2.

Exams 1, 2, 3, 4, 5 and 7 and the final exam were all computer-based exams held in a proctored computer lab with computers that had limited network access [25]. Exams consisted of a mixture of at most one coding question (Exam 7 and the final exam had no coding questions) and homework problems selected from the online homework, except with randomized parameters to minimize cheating.

For coding questions, students were given access to development tools such as compilers and debuggers. Code was auto-graded upon submission with partial credit awarded based on the number of test cases the students' code passed. Short-answer questions were selected from students' online homework problems but with new, randomized parameters to minimize cheating.

Exam 6 is the only paper-based exam in the course. It is the culmination of the course's main learning objectives (an understanding of performance considerations for single-core architectures). The exam requires students to reason about pipelined datapaths they have not seen before, identify performance issues related to pipeline stalls and flushes, reason about the cache hit and miss rates of some code given a variety of cache configurations, and re-write code to improve performance by being more cache conscious. Answering questions about these topics requires students to have a deep understanding of all prerequisite material. Exam 7 and the final exam are lower stakes, cumulative exams consisting only of questions asked on the homework problems. They focus on testing teaser topics for future courses such as parallelism and virtual memory.

Exam 6 is the hardest exam in the course and was chosen as the primary anchoring comparison measure between offerings of the course. Because the exam is paper-based, we can maintain slightly better exam security by keeping all paper copies. We administered the same Exam 6 in both the FA18 and SP19 course offerings with only minor tweaks to obfuscate similarities. Tweaks were minor such as renaming variables or scaling all numerical values by a common factor. No change fundamentally changed any solution strategy and we used the same grading rubrics. Students were not informed of this decision to re-use the exam and exam questions had never been re-used previously.

4 RESEARCH MEASURES

We measured students' cognitive and non-cognitive (perceptions of stress, difficulty, interest, competence, test anxiety, and time spent) outcomes. Cognitive outcomes were measured by the course exams

Table 1: Comparison of controls and variants between offerings of Computer Architecture. Merged columns indicate things that were held constant across course offerings

<table>
<tr><th>Lecture w/ clickers (FA18, N = 363)</th><th>Peer Instruction (SP19, N = 179)</th><th>Collaboration (SP19, N = 73)</th></tr>
<tr><td>276 Male, 82 Female, 5 Unknown</td><td>130 Male, 49 Female</td><td>55 Male, 17 Female, 1 Unknown</td></tr>
<tr><td colspan="2">3 lectures per week</td><td>3 problem solving sessions per week</td></tr>
<tr><td>Average 3 MCQ per lecture, course content introduced</td><td>Average 6 MCQ per lecture, course content reviewed</td><td>Average 4 open-ended questions per discussion, course content reviewed</td></tr>
<tr><td colspan="3">Points awarded for attendance (3% of grade)</td></tr>
<tr><td>Video lectures available</td><td colspan="2">Video lectures required before class</td></tr>
<tr><td>Online homework assignment due 36 hours after each lecture (5% of grade)</td><td colspan="2">Same online homework assignment split into two parts: part 1 due before class, part 2 due 36 hours after each lecture (5% of grade)</td></tr>
<tr><td colspan="2">1 discussion/lab section per week</td><td>1 lecture/lab section per week</td></tr>
<tr><td colspan="2">Students work in teams on small exercises to prepare for the weekly machine problem (extended coding problem)</td><td>Review concepts with MCQ and prep students for weekly machine problem</td></tr>
<tr><td colspan="3">Short (10 minute) weekly quiz (2% of grade)</td></tr>
<tr><td colspan="3">Weekly machine problem (extended coding problem completing in pairs - 25% of grade)</td></tr>
<tr><td colspan="3">Parameterized exams (questions chosen from homework pool with random question parameters), 7 midterm exams with second-chance mastery exams (55% of grade), one final exam (10% of grade)</td></tr>
</table>

and attendance, particularly exam 6 (See Section 3.4). Non-cognitive outcomes were measured by an end-of-semester survey.

Students' attendance was recorded using clickers. To isolate the effect of attendance on course performance, we measured students' course performance omitting attendance points.

The end-of-semester comprised validated surveys including the Sense of Belongingness Survey (11 items) [14], the Motivated Strategies for Learning Questionnaire (MSLQ - 9 items) [9], and students' self-reported perceptions of course difficulty, stress, and time spent on the course.

The sense of belongingness survey measures how much students believe they have meaningful relationships in a course, using 5-point Likert scale items (0 - Strongly Disagree to 4 - Strongly Agree) such as "I feel that an instructor would take the time to talk to me if I needed help" and "I feel comfortable asking a question in class." A student's perception of her or his 'fit' into the college environment, their "Sense of Belonging," plays a critical role in whether students persist in college [14]. Large lecture courses can depress students' sense of belonging as they may feel like just a number [14]. It is expected that increased faculty-student interactions may improve students' sense of belonging.

From the MSLQ, we used the intrinsic motivation (3 items), competence beliefs (3 items), and test anxiety (3 items) sub-scales [9]. Intrinsic motivation measures students interest in learning the course material (e.g., "It is important for me to learn the course material"). Competence beliefs measure beliefs that they can learn the material if they exert sufficient effort (e.g., "If I try hard enough, then I can understand the course material"). Test anxiety measures negative emotions that may hinder students' performance on exams (e.g., "I have an uneasy, upset feeling when I take an exam in this course"). All items use the same 5-point Likert scale (0 - Strongly Disagree to 4 - Strongly Agree).

Students were asked to rate the difficulty of the course and amount of stress experienced during the course relative to other courses (0 - Much less to 4 - Much More).

Students also reported how much time (in hours) they spent on the course during a typical week.

4.1 Data collection and analysis

The survey was collected anonymously and analyzed according to IRB protocols by a graduate research assistant who was unaffiliated with the course. All course offerings were statistically compared using one-way ANOVAs with $\alpha = 0.05$ for significance testing unless otherwise specified. The relationship between students' performance and attendance was calculated using Pearson's correlations and Fisher's Z-test with $\alpha = 0.05$.

5 RESULTS

We present results of comparisons for both cognitive outcomes and non-cognitive outcomes.

5.1 Cognitive Outcomes

To measure learning, we compared students' scores on Exam 6 scores (Table 2) There was not a significant effect on cognitive outcomes [F(2, 606) = 0.887, p = 0.412].

Table 2: Exam 6 scores for each course offering

	FA18 $\mu(\sigma)$	SP19PI $\mu(\sigma)$	SP19CP $\mu(\sigma)$	p-vlaue
All students	68.2 (17.7)	69.4 (19.3)	71.1 (18.4)	0.412
Male	69.5 (16.8)	72.2 (18.4)	70.6 (19.8)	0.342
Female	63.7 (19.6)	62.0 (19.8)	72.8 (12.8)	0.130

To examine whether the different pedagogies affected men and women differently, we compared the performance of men and women on Exam 6 in each of the pedagogies (Figure 1) There was a significant difference in performance by gender during both FA18 [F(1, 356) = 6.801, p = 0.009] and SP19 PI [F(1, 177) = 10.554, p = 0.001]. A post-hoc Tukey test revealed women underperform the men in both FA18 and SP19PI (p = 0.009 and p = 0.001 respectively). This difference represents a Cohen's d effect size of 0.995 (FA18) and 1.329 (SP19 PI). There were no significant difference in performance between genders for SP19 CP.

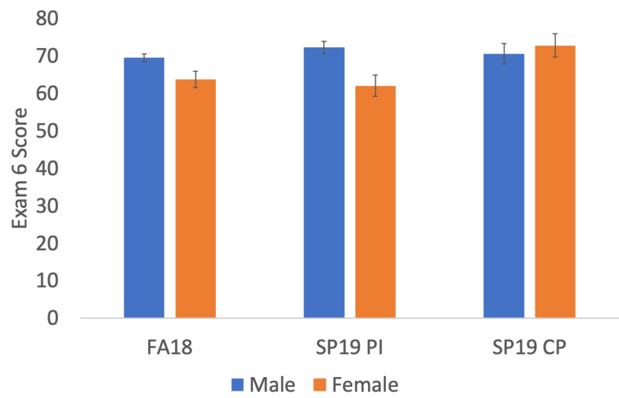

Figure 1: Exam 6 scores by gender. FA18 and SP19PI have significant differences between genders. SP19CP does not.

Pearson correlations revealed significant correlations (p <0.01) between attendance and course performance for all offerings of the course (see Table 3). Fisher's z-test revealed a significant pairwise difference between FA18 and SP19CP (see Table 3), with SP19CP having the strongest correlation between attendance and course performance. There were no other pairwise significant differences.

Table 3: Correlation co-efficients between attendance and course performance and Fisher Z-test p-values, comparing strengths of correlation

	SP19PI (r = 0.611)	SP19CP (r = 0.727)
FA18 (r = 0.507)	p = 0.097	p = 0.136
SP19PI (r = 0.611)		p = 0.005

To further explore this finding, we performed a follow-up test to examine whether there were differences in how important students perceived class attendance to be for their learning. Students rated the importance of attending class on a 4-point scale (0 - not important to 4 - essential). Students' reported perception of the value of class attendance revealed a significant effect [F(2, 423) = 3.687, p = 0.026]. A post-hoc Tukey test revealed that attendance in SP19CP was perceived to be more important than in both FA18 and SP19PI (p = 0.028 and p = 0.030 respectively). This difference represents a Cohen's d effect size of 0.45. There were no other significant pairwise differences.

5.2 Non-Cognitive Outcomes

A summary of non-cognitive outcomes can be found in Table 4.

Students' reported average time spent per week on the course (Figure 2) revealed a significant effect [F(2, 429) = 3.593, p = 0.028]. A post-hoc Tukey test revealed that the SP19PI version required less time than the FA18 version (p = .048). This difference represents a Cohen's d effect size of 3.62. There were no other significant pairwise differences.

Table 4: Non-cognitive survey results for each course offering. All comparisons are on a 4-point Likert scale except time spent. Higher is worse for difficulty, stress, and test anxiety. Higher is better for interest, competence, and belonging. A response value of 2 is a neutral response (neither agree nor disagree) for each scale. Pairwise significant differences bolded

	FA18 N = 310 $\mu(\sigma)$	SP19PI N = 84 $\mu(\sigma)$	SP19CP N = 38 $\mu(\sigma)$
Time (hours)	**11.24 (0.32)**	**9.61 (0.55)**	12.1 (1.11)
Difficulty	**2.56 (0.04)**	2.36 (0.08)	**2.24 (0.13)**
Stress	**3.79 (0.05)**	**3.51 (0.10)**	3.47 (0.16)
Test Anxiety	2.52 (0.96)	2.32 (0.99)	2.49 (0.57)
Interest	1.71 (0.60)	1.69 (0.68)	1.54 (0.47)
Competence	2.74 (0.73)	2.75 (0.72)	2.70 (0.58)
Belonging	**2.41 (0.60)**	**2.61 (0.60)**	**3.02 (0.62)**

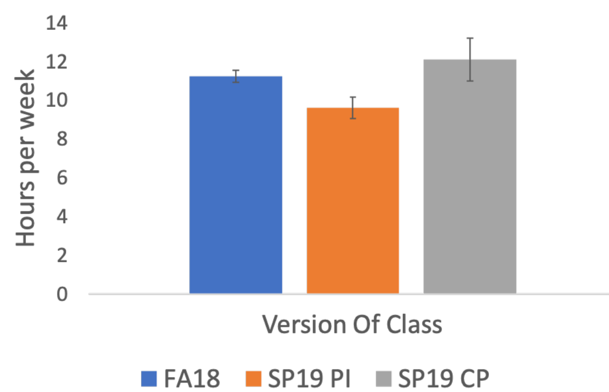

Figure 2: Time spent on the course. SP19PI took significantly less time than FA18.

Students' reported perception of course difficulty (Figure 3) revealed a significant effect [F(2, 429) = 4.34, p = 0.013]. A post-hoc Tukey test revealed that the SP19CP version was perceived to be less difficult than the FA18 version (p = .046). This difference represents a Cohen's d effect size of 1.07. There were no other significant pairwise differences.

Students' perceptions of course stress (Figure 3) revealed a significant effect [F(2, 429) = 4.880, p = 0.008]. A post-hoc Tukey test revealed that SP19PI was less stressful than FA18 (p = .026). This difference represents a Cohen's d effect size of 3.54. There were no other significant pairwise differences.

Students' perceptions of test anxiety [F(2, 429) = 1.54, p = 0.215], intrinsic motivation to learn the course material [F(2, 429) = 1.17, p = 0.312], and competence beliefs about ability to learn the material [F(2, 429) = 0.06, p = 0.939] did not reveal any significant differences.

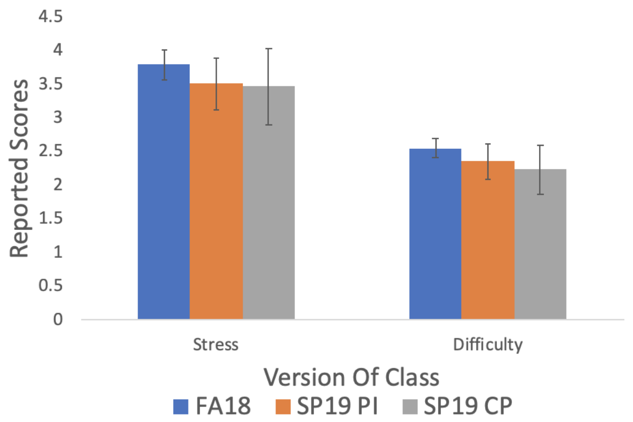

Figure 3: Perceived stress and difficulty of the course. SP19PI was perceived as significantly less stressful than FA18. SP19CP was perceived as significantly easier than FA18.

Students' sense of belonging (Figure 4) revealed a significant effect [$F(2, 429) = 18.92$, $p < 0.001$]. A post-hoc Tukey test revealed significant pairwise differences (Table 5) between all offerings.

Table 5: Pairwise p-values and Cohen's d effect sizes for post-hoc Tukey test for students' sense of belonginginess

	SP19PI	SP19CP
FA18	p = 0.02 (d = 0.33)	p = 0.001 (d = 1.0)
SP19PI		p = 0.002 (d = 0.67)

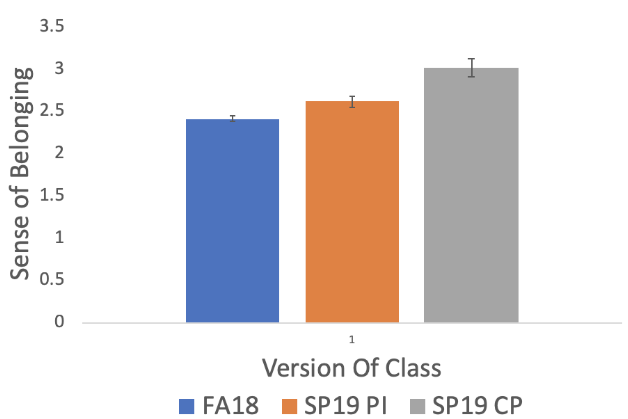

Figure 4: Sense of belonging felt by students. Significant differences observed between all offerings of the course

6 DISCUSSION AND CONCLUSIONS

This study revealed that using evidence-based pedagogies improved student outcomes above and beyond using ad hoc components of active learning in a large enrollment, computing course. The different pedagogies provided different benefits to the students.

While students did not perform better on exams, they reported spending less time and found the course to be less stressful when taught using peer instruction. Peer instruction seemed to make students' learning more efficient and afforded them more time to devote to other responsibilities, reducing their stress. Students saved a substantial amount of time, reducing time spent by over an hour and a half per week (~15%). Given the high demands on students' time in a computing curriculum, this outcome is non-trivial. Students also reported a slightly higher sense of belonging (small effect size) in the course, perhaps because of the increased exchanges with the instructor.

For collaborative learning, the students did not perform better on the exams nor did they report spending less time on the course. They did, however, report feeling like the course was less difficult, that they had a substantially higher sense of belonging, and attendance had a significantly greater effect on their learning both actual and perceived. We believe that collaborative problem solving forces students to struggle through learning the content with less direct assistance from instructors, so their learning is not more efficient. However, the students had better social support networks to help them through those struggles. These social supports can help students persist through the course [14].

Additionally, collaborative learning did not have a "gender gap" in students' grades, whereas the other course offerings did have a "gender gap." While women performed better in the collaborative learning context, men performed about the same. Together, with the finding that students experienced a greater sense of belonging, our findings suggest that the context-rich collaborative problem solving style of course may help improve retention of women in computing. Future studies could explore this dynamic more, both examining the robustness of this finding and the long term cohort and curricular effects on retention.

Our findings are limited to one course and may not generalize to other contexts. Differences and effects on learning may be understated as the course only increased the amount and style of active learning. The size of effects and how differences manifest will depend on the quality of implementation of both the learning resources and the instructors' ability with the pedagogy. Prior efforts to flip courses and increase active learning have reported similar findings though: students do not learn more but have improved affective outcomes [23], suggesting a degree of robustness in the findings. Additional replication in other courses by other instructors is needed to understand the robustness of these findings.

The increased efficiency of learning in peer instruction and increased social connectedness of collaborative problem solving can both potentially play critical roles for improving student outcomes and well-being in computing and should be explored further. These differences in outcomes also underscore the need to further explore the relative benefits of different active learning pedagogies to better help instructors choose the best pedagogies for their context.

ACKNOWLEDGMENTS

Thank you to the graduate teaching assistants and undergraduate course assistants who made this possible.

REFERENCES

[1] [n. d.]. PrairieLearn. https://prairielearn.engr.illinois.edu. Accessed: 2019-11-19.

[2] A. Beach, C. Henderson, and N. Finkelstein. 2012. Facilitating change in undergraduate STEM education: Implications from an analytic review of literature. *Change: The Magazine of Higher Learning* 44, 6 (2012), 52–59.

[3] J. L. Bishop and M. A. Verleger. 2013. The flipped classroom: A survey of the research. In *Proceedings of the ASEE Annual Conference & Exposition*. Atlanta, GA. https://peer.asee.org/22585

[4] M. Borrego, S. Cutler, M. Prince, C. Henderson, and J. E. Froyd. 2013. Fidelity of implementation of Research-Based Instructional Strategies (RBIS) in Engineering Science Courses. *Journal of Engineering Education* 102, 3 (2013), 394–425.

[5] C. H. Crouch and E. Mazur. 2001. Peer instruction: Ten years of experience and results. *American Journal of Physics* 69, 9 (2001), 970–977.

[6] P. A. Ertmer and T. J. Newby. 1993. Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly* 6, 4 (1993), 50–72.

[7] Cynthia J. Finelli, Kevin Nguyen, Matthew DeMonbrun, Maura Borrego, Michael Prince, Jennifer Husman, Charles Henderson, Prateek Shekhar, and Cynthia K. Waters. 2018. Reducing student resistance to active learning: Strategies for instructors. *Journal of College Science Teaching* 47, 5 (2018), 80–91.

[8] S. Freeman, S. L. Eddy, M. McDnough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America* 111, 23 (2014), 8410–8415.

[9] T. Garcia and P. R. Pintrich. 1996. *Assessing students' motivation and learning strategies in the classroom context: The Motivated Strategies for Learning Questionnaire.* Kluwer Academic/Plenum Publishers, New York, NY, 319–339.

[10] P. Heller and M. Hollabaugh. 1992. Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics* 60 (1992), 637–644.

[11] R. Heller, P. Keith and S. Anderson. 1992. Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics* 60 (1992), 627–636.

[12] G. Herman, K. Varghese, and C. Zilles. 2019. Second-chance testing course policies and student behavior. In *Proceedings of the 2019 IEEE Frontiers in Education Conference (FIE).*

[13] G. L. Herman. [n. d.]. CS 233 Video Lectures. https://mediaspace.illinois.edu/channel/CS%2B233/99753341. Accessed: 2019-11-19.

[14] M. Hoffman, J. Richmond, J. Morrow, and K. Salomone. 2002. Investigating "sense of belonging" in first-year college students. *Journal of College Student Retention: Research, Theory and Practice* 4 (2002), 227–256.

[15] A. Karabulut-Ilgu, N. Jaramillo Cherrez, and C. T. Jahren. 2018. A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology* 49, 3 (2018), 398–411. https://doi.org/10.1111/bjet.12548

[16] R. E. Mayer. 2009. *Multimedia Learning Theory.* Cambridge University Press.

[17] R. E. Mayer and R. Moreno. 2010. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist* 38, 1 (2010), 43–52.

[18] E. Mazur. 1997. *Peer Instruction: A User's Manual.* Prentice Hall, Upper Saddly River, NJ.

[19] C. P. Nemiec and R. M. Ryan. 2009. Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education* 7, 2 (2009), 133–144.

[20] J. Piaget. 1954. *The construction of reality in the child.* Basic Books, New York, NY.

[21] Leo Porter, Dennis Bouvier, Quintin Cutts, Scott Grissom, Cynthia Lee, Robert McCartney, Daniel Zingaro, and Beth Simon. 2016. A Multi-institutional Study of Peer Instruction in Introductory Computing. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16).* ACM, New York, NY, USA, 358–363. https://doi.org/10.1145/2839509.2844642

[22] A. A. Rahman, H. Mohamed, B. Aris, and N. M. Zaid. 2014. The influences of flipped classroom: A meta analysis. In *IEEE Conference on Engineering Education.* Kuala Lumpur, Malaysia. https://doi.org/10.1109/ICEED.2014.7194682

[23] Timothy Stelzer, David T. Brookes, Gary Gladding, and JosÃl' P. Mestre. 2010. Impact of multimedia learning modules on an introductory course on electricity and magnetism. *American Journal of Physics* 78, 7 (2010), 755–759. https://doi.org/10.1119/1.3369920 arXiv:https://doi.org/10.1119/1.3369920

[24] Timothy Stelzer, Gary Gladding, Jose P. Mestre, and David T. Brookes. 2009. Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content. *American Journal of Physics* 77, 2 (Feb 2009), 184–190. https://doi.org/10.1119/1.3028204 arXiv:physics.ed-ph/0806.0405

[25] C. Zilles, M. West, D. Mussulman, and T. Bretl. 2018. Making testing less trying: Lessons learned from operating a Computer-Based Testing Facility. In *Proceedings of the 2018 IEEE Frontiers in Education Conference (FIE).*

[26] Daniel Zingaro and Leo Porter. 2014. Peer Instruction in Computing: The Value of Instructor Intervention. *Comput. Educ.* 71 (Feb. 2014), 87–96. https://doi.org/10.1016/j.compedu.2013.09.015