Martin Barrett, CARNEGIE MELLON UNIVERSITY, martinba@cmu.edu
Chad Hershock, CARNEGIE MELLON UNIVERSITY, hershock@cmu.edu
Michael McCarthy, CARNEGIE MELLON UNIVERSITY, mm6@andrew.cmu.edu
Michael Melville, CARNEGIE MELLON UNIVERSITY, mmelvill@andrew.cmu.edu
Joe Mertz, CARNEGIE MELLON UNIVERSITY, joemertz@andrew.cmu.edu

What Type of Debrief is Best for Learning during Think-Pair-Shares?

ABSTRACT

Copious research demonstrates the benefits of adding active learning to traditional lectures to enhance learning and reduce failure/withdrawal rates. However, many questions remain about how best to implement active learning to maximize student outcomes. This paper investigates several "second generation" questions regarding infusing active learning, via Think-Pair-Share (TPS), into a large lecture course in Computer Science. During the "Share" phase of TPS, what is the best way to debrief the associated course concepts with the entire class? Specifically, does student learning differ when instructors debrief the rationale for every answer choice (full debrief) versus only the correct answer (partial debrief)? And does the added value for student outcomes vary between tasks requiring recall versus deeper comprehension and/or application of concepts? Regardless of discipline, these questions are relevant to instructors implementing TPS with multiple-choice questions, especially in large lectures. Similar to prior research, when lectures included TPS, students performed significantly better (~13%) on corresponding exam items. However, students' exam performance depended on both the type of debrief and exam questions. Students performed significantly better (~5%) in the full debrief condition than the partial debrief condition. Additionally, benefits of the full debrief condition were significantly stronger (~5%) for exam questions requiring deeper comprehension and/or application of underlying Computer Science processes, compared to simple recall. We discuss these results and lessons learned, providing recommendations for how best to implement TPS in large lecture courses in STEM and other disciplines.

KEYWORDS

active learning, think-pair-share, large courses, explanation feedback

INTRODUCTION

Copious research demonstrates the positive impacts of active learning during class sessions on both learning outcomes and student persistence (Freeman et al. 2014; Prince 2004). In this context, active learning refers to intentional (individual or collaborative) constructivist activities during which students are leveraging the cognitive process known as the generation effect (Bertsch et al. 2007). This is in contrast to passive, exposition-centered activities in which students are primarily listening and/or taking notes, or activities occurring between class meetings such as homework assignments. Prince

(2004) reviewed classroom research studies across several types of active learning in Engineering disciplines. Although effect sizes varied among different forms of active learning, overall, his meta-analysis found empirical support (i.e., positive effect sizes) for the types of active learning reviewed. Similarly, in a meta-analysis of 225 classroom research studies spanning eight STEM disciplines (Freeman et al. 2014), on average, embedding active learning in lecture courses increased student performance on exams and/or concept inventories by 0.5 standard deviations (~6%) and decreased student drop, failure, and withdrawal rates by a factor of 1.5 (~12%). Furthermore, active learning positively impacted student outcomes across a range of course sizes, including large courses (>100 students). Consequently, the authors argued that studying the presence versus absence of active learning during lectures (i.e., whether or not active learning "works") is no longer an interesting research question, per se.

However, because many other practical questions about active learning remain unanswered, Freeman and colleagues (2014) encouraged future researchers to investigate what they referred to as "second generation" research questions. Conspicuous examples of second generation questions include the following. For whom does active learning work? Do all students benefit, or do benefits differ among demographic groups of students? And, can active learning mitigate any performance gaps across demographic groups of students? Does the frequency of active learning matter for student outcomes? Is there a dose-response effect? What types of active learning work best for particular learning objectives (e.g., recall vs. comprehension vs. application)? What is the best way for instructors to implement a particular active learning technique in a particular teaching context? Although these questions may lead to quite different lines of research, they do share a common denominator; the answers to any second generation questions could help instructors better understand how to optimize student outcomes via active learning during their class sessions.

Some of the aforementioned second generation questions have already received some attention and are accruing nascent bodies of research. For example, multiple studies suggest that active learning, in combination with other targeted interventions, can decrease achievement gaps for underrepresented populations (Eddy and Hogan 2014; Gavassa et al. 2019; Goeden et al. 2015; Haak et al. 2011; Webb 2017; Winkelmes et al. 2016). Other studies have explored how the specific implementation of active learning techniques drive observed learning gains. For instance, student outcomes tend to be higher when active learning includes a collaborative component, rather than individual work alone (e.g., Linton, Farmer, and Peterson 2014; Smith et al. 2009). In these studies, learning gains also tend to be greater when active learning targets difficult concepts and higher-order learning objectives, compared to simple recall of basic concepts. Although these findings are undoubtedly useful in some contexts, there still remain many interesting and unexplored applications of second generation questions that could benefit from additional study.

For example, how should instructors facilitate whole-class debriefs following active learning to maximize student learning? Active learning works, in part, because it provides low stakes practice and feedback to students, both of which are critical components of learning (Ambrose et al. 2010). Feedback, in particular, helps students correct misconceptions and errors in knowledge and reinforces correct knowledge. Unsurprisingly, the characteristics of the feedback students receive also impact the quality of learning (Hattie and Timperley 2007). In studies using multiple-choice questions as practice exercises, students often retain more when feedback clearly indicates the correct answer, not just whether students' responses are right or wrong (Pashler et al. 2005). Moreover, when the rationale for

the correct answer is also explained (i.e., "explanation feedback"), students are better able to transfer and apply their prior conceptual learning to unfamiliar scenarios (Butler, Godbole, and Marsh 2013). Therefore, instructors' facilitation choices may impact both the depth and quality of the feedback students receive via active learning during class. In other words, student learning may in part depend on the extent to which correct answers are explicitly signaled to students during active learning, assuming exercises are closed-ended (i.e., have a single correct answer, but this is certainly not the only option for implementing active learning, see Bruff 2009). In addition, it is possible that the *amount* of explanation feedback provided as part of an active learning exercise may influence its impact on learning. We are not aware of classroom research directly testing either of these predictions.

In this paper, we report on classroom research regarding the implementation of active learning in the form of Think-Pair-Share (TPS) exercises (see Lyman 1987) in a large, lecture-based course in Computer Science. Our study sought to answer the following specific, practical research questions:

- Research Question 1 (RQ1): Does adding brief TPS exercises to lecture sessions enhance student outcomes?
- Research Question 2 (RQ2): During the "Share" phase of TPS, what's the best way to debrief the exercise with the entire class (in terms of learning and efficiency)? Specifically, do student outcomes differ when instructors debrief the rationale for all the answer choices provided (full debrief) versus only the correct answer choice (partial debrief)?
- Research Question 3 (RQ3): Do the effects of TPS and debriefing methods differ when learning objectives increase in cognitive demand from simple recall to deeper comprehension and/or application of concepts?

RQ2 and RQ3 directly investigate two of the second generation questions discussed above. We believe RQ2 and RQ3 are relevant to instructors, regardless of STEM discipline, who wish to infuse active learning into lectures via the TPS method and by leveraging multiple-choice questions (see Bruff 2009), particularly for questions with a single, best answer. To investigate RQ2, we intentionally manipulated the depth of explanation feedback provided by instructors during active learning. In practical terms, if there is no evidence of a difference in learning as a result of more in-depth debriefing, then instructors could limit the debrief of active learning to identifying and explaining only correct answers, and thus save time. Additionally, answering RQ2 has implications for the design of asynchronous, online learning experiences embedding active learning because they frequently employ explanation feedback consistent with a partial or full debrief of answer choices (e.g., Lovett, Meyer, and Thille 2008). Investigating RQ3 can help instructors target how to best leverage active learning. Faculty may perceive multiple barriers to adopting active learning, including constraints on time for preparing activities and teaching during class sessions (e.g., Michael 2007; Miller and Metz 2014). We call the latter source of instructor resistance "the coverage conundrum," the belief that implementing active learning requires an instructor to sacrifice content coverage due to time constraints. Theoretically and practically, we disagree with this premise. One can often approach active learning as a potential one-to-one temporal replacement of didactically delivered content. Nevertheless, if an instructor wishes to strategically prioritize use of active learning to maximize the "bang for the buck," then data regarding RQ3 can help inform instructor choices. To explore RQ3, each instance of active learning contained one task requiring recall of fundamental concepts and one task requiring deeper comprehension and/or application of those concepts. For the concepts targeted by active learning during lectures, we then

compared student outcomes on exam questions requiring recall to questions requiring comprehension and/or application.

Our classroom research was motivated by both institutional constraints on teaching as well as the aforementioned, practical considerations regarding how to implement active learning effectively and efficiently at scale. Like many universities, we face burgeoning enrollments in Computer Science courses. Our particular course was taught in two, 80-minute lectures per section per week. Adding numerous additional course sections did not scale well in terms of faculty effort nor classrooms available. And, we feared significantly larger lecture sections would lead to less learning. Although we did increase the size of the two existing sections somewhat and added a third section, we took the opportunity to integrate more active learning into our pedagogy. We thought that highly scripted lab exercises, overseen by Teaching Assistants, could be a scalable way to increase experiential learning in Computer Science courses with burgeoning enrollments. First, we moved to a single lecture per section each week. Then, we converted the second, weekly class meeting into a lab session. We split each 60-student lecture section into four lab sections. Each lab section met in a different portion of the same room and was advised by a Teaching Assistant. Instead of a lecture on the week's topic, each lab activity was a highly scripted set of instructions that guided students through similar content with activities such as programming a simple (GET only) HTTP server or comparing the speed of XML web services versus binary Java RMI. The labs were well-received by students, but because the lab exercises were highly scripted and not discussed at all during lecture sessions with the faculty instructors, we were concerned about whether students were completing the labs without learning key concepts. As we feared, students' exam performance on lab content was lower than expected, consistent with a C grade (figure 1, see data for "Lecture (F15)"). In contrast, students were not systematically struggling with other course concepts. Consequently, we set out to understand how we might improve students' learning of lab material via adding active learning exercises on lab concepts during subsequent lecture sessions. And, we wondered about the extent to which the details of our implementation strategy for active learning during lectures mattered for student outcomes (see RQ2 and RQ3 above).

During each lecture session following a lab, we implemented an active learning technique known as TPS that has been used in many areas of education, including fields as disparate as Biology (Smith et al. 2009), Computer Science (Porter et al. 2011), and Comparative Literature (Bruff 2009; Tinkle et al. 2013). TPS combines an individual activity ("Think") with a low-stakes peer discussion ("Pair"), followed by a whole-class debrief ("Share"), encouraging all students to engage cognitively and summarize a concept in their own words (Lyman 1987). Although active learning can include both individual and cooperative exercises, TPS is a form of active learning that intentionally leverages peer instruction (Crouch and Mazur 2001) during the "Pair" phase. In addition to the possible added value of collaborative learning (Linton et al. 2014; Smith et al. 2009), TPS is particularly amenable to large lecture contexts because it can be implemented flexibly at scale, regardless of the classroom infrastructure, with or without educational technology (e.g., personal response systems or "clickers," see Bruff 2009).

This paper discusses how we leveraged TPS with multiple-choice questions during lectures to reinforce learning of fundamental concepts introduced during the lab sessions. This approach leverages spaced, low stakes practice and feedback in a way that is potentially pedagogically effective, yet requires only a small amount of class time. We emphasize empirical lessons learned regarding the details of how best to implement TPS in a large lecture setting to maximize student outcomes. Our results are relevant

to instructors across disciplines, especially those teaching large lectures and grappling with how best to design and implement short, active learning exercises to target particular learning outcomes.

While our interventions during lectures targeted lab content, our research questions and experimental manipulations do not investigate how best to design or teach labs in STEM, per se, and should not be construed as such. We could have targeted any fundamental course concepts using the approach reported here. We simply focused on the content most challenging to our students, based on past exam performance, which happened to be concepts initially covered during lab sessions.

METHODS

Our course, Distributed Systems for Information Systems, is a master's level required course at a medium-sized, research-intensive, midwestern university. We currently enroll 150-180 students in three sections per semester. Each section meets twice per week: a lecture session on new material, followed by a lab session covering the lecture material, and sometimes additional material. The course is typically taught by two instructors who divide up the topics and then each lecture in all sections. The lab sessions provide highly scaffolded, "hands-on" programming exercises facilitated by a teaching assistant. Despite good overall exam results and positive student feedback on the labs, students were not performing well on exam questions related to the lab content. A comprehensive, retrospective review of the lab content, either at the end of the lab session or the beginning of the next lecture, would require too much time. Instead, we tried a simpler, more efficient method using short, targeted active learning exercises: asking two multiple-choice questions, one at a time, about key concepts from the lab content at the beginning of the next lecture session, but using TPS to actively engage students in a discussion about the questions. Students individually answered the questions ("Think") and then discussed their answers and their underlying rationale with a peer ("Pair"). Finally, instructors facilitated a debrief of the questions, including calling on students to identify and justify their answer choices ("Share"). We did not employ instructional technology, such as clickers or personal response systems (Bruff 2009) during active learning exercises. On average, TPS activities lasted a total of 5 minutes.

Procedure

In Fall 2018, we implemented TPS in two of the three course sections during the lectures following each lab session. Each occurrence of TPS included two types of multiple-choice questions: a simple identify-style question (I-type) requiring basic recall of concepts, and an applied question (A-type) requiring a deeper understanding of the underlying computer science principles and processes (see table 1 for examples). Before the end of each TPS, instructors clearly identified correct answers (visually and verbally) and their underlying rationales (verbally). However, instructors discussed the incorrect answer choices differently across course sections (conditions randomly assigned to sections). One section used a full debrief, also discussing why each incorrect answer choice provided was wrong. The other section used a partial debrief and did not discuss the rationale for any incorrect answer choices.

For comparison, the Fall 2015 version of the course served as a control (i.e., the course contained labs, but active learning was absent from lectures and lectures did not review lab content). To measure the impact of TPS, we compared students' performance on lab-related exam items in Fall 2015 to the two sections receiving TPS in Fall 2018. To measure the effect of full versus partial debriefs on learning outcomes, we compared performance on lab-related exam items between the two sections

receiving TPS in Fall 2018. Because exams included both I-type and A-type questions, we could also explore whether effects depended upon the type of learning objective (i.e., recall vs. deeper conceptual understanding and/or application). In Fall 2018, exam questions differed from those used during TPS exercises, but were isomorphic, testing the same concepts (table 1). To account for potential differences between student cohorts in our program or between sections in Fall 2018, we statistically controlled for student's incoming GPA (see Results).

Table 1. We used two types of questions during class activities and on exams, I-type required only recall of information, A-type required application and/or comprehension of an underlying principle or process

	Example questions from active learning exercises and exams

I-type question	In Raft, each node is in one of three states. These states are:
	   a. Zero, one, two
	   b. Term 1, Term 2, Term 3
	   c. Heartbeat, election, replicated log
	   d. Follower, candidate, leader
	   e. Senator, president, judge

A-type question	In Raft*, if a leader fails, then the other nodes will miss the heartbeat. The next node to become a candidate is selected:
	   a. By a majority
	   b. At random
	   c. By its node number – higher nodes are selected first
	   d. By a super majority of 2/3
	   e. By the previous leader

* In this case, students had to understand a key component of Raft and be able to describe the next step in the protocol.

Participants

The Fall 2018 course had 120 students who self-identified as Asian (79%), White (4%), Hispanic (2%), Black (1%), and unreported (14%). The Fall 2015 course had 171 students who self-identified as Asian (89%), White (4%), Hispanic (1%), and unreported (6%). The mean incoming cumulative GPA for the students was 3.61 ($SD = 0.25$) for the fall 2015 students, and 3.55 ($SD = 0.27$) for the fall 2018 students. An independent samples *t* test revealed that this GPA difference is statistically significant, $t(289) = 1.975$, $p < .05$, $d = .24$. Consequently, we chose to control for this difference in our analyses by including student GPA as a covariate to account for any influence that variable may have on our results. The University Registrar supplied all demographic and GPA data, consistent with approved IRB protocols from our institution.

RESULTS

To address our first research question, students were evaluated on instructor-designed, multiple-choice questions from three exams—two midterms and a final—to determine whether the active learning intervention improved student performance compared to lecture. An analysis of covariance

(ANCOVA) was conducted using condition (i.e., active learning or no active learning) as the independent variable, student GPA as a covariate, and mean student exam score on lab-related content as the outcome. Results showed that, when controlling for differences in GPA, students performed significantly better in Fall 2018 on questions covered by the TPS intervention compared to performance during Fall 2015 on the same items when students received lecture only, $F(2,288) = 49.22$, $p < .001$, $\eta_P^2 = .146$. Figure 1 and table 2 depict estimated marginal means for exam performance with versus without active learning.

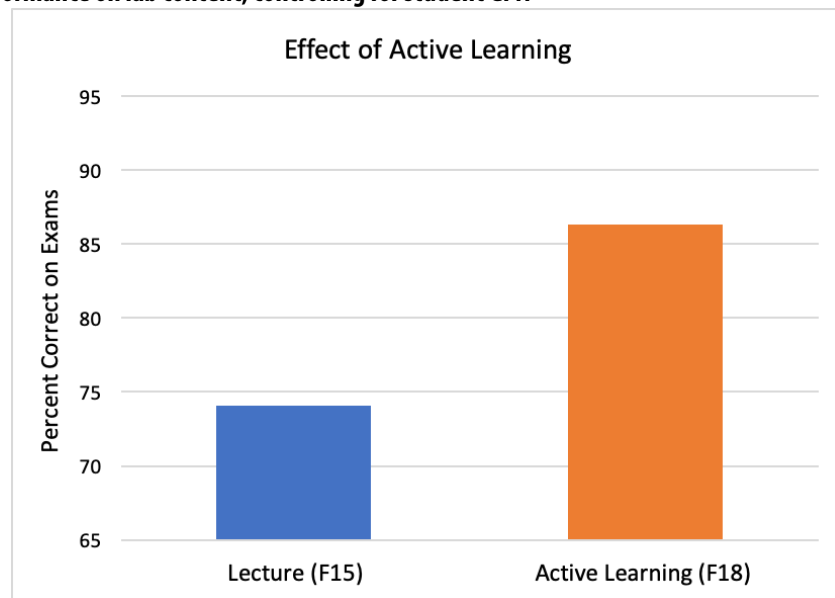**Figure 1. Exam performance on lab content, controlling for student GPA**

Table 2. Exam performance on lab content in Fall 2015 and Fall 2018

Lab intervention	N	Mean	Std.Error
Active learning (F18)	120	86.30	1.32
Lecture (F15)	171	74.10	1.11

In addition, we compared student performance between the two semesters on all other exam content not related to the TPS intervention. These analyses showed that, while controlling for student GPA, students in the Fall 2018 semester did score significantly higher than students in the Fall 2015 semester, $F(2,288) = 4.48$, $p < .05$, $\eta p2 = .038$. This difference was substantially smaller than the difference found for the TPS-related content. Estimated marginal means are shown in table 3.

Table 3. Exam performance on non-lab content in Fall 2015 and Fall 2018

Non-lab intervention	N	Mean	Std.Error
Active learning (F18)	120	78.75	0.75
Lecture (F15)	171	76.65	0.63

For the second research question regarding how to best debrief the TPS exercises, student exam performance was compared across the two treatment groups during Fall 2018. Another ANCOVA was conducted using debriefing type (i.e., full or partial) as the independent variable, student GPA as the covariate, and mean student exam score on the lab-related content as the outcome. Results showed that, controlling for GPA, students' mean performance on all lab-related exam content was significantly better for students in the full debrief group compared to the partial debrief group, $F(1,117) = 4.94$, $p < .05$, $\eta_P^2 = .041$. figure 2 and table 4 depict estimated marginal means for exam performance on content that was partially or fully debriefed during active learning exercises.

In addition, we compared student performance on exam content that was *not* related to the TPS lab exercises between the partial and full debrief groups. These analyses showed that, controlling for student GPA, there was no significant difference in student performance between the two groups, $p = .327$.

Figure 2. Exam performance on lab content when partially or fully debriefed active learning, adjusted for student GPA
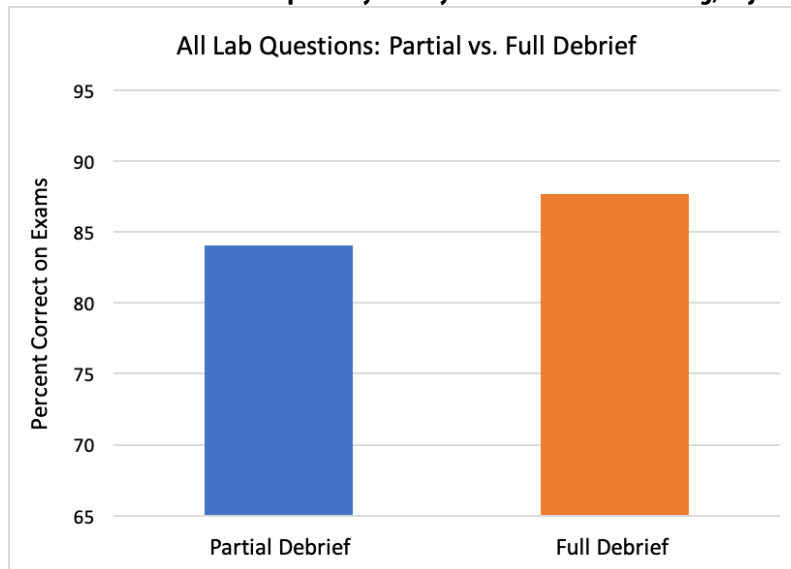

Table 4. Exam performance on lab content in the partial and full debriefing conditions during Fall 2018

All lab questions	N	Mean	Std.Error
Partial debrief	59	84.29	1.21
Full debrief	61	87.89	1.19

Next, to address our third research question, we tested to see if the effect of debriefing strategy on lab-related exam performance depended on the type of exam questions being answered (i.e., I-type or A-type). A repeated measures ANOVA revealed a statistically significant interaction between debriefing condition and question type, $F(1,117) = 5.43$, $p < .05$, $\eta_P^2 = .044$, suggesting that the beneficial effect of the full debriefing intervention varies depending on which type of exam questions are being used. Indeed, simple effects analysis examining student performance on only A-type questions reveals that students who received the full debriefing strategy scored significantly higher than students who had received only the partial debriefing intervention, $b = .07$, $t(118) = 3.08$, $p < .05$, $\eta_P^2 = .075$. In contrast,

however, no significant difference was found between the conditions for I-type questions, $p = .547$. Estimated marginal means for all combinations of question type and debrief are shown in figure 3 and table 5.

Figure 3. Exam performance on A-type and I-type questions when partially or fully debriefing active learning, controlling for student GPA
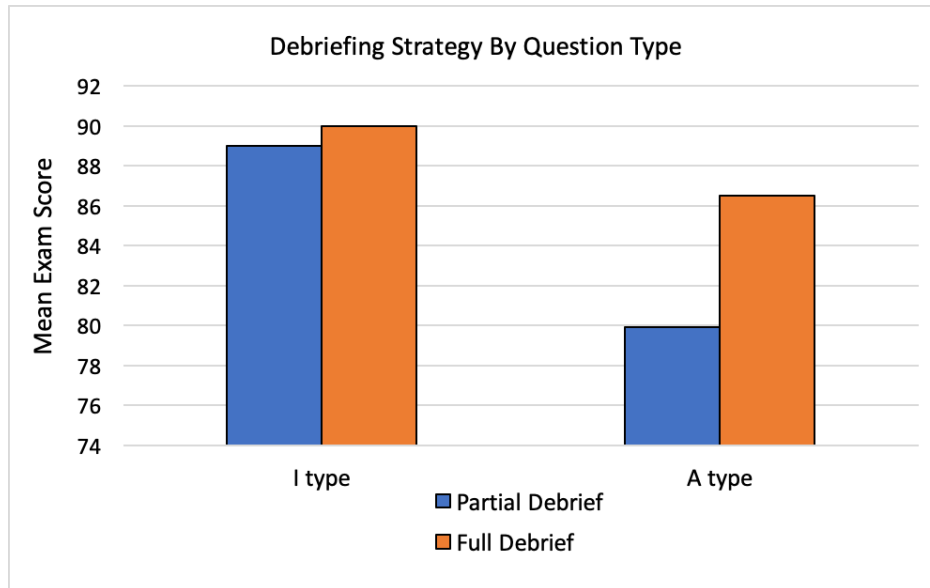

Table 5. Exam performance on A-type questions in the partial and full debriefing conditions during Fall 2018

	N	I-type questions		A-type questions	
		Mean	Std.Error	Mean	Std.Error
Partial debrief	59	88.99	1.17	79.92	1.52
Full debrief	61	89.98	1.15	86.48	1.49

DISCUSSION

We evaluated three practical research questions to inform our future teaching. First, does adding brief, active learning exercises (using the TPS method with multiple-choice questions) during lectures enhance student outcomes in a large lecture course? Second, during these TPS exercises, what is the best way to debrief the associated course concepts with the entire class? Specifically, does student learning differ when instructors debrief the rationale for all the answer choices versus only the correct answer? Third, does the added value TPS for student outcomes vary between comparatively lower and higher-level learning objectives (recall versus comprehension and/or application)? Below, we highlight lessons learned for maximizing the effectiveness of TPS in STEM disciplines, including, but not limited to, Computer Science, as well as large, lecture-based courses, regardless of discipline.

Did adding active learning via TPS to lectures increase student outcomes?

Student retention of lab material covered outside lecture improved when we added short TPS exercises to the subsequent lecture (figure 1, table 2). Furthermore, this increase in performance was not seen for content that was not augmented with those exercises. When using TPS with two multiple-choice questions per lecture, we observed performance gains (0.83 standard deviations or approximately 13% on average) comparable to other studies measuring the impacts of adding active learning to lectures (Crouch and Mazur 2001; Freeman et al. 2014). Our observed student outcomes further illustrate the return on investment from small, targeted uses of active learning to improve conceptual learning in general, but in Computer Science, specifically.

While many studies of active learning target concepts introduced during lecture readings assigned prior to lectures (see Freeman et al. 2014), we intentionally targeted content introduced through lab exercises occurring outside of lectures because students' exam performance was lower than expected on this material. Therefore, our results are also noteworthy for STEM instructors designing curricula in which lab and lecture sections are traditionally separate learning experiences. While labs and lecture sections are often related conceptually, they often differ in their primary learning objectives, teaching methods and assessments, or the timing of content delivery during a semester. Consequently, learning from labs and lectures may not be effectively integrated or connected for students, especially when instructors differ (e.g., Teaching Assistants versus faculty, respectively). For instructors concerned about this issue, we demonstrate that brief, active learning interventions during lectures—focused on lab content—can dramatically enhance learning outcomes from labs that relate to lecture content.

Several cognitive mechanisms may have contributed to our observed results. Normally, benefits of active learning targeting "lecture concepts" may be attributed to the generation effect (Bertsch et al. 2007). Students who actively generate information as part of the learning experience encode and remember the information better (compared to passively receiving information, as through reading, or by extension, traditional lecture). However, in our case, TPS was reinforcing previous exposure to content from the previous week's lab exercise. Previous research illustrates the positive impacts of both retrieval practice (e.g., Butler 2010, Roediger and Karpicke 2006; Trumbo et al. 2016) and spaced practice (e.g., Cepeda et al. 2008; Kapler, Weston, and Wiseheart 2015; Rohrer and Taylor 2006) on student outcomes. Students remember more information when they: (1) practice retrieving previously encoded information compared to passively re-reading or studying information (the testing effect), and (2) space practice over time compared to massing the same amount of practice in one episode of practice (the spacing effect). In addition to the generation effect, our implementation of TPS may have leveraged both the testing effect and spacing effect. Unfortunately, our study design does not allow us to parse the relative contribution of the generation, testing, and spacing effects to the observed increase in student performance. Regardless, it is worth noting that one's implementation of active learning may synergistically leverage several, well-supported learning principles.

Additionally, our results suggest that TPS, a specific active learning method, is an effective, efficient way to incorporate active learning into large lectures, especially if instructors wish to integrate and reinforce learning that occurs outside of lectures in their courses. The steps of TPS are easy for both students and instructors, and TPS improved learning outcomes with a relatively small investment of class time—less than the time it would take to review all the concepts and lab examples thoroughly by lecturing. Because it encourages interactions among students during the Pair phase, TPS also has the

added potential benefit of promoting socialization among students (Porter et al. 2011) and creating a more collegial, inclusive atmosphere, both of which are challenges in a large lecture course.

Because TPS is easy to use, flexible, and requires very little time, it is surprising that use cases have not been reported on more frequently in Computer Science (beyond those reported in Freeman et al. 2014). For instance, in an introductory Computer Science course, researchers quantified the specific behaviors of students participating in TPS activities, finding that, on average, 83% of students were actively engaged in productive behaviors across the three phases (Kothiyal et al. 2013). Other computer scientists have used TPS to encourage students to work together using a collaborative code tracing tool (Wormeli 2004) and in a computer graphics course when introducing new tools (Schweitzer, Boleng, and Scharff 2011). Pair programming is related to TPS as two programmers take turns coding and observing. Pair programming in computer science education has been reported on more extensively (Radermacher and Walia 2011; Zheng, Kang, and Harrington 2019). Here, we used TPS to explore the extent to which short, low stakes practice and feedback exercises can reinforce conceptual learning from labs, and how varying the implementation of TPS impacts learning outcomes. Our implementation and the use cases described above indicate the potential to apply active learning strategically in many Computer Science teaching contexts.

What is the best way to debrief during the "Share" portion of TPS exercises?

We implemented active learning using TPS with multiple-choice questions. This approach is common across large lecture courses in STEM, Social Sciences, and Humanities (see examples in Bruff 2009). While we did not use "clickers" to debrief our multiple-choice questions, instructors frequently use instructional technologies to implement TPS and other forms of peer instruction (Mazur and Couch 2001; Bruff 2009; Smith et al. 2009). When using multiple-choice questions with a single best answer, numerous studies suggest that students learn best when the correct answer is clearly communicated as part of their feedback, rather than only receiving feedback on whether their response is correct or incorrect (reviewed in Butler et al. 2013). Furthermore, previous research on feedback supports the added value of communicating the rationale for the correct answer ("explanation feedback"), rather than simply indicating which answer is correct (Butler et al. 2013). We questioned whether expanding explanation feedback to include the rationale for each answer choice provided would further enhance student outcomes.

When using multiple-choice questions during TPS, our data suggest including a full debrief of all possible answer choices may have added value (figure 2, table 4). Our students who received a full debrief of both correct and incorrect answers showed better depth of knowledge than students debriefing only the rationale for the correct answer, as evidenced by performance on related exam questions, especially those requiring deeper comprehension or application of concepts rather than recall alone. Interestingly, Butler and colleagues (2013) found that the impacts of explanation feedback were greatest for application questions, i.e., questions requiring the transfer of conceptual learning to new situations requiring an inference. Our study expands on this finding, demonstrating that augmenting explanation feedback by also communicating the rationale for incorrect answer choices can further enhance learning outcomes on application tasks, or when a deeper understanding of underlying principles is required. This result is potentially relevant to instructors in Humanities, Social Sciences, and STEM leveraging multiple- choice questions with a single best answer for active learning during large, introductory lecture courses.

However, TPS with multiple-choice questions is certainly not limited to questions with a single best answer, in STEM or other disciplines (Bruff 2009; Tinkle et al. 2013). For example, in the Humanities one can practice "close reading" and critical thinking during class, even in large lectures, by providing students with a text passage or work of art and multiple-choice questions about how best to interpret or analyze it. The multiple-choice questions provided may not have a single correct answer because a passage can be interpreted multiple ways using the evidence provided by the text/artwork itself and various disciplinary lenses. In such cases, the potential value of the active learning exercise lies in practicing a disciplinary skill or habit of mind, such as formulating an argument (or comparing and contrasting arguments) based on close reading of a text or image, rather than selecting the "best" answer, per se. The targeted skills can then be assessed on subsequent tests or via other types of assignments that require their demonstration. One can easily imagine similar approaches in STEM and the Social Sciences related to critical reading, application of fundamental concepts, problem solving, or programming. When multiple-choice questions do not have a single correct answer, logically, it is likely that a full debrief of answer choices following TPS may similarly enhance student outcomes. We hope that future research will attempt to explore how best to implement such "open-ended," multiple-choice questions, in STEM and other disciplines.

Does TPS work better for some learning outcomes than others?

While copious research demonstrates that active learning enhances student outcomes in large courses (Freeman et al. 2014), whether or not active learning works better for certain types of learning objectives compared to others remains an open question. Limited prior research addresses this question directly. Smith and colleagues (2009) ranked concepts targeted during active learning (peer instruction) as hard, moderate, or easy in difficulty. Gains from active learning were largest for the concepts instructors ranked as the hardest difficulty. Likewise, Linton, Farmer, and Peterson (2014) found that active learning in small groups benefited students on assessment items described as "higher-level, extended response questions," but not on "low-level, multiple-choice questions."

To compare the benefits of TPS for learning objectives focused on recall versus comprehension/application of fundamental concepts, we intentionally included questions differing in cognitive difficulty during classroom activities and on exams (table 1). Similar to the aforementioned studies, our results suggest that the effects of peer instruction are strongest for higher-level learning outcomes requiring comprehension and/or application, rather than recall alone (figure 3, table 5). While this body of research is relatively small, the available data provide consistent, albeit preliminary, practical guidance for instructors. If one perceives class time (and thus opportunities for active learning) as limited, or one is considering how best to employ active learning, it may be best to avoid using TPS for the most basic or "easy" concepts, or to practice recall. Instead, instructors should consider strategically designing TPS to target difficult concepts or the application of concepts. Do students routinely struggle with a particular concept or misconception? Perhaps this is the place to prioritize.

Limitations

Our study has several limitations that could influence our observed results. First, it is possible that the quality of instruction improved from 2015 to 2018. One instructor was added to the team after 2015, and perhaps the other instructors, who taught in both years, simply became better instructors over time. However, we doubt either possibility accounts for the observed effect size. Second, it is possible

that exam questions were easier in 2015 than 2018, although after inspection of questions, this is also unlikely. While these limitations could explain some of the differences observed for RQ 1 (figure 1), given that our results replicate findings from a large body of previous research on active learning, we think the above concerns are relatively minor. Additionally, our study design did not explicitly control for time on task between full and partial debrief conditions. The full debrief condition involved slightly more discussion time than the partial debrief condition. Rather than discussing incorrect answer choices during TPS, it is possible that simply spending 1-2 minutes longer engaging with lab content during each lecture, in any format, would produce the same results for RQ 2 (figures 2 and 3). Nevertheless, our results are consistent with prior research and theoretical predictions regarding the impacts of explanation and elaboration feedback (reviewed in Butler et al. 2013).

CONCLUSIONS

Our data suggest that active learning implemented as TPS with multiple-choice questions: (1) enhances student outcomes when implemented in large lectures in Computer Science; (2) can improve students' learning of concepts introduced outside of lectures, for example, during lab sections of a course; (3) is more impactful when focused on deeper comprehension and/or application of concepts, compared to simple recall; and (4) is more effective when instructors include feedback regarding why each answer choice they provide is correct or incorrect, rather than signaling the correct answer choice alone, especially when deeper comprehension and/or application of concepts is required.

Additionally, our approach illustrates that active learning can be short in duration, yet impactful on student outcomes, without significantly sacrificing lecture content. Instructors often exhibit resistance to adopting active learning, despite the large evidence base, sometimes due to the perception that using active learning necessitates teaching less content (Michael 2007; Miller and Metz 2014). We don't subscribe to the view that active learning, at least via the TPS method, requires an instructor to sacrifice content coverage and, by extension, student learning. Often, one can design active learning to be a one-to-one replacement of a short segment of lecture in STEM disciplines. For example, instead of lecturing about the results of a classic experiment, an instructor can challenge students to interpret a figure from the published paper themselves and connect it to course concepts. Instead of describing multiple examples to illustrate a concept for students, an instructor can challenge students to find, generate, or compare and contrast their own examples. Or, instead of lecturing on the application of a concept, students can actively discuss an application question provided by the instructor. All of these examples (and more) can occur in approximately the same amount of time it takes to deliver the lecture equivalent. Our TPS interventions used two multiple-choice questions (including up to five answer choices each) per lecture and averaged five minutes total per intervention, including those with a full debrief of all answer choices. In our opinion, this represents an extremely minimal time investment of class time in active learning, both within a lecture and across the entire course. And, our observed gains in student outcomes far outweighed any concerns regarding potential sacrifices of content coverage resulting from either the TPS exercises or the extra time (~1-2 minutes per session) required to debrief all of the answer choices. Nevertheless, we are sensitive to the question of how best to strategically target the design and implementation of active learning, given the perception of limited time during class sessions. We hope our results, and the research cited above, will help instructors make those difficult choices regarding how to implement active learning to maximize student learning and teaching efficiency within a lecture and course.

There are several second generation questions about how best to implement and debrief active learning, in general, and TPS more specifically, that we have not addressed here. Is there a dose-response effect for active learning? For example, are more instances of active learning per unit time better, or do positive impacts asymptote with increased frequency of active learning? One could investigate this question at two distinct timescales, within and across class sessions, both of which would be informative for instructors. Are open-ended questions, or questions with more than one possible answer, more or less effective than multiple-choice questions with a single, best answer? In addition to how one facilitates the debrief through active learning, gains from active learning may depend on the design of the questions used to provide low stakes practice and feedback. Would making the Think and/or Share portion of TPS a low-stakes, individual, written activity (e.g., articulating the rationale for correct and/or incorrect answers), collected by instructors, improve the results? Trumbo, Leiting, McDaniel, and Hodge (2016) found that low stakes retrieval practice leveraging the testing effect learning principle (Roediger & Karpicke 2006) was most effective when students were required to complete online practice quizzes, compared to when they were optional. Perhaps this incentivization applies to active learning during class sessions as well. Would explanation feedback be more impactful if generated by the students, rather than provided by the instructor? In our study, TPS leveraged the generation effect (Bertsch et al. 2007) for students during the Think and Pair phases of exercises. However, for efficiency and accuracy, instructors provided the explanation feedback for answer choices at the conclusion of the Share phase. Perhaps student outcomes would be even greater if all three phases of the TPS method intentionally leveraged the generation effect. We hope that future work will explore these questions and more.

Martin Barrett is an Associate Teaching Professor in the Institute of Software Research and Heinz College of Public Policy and Information Management at Carnegie Mellon University (USA). His teaching focuses Information Systems and Computer Science.

Chad Hershock is the Director of Faculty & Graduate Student Programs at the Eberly Center for Teaching Excellence & Educational Innovation at Carnegie Mellon University (USA). His work supports the adoption of evidenced-based pedagogy and SoTL.

Michael McCarthy is an Associate Teaching Professor in the Heinz College of Public Policy and Information Management at Carnegie Mellon University (USA). His teaching focuses Information Systems.

Michael Melville is a Data Science Research Associate at the Eberly Center for Teaching Excellence & Educational Innovation at Carnegie Mellon University (USA). His work supports the Scholarship of Teaching & Learning conducted by instructors.

Joe Mertz is a Teaching Professor in the Dietrich College of Humanities and Social Sciences and Heinz College of Public Policy and Information Management at Carnegie Mellon University (USA).

ETHICS

The human subjects research in this article was covered by the IRB approved Exempt protocols for STUDY2016_00000148 - Studying Impacts of Educational Interventions at Carnegie Mellon University.

REFERENCES

Ambrose, Susan A., Marsha C. Lovett, Michael W. Bridges, Michele DiPietro, and Marie K. Norman. 2010. *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco, CA: Jossey-Bass.

Bertsch, Sharon, Bryan J. Pesta, Richard Wiscott, and Michael A. McDaniel. 2007. "The Generation Effect: A Meta-Analytic View." *Memory & Cognition* 35, no. 2: 201–10. https://doi.org/10.3758/BF03193441.

Bruff, Derek. 2009. *Teaching with Classroom Response Systems*. San Francisco, CA: Jossey-Bass.

Butler, Andrew. C. 2010. "Repeated Testing Produces Superior Transfer of Learning Relative to Repeated Studying." *Journal of Experimental Psychology* 36, no. 5: 1118–33. https://doi.org/10.1037/a0019902.

Butler, Andrew C., Namrata Godbole, and Elizabeth J. Marsh. 2013. "Explanation Feedback is Better than Correct Answer Feedback for Promoting Transfer of Learning." *Journal of Educational Psychology* 105, no. 2: 290–98. https://doi.org/10.1037/a0031026290.

Cepeda, Nicholas J., Edward Vul, Doug Rohrer, John T. Wixted, and Harold Paschler. 2008. "Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention." *Psychological Science* 19, no. 11: 1095-102. https://doi.org/10.1111/j.1467-9280.2008.02209.x.

Crouch, Catherine H., and Eric Mazur. 2001. "Peer Instruction: Ten Years of Experience and Results." *American Journal of Physics* 69, no. 9: 970-77, Retrieved from https://doi.org/10.1119/1.1374249.

Eddy, Sarah L., and Kelly A. Hogan. 2014. "Getting under the Hood: How and for Whom Does Increasing Course Structure Work?" *CBE-Life Sciences Education* 13, no. 3: 453-68. https://doi.org/10.1187/cbe.14-03-0050.

Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences* 111, no. 23: 8410-15. https://doi.org/10.1073/pnas.1319030111.

Gavassa, Sat, Rocio Benabentos, Marcy Kravec, Timothy Collins, and Sarah Eddy. 2019. "Closing the Achievement Gap in a Large Introductory Course by Balancing Reduced In-Person Contact with Increased Course Structure." *CBE—Life Sciences Education* 18, no. 1. https://doi.org/10.1187/cbe.18-08-0153.

Goeden, Terrah J., Martha J. Kurtz, Ian J. Quitadamo, and Carin Thomas. 2015. "Community-Based Inquiry in Allied Health Biochemistry Promotes Equity by Improving Critical Thinking for Women and Showing Promise for Increasing Content Gains for Ethnic Minority Students." *Journal of Chemistry Education* 92, no. 5: 788-96. http://dx.doi.org/10.1021/ed400893f.

Haak, David C., Janneke HilleRisLambers, Emile Pitre, and Scott Freeman. 2011. "Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology." *Science* 332, no. 3: 1213-16. https://doi.org/10.1126/science.1204820.

Hattie, John, and Helen Timperley. 2007. "The Power of Feedback." *Review of Educational Research,* 77, no. 1: 81-112. https://doi.org/10.3102/003465430298487.

Kapler, Irina V., Tina Weston, and Melody Wiseheart. 2015. "Spacing in a Simulated Undergraduate Classroom: Long-Term Benefits for Factual and Higher-Level Learning." *Learning and Instruction,* 36: 38-45. https://doi.org/10.1016/j.learninstruc.2014.11.001.

Kothiyal, Aditi, Rwitajit Majumdar, Sahana Murthy, and Sridhar Iyer. 2013. "Effect of Think-Pair-Share in a Large CS1 Class: 83% Sustained Engagement." *Proceedings from ICER '13: The Ninth Annual International ACM Conference on International Computing Education Research*, *San Diego, CA, 12-13 August 2013*, ACM Press, New York, NY, 226-36. https://doi.org/10.1145/2493394.2493408.

Linton, Debra L., Jan Keith Farmer, and Ernie Peterson. 2014. "Is Peer Interaction Necessary for Optimal Active Learning?" *CBE—Life Sciences Education* 13, no. 2: 243-52. https://doi.org/10.1187/cbe.13-10-0201.

Lovett, Marsha, Oded Meyer, and Candace Thille. 2008. "The Open Learning Initiative: Measuring the Effectiveness of the OLI Statistics Course in Accelerating Student Learning." *Journal of Interactive Media in Education* 2008, no. 1. https://doi.org/10.5334/2008-14.

Lyman, Frank. 1987. "Think-Pair-Share: An Expanding Teaching Technique." *MAA-CIE Cooperative News* 1: 1-2.

Michael, Joel. 2007. "Faculty Perceptions about Barriers to Active Learning." *College Teaching* 55, no. 2: 42-47. https://doi.org/10.3200/CTCH.55.2.42-47.

Miller, Cynthia J., and Michael J. Metz. 2014. "A Comparison of Professional-Level Faculty and Student Perceptions of Active Learning: Its Current Use, Effectiveness, and Barriers." *Advances in Physiological Education* 38: 246–52. https://doi.org/10.1152/advan.00014.2014.

Pashler, Harold, Nicholas J. Cepeda, John T. Wixted, and Doug Rohrer. 2005. "When Does Feedback Facilitate Learning of Words?" *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31, 3-8. https://doi.org/10.1037/0278-7393.31.1.3.

Porter, Leo, Cynthia Bailey Lee, Beth Simon, and Daniel Zingaro. 2011. "Peer Instruction: Do Students Really Learn from Peer Discussion in Computing?" Proceedings from ICER '11: *The ACM SIGCSE 2011 International Computing Education Research Workshop, Providence, RI, 8-9 August 2011*, ACM Press, New York, NY, 45-52. https://doi.org/10.1145/2016911.2016923.

Prince, Michael. 2004. "Does Active Learning Work? A Review of the Research." *Journal of Engineering Education* 93, no. 3: 223-31. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x.

Radermacher, Alex D., and Gursimran S. Walia. 2011. "Investigating the Effective Implementation of Pair Programming: An Empirical Investigation." Proceedings from SIGCSE '11: *The 42nd ACM Technical Symposium on Computer Science Education, Dallas, TX, 9-11 March 2011*, ACM Press, New York, NY, 655-60. https://doi.org/10.1145/1953163.1953346.

Roediger, Henry L., and Jeffrey Karpicke. 2006. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science* 17, no. 3: 249-55. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Rohrer, Doug, and Kelli Taylor. 2006. "The Effects of Overlearning and Distributed Practise on the Retention of Mathematics Knowledge." *Applied Cognitive Psychology* 20, no. 9: 1209-24. https://doi.org/10.1002/acp.1266.

Schweitzer, Dino, Jeff Boleng, and Lauren Scharff. 2011. "Interactive Tools in the Graphics Classroom." Proceedings from ITiCSE '11: *The 16th Annual Joint Conference on Innovation and Technology in Computer Science Education, Darmstadt, Germany, 27-29 June, 2011*, ACM Press, New York, NY, 113-17. https://doi.org/10.1145/1999747.1999781.

Smith, Michelle K., William B. Wood, Wendy K. Adams, C. E. Wieman, Jennifer K. Knight, Nancy Guild, and Tin Tin Su. 2009. "Why Peer Discussion Improves Student Performance on In-Class Concept Questions." *Science* 323, no. 5910: 122-24. https://doi.org/10.1126/science.1165919.

Tinkle, Theresa, Daphna Atias, Ruth M. McAdams, and Cordelia Zukerman. 2013. "Teaching Close Reading Skills in a Large Lecture Course." *Pedagogy: Critical Approaches to Teaching Literature, Language, Composition, and Culture* 13, no. 3 505-35. https://doi.org/10.1215/15314200-2266432.

Trumbo, Michael C., Kari A. Leiting, Mark A. McDaniel, and Gordon K. Hodge. 2016. "Effects of Reinforcement on Test-Enhanced Learning in a Large, Diverse Introductory College Psychology Course." *Journal of Experimental Psychology: Applied* 22, no. 2: 148-60. https://doi.org/10.1037/xap0000082.

Webb, David J. 2017. "Concepts First: A Course with Improved Educational Outcomes and Parity for Underrepresented Students." *American Journal of Physics* 85, no. 8: 628-32. https://doi.org/10.1119/1.4991371.

Winkelmes, Mary-Ann, Matthew Bernacki, Jeffrey Butler, Michelle Zochowski, Jennifer Golanics, and Kathryn Harriss Weavil. 2016. "A Teaching Intervention that Increases Underserved College Students' Success." *Peer Review* 18, no. 1: 31-36.

Wormeli, Rick. 2004. *Summarization in Any Subject: 50 Techniques to Improve Student Learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Zheng, Jun, Sohee Kang, and Brian Harrington. 2019. "Immediate Feedback Collaborative Code Tracing." Proceedings from WCCCE '19: *The Western Canadian Conference on Computing Education, Calgary, Canada, 3-4 May 2019,* ACM Press, New York, NY, 12:1-12:2. https://doi.org/10.1145/3314994.3325087.