Examining the Effectiveness of a Support Model for Introductory Statistics

Seth Chart Melike Kara Felice Shore Sandy Spitzer
Towson University

This paper describes a pilot program, aimed at improving outcomes in Introductory Statistics, in which undergraduate peer coaches led teams of students in activities designed to address common misconceptions about statistics during weekly sessions. Preliminary analysis suggests that introducing these sessions may reduce the percentage of students that finish Introductory Statistics with a grade of D, F or W, although the small number of students in the pilot program did not provide sufficient power to detect statistical significance. We also observe that the population of students who attend most of the optional sessions seems to be a mixture of high performing students and lower performing students. Participants in the program reported mainly positive perceptions of the program's usefulness. We intend to continue investigating these observations in future iterations of the program where we hope to improve participation and refine the session activities.

Keywords: Introductory Statistics, Peer-Led Team Learning, Collaborative Learning, Statistics

An ongoing concern in undergraduate mathematics education is students' struggle in "gateway" mathematics courses, such as College Algebra or Basic Statistics. Research suggests that student success in these gateway courses strongly correlates with retention and degree completion (Adelman, 2006). Thus, these courses offer a significant opportunity to boost overall retention. In this paper, we describe one attempt to increase course success in such a gateway course: Introductory Statistics. The choice of Introductory Statistics was motivated by several factors. First, increasing numbers of students are taking an introductory Statistics course to satisfy either a major or general education requirement (Blair, Kirkman, & Maxwell, 2013). This increasing enrollment is in response to calls to increase access to college level courses (e.g. Treisman, 2015), and boosted by evidence that success in a first Statistics course does not require algebra-intensive preparation (Charles A. Dana Center, 2015). However, Introductory Statistics can only fulfill the vision of improving access to college mathematics if students master course material and successfully complete the course. At our own institution, historical failure rates (including students earning D, F, and W) range between 0-65% by section, with a mean failure rate of 24%. Because we offer a large number of sections (~ 20 per semester), reducing this failure rate (even by a modest amount) could meaningfully improve overall retention.

With the goal of improving student success in this course, we implemented a Peer-Led Team Learning (PLTL) model, in which students attend collaborative sessions to improve their prerequisite skills and conceptual understanding. Chemistry programs have used this model extensively (see, e.g., Grosser et al., 2008), and a recent meta-analysis indicated significant evidence that this model is effective across the STEM disciplines (Wilson & Varma-Nelson, 2016). Indeed, PLTL programs have demonstrated success in improving mathematics students' understanding of content (Merkel et al., 2015), course passing rates (Hooker, 2011), retention (Quitadamo et al., 2009), and attitudes toward mathematics (Curran, Carlson, & Celotta, 2013).

In this paper, we aim to describe our methodology, including a detailed description of program structure, in order that other institutions might learn from our successes and challenges. We also present data on the effectiveness of the program, focusing on three central questions:

A) Who attended the PLTL sessions?

B) What effect did the PLTL sessions have on course outcomes?
C) How did participants perceive the PLTL sessions?

In our presentation, we aim to begin a conversation, soliciting ideas from others about how the PLTL program might be improved, as well as how such complex interventions can be studied and replicated.

Methodology

The implementation and study of our PLTL program was informed in part by descriptions of similar programs that reported some success (e.g. Carlson et al., 2016). In order to test the effectiveness of the program, both in terms of course outcomes and participants' perceptions of the program, we used a quasi-experimental, control group design with matched pairs of course sections.

Participants and Measures

Study participants consisted of students enrolled in ten sections of introductory statistics, chosen among a total of 20 sections offered at a large, comprehensive, regional public university in the Mid-Atlantic. We selected five pairs of sections, such that a common instructor taught each pair of sections. From each pair of sections, we randomly selected a treatment section. Thus, each of the five treatment sections had a corresponding control section taught by the same instructor. Total enrollment was 134 students in the treatment sections and 136 students in the control section. To investigate the effectiveness of the program, we examined students' final course grades, as well as performance data (e.g. exam scores) provided to us by the course instructors.

In order to gauge students' thoughts about the recitation sessions (and answer Research Question C), we created an online survey and invited all treatment group students to participate in the survey. The survey was administered after the end of semester. The participation was voluntary and anonymous. Students were able to leave any question blank. The survey consisted of both open-ended and multiple-choice items students' perception of the sessions in terms of content, instructional techniques, and schedule. Students were also asked to give feedback for further improvements of the program. Further details about the survey are shared alongside the results.

Program Structure and Content

Students from the five treatment sections (each of which enrolled approximately 28 students) of Introductory Statistics were invited to attend weekly 2-hour collaborative problem-solving sessions, facilitated by paid peer coaches. A total of 72 students could be accommodated each week, choosing from among six different sessions, each with a different coach and at a different day and time. PLTL sessions began about one week prior to the halfway point of the semester and one or two weeks after students' first midterm exam was returned. The hope was for struggling students to attend voluntarily, encouraged by a modest grade incentive.

Coaches were undergraduate students who had received an "A" in the same course two semesters prior and who were hired based on an interview to ascertain their relatability and confidence to lead sessions. Coaches attended Friday training meetings at which they went through and provided feedback on the following week's session activities. Two of the authors, one a mathematician and one a mathematics educator, both faculty in the Department of Mathematics, devised the weekly session activities and ran the training meetings. The goal was to prepare the coaches to facilitate the students' work through the activities and especially to stimulate discussion of the focal ideas. Thus, besides becoming familiar with the specific activities, coaches were given lesson plans that included discussion-generating questions to help focus attention on the most crucial concepts.

The intent of the program was to deepen students' conceptual understanding of key ideas in statistical thinking outlined in the GAISE College Report (2016). Additionally, although the sessions were constructed so as not to feel like another "lecture class," our hope was for students to engage with those ideas through activities recommended by the report as much as possible. Finally, we wanted weekly topics to be timely relative to recently covered class material. That was accomplished to the extent possible, given that individual instructors have some flexibility in the sequence and timing of topics on the course outline. Ultimately, a schedule of 8 sessions was constructed to best meet the aforementioned goals. Brief descriptions of the content of each session follow.

Session 1: Histograms, standard deviations, and normality. Students created and examined histograms for six different large sample data sets, matching each to its respective population description based on shape. Students revisited standard deviation calculation, then determined the percentage of data falling within one, two, and three standard deviations of the mean.

Session 2: Normal distribution and the empirical rule. Students examined a uniform distribution as a comparison to the normal. They used the "empirical rule" on a contextualized normal distribution to identify scores associated with different relative frequencies and vice versa, and to compare normal curves on same axes, given means and standard deviations.

Session 3: Meaning of z-scores and associated probability statements for a random score from a normal population. Students developed meaning of a z-score and reasoned out the formula based on its meaning. They standardized scores from a contextualized distribution, locating them along the x-axis using the empirical rule as a reference. They determined probabilities associated with population values falling in various intervals and vice versa.

Session 4: Sampling distributions 1 – sample proportions. Students physically tossed coins to first predict, then examine the distribution of proportions of heads from samples of size five and then 10, comparing results. Next, they observed computer simulations for proportions of heads for samples of size 10, then 30, then 100, where dot plots for 1000 samples were generated. Discussion focused on effect of sample size on variability.

Session 5: Sampling distributions 2 – sample means. Students physically rolled dice to first predict, then examine the distribution of mean rolls from samples of size one (uniform distribution) and size three (more normal). Key features of the distributions were discussed. Next, they observed computer simulations for mean dice rolls for samples of size 10, then 30, then 100, where dot plots for 1000 samples were generated. Students reviewed z-scores in light of CLT and answered questions, relating questions and answers to confidence intervals.

Session 6: Confidence intervals and hypothesis tests on real data. Students gathered data on Hershey's Easter candy to check the reasonableness of company claims found on product packaging using confidence intervals and hypothesis tests on means and proportions.

Session 7: Connecting concepts and representations. Students played matching games and completed sorting tasks that had them reviewing big ideas of the course, particularly those central to the inferential techniques learned thus far.

Session 8: Correlation and linear regression. Students played matching games and used computer applets to combine their intuition and knowledge to answer questions about how two numerical variables were linearly related.

Results

Our results indicate that the PLTL intervention had somewhat complicated and mixed results; overall, the small number of participants limits our ability to detect statistically significant relationships. We present our results in terms of each of our initial research questions.

Research Question A: Who attended the PLTL Sessions?

We invited all 134 students enrolled in treatment sections to participate in the eight PLTL sessions. The number of students who chose to participate is presented in Table 1. Out of the 134 students enrolled in a treatment section, 40 (29.9%) participated in at least one session, 27 (20.1%) participated in three or more sessions, and 15 (11.2%) participated in six or more sessions. The percentage of students participating in one or more sessions ranged from 18.5% to 42.3% in treatment sections. Similarly, the percentage of students participating in three or more sessions ranged from 10.7% to 34.6%. Although Table 1 indicates that many students chose to participate in the PLTL sessions, we plan to make attempts in future semesters to both recruit students to attend a session and retain those students to additional sessions.

Table 1. Number of students participating in one or more and six or more sessions organized by instructor.

Instructor	Enrollment: Control Section	Enrollment: Treatment Section	Students attending one or more sessions	Students attending three or more sessions
1	24	28	6 (21.4%)	3 (10.7%)
2	29	26	9 (34.6%)	8 (30.8%)
3	27	27	9 (33.3%)	3 (11.1%)
4	28	27	5 (18.5%)	4 (14.8%)
5	28	26	11 (42.3%)	9 (34.6%)
Totals	136	134	40 (29.9%)	27 (20.1%)

We were also interested in the achievement levels of the students who chose to participate in the PLTL sessions. We recorded the overall GPA (for all collegiate courses that students had completed at the end of the semester they enrolled in Introductory Statistics) for students in the control group, and for those attending different number of PLTL sessions. These results are presented in Figure 1. Examining this figure, we can observe from the first graph that the entire treatment group and control group had similar distributions of GPA. In the second graph, we notice that the distribution of GPA for students attending three or more sessions is slightly bimodal, with higher frequencies of both students with overall grades in the A and C range. This bimodal tendency is even stronger for students attending six or more sessions.

We are encouraged by the fact that a mix of students of different overall achievement levels chose to participate in the PLTL sessions. In future semesters, our goal is to maintain a diversity of achievement levels. However, given our goal of reducing course failure (DFW) rates, it might be prudent to target recruitment toward students performing at the overall level of "C" or lower, and/or scoring below average on the first course exam.

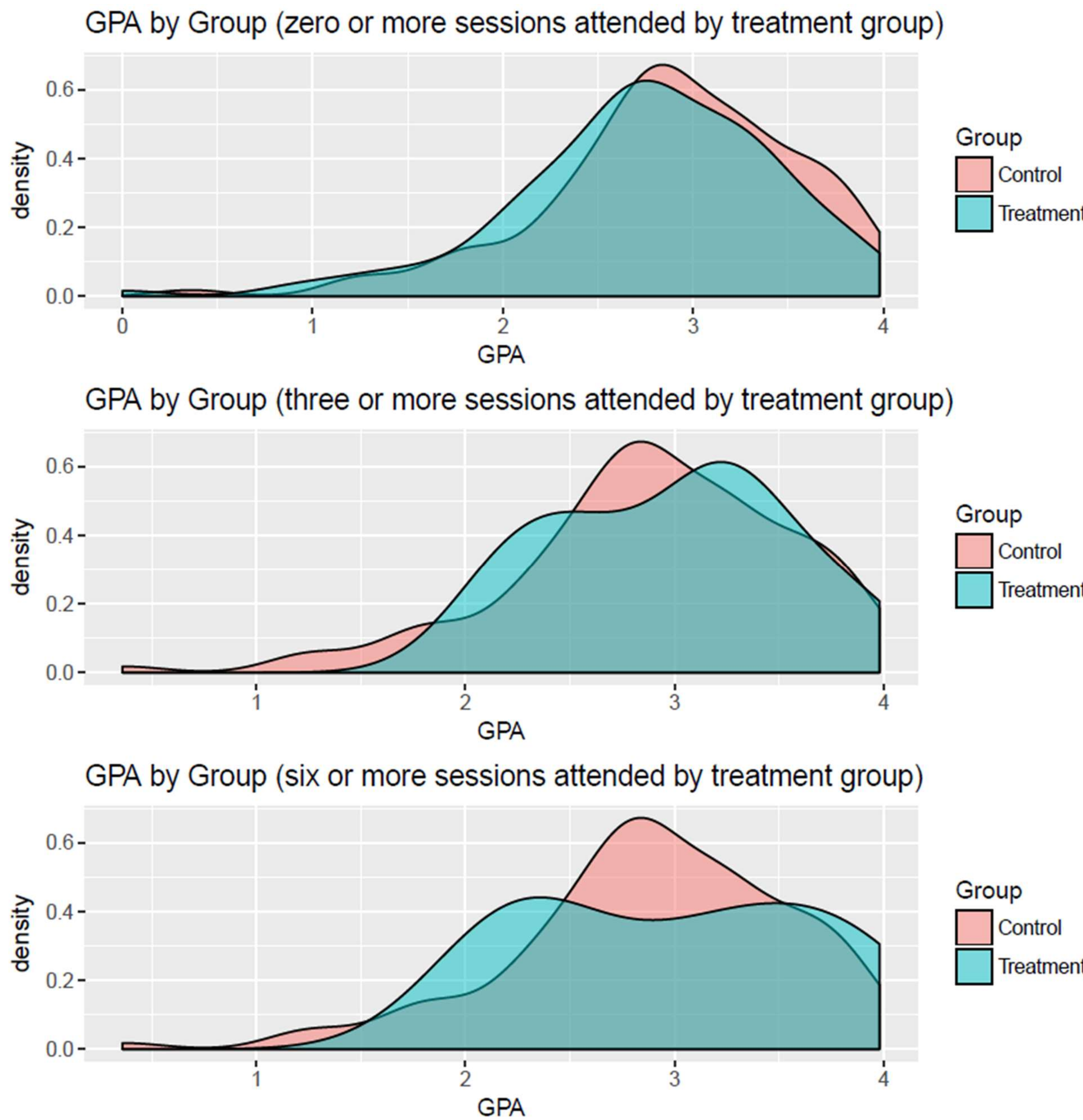

Figure 1. Density plots of overall college GPA for students in the control group and attending PLTL sessions.

Research Question B: What effect did the PLTL sessions have on course outcomes?

Because our primary goal in instituting the program was to increase student success rates, we began by analyzing its effects on students' course passage. We recorded the number of students obtaining a grade of D, F or W (that is, students not completing the course successfully) and compared the percentage of DFW's in the control group to students who participated in three or more sessions, and to students who participated in six or more sessions. The students who participated in three or more sessions were well matched to students in the control group based on first exam performance (Hedge's g=0.007). This indicates that, on average, students who chose to participate in three or more sessions and students in the control group did not perform differently on the first exam, which occurred before sessions began. We observe that the percentages of DFWs in the participating groups (32.5% and 33.3%) are approximately 5% lower than the percentage of DFWs in the control group (37.5%), but this difference is not statistically significant (p = 0.347 for students who attended one or more sessions, and p = 0.49 for students who participated in six or more sessions).

We also examined the effects of the intervention on course grade. Figure 2 presents a histogram displaying the course grades for the control group (in order to provide a comparison) and students who attended three or more PLTL sessions. The mean course grade (with A=4, B=3, etc, and W grades removed) was 2.20 (SD = 1.30) for the control group and 2.11 (SD = 1.16) for students attending three or more sessions; this difference was not significant. This finding, along with the data in Figure 2, demonstrates that attending three or more sessions did not appear to have a strong influence on students' success in the course.

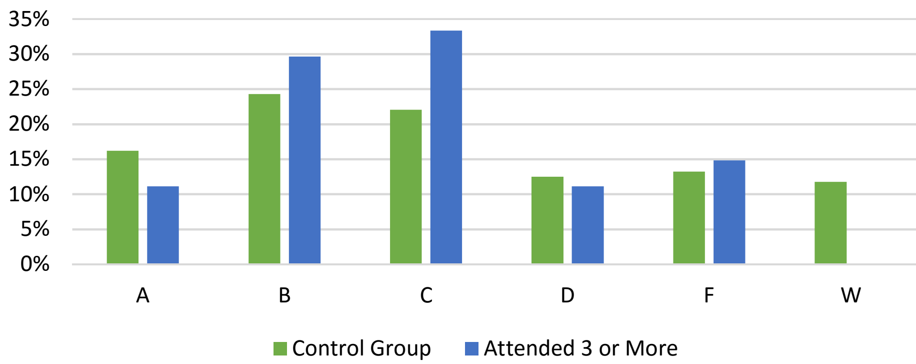

Figure 2. Course grades for control group and students attending three or more sessions.

While the observed data does not allow us to draw any firm conclusions about the effectiveness of the PLTL at lowering the DFW rate and increasing course success in introductory statistics, we plan to collect data during the Fall 2017 and Spring 2017 semesters. Program improvements might enhance the effects on student achievement, while increased participation may allow us additional power to observe such effects.

Research Question C: How did participants perceive the PLTL sessions?

All students who attended one or more PLTL sessions (N=40) were invited to complete a survey about their perception of the PLTL intervention; 20 students responded (a response rate of 50%). Table 2 shows the number of sessions attended by each of the 20 respondents. We asked why participants attended sessions and how useful they found the sessions to be. We further asked about the most helpful characteristics of the sessions and which topics they found most valuable. Finally, we asked a variety of questions to ascertain participants' perceptions of the coaches, including their helpfulness, preparedness, and enthusiasm.

Table 2. Number of sessions each of the 20 survey respondents attended

Number of sessions attended	1	2	3	4	5	6	7	8
Number of students responding	1	1	2	3	2	2	2	7

The most common reasons for attendance were "to earn the incentive" (nine students) and "to hopefully do better in the course by learning the content better" (nine students). When asked "If you didn't come to the sessions regularly or if you stopped coming, what was the reason?", six out of nine students chose “Session content did not match up with my class,” and three students stated that “Two hours for sessions was too long.”

For the most helpful and valuable characteristics of the sessions, common responses included “The sessions reviewed ideas and gave me practice on material I needed” (nine students) and “The sessions taught me something I hadn't understood before” (five students). According to the responses, coaches were viewed positively overall. On a 5-point scale that included “neutral,” 80% or more of the responding attendees agreed or strongly agreed that coaches explained concepts clearly, seemed knowledgeable, and were helpful, organized, enthusiastic, and encouraged discussion.

For each specific topic we covered (see Program Structure, above), we asked students: “How much did the session improve your understanding on the following topics?” 60% or more respondents considered sessions on the following topics at least somewhat helpful: standard deviation, z-score, probabilities associated with the normal curve, hypothesis testing, and meaning of distribution shapes. Eleven students (58%) found half or more of the sessions to be useful. Moreover, the majority of the students stated that going to the sessions was a positive experience and they would recommend these sessions to other students.

Responding attendees made several suggestions for improving the program. More than one student expressed the wish: 1) that sessions started earlier in the semester; 2) for more problems from the textbook and 3) for more hands-on activities. An unsolicited email came to us after the semester in which the student expressed that whereas the tutoring center was not a helpful support service for the course, these “extra classes” were. The student elaborated, “[Coach’s name] ran the classes I went to, and I got a good bit from them. I really got the Empirical Rule, and could estimate some answers just based on my understanding of it. Based on my experience and talking to the other students in [Coach’s name]'s group, I recommend the program.”

Discussion

In this pilot project, we were unable to document statistically significant changes to students’ success in Introductory Statistics. However, some data, such as the positive perception that students had of the program and the anecdotal finding that shows a small drop in DFW rates, motivate us to offer this program again. Looking forward, we aim to conduct further research on Peer-Led Team Learning structures by conducting a second iteration of this program in a future semester. We plan to make some important adjustments, including starting earlier in the semester and redesigning activities to make them more engaging. Our goal is to increase the program’s impact on student learning and course success. In addition, we aim to offer the intervention to more course sections and increase recruitment efforts. Additional participants will allow a more robust statistical analysis of the effect of sessions on student outcomes.

Implications for undergraduate mathematics education point to the importance and difficulty of creating scalable extra-curricular programs that can support student success in “gateway” mathematics courses. Using knowledgeable peers (as we did in our PLTL program) requires less faculty involvement, but introduces the possibility of misconceptions being passed from coaches to students. We also experienced tensions between our goal of deepening students’ conceptual understanding and students’ goals of improving procedural fluency in answering homework and exam questions, as well as difficulty in providing content that was applicable to students across different sections and instructors. Addressing these challenges will take collaboration and conversation across Mathematics Departments nationally; we aim to add to that conversation through our presentation and this paper.

References

Adelman, C. (2006). The Toolbox Revisited: Paths to Degree Completion from High School Through College. *US Department of Education*.

Blair, R., Kirkman, E. E., & Maxwell, J. W. [for the Conference Board of the Mathematical Sciences]. (2013). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2010 CBMS survey.* Washington, D.C.: American Mathematical Society. Retrieved July 30, 2017, from www.ams.org/profession/data/cbms-survey/cbms2010-Report.pdf.

Carlson, Kerri; Turvold Celotta, Dayius; Curran, Erin; Marcus, Mithra; and Loe, Melissa, Assessing the Impact of a Multi-Disciplinary Peer-Led-Team Learning Program on Undergraduate STEM Education, *Journal of University Teaching & Learning Practice*, 13(1), 2016.
Available at: http://ro.uow.edu.au/jutlp/vol13/iss1/5

Charles A. Dana Center at the University of Texas (2015). "A Call to Action to Expand Access to Statistics." Retrieved July 30, 2017 from https://dcmathpathways.org/resources/call-action-expand-access-statistics.

Curran, E. M., Carlson, K., & Celotta, D. L. T. (2013). Changing attitudes and facilitating understanding in the undergraduate statistics classroom: A collaborative learning approach. *Journal of the Scholarship of Teaching and Learning*, *13*(2), 49-71.

GAISE College Report ASA Revision Committee, "Guidelines for Assessment and Instruction in Statistics Education College Report 2016." Retrieved July 30, 2017 from http://www.amstat.org/education/gaise.

Gosser Jr, D. K., Kampmeier, J. A., & Varma-Nelson, P. (2010). Peer-led team learning: 2008 James Flack Norris award address. *Journal of Chemical Education*, *87*(4), 374-380.

Hooker, D. (2011). Small peer-led collaborative learning groups in developmental math classes at a tribal community college. *Multicultural Perspectives*, *13*(4), 220-226.

Merkel, J. C., & Brania, A. (2015). Assessment of Peer-Led Team Learning in Calculus I: A Five-year Study. *Innovative Higher Education*, *40*(5), 415-428.

Quitadamo, I. J., Brahler, C. J., & Crouch, G. J. (2009). Peer-led team learning: A prospective method for increasing critical thinking in undergraduate science courses. *Science Educator*, *18*(1), 29.

Treisman, U. (2015). "National trends in collegiate mathematics: The structural forces shaping our future." Retrieved July 30, 2017 from: http://www.utdanacenter.org/wp-content/uploads/national_trends_in_collegiate_mathematics.pdf.

Wilson, S. B., & Varma-Nelson, P. (2016). Small Groups, Significant Impact: A Review of Peer-Led Team Learning Research with Implications for STEM Education Researchers and Faculty. *Journal of Chemical Education*.