Using Simulation and Structured Group Work to Address Statistical Misconceptions

Scott Streiner
Department of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA, USA
scs42@pitt.edu

Mary Besterfield-Sacre
Department of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA, USA
mbsacre@pitt.edu

Sam Donovan
Department of Biological Sciences
University of Pittsburgh
Pittsburgh, PA, USA
sdonovan@pitt.edu

***Abstract*— There is significant interest in research regarding student understanding and performance, especially in probability and statistics. Past research has focused on misconceptions in statistical inference, but, there is little research regarding statistical misconceptions for undergraduate engineering students. Additionally, engineering educators recognize that active-learning strategies can improve undergraduate STEM education, but unfortunately intervention-based research on reducing statistical misconceptions is not prevalent. This research aims to address these literature gaps by employing a simulation-based structured group work activity whose goal was to increase awareness of and help students overcome misconceptions regarding the Central Limit Theorem (CLT). The CLT was chosen based on its abstract, non-intuitive nature, prevalence in the literature, and its foundational importance to the field of probability and statistics.**

Informed by the work of Schwartz and Bransford, this study draws on contrasting cases in conjunction with a simulation-based group assignment given to undergraduate industrial engineering students enrolled in an intermediate-level probability and statistics course at the University of Pittsburgh. Through this active-learning intervention, the following research questions are addressed: (1) How can active-learning strategies help students overcome misconceptions in statistics? (2) How do active-learning strategies affect the retention of statistical concepts across a curriculum?

Keywords—statistics; misconceptions; conceptual change; active learning; simulation

I. INTRODUCTION

The teaching and learning of probability and statistics (among other STEM fields) has been an area of research that has garnered much attention over the past two decades. This area continues to be an integral part of the engineering curriculum, with the ability to understand, interpret, and evaluate findings becoming an essential skill for the future engineering workforce [1]. For this reason, it's important for engineering students to understand some of the core tenants of the statistics field. There is significant interest in research regarding student understanding and performance in probability and statistics. Past research has focused on misconceptions, biases, and faulty heuristics in statistical inference, but, there is little research regarding statistical misconceptions for undergraduate engineering students. There are many misconceptions and faulty intuitions used by students that are stubborn and difficult to overcome, despite even the best statistics instruction [2]. Engineering educators recognize that active-learning strategies can improve undergraduate STEM education, but unfortunately intervention-based research on reducing statistical misconceptions is not prevalent [3].

There are many areas for which students have misconceptions regarding statistical inference, including sampling variability, the interpretation of the numerical value of the p-value, and confidence intervals. But this intervention research study focuses on the Central Limit Theorem (CLT). The CLT was chosen based on its abstract, non-intuitive nature, prevalence in the literature, and its foundational importance to the field of probability and statistics. A common pedagogical decision to clarify abstract and difficult concepts of statistics is to use computer simulation methods [1]. Many researchers have advocated the use of simulation methods to reinforce students' understanding of the concepts involving the CLT. However, there is a lack of research on the use of simple simulation assignments to augment and reduce misconceptions engineering students may have about the CLT. This research aims to address these literature gaps by employing a simulation-based structured group work activity whose goal is to increase awareness of and help students overcome misconceptions regarding the CLT.

Informed by the work of Schwartz and Bransford on conceptual change, this study draws on contrasting cases in conjunction with a simulation-based assignment given to undergraduate industrial engineering students enrolled in an intermediate-level probability and statistics course at the University of Pittsburgh. Students were tasked to work individually on personalized simulation assignments that forced them to think about the CLT and its implications in practice. After the initial assignment was completed, instructor-assigned groups of students shared simulation strategies, compared and contrasted results of their simulations, and discussed the significance of the differences they discovered. A short lecture on the CLT was given after the structured group work was completed.

A pre-and post- assessment was given to test students' understanding of the CLT. Retention was also tested by administering the same assessment on another cohort of students, some who've participated in the activity a year ago. Through this simulation-based intervention, the following research questions are addressed:

(1) How can simulation-based instruction reduce misconceptions about the Central Limit Theorem? What are the short-term and immediate learning gains for this type of pedagogy?

(2) How can simulation-based instruction affect the long-term retention of statistical concepts across a curriculum?

II. LITERATURE REVIEW

A. The Nature of Misconceptions and the Theory of Conceptual Change

Lecturing in university classrooms has become a mainstay of instruction for as long as universities were around [3]. Instructors can be astonished to learn that, despite their best efforts, students do not always grasp fundamental ideas covered in class. Many students have not developed an appropriate understanding of fundamental concepts from prior studies, and this shortcoming interferes with subsequent learning [4]. There are many types of 'misconceptions' to consider, including preconceived notions, nonscientific beliefs, conceptual misunderstandings, vernacular misconceptions, and factual misconceptions. Of particular importance in this study are conceptual misunderstandings. The constructivist view on learning suggests that learning may involve changing a person's conceptions rather than simply adding new knowledge to what was already there [5],[6]. However, this can be challenging. Research has shown that new concepts cannot be learned if alternative models already exist in the learners mind. Therefore, many misconceptions persist even when they are confronted by the most innovative forms of instruction [7]. Before a student can embrace and understand the new concept, they must confront their own beliefs and attempt to reconstruct this knowledge. This requires instructor to identify students' misconceptions, provide a platform for students to confront those misconceptions, and then help students reconstruct and internalize their knowledge, based on scientific empirical models [4]. This theory of conceptual change was leveraged to create an intervention that presented contrasting cases (via simulation) to help students develop more differentiated knowledge that would guide their subsequent learning during lectures regarding the CLT [8], [9] .

B. Statistical Misconceptions

Like many other STEM disciplines, research has been conducted to identify common faulty heuristics, biases, and misconceptions found in college students and adults and then training these individuals to reason more correctly [2]. A problem that arises in the encounter with a new statistical concepts is that of a change in the existing schema [10]. As such, many statistical misconceptions exist. Castro Sotos et al. reviewed the literature and identified several, as seen in Table 1. Many other researchers have sought to identify and remedy varying statistical misconceptions, including center and variability, hypothesis testing, p-values, data analysis, histograms, central limit theorem, and others [11]–[16]. However, the central limit theorem was chosen for this intervention due to instructor observation and foundational importance in the field of probability and statistics.

TABLE 1. Statistical Misconceptions [15]

Statistical Misconceptions	Topics
Sampling Distributions	Law of small numbers; different distributions; central limit theorem
Hypotheses Tests	Approaches to hypothesis testing; definition of hypotheses; significance levels; interpretation of the p-value; evaluation of statistical significance
Confidence Intervals	Plausible values of sample mean; range of scores; relative width of confidence interval

There have been a few tools developed to better understand misconceptions students have about statistics. One such tool is the Statistics Concept Inventory (SCI), developed by Kirk Allen et al. The SCI is a multiple choice instrument that assesses student understanding of fundamental statistics concepts [17], [18]. Shuman et. al have taken a different approach to understanding and improving learning in statistics with the use of model-eliciting activities (MEA). MEAs elicit a mathematical or conceptual system as part of its procedural requirements. To resolve an MEA, ,students need to make new connections, combinations, manipulations and predictions [19]. However, the most prevalent method for reducing misconceptions in statistics has been the utilization of computer simulation methods, all of which had promising results [1], [20]–[22]. This was the primary instructional activity chosen for this study, combined with the think, pair, share methodology. Think, Pair, Share is a cooperative learning technique that has students think about questions independently, then grouped together to discuss their thoughts, and finally share their ideas with the class. In addition to the theory of conceptual change, this part of the intervention is hypothesized to have impact on reducing statistical misconceptions and enhance learning by giving students opportunities to vocally reason through their ideas, internally process, organize, and retain those ideas [6]. The methodology of the intervention is described in detail below.

III. METHODOLOGY

This study was conducted at the University of Pittsburgh. The University's Institutional Review Board approved this study as exempt (IRB #PRO016020423). The study was also approved by the instructor of the course. The intervention was carried out over two semesters in two different academic years, covering two cohorts of sophomore industrial engineering students which we will label Cohort 1 and Cohort 2, respectively. The intervention was carried out across different groups of students in two different years for two reasons: refinement and retention. After completing the intervention on Cohort 1, we noticed areas of improvement in terms of assignment difficulty and clarity, as well as ways to improve the assessment. Additionally, we assessed Cohort 1 again one year later (the students now taking a simulation class) to see how much they've retained regarding the CLT. Figure 1 outlines the study process.

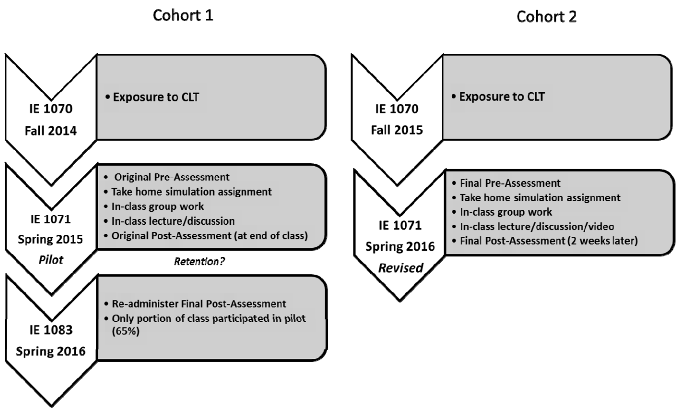

Figure 1. Intervention Study Process

The 69 students in Cohort 1 were given the intervention in Spring 2015 in an advanced undergraduate statistics course that covered topics such as regression, ANOVA, and design of experiments. The 66 students in Cohort 2 were given the same intervention in Spring 2016 in the same course. Because this study was carried out over two academic semesters, two versions of the assessment tool and the intervention were used. While the differences in the actual intervention were minor focusing mainly on improving the instructions, the original assessment tool differs from the final assessment tool by improving one question. The differences in these two versions, as well as the design of the intervention itself, will be covered more thoroughly below.

A. Design of Intervention

The CLT intervention can be broken down into three major components:

1) Individual simulation assignment
2) In-class group work
3) Lecture and discussion

Each component of the intervention is discussed in more detail below.

Individual Simulation Assignment: The first part of the intervention for both cohorts was to assign a simulation-based, take home assignment that had students confront their current misconceptions about the CLT. To reduce the misconceptions students had about the CLT, we gave the students the opportunity to generate and plot distributions of the individual observations and the sample means in Excel. Each student was given one of the following underlying distributions based on their last name: Exponential, Normal, Poisson, and Uniform. These distributions were chosen in part because the students were familiar with them from previous statistics courses and also have extensive use in their future courses. The assignment was broken into three parts. The first part had the students generate different number of random observations (10,100, and 500) from their respective distributions and plot histograms of each. The second part had the students generate samples of random observations in different sizes (e.g. 500 samples of sizes 5, 50, and 100) and plot the histograms of the sample means. The third part resembles the second, except now the students were told to change the number of samples, and keep the size of each sample the same (e.g. 10,100, and 500 samples of size 50). The purpose of the last two parts was for students to get an understanding about the effect of sample size and the number of samples in relation to the CLT and its applications. The students, therefore, had to produce 9 histograms, with each histogram having different parameters.

Because the students are not yet familiar with simulation techniques using Excel such as the inverse transform technique, instructions on how to generate random numbers using the Data Analysis Toolpak were given. The assignment can be viewed in its entirety at the authors request.

The students completed the assignment individually, and out of class. The students were given one week to complete the assignment, and to bring their results to class so that they could participate in the next part of the intervention: in-class group work.

In-class Group Work: The second part of the intervention was designed to give the students the opportunity to discuss the results of their individual simulation assignment with their peers. The main objective of this part of the intervention was for students to compare and contrast their results with others, with the main difference between students being the underlying distribution. It also helped reinforce the findings the students were already starting to recognize in their individual assignment, and a way for confusing phenomena to be clarified by their peers.

The classroom consists of 12 tables, with 6 students sitting at each table. Once the students arrived and sat at their tables (mostly with their friends), the students were instructed to have their simulation assignment results readily available. Each table was considered a group, and there was no group that all had the same underlying distribution. Each group was given a handout that they were to complete, and asked how and why the results for each section of the simulation assignment were either different or the same. The groups also discussed the implications of these differences and how it relates to the CLT. Finally, the groups discussed the following factors in relation to the CLT in preparation for the class discussion:

1) Underlying distribution
2) Sample means vs individual observations
3) Sample size
4) Number of samples
5) CLT importance and key takeaways

The group work activity took 30 minutes of class time. The group work activity sheet can be vied at the authors request. After the students completed this part of the intervention, we moved on to the lecture and discussion about the CLT.

Lecture and Discussion: Now that the students have a better understanding of the CLT through hands on experience with the phenomena, and have had a chance to discuss their thoughts with their peers, the third and last part of the intervention involved a short lecture about the CLT and a class discussion about the simulation assignment and the group work. The class first discussed, in part, what they have learned about the CLT so far, and what their key takeaways were. Other discussion points revolved around the 5 factors listed above. We allowed the students to drive the discussion and invited students to respond to each other (either agreeing or dissenting). When the discussion subsided, a lecture was given by one of the researchers (not the instructor of the course). In the first iteration of the intervention (Cohort 1), the researcher prepared a short lecture ahead of time that went over the CLT from a conceptual standpoint. However, the second iteration was shown a Khan Academy video about the CLT [23]. The lecture and discussion lasted roughly 20 minutes.

B. Assessment

Each cohort was given a pre-assessment and a post-assessment for conceptual understanding of the CLT. When the students were assessed and how they were assessed differ slightly among the two cohorts. We will first discuss the differences in the two versions of the assessment, and then describe the nature of the assessment.

The original as well as the final assessment tool asked three different types of questions. The first question was a visual question about the CLT, the second question asked about the distribution of individual observations from a distribution, the third question asked about the distribution of sample means from a distribution, and the fourth question was a conceptual question about the implications of the CLT. Both versions of the assessment were approved by the instructor of the course as being appropriate. The final assessment tool can be viewed at the authors request

The final assessment tool differed from the original in two ways. First, question four was changed from a T/F question, to a multiple choice question for a more nuanced understanding of student beliefs. The two versions of the question are presented in Table 2

TABLE 2. Assessment – Question 4 Versions

Original Assessment Question 4	**Final Assessment Question 4**
T/F: The Central Limit Theorem states that any distribution will approach a normal as long as the sample size is sufficiently large. Explain your answer.	The Central Limit Theorem implies that a. All variables have approximately bell-shaped sample distributions if a random sample contains at least 30 observations. b. Population distributions are normal whenever the population size is sufficiently large. c. For large random samples, the sampling distribution of the means is approximately normal, regardless of the shape of the population distribution d. The sampling distribution looks more like the population distribution as the sample size increases. e. All of the above.

Second, a fifth question was added (for the post-assessment only) to ask the students what part of the intervention was most helpful in understanding the CLT. The open-responses were later coded into four categories: Assignment, Group Work, Lecture/Discussion, and Video.

The two cohorts were also assessed at different times. We wanted to explore the short-term, intermediate, and long-term impacts of the intervention. Cohort 1 was not given question 4 on the pretest, but did answer the first three questions. Question 4 on the original assessment was instead given to an upper-level simulation class, to get a baseline for that particular question. However, question 4 was given to cohort 1 on the posttest. Cohort 1 was given the post-assessment immediately following the intervention, which tested for short-term impact. Cohort 1 was also given the post-assessment 1 year later (when a portion of the students were taking a simulation class), which tested for long-term impact and retention. It's worth noting that 65% the students in the simulation class participated in the intervention 1 year prior. Cohort 2 was given the post-assessment 2 weeks after the intervention, which tested for intermediate impact. Table 2 depicts the nature of the assessment of both cohorts.

TABLE 2. Time and Versions of Assessment

	Baseline Pre-Assessment	**Short-term Impact Assessment**	**Intermediate Impact Assessment**	**Long-term Impact Assessment**
Cohort 1	Original*	Original; immediately after	--	Final; one year later
Cohort 2	Final	--	Final; two weeks later	--

* Only given first three questions. Question 4 given to upper-level simulation class as a baseline

IV. RESULTS

This section outlines the results obtained from the intervention, broken down by cohort. Two types of analyses were constructed: scores on the pre-and post-assessments broken out by each question and improvement for each student broken out by each question. The pre-and post-assessment scores reflect the proportion of students who answered each question correctly, and the overall score is the average across all questions. Since we have paired data, we also explored how the intervention affected the marginal proportions of correct answers on each question using McNemar's Test [24]. The effect sizes were calculated for each question using Cramer's phi (ϕ), where any ϕ greater than 0.50 is considered a large effect size, between 0.30 and 0.50 a moderate effect size, and anything less than 0.30 a small effect size. This is statistical test used on paired nominal data to determine whether row and column marginal frequencies are equal. It's similar to a chi-squared test for homogeneity, but for paired data. For Cohort 1 only, a 2-proportion z-test was done to test if there was a significant difference in the proportion of students who got Question 4 correct (the T/F version) since this is the only question where we don't have paired data. The results obtained are mapped to areas of short-term, intermediate, and long-term impact on student learning regarding the CLT.

A. Cohort 1

TABLE 3. Scores for Cohort 1 by Percentage Correct

Cohort 1 Spring 2015	Pre-Assessment (n=69)	Post-Assessment (n=69)	McNemar's Test (ϕ)
Question 1	20%	94%	0.82
Question 2	25%	80%	0.66
Question 3	76%	96%	0.37
Question 4	6%**	74%	---
OVERALL	40%*	86%*	---
OVERALL (first 3 questions)	40%	89%	---

* Overall pretest calculated from 3 questions; overall posttest calculated from 4 questions
** Given to simulation class, not to statistics class

As seen in Table 3 and Table 4, the proportion of correct answers increased across all questions as a result of the intervention. The magnitudes of these gains differ considerably based on the type of question. The largest gains occurred on Question 1 (Visual Question) that had students analyze pictures of distributions and match the correct distribution to the sampling scheme given. Question 1 was most similar to their homework, as it provided visual clues, instead of just words. Question 3 (sampling distribution of the means) had the smallest gain and only moderate effect size. Many students got this question correct, and students tend to err on the side of a distribution being normal, if no other alternative seemed possible, which is in line with current CLT misconceptions. The learning gains for Question 2 were much higher, as this was a main learning objective of the individual assignment and subsequent class group work. The results also indicate that the proportion who got Question 4 correct after the intervention is significantly higher than the proportion of students who got the question correct in the non-intervention group ($z=8.33$, $p < 0.001$).

TABLE 4. Scores on Post-Assessment for Intervention Non-Participants

Cohort 1 and Non-Participants Spring 2016	Non-Participants (n=25)	Cohort 1 (n=46)
Question 1	16%	52%
Question 2	44%	67%
Question 3	72%	96%
Question 4	28%	43%
OVERALL	40%	65%
OVERALL(first 3 questions)	44%	72%

It's important to remember that Cohort 1 was assessed immediately after the intervention, which could be a reason for the strikingly high learning gains. The question becomes: are these learning gains retained? To answer this, Cohort 1 was re-assessed 1 year later in a simulation class. The results show a decline in scores across all questions (with the exception of Question 3). The scores obtained are still much higher than the original pre-assessment scores and more importantly, are higher than the students in the simulation class who did not participate in the intervention. In fact, the scores on the post-assessment for the non-participants are roughly the same as the pre-assessment scores for the participants. This indicates that the intervention has had an effect on the retention of the conceptual nature of the CLT, ruling out natural maturation as a potential explanation for higher scores.

So by utilizing contrasting cases via simulation, we reduce misconceptions about the CLT and improve understanding. There were both immediate and long-term learning gains. However, we wanted to refine the intervention and explore the robustness of these initial findings. To do this, the intervention was carried out one year later on another cohort of students. The results for this cohort are presented below.

B. Cohort 2

TABLE 5. Scores for Cohort 2 by Percentage Correct

Cohort 2 Spring 2016	Pre-Assessment (n=66)	Post-Assessment (n=66)	McNemar's Test ϕ
Question 1	18%	50%	0.44
Question 2	38%	67%	0.38
Question 3	89%	92%	0
Question 4	21%	41%	0.30
Overall	42%	64%	---
OVERALL (first 3 questions)	42%	70%	---

As mentioned previously, the main differences in the intervention were a change in Question 4 and when the cohort of students was assessed. First, the pre-assessment scores of Cohort 2 are in line with the pre-assessment scores of Cohort 1, and also the assessment scores of the non-participants. This provides a stable baseline level of knowledge and understanding about the CLT. After two weeks, there were significant gains across all of the questions (except Question 3), but are much lower than the gains from Cohort 1 when assessed immediately. But what's more striking are the results from the post-assessment are very similar to the post-assessment results from Cohort 1 after one year had passed. This suggests that there is a noticeable drop in learning gains after 2 weeks, but this drop stabilizes after one year. It also helps answer the question about the robustness of the intervention. The results from both assessments, even with the changes, are similar across cohorts.

The final post-assessment also included a question that asked the students what parts of the intervention were most helpful in understanding the CLT (see Figure 2). This was an open-ended question, but the question listed the four main elements of the intervention for students to comment on. Overall, the students found the individual simulation assignment the most helpful, and all of the other elements equally as helpful. This result is promising, as the design of the intervention was based on the theory of conceptual change using contrasting cases to reduce misconceptions. The individual simulation assignment leveraged this theory.

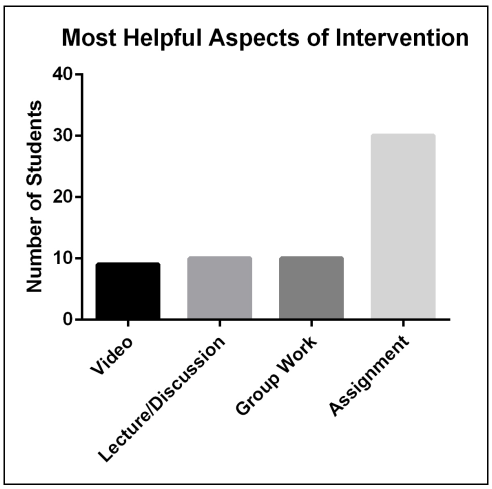

Figure 2. Most Helpful Aspects of Intervention

V. IMPLICATIONS

A. Impact on Learning: Short-Term, Intermediate, and Long-Term Outcomes

There were significant increases in short-term, intermediate, and long-term learning outcomes. When the students were assessed immediately after the intervention, the learning gains were quite high and varied amongst different question types. These outcomes were consistent across different groups of students and in different classes (see Figure 4). To reinforce the learning gains, each cohort had similar baseline levels of knowledge about the CLT. There was a drop in learning gains after 2 weeks (which was expected), but what was a bit surprising was that these learning gains remained stable even after 1 year. The majority of students indicated that the individual simulation assignment was the most helpful part of the intervention. The study also demonstrated the robustness of such an intervention.

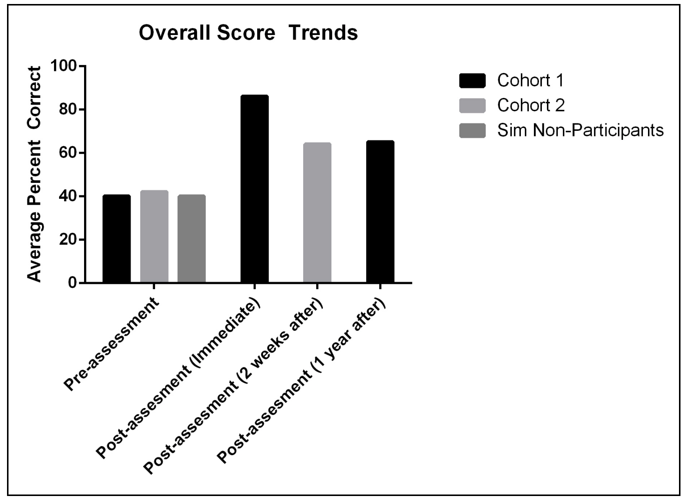

Figure 3. Overall Assessment Trends

This study adds to the increasing amount of evidence that active-learning strategies improve student achievement. Much of the research on instructional strategies has focused on entire classroom changes or larger strategic visions for how the course should be taught in general. There are many cases where STEM instructors, and in this case, statistics instructors that want to target a specific student misconceptions, without changing the nature in which they teach the class, for better or for worse. This study demonstrates an easily implementable active-learning intervention, based on the theory of conceptual change, which enhances how a particular foundational concept is taught. The results from this study should encourage statistics instructors to implement more active-learning strategies throughout the probability and statistics curricula. More generally, adopting pedagogical strategies that give students a chance to get hands-on experience with data and how conceptually abstract phenomena operate in practice can be helpful in the many other statistical areas where misconceptions are present. While this study outlines a way for misconceptions about the CLT to be reduced, we imagine that interventions grounded in similar theories of action can reduce misconceptions about other important areas of probability of statistics.

B. Limitations and Future Directions

The results of this intervention study on reducing statistical misconceptions are promising, but there are various limitations that must be stated. First, while the learning gains for the CLT are evident, the overall scores on the assessment after 2 weeks and 1 year are still quite low. Roughly one-third of the students in both cohorts still have misconceptions about the CLT. More needs to be done to reinforce the conceptualization and application of the CLT in practice over the curriculum. Second, the assessment instrument only consisted of four questions that all required either the interpretation of the CLT or the application of it. It is difficult to conclude from a four question assessment whether a student truly understands the CLT. Moreover, this study would have benefited from a more in-depth understanding of exactly how the students conceptualized the CLT. This would require talking to the students who did poorly on the assessment to better understand their prior knowledge and mental models of this phenomena. Lastly, this study utilizes the theory developed by Schwartz and Branford, but doesn't follow their analytic process exactly. Future work could be to look at the sequence of instructional activities, similar to the work conducted in Time for Telling [25]. What if the lecture on the CLT occurred before the simulation assignment? How does the simulation assignment compare to other methods of learning out of the classroom (problem sets, reading, etc.).

Reducing misconceptions about the CLT was a good starting point to fill in gaps between research and practice of teaching and learning statistics. However, as noted previously, many misconceptions about probability and statistical inference still exist. Future work will consist of developing a portfolio of easily implementable interventions to help students understand the more difficult concepts within the discipline, including the interpretation of the p-value, the conditional nature of significance levels, the difference between population and sampling distributions, and the understanding and application of different distributions in practice.

ACKNOWLEDGMENT

This study was approved by the University Institutional Review Board ((IRB #PRO016020423). Many thanks for the Center for the Integration of Research, Teaching, and Learning (CIRTL), as well as the feedback from the local Pitt-CIRTL chapter.

REFERENCES

[1] J. D. Mills, "Using computer simulation methods to teach statistics : A Review of the Literature," *J. Stat. Educ.*, vol. 10, no. 1, 2002.

[2] J. Garfield and D. Ben-Zvi, "How students learn statistics revisited : a current review of research on teaching and learning statistics," *Int. Stastical Rev.*, vol. 75, no. 3, pp. 372–396, 2007.

[3] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, "Active learning increases student performance in science, engineering, and mathematics.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 23, pp. 8410–8415, Jun. 2014.

[4] C. Moore and I. Abella, "Overcoming misconceptions : misconceptions as barriers to understanding science," in *Science Teaching Reconsidered: A Handbook*, Washington, D.C.: National Academies Press., 1997.

[5] P. W. Hewson, "The role of conceptual conflict in conceptual change and the design of science instruction," *Instr. Sci.*, vol. 13, pp. 1–13, 1984.

[6] S. Ambrose, M. Bridges, M. DiPetrio, M. Lovett, and M. Norman, *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley and Sons, 2010.

[7] M. T. H. Chi and R. D. Roscoe, "The processes and challenges of conceptual change," *Reconsidering Conceptual Change. Issues in Theory and Practice*. pp. 3–27, 2002.

[8] D. L. Schwartz, J. D. Bransford, and D. Sears, "Effiency and innovation in transfer," in *Transfer of learning from a modern multidisciplinary perspective*, no. 3, 2005, pp. 1–51.

[9] D. L. Schwartz, C. C. Chase, M. a. Oppezzo, and D. B. Chin, "Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer.," *J. Educ. Psychol.*, vol. 103, no. 4, pp. 759–775, 2011.

[10] Z. R. Mevarech, "A deep structure model of students' statistical misconceptions," *Educ. Stud. Math.*, vol. 14, no. 4, pp. 139–146, 1983.

[11] L. L. Cooper and F. S. Shore, "Students ' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots," *J. Stat. Educ.*, vol. 16, no. 2, pp. 324–326, 2008.

[12] S. Krishnan and N. Idris, "Students ' misconceptions about hypothesis test," *REDIMAT*, vol. 3, no. 3, pp. 276–293, 2014.

[13] S. Goodman, "A dirty dozen: twelve p-value misconceptions.," *Semin. Hematol.*, vol. 45, no. 3, pp. 135–40, Jul. 2008.

[14] C. L. Aberson, D. E. Berger, M. R. Healy, D. J. Kyle, and V. L. Romero, "Evaluation of an interactive tutorial for teaching the central limit theorem," *Teach. Psychol.*, vol. 27, no. 4, pp. 289–291, 2000.

[15] A. E. Castro Sotos, S. Vanhoof, W. Van den Noortgate, and P. Onghena, "Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education," *Educ. Res. Rev.*, vol. 2, no. 2, pp. 98–113, Jan. 2007.

[16] H. J. Motulsky, "Commentary common misconceptions about data analysis and statistics," *J. Pharmacol. Exp. Ther.*, vol. 351, pp. 200–205, 2014.

[17] K. Allen, "The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics," *Diss. Univ. Oklahoma*, 2006.

[18] K. Allen, T. Reed-Rhoads, R. Terry, T. Murphy, and a Stone, "Coefficient alpha: An engineer's interpretation of test reliability," *J. Eng. Educ.*, no. 1937, pp. 87–95, 2008.

[19] L. J. Shuman, R. Clark, M. Besterfield-sacre, and T. P. Yildirim, "The Model Eliciting Activity (MEA) Construct: Moving engineering education research into the classroom," in *Proceedings of the 9th Biennial ASME Conference on Engineering Systems Design and Analysis*, 2008, pp. 1–9.

[20] C. Robert, J. Garfield, and B. L. Chance, "A model of classroom research in action : Developing simulation activities to improve students ' statistical reasoning," *J. Stat. Educ.*, vol. 7, no. 3, 1999.

[21] D. M. Lane and Zhihua Tang, "Effectiveness of simulation training on transfer of statistical concepts," *J. Educ. Comput. Res.*, vol. 22, no. 4, pp. 383–396, 2000.

[22] J. D. Mills, "Learning abstract statistics concepts using simulation," *Educ. Res. Q.*, vol. 28, no. 4, pp. 18–33, 2004.

[23] S. Khan, "Central Limit Theorem," *Khan Academy*, 2016. [Online]. Available: https://www.khanacademy.org/math/probability/statistics-inferential/sampling_distribution/v/central-limit-theorem. [Accessed: 15-Feb-2015].

[24] A. Agresti, *Categorical Data Analysis*. 2013, 3rd edition, Hoboken, NJ, USA: John Wiley & Sons.

[25] D. L. Schwartz and J. D. Bransford, "A time for telling," *Cogn. Instr.*, vol. 16, no. 4, pp. 475–522, 1998.