STUDENTS' CONCEPTUAL UNDERSTANDING OF THE RELATIONSHIP BETWEEN STUDY DESIGN AND CONCLUSIONS IN AN INTRODUCTORY STATISTICS COURSE

Elizabeth Brondos Fry
University of Minnesota
fryxx069@umn.edu

Recommended learning goals for introductory statistics courses include understanding the role of random sampling in allowing for generalizations to a population, and the role of random assignment in enabling causal claims. However, these can be difficult topics for students to learn, sometimes resulting in confusion between the purposes of random sampling and random assignment. A two-and-a-half week unit on study design was created and implemented in an undergraduate introductory statistics course. A brief description of the unit is presented, along with a selection of results from assessments. Although results showed overall improvement, some difficulties remained, such as misconceptions related to sample size and some lingering confusion between random sampling and random assignment.

BACKGROUND AND MOTIVATION

Learning about statistical inference in an introductory statistics course often includes going beyond the data at hand to make a wider conclusion, which involves consideration of study design. According to guidelines for what students should learn in introductory statistics courses (e.g., GAISE, 2016), students should be able to recognize and explain the role of randomness in study design and conclusions that can be made. This includes understanding why random sampling allows the results of statistical studies to be extended to the population from which the sample was generated, and understanding why random assignment allows cause-and-effect conclusions to be made from comparative experiments.

However, these concepts can be difficult for introductory statistics students. For example, students have shown confusion between the distinct roles that random sampling and random assignment each play in the design of studies (Derry et al., 2000). Students can also have misconceptions about study design topics, such as preferring a large volunteer sample over a smaller random sample, or preferring systematic assignment over random assignment (Wagler & Wagler, 2013). Students sometimes express disbelief in the tendency of random assignment to balance out confounding variables between groups (Sawilowsky, 2004).

METHODS

A study design unit for an undergraduate introductory statistics course was created and implemented in order to build students' conceptual understanding and address potential misconceptions. The development of this unit was guided by recommendations from cognition literature on conceptual change, such as addressing students' existing conceptions and engaging them in activities that allow them to gradually build their conceptual understanding (e.g., Vosniadou, 2013). The research question addressed in this study is: *How does introductory statistics students' conceptual understanding of study design and conclusions (in particular, unbiased estimation and establishing causation) change after participating in a learning intervention designed to promote conceptual change in these areas?*

Course and Audience

The curriculum was implemented in four sections of an undergraduate introductory statistics course that was taught using simulation-based inference methods. All four sections were taught by advanced graduate students in a statistics education PhD program who were experienced instructors. Three of the course sections met twice a week for 75 minutes each, and the fourth section was taught entirely online. Students came from a variety of different majors, and most took the course in order to fulfill a basic mathematical thinking requirement. In this course, content was presented mostly through activities using technology, and the study design unit was created to be consistent with this method of teaching that relied heavily on active learning.

Unit and Assessments

The unit occurred during weeks 10-12 of a 16-week semester. Activities were either developed or modified from previous course activities (Zieffler et al., 2013). In the first activity, students were introduced to sampling methods and unbiased estimation by contrasting estimates from biased samples with estimates from random samples. In the second activity, students were introduced to methods of assignment to groups in an experiment by contrasting a purposeful assignment to groups with random assignment to groups, and observing the tendency of random assignment to balance out confounding variables. The third activity involved students carrying out a randomization hypothesis test in the context of an observational study and reasoning about conclusions that could or could not be made. After these three activities, students took a group quiz about study design concepts. The final activity in the unit involved students carrying out both random sampling and random assignment in a context where both study designs were possible, and distinguishing between the purposes and implications of each of these two study designs.

A 22-item forced-choice assessment, the *Inferences from Design Assessment* (IDEA) was developed to use as a pretest and posttest. The items were originally taken or modified from assessments that have been used previously in statistics education (e.g., CAOS, delMas et al., 2007; ARTIST, Garfield et al., 2002). The IDEA was then modified based on feedback from three external reviewers. Students took IDEA as a pretest just before the unit began and again as a posttest just after the unit concluded. A group quiz and lab assignment were also created as short-answer assessments to gather qualitative data.

RESULTS

A total of n = 125 students took the IDEA pretest and posttest. There was significant and substantial improvement in total score from pretest to posttest (t = 12.57, p < .001, Cohen's d = 1.12). For all but one of the 22 items, at least 60% of students answered the item correctly on the posttest. There were nine items with statistically significant improvement, even after adjusting for multiple comparisons (adjusted α = .002). The remaining items did not show statistically significant changes.

Areas of most improvement

Of the nine items with significant gains from pretest to posttest, four items had an increase of approximately 40% or more in percentage correct from pretest to posttest. The measured learning outcomes and results from these items are shown in Table 1. Two of these items had learning goals related to sampling and generalization, and the other two related to assignment and causation.

Table 1. IDEA items with the largest improvement from pretest to posttest

Item	Measured Learning Outcome: Ability to understand…	n	% of Students Correct: Pretest	% of Students Correct: Posttest	McNemar's test p
3	What it means to make an appropriate generalization	125	23.2	63.2	<.0001
6	Small random sample is preferable to larger, biased sample	125	46.4	85.6	<.0001
16	Correlation does not imply causation.	125	28.0	77.6	<.0001
18	The purpose of random assignment in an experiment	125	32.0	77.6	<.0001

Two of the IDEA items with the most improvement had been modified from items on the Comprehensive Assessment of Outcomes in Statistics (CAOS; delMas, Garfield, Ooms, & Chance, 2007) test. Performance on these IDEA items was compared to performance on similar CAOS items for a large national sample of students across different introductory statistics curricula (delMas et al., 2007), and for two samples of introductory statistics students at a small college in a randomization-based curriculum and in a consensus curriculum (Tintle, Topliff, VanderStoep, Holmes, & Swanson, 2012). (Note that the IDEA was taken just before and just after the unit, while the CAOS test was taken at the beginning and at the end of an entire course for the other samples.) The comparisons are seen in Table 2. For the learning outcome related to understanding that correlation does not imply causation, students taking the IDEA test performed noticeably worse on the pretest, but better on the posttest. The item related to understanding the purpose of random assignment was one of the most difficult CAOS items on both pretest and posttest for the delMas et al. and Tintle et al. samples, but by the end of the study design curriculum, more than three-quarters of students who took the IDEA answered this item correctly.

Table 2. Comparison of performance on IDEA items to performance on similar CAOS items

	Learning Outcome: Understand that correlation does not imply causation		Learning Outcome: Understand the purpose of random assignment in an experiment	
	% of Students Correct		% of Students Correct	
Sample	Pretest	Posttest	Pretest	Posttest
delMas et al. (2007)	54.6	52.6	8.5	12.3
Tintle et al. (2012) Randomization-based curriculum	47.4	59.2	1.3	18.4
Tintle et al. (2012) Consensus curriculum	57.7	62.8	7.7	14.1
IDEA	28.0	77.6	32.0	77.6

Areas of remaining difficulties

Despite the fact that the curriculum emphasized distinguishing between random sampling and random assignment, a small, but noticeable, portion of students continued to show confusion between the two types of study design. For example, qualitative data analysis of the lab assignment revealed that over 10% of students brought up only random assignment, not random sampling, when asked about whether or not they could use the results of a study to generalize about a population. Also, several items on the IDEA test had incorrect answer options indicating a misunderstanding related to confusing random sampling with random assignment. For three of these items, on the posttest, more than 10% of students chose the incorrect option related to that confusion. The results for these three items are shown in Table 3. However, on the posttest, less than 10% of students chose two of these three incorrect options simultaneously, and 0% chose all three incorrect options simultaneously.

Table 3. A selection of misconceptions or misunderstandings shown on IDEA test

		% choosing incorrect option (n = 125)		
Item	Misconception or misunderstanding	Pretest	Posttest	McNemar's test p
9	Cannot generalize due to lack of random *assignment*	9.6	23.2	.0046
16	Random *selection* enables causation	24.8	12.0	.0090
18	Purpose of random assignment: To ensure participants are *representative* of population	40.0	14.4	<.0001

Other difficulties understanding study design topics were also explored. For example, on some IDEA items, a noticeable portion of students tended to choose incorrect answer options indicating that sample size was the most relevant factor for making generalizations or causal claims, rather than choosing the correct option related to the design of the study. However, the tendency to over-emphasize sample size rather than sampling method declined significantly from pretest to posttest. Another problem observed was that on the group quiz, students' answers revealed difficulty recognizing whether or not statements were making a generalization, and whether or not statements were making a causal claim. More details on the methods and results of this study can be found in Fry (2017).

DISCUSSION

Overall, there was evidence of learning gains in concepts of study design and conclusions after the study design unit. The effect size for the change in IDEA score was just over one standard deviation, and for all but one item, the majority of students answered correctly on the posttest. Results suggest that a unit built on active learning and targeting potential misconceptions can help students learn about the purposes of random sampling and random assignment, and distinguish between the two. However, results also suggest that this learning can take time, because even after the unit, a small but noticeable portion of students displayed lingering confusion between random sampling and random assignment. It is also important to note that the instructors teaching the curriculum were highly experienced teachers and were accustomed to using active learning methods and technology in the course curriculum. It is therefore unclear what aspects of the unit (e.g., content, active learning, technology, instructor scaffolding) were most helpful for building students' understanding of study design and conclusions.

One major limitation of this study is that the study design curriculum was implemented in all sections of the introductory statistics course. Because all sections of the course typically share a common curriculum, it was not possible to teach different curricula to different sections. Therefore, no comparisons were made to other possible units that teach study design. Future research could compare different curricula, such as varying the place in which study design topics are taught in the curriculum, or varying the amount of time spent on these topics. Another limitation is that due to time constrictions, the pretest was given just before the unit and the posttest was given just after the unit. Future research could examine students' understanding of study design topics at the beginning and end of an academic term, or their retention of conceptual understanding even after they leave the course.

REFERENCES

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28–58.

Derry, S. J., Levin, J. R., Osana, H. P., Jones, M. S., & Peterson, M. (2000). Fostering Students' Statistical and Scientific Thinking: Lessons Learned from an Innovative College Course. *American Educational Research Journal*, *37*(3), 747–773.

Fry, E. (2017). *Introductory Statistics Students' Conceptual Understanding of Study Design and Conclusions* (Doctoral dissertation). http://iase-web.org/Publications.php?p=Dissertations

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

Sawilowsky, S. S. (2004). Teaching random assignment: Do you believe it works? *Journal of Modern Applied Statistical Methods, 3*(1), 221-226.

Tintle, N., Topliff, K., VanderStoep, J., Holmes, V. L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, *11*(1), 21–40.

Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. In Vosniadou, S. (Ed.), *International handbook of research on conceptual change* (2nd ed., pp. 11-30). New York: Taylor & Francis.

Wagler, A. & Wagler, R. Randomizing roaches: Exploring the "bugs" of randomization in experimental design. *Teaching Statistics, 36*(1), 13-20.

Zieffler, A., & Catalysts for Change. (2013*). Statistical Thinking: A simulation approach to uncertainty* (2nd ed.). Minneapolis, MN: Catalyst Press.