Abstract

Simulations have played an increasing role in introductory statistics courses, as both content and pedagogy. Empirical research and theoretical arguments generally support the benefits of learning statistical inference with simulations, particularly in place of traditional formula-based methods with which introductory students typically struggle. However, the desired learning benefits of simulations have not been consistently observed in all circumstances. Moreover, students in introductory courses have exhibited several types of misconceptions specific to simulations. One theme common to several of these misconceptions is conflating the hypothetical nature of simulations with the real world. These misconceptions, however, have only been discussed in the context of null-hypothesis significance testing (NHST), typically with a randomization-test simulation. Misconceptions about bootstrapping for the purposes of statistical estimation, a common component of simulation-based curricula, have remained unexplored.

The purpose of this study was to explore introductory statistics students' real-world interpretations of hypothetical simulations. The research questions driving this study were the following: (1) *To what extent are there quantitative differences in student understanding of the hypothetical nature of simulations when working with null-hypothesis significance testing vs. estimation?* and (2) *What typical themes emerge that indicate students are conflating the hypothetical nature of simulations with the real world?* The Simulation Understanding in Statistical Inference and Estimation (SUSIE) instrument was created to evaluate student interpretations about the properties of simulations throughout the entire statistical analysis process. The final instrument consisted of eighteen constructed-response items interspersed throughout descriptions of two different statistical research contexts. One context presented the randomization test for the purpose of NHST, and the second context presented bootstrapping for the purpose of statistical estimation. The instrument was developed, piloted, and updated over eight months and then administered to 193 introductory statistics students from one of two simulation-based curricula. Responses to the instrument were quantitatively scored for accuracy and qualitatively classified for clear examples of conflating the hypothetical nature of simulations with the real world. Quantitative scores were analyzed with descriptive statistics, inferential statistics, and several linear models. Qualitative classifications were analyzed by identifying the primary themes emerging from the responses to each item.

Results from the quantitative analysis suggest that there was no meaningful difference in the aggregate performance between interpreting the randomization simulation vs. the bootstrap simulation (average within-participant instrument section score difference = 0 points, 95% CI: -0.3 to 0.2 points, out of a possible 18 points). However, there was evidence of some differences in performance in parallel items between the NHST and estimation instrument sections. This indicates that participants inconsistently struggled with correctly interpreting the randomization test for NHST vs. bootstrapping for estimation, across the steps of a statistical analysis. Moreover, performance on the instrument overall (average total score = 9.0 points, SD = 3.7 points) and on a per-item basis indicates that several topics were difficult for participants. Bolstering this outcome, results from the qualitative analysis indicate that the participants held a large variety of misconceptions about simulations that pertain to real-world properties and that these misconceptions may vary by the type of simulation used. This includes thinking that simulations can improve several real-world aspects of studies, can increase the sample size of a study after the data are collected, and are a sufficient replacement for a real-world process such as study replication. Implications from these results suggest the need for real-world conflations to be better addressed in the classroom, a clearer framework to define conflations, and new assessment to efficiently identify the prevalence of conflations and how they emerge when learning statistics with simulations.