With scholars decrying a bastardization of the classical statistical testing procedure that has promulgated throughout common practice, many are turning away from statistical tests and their misunderstood null hypotheses and *p*-values. The issue at the heart of this matter is that the logic of statistical tests is confusing, the manner in which statisticians formulate their expectations as a probability distribution is poorly understood (e.g., Nickerson, 2000), and more generally, reasoning about distributions is difficult.

Over the years, statistics educators have tried many ways to improve their teaching and communication of statistical theory and practice. Starting in the 1980s, they began to utilize simulation. The statistics education community generally believed that this pedagogy was more apt to developing students' conceptual understanding of statistics, by placing the logic of statistical inference at the core of instruction and eschewing units on probability that was challenging for many students.

Considering the case of graduate students, as researchers applying statistical methods or practitioners interpreting statistical results, it's clear that they need to understand the logic of statistical inference. As science is fundamentally about theory generation and theory testing, it's also clear that they need to be fluent in at least one method of statistical testing. Might SBI curricula be able to support graduate students' development of an understanding of the core logic of statistical testing? How do graduate students, who have completed an SBI course, think while conducting statistical tests?

To answer these questions, a multi-model multiple descriptive case study of graduate students was conducted. Six graduate students in the educational sciences were recruited to complete the study approximately 7 months after they had completed an introductory SBI statistics course. Data sources included audio, video, and gaze recordings, analytic memos generated by the researcher, as well as written artifacts generated by the participants. Participants generated concept maps, conducted statistical tests, interpreted statistical results from tests, and participated in a video-cued retrospective interview. Data analysis was conducted through an interpretivist epistemological stance and employed the constant comparative method to identify relevant moments across all data artifacts.

Preliminary analyses suggest that students' statistical reasoning (i.e., their understanding of why a specific statistical procedure should be conducted and what can be learned from it) was generally quite good, although there were many gaps in their statistical thinking (i.e., their ability to determine which statistical procedures should be done at what time and how to enact those procedures). Students generally struggled in thinking about null hypotheses and the probability distributions they specified, as well as with thinking about *p*-values, instead focusing on point and interval estimates for statistics of interest. In particular, students struggled to contextualize the null hypothesis and *p*-value, but readily saw the connection between point and interval estimates and the original problem context. When students did consider *p*-values, they seemed to only vaguely remember that "< .05 is something".

This study is one of the first to examine graduate students' statistical thinking several months after the completion of an SBI introductory course. It is expected that students will forget many of the details taught. What they did remember, high level reasoning and a focus on variability through the examination of point and interval estimates, suggests that statistics instructors might anchor instruction about statistical inference and tests to descriptive statistics and their interpretation and contextualization. These results allow us to evaluate the potential benefits of an SBI curriculum for these students, identify conceptual difficulties that can be addressed with pedagogical reform, and inform reformation in the practice of statistics to address extant controversies and crises.