Abstract

Though statistical testing is commonly practiced, the logic of statistical tests is confusing, thinking about distributions is difficult, and the way statisticians formulate expectations as probability distributions is poorly understood. To support instruction, the statistics education community has increasingly utilized simulation-based pedagogies that place the logic of statistical inference at the core of instruction. Might this approach support and sustain the development of graduate students' statistical thinking, especially during statistical testing? How do graduate students, who have completed a simulation-based course, think while conducting statistical tests, months after completing the course?

To answer these questions, a multi-modal multiple descriptive case study of six graduate students in the educational sciences was conducted. 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 for the logic of statistical tests, conducted statistical tests using statistical software, interpreted results from statistical tests, and participated in a retrospective video-cued interview. Data were analyzed through an interpretivist epistemological stance and employed the constant comparative method to identify relevant moments across all data artifacts to credibly describe participants' thinking.

Results suggest that students' planning (i.e., deciding what to do and when to do it) was generally quite good. However, students generally struggled in monitoring and evaluating their plan (i.e., ensuring that the plan was being executed correctly, and that no changes to the plan were needed). Furthermore, they generally did not seem to think about null models, core to the logic of statistical testing. Instead, they focused on point and

interval estimates for statistics of interest, and primarily thought about sampling variability in terms of a bootstrap dot plot, if at all.

This study is one of the first to examine graduate students' statistical thinking several months after the completion of a simulation-based introductory course. How students were thinking – generally able to reproduce a plan for analyzing the data consistent with what they were taught, and with a focus on variability through the examination of a bootstrap dot plot – suggests that statistics instructors might anchor instruction about statistical tests to descriptive statistics and their interpretation and contextualization. Furthermore, it suggests that the likelihood approach to statistical inference, evaluating hypotheses against given data, may be conceptually easier for students to think about.