Abstract

Extensive research has documented students' difficulty understanding and applying the Empirical Law of Large Numbers, the statistical principle that larger random samples result in more precise estimation. However, existing interventions appear to have had limited success, perhaps because they merely demonstrate the Empirical Law of Large Numbers rather than support students' conceptual understanding of *why* this phenomenon occurs. This dissertation developed a sequence of activities, *Growing Certain,* which provided support for two mechanistic explanations of the Empirical Law of Large Numbers for students in a simulation-based introductory statistics course: *swamping*, the decreasing influence of extreme values on the mean as sample size increases, and *heaping*, the increasing concentration of possible sample means around the population mean. Five students participated in over six hours of one-on-one clinical interviews, with analysis focused on one focal participant, "S". S's responses were analyzed using a detailed coding of S's articulation of mechanism components. S already displayed strong inclination towards swamping in the pre-interview questions, and their articulation of swamping became more sophisticated as they progressed in Growing Certain. However, S's understanding of the connections between population and sample were weak throughout, and S had a lot of difficulty reasoning about multiple sample means simultaneously in a sampling distribution. S's lack of abstraction of the sample mean appeared to support them in attending to the dynamics of swamping, but hindered them in being able to reason about heaping. Future research could examine representations that bridge swamping and heaping, and to examine individual differences in attention to the mechanistic components of the Empirical Law of Large Numbers.