Investigating Students' Ways of Reasoning When Estimating an Unknown Population Parameter

Claire Miller
University of Georgia

To examine the forms of reasoning novice statistics students employed when reasoning about sampling distributions, I used Charles S. Peirce's three classic forms of inferential reasoning—deduction, induction, and abduction—defined by case, rule, and result. In this paper, I report on a subset of these findings by describing and comparing the reasoning of seven undergraduate students. I engaged each student in a statistical task that I designed to investigate their reasoning in a repeated sampling environment when given a population with an unknown parameter. Findings indicate that abductive reasoning was powerful in making inferences from sample data to the unknown population.

Keywords: statistics education, reasoning, sampling distribution

Data are everywhere. Data are often reported in the form of polls, medical studies, and advertisement information. Introducing students to the basic language and fundamental ideas of statistics will help them develop statistical reasoning skills to evaluate data-based claims, enabling them to become better educated, well-informed members of society (Bargagliotti et al., 2020; Ben-Zvi & Garfield, 2004; GAISE College Report ASA Revision Committee, 2016). A major component of statistical reasoning is inferential reasoning, or drawing conclusions from data. Noll and Hancock (2015) noted that inferential reasoning is "a fundamental skill for participation in a global, technological, and data-driven society" (p. 362). Inferential reasoning includes many statistical ideas, such as repeated sampling, variability, and randomness. Understanding these ideas is important for improving students' statistical literacy and productive citizenship (Franklin et al., 2007; Saldanha & Thompson, 2007).

Despite the importance of understanding these statistical ideas and being able to reason inferentially, research suggests that students struggle with these ideas (Saldanha & Thompson, 2002, 2014; Sotos et al., 2007). For example, many students may be able to carry out the calculations involved in formal inference procedures, such as confidence intervals and hypothesis tests, but they often struggle to understand the underlying process and logic behind statistical inference (Chance et al., 2004). One reason for this difficulty stems from the complex and abstract concept of sampling distribution, which requires students to coordinate multiple ideas such as sample, population, distribution, variability, and repeated sampling (Chance et al., 2004; Kadijevich et al., 2008; Noll & Shaughnessy, 2012; Saldanha & Thompson, 2002, 2007). Another source of difficulty is that the forms of reasoning that are expected and highlighted in statistics differ from those that are standard in mathematics (Groth, 2015).

Theoretical Framework

Definitions of reasoning include some process of transforming given information to develop a conclusion or make an inference (e.g., Galotti, 1989; National Council of Teachers of Mathematics, 2009). Deduction, induction, and abduction are three classic forms of inferential reasoning. I adopt the definitions of deductive, inductive, and abductive reasoning from Peirce (1878) and others (e.g., Reid & Knipping, 2010). *Deductive reasoning* is the process of drawing logically valid conclusions based on stated assumptions or rules and is thought to be the only form of reasoning that establishes a conclusion with certainty. *Inductive reasoning* involves generalizing a rule from several observations. *Abductive reasoning* starts with an observation and involves *hypothesizing* what rule might lead to that particular observation.

Peirce (1878) defined deduction, induction, and abduction in terms of a triadic structure involving a *case*, *rule*, and *result* that are linked in a particular order. Reid & Knipping (2010) interpreted Peirce's definitions and defined a *case* as a specific observation that a condition holds. A condition describes an attribute of something, or a relation between things. A *rule* is a general proposition that states that if one condition occurs then another one will also occur. A *result* is a specific observation, similar to a case, but referring to a condition that depends on another condition linked to it by a rule. The order in which one links a case, rule, and result determines the kind of inference needed to gain more knowledge about the situation. In deductive reasoning, a rule and a case conclude a result. In inductive reasoning, a case and a result (or often, many cases and many results) lead to a rule. In abductive reasoning, a result and a rule infer a case. Consider the different ways case, rule, and result can be linked using Peirce's (1878) original example about a pile of beans found on a table: (1) If I know that all of the beans in the bag on the floor are white (the rule) and that a pile of beans on the table is from the bag on the floor (the case), then I can *deduce* (with certainty) that all of the beans in the pile on the table are white (the result); (2) If I know that a pile of beans on the table is from the bag of beans on the floor (case) and that all of the beans in the pile on the table are white (result), then I can *induce* that (probably) all of the beans in the bag on the floor are white (rule); and (3) If I know that all of the beans in the pile on the table are white (result) and all of the beans in the bag on the floor are white (rule), then I may *abduce* that (possibly) the pile of beans on the table is from the bag of beans on the floor (case).

In both mathematics and statistics, we see all three forms of reasoning. Conjectures and generalizations through inductive reasoning occur in both mathematics and statistics. Abduction occurs in both mathematics and statistics when hypotheses are constructed to explain some observation or event. Although mathematics and statistics share similar forms of reasoning, the emphases in each field differ. In mathematics there is an emphasis on deduction and proof, arriving at conclusions with certainty. The emphasis in statistics is on reasoning with uncertainty, a feature of both inductive and abductive reasoning. This paper reports on a subset of my findings from a larger project in which I investigated how novice statistics students understand sampling distributions. For this paper, I studied the research question: When given a population with an unknown parameter, what forms of reasoning (i.e., deduction, induction, and abduction) do novice statistics students employ when making inferences from one sample outcome to the population from which it was drawn, and what do these forms of reasoning reveal about their understanding of sampling distributions?

Methods

Participants for the larger study were 11 undergraduate students from a large university who had recently completed an introductory statistics course. I engaged each student in two clinical interviews, each 60-75 minutes in length. Across both interviews, the context was about the true population proportion of intended business majors at two large universities. North University's website states that 20% of their undergraduate students are intended business majors while South University's website does not have this information posted. In the first interview, students reasoned about sample outcomes drawn from North University, a population with a known parameter. They made predictions for sample outcomes, physically drew samples from a large box of beads with proportions corresponding to the population of interest, and simulated drawing many samples using a web-based applet (Rossman & Chance, 2021). Throughout the interview, they updated their predictions for sample outcomes as they continued to draw more samples. In the second interview, I tasked students with estimating the unknown population parameter at South University based on the result of one sample outcome. They drew one sample from a box of beads that modeled the population of interest, made an initial prediction for the unknown parameter, tested and refined their prediction using the web-based applet, and ultimately provided a range of plausible values they believed captured the true population proportion. This paper reports on data collected from the second clinical interview.

Primary data for this paper include the interview transcripts. After checking and timestamping interview transcripts against the audio and video recordings of the interviews, I adapted Simon's (2019) three-level approach to analyzing qualitative data. In the first level, I analyzed the interview transcripts line-by-line and identified segments of data in which students provided an explanation or a justification. Taking a convergent approach (Clement, 2000) in the second level of analysis, I applied Peirce's (1878) case-rule-result framework to each segment to infer the form of reasoning students used. This involved identifying what the student used as a rule and how they either applied the rule to a particular case (deductive reasoning), developed the rule from multiple cases (inductive reasoning), or hypothesized a rule to explain a particular case (abductive reasoning). In the third level of analysis, I identified emergent themes that provided additional insight into students' understanding of sampling distributions beyond the three broader forms of reasoning.

Results

This paper reports on a subset of the findings of the larger study. By describing and comparing the reasoning of seven students, I highlight a critical difference in two forms of reasoning. All seven students used the applet to continually make and test predictions for the unknown population parameter. In addition, all seven students used the likelihood of their sample outcome to assess their predictions for the unknown population parameter. However, only three students related one or more sampling distributions that they constructed using the applet to a hypothetical population. These three students reasoned abductively by hypothesizing multiple populations that could have produced their one sample outcome. In contrast, students who did not consider a hypothetical population developed a rule—using the likelihood of their sample outcome—to deduce a range of plausible values for the unknown population parameter. I identified a subtle yet critical distinction in students' reasoning across these two groups; those who reasoned abductively coordinated two levels of distribution—population distribution and sampling distribution—and those who reasoned deductively did not show evidence of this coordination.

Abducing Plausible Values for the Unknown Population Parameter

Three students related one or more sampling distributions that they constructed using the applet to a hypothetical population that could have produced the collection of sample outcomes. These students abduced a range of values based on the likelihood of the outcome of their one random sample that they drew from the box of beads that modeled the South University context. They repeatedly predicted and tested multiple values for the unknown parameter and assessed the plausibility of each value by considering how often their sample occurred in the many samples they drew using the applet (see Figure 1). Each time they determined that a hypothesized value was plausible, they abduced a case from a result and a rule. Starting with their one sample outcome (result), they hypothesized a value for the unknown population parameter, then constructed a sampling distribution using the applet with the probability of success equal to the hypothesized value and drew many samples. Because their sample outcome was likely to have occurred under the hypothesized value (rule), they abduced that their sample could have been drawn from a population whose parameter was equal to the hypothesized value (case).

For example, when testing her prediction of 60% for the unknown population parameter, Lorraine drew 500 random samples with the applet using a probability of success of 0.6 (see Figure 1). Using the results of her simulation, Lorraine assessed the likelihood of her sample outcome of 52% when the probability of success was 0.6. She said,

> It looks like 52 would be reasonable…it looks like there's a pretty significant number of results that are…52. And then past that is where it tends to kind of taper out. So maybe if the true proportion were 60%, the lower bound would be like 48 or so.

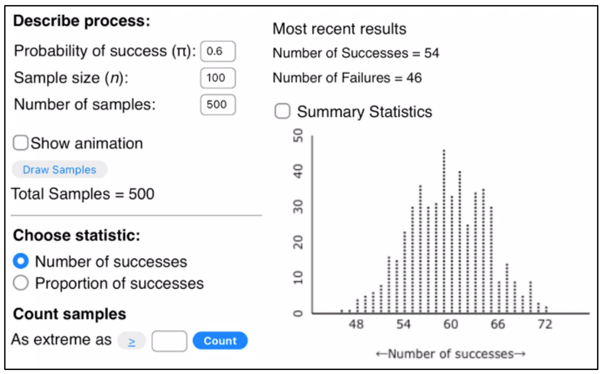

Figure 1. The Result of Lorraine's Simulation When Testing 60%

Given her sample outcome of 52% (result), Lorraine assessed the likelihood of her sample outcome, determining that her sample outcome "would be reasonable" "if the true proportion were 60%" (rule). She abduced that her sample outcome could have come from a population whose true parameter was 60% (case) and included 60% in her range of values for South University's true parameter. Lorraine related the sampling distribution she constructed to a hypothetical population with a "true proportion" of 60%, indicating she understood that the sampling distribution represented the outcomes of 500 random samples drawn from a population whose parameter was 60%. Lorraine continued to use the applet in this way to abduce multiple values that were plausible for the true percentage of undergraduates at South University who were intended business majors, ultimately providing a range of 42% to 62% for the unknown population parameter.

Like Lorraine, Eric and Lyla repeatedly predicted and tested multiple values for the unknown population parameter, assessing the plausibility of each value based on how often their sample outcome occurred. Furthermore, both showed evidence of relating one or more sampling distributions they constructed using the applet to a hypothetical population that could have produced that collection of sample outcomes. When Eric tested his prediction of 40%, he explained that the 40% represented "the assumption that if South University had their proportion online…that proportion is 40%. It's like the 20% from North University," indicating that he related the sampling distribution he constructed with a probability of success of 0.40 to a hypothetical population whose true parameter was equal to 40%. To assess the plausibility of his prediction, he considered the likelihood of his sample outcome under this assumption. He said, "If you say it's 40, then 50 could be a possibility," pointed to a sample outcome of 50 on the dot plot, and said, "So we could say, oh, that one [sample] that I drew was that one." When Lyla tested her prediction of 50%, she explained that she "made a distribution based on this 0.5" and looked for "the probability of drawing 46 students in a sample of 100 from this larger population." She related the sampling distribution she constructed with a probability of success to a "larger population" and considered the likelihood of her sample outcome of 46. Like Lorraine, both Eric and Lyla repeatedly reasoned in this way, abducing a case from a result and a rule. They abduced that their sample outcome (result) could have been drawn from a population (case) whose parameter was equal to their prediction (rule).

The conclusion of this abductive inference is that the one random sample these students drew could have been drawn from a population with some parameter, p. Without connecting a collection of sample outcomes to the population from which they were drawn, this conclusion is not possible. Without this connection, considering the likelihood of the outcome of the one sample can be used as a procedure to deduce a range of values for the unknown population parameter. This was the case for Tami, Tyra, Mindy, and Corrina.

Deducing Plausible Values for the Unknown Population Parameter

Four students used the likelihood of their sample outcome to assess the plausibility of multiple values for the unknown parameter, without relating each sampling distribution they constructed to a hypothetical population that could have produced those outcomes. Rather than hypothesizing what population could have produced their sample outcome, they used the likelihood of their sample outcome as a rule to determine whether they should include their prediction in the range of values for the unknown population parameter. These students reasoned that if their sample outcome occurred often when the probability of success was equal to p, then p should be included in their range of values, without connecting the collection of sample outcomes to a population whose parameter is equal to p. This connection is the subtle yet critical distinction between the abductive inference I described in the previous section and these four students' deductive inference. By considering the likelihood of their sample outcome each time they constructed a new distribution using the applet, they understood that the distribution they constructed represented a collection of sample outcomes to which they could compare their one sample outcome. However, they did not show evidence of understanding that each sampling distribution they constructed with some fixed parameter, p, represented a collection of possible outcomes from a population whose true parameter was equal to p.

Corrina's reasoning is representative of the group. When testing her prediction of 40%, Corrina considered the likelihood of her sample outcome of 51 (see Figure 2a). She said, "It's kind of on the higher end. It's not as likely, but it totally could happen" and included 40% in her range of values for South University's true parameter. When she tested 60%, she said, "Yeah, so that 51, again, is kind of on the lower end, but still likely, so I think it could be 60" (see Figure 2b) and, again, included 60% in her final range of 38% to 62% at the end of the task.

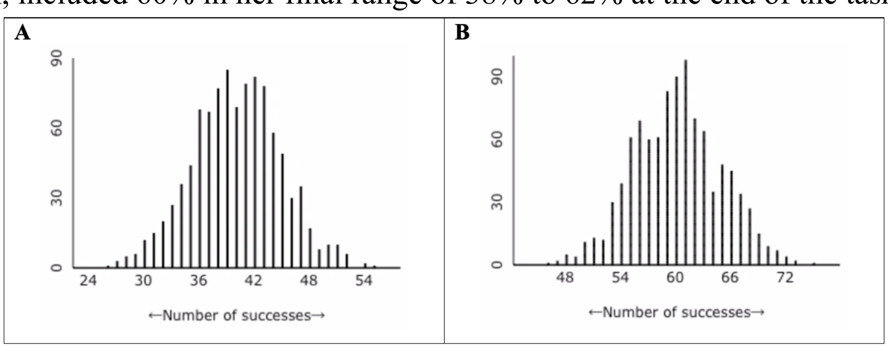

Figure 2. The Results of Corrina's Simulations When Testing 40% and 60%

Corrina applied a general rule to each of her predictions to deduce whether she should include the prediction in her range of values for the true parameter at South University: If my sample outcome occurred in a large collection of outcomes when the probability of success is equal to *p*, then *p* should be included in my range of values. She applied her rule to her prediction of 40% and reasoned that because her sample outcome of 51 "could totally happen" when the probability of success was equal to 40% (case), then 40% was a plausible value for the unknown parameter (result). Similarly, because her sample outcome of 51 was "on the lower end, but still likely" when the probability of success was 60% (case), then 60% was a plausible value for the unknown parameter (result).

Students' Summaries

At the end of the task, students summarized how they determined their range of plausible values for the true percentage of intended business majors at South University. These summaries provided additional evidence of the difference in students' understanding across the two groups. Table 1 shows the summaries of a subset of students who are representative of each group.

Table 1. Students' Summaries

Form of Reasoning	Summary
Abductive	Lorraine: If I put in the probability of success as 42, or if I put 0.42 in the center, that would be the proportion of business students. Then it'll tell me, or it won't tell me, but I can guess how likely it is that I would pull the result that I pulled (motions to her sample of 100 beads on the scooper) from that [box of beads]…If it was likely enough that I could pull a sample of 100 students where 0.52 was 52 is the percent that were business majors, then it would be likely that that number could be within the range of the actual proportion…It's kind of a messy way to explain it, but since we put in 0.42 and 52 was a reasonable, or it wasn't an uncommon result, yeah, it wasn't significantly high or significantly low, then it's likely that this [0.42] could be the actual proportion. So, it should be included in the range.
	Lyla: We started by drawing a sample of undergraduates without knowing the true proportion of students that are intended business majors, which my value was 0.46. And then I just made a prediction what the true population may look like based on this sample. I didn't stray much from the 0.46, I did 0.5 and then made a distribution based on this 0.5 and saw where the probability of drawing 46 students in a sample of 100 from this larger population fell on that success rate. And then based on that, tried to find values where drawing 46 was the very furthest extremity and then the lowest extremity to try to find cases in which it seems nearly impossible to get 0.46, both it being the smallest number and the largest number. And from that, knowing the proportions in which it was basically impossible to get the number 46, we know where it is possible to have 46.
Deductive	Corrina: So, when looking at the dot plot, I would put in our prediction of maybe it's 50% and then look for that 51 on the dot plot. And then just obviously we looked at the other numbers, but if you were really looking for, is it this or not, you could just kind of look at 51 and see how likely that happened. So, normally I would just kind of go straight for that 51 and see on this one it's this teeny tiny dot, but on one of the other ones it was the most probable thing on there…Just kind of where 51 was sitting on the scale. The number that we put in here (points to the probability of success on the applet) kind of is usually towards the middle, just kind of how close it's getting to this sample [outcome of 51] was kind of deciding whether or not it ends up in the scale.
	Mindy: So, this [first] one was just kind of a random prediction, because it's super close to 49. And then we plugged in different ranges, and I determined whether or not 49 would fall on the graph. And because of that you can kind of determine whether or not you think 20 or 30 would actually be in the range…When we saw 20, 49 wasn't super likely to show up.

There are two important differences between the summaries of students who reasoned abductively and those who reasoned deductively. First, those who reasoned abductively mentioned the context of the problem multiple times when they referred to "business students," "business majors," "a sample of undergraduates," and "a sample of 100 students." Second, they related the context to a hypothetical population when they referred to "the proportion of business students," "the true proportion of students that are intended business majors," and "students in a sample of 100 from this larger population." In addition to the explicit connections these students made to a hypothetical population in their explanations when testing multiple predictions, these key differences in their summaries provide additional evidence that this group of students coordinated two levels of distribution: population distribution and sampling distribution. In contrast, students who reasoned deductively referred to "the dot plot," "this teeny tiny dot," and "the graph," without ever referring to the context of the problem or using words or phrases that suggested they were thinking about a hypothetical population in relation to the samples they drew using the applet.

Discussion and Implications

This study examined the forms of inferential reasoning seven novice statistics students employed when reasoning about sampling distributions. When given a population with an unknown parameter, students used abductive or deductive reasoning when making inferences from one sample outcome to the population from which it was drawn. Those who reasoned abductively hypothesized multiple populations (and corresponding parameters) that could have produced their sample outcome. In doing so, they reasoned about the multi-tiered repeated sampling process (Saldanha & Thompson, 2002) and considered multiple levels of distribution in their reasoning—a critical understanding to reason productively from sample data to the population from which it was drawn. In contrast, students who reasoned deductively used the likelihood of their sample outcome in their rule for determining whether to include a particular prediction in their range of values for the unknown parameter. These distinct ways of reasoning have important implications for research and teaching.

Using Peirce's (1878) framework illustrates how these three broad forms of inferential reasoning are useful in distinguishing between the critical aspects of statistical reasoning and mathematical reasoning. In particular, I defined statistical inference in terms of abductive reasoning, which enabled me to understand how students coordinated multiple levels of distribution to draw inferences from sample data to a larger population. This approach of using Peirce's (1878) forms of inferential reasoning could be adopted by other researchers to investigate students' understanding of other important statistical concepts and statistical reasoning more broadly.

The focus of many introductory statistics courses at the undergraduate level is on statistical inference, whether informal or formal. The results of this study showed that abductive reasoning was powerful in making inferences from sample data to a larger population. Providing students with opportunities to reason abductively can support them in understanding the underlying logic and processes in estimating an unknown population parameter. For example, designing tasks that provide students with opportunities to continually hypothesize multiple populations from which a sample may have been drawn can promote abductive reasoning. Engaging with simulation tools can promote these critical forms of reasoning, but more research is needed to understand the meanings students construct when using these tools.

References

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseiiprek-12_full.pdf

Ben-Zvi, D., & Garfield, J. B. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), The challenge of developing statistical literacy, reasoning, and thinking (pp. 3-15). Springer Dordrecht. https://doi.org/10.1007/1-4020-2278-6_1

Chance, B., delMas, R., & Garfield, J. B. (2004). Reasoning about sampling distributions In D. Ben-Zvi & J. B. Garfield (Eds.), The challenge of developing statistical literacy, reasoning and thinking (pp. 295-323). Springer Dordrecht. https://doi.org/10.1007/1-4020-2278-6_13

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In A. E. Kelly & R. A. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 547-589). Lawrence Erlbaum Associates.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseprek-12_full.pdf

GAISE College Report ASA Revision Committee. (2016). Guidelines for assessment and instruction in statistics education (GAISE) college report 2016. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. Psychological Bulletin, 105(3), 331-351. https://doi.org/10.1037/0033-2909.105.3.331

Groth, R. E. (2015). Working at the boundaries of mathematics education and statistics education communities of practice. Journal for Research in Mathematics Education, 46(1), 4–16. https://doi.org/10.5951/jresematheduc.46.1.0004

Kadijevich, D., Kokol-Voljc, V., & Lavicza, Z. (2008). Towards a suitable designed instruction on statistical reasoning: Understanding sampling distribution with technology. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.

National Council of Teachers of Mathematics. (2009). Reasoning and sense making.

Noll, J., & Hancock, S. (2015). Proper and paradigmatic metonymy as a lens for characterizing student conceptions of distributions and sampling. Educational Studies in Mathematics, 88(3), 361-383. https://doi.org/10.1007/s10649-014-9547-1

Noll, J., & Shaughnessy, M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. Journal for Research in Mathematics Education, 43(5), 509-556. https://doi.org/10.5951/jresematheduc.43.5.0509

Peirce, C. S. (1878). Illustrations of the Logic of Science VI: Deduction, Induction, and Hypothesis. The Popular Science Monthly, 13(August), 470-482.

Reid, D. A., & Knipping, C. (2010). Proof in mathematics education: Research, learning and teaching. Sense.

Rossman, A., & Chance, B. (2021). *Rossman/Chance applet collection: One proportion.* https://www.rossmanchance.com/applets/2021/oneprop/OneProp.htm

Saldanha, L. A., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. Educational Studies in Mathematics, 51(3), 257-270. https://doi.org/10.1023/A:1023692604014

Saldanha, L. A., & Thompson, P. W. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of students' engagement and thinking in the context of instruction involving repeated sampling. International Electronic Journal of Mathematics Education, 2(3), 270-297. https://doi.org/10.29333/iejme/213

Saldanha, L. A., & Thompson, P. W. (2014). Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments. The Journal of Mathematical Behavior, 35, 1-30. https://doi.org/10.1016/j.jmathb.2014.03.001

Simon, M. A. (2019). Analyzing qualitative data in mathematics education. In K. R. Leatham (Ed.), Designing, conducting, and publishing quality research in mathematics education (pp. 111-122). Springer International Publishing. https://doi.org/10.1007/978-3-030-23505-5_8

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. Educational Research Review, 2(2), 98-113. https://doi.org/10.1016/j.edurev.2007.04.001