STUDENTS' "MULTI-SAMPLE DISTRIBUTION" MISCONCEPTION ABOUT SAMPLING DISTRIBUTIONS

Kiya L. Eliason
Brigham Young University
kiya.eliason@gmail.com

Steven R. Jones
Brigham Young University
sjones@mathed.byu.edu

The sampling distribution (SD) is a foundational concept in statistics, and simulations of repeated sampling can be helpful to understanding them. However, it is possible for simulations to be misleading and it is important for research to identify possible pitfalls in order to use simulations most effectively. In this study, we report on a key misconception students had about SDs that we call the "multi-sample distribution." In this misconception, students came to believe that a SD was composed of multiple samples, instead of all possible samples, and that the SD must be constructed by literally taking multiple samples, instead of existing theoretically. We also discuss possible origins of this misconception in connection with simulations, as well as how some students appeared to resolve this misconception.

Keywords: Statistics, Sampling Distribution, Process and Object, Multi-sample Distribution

It is important to help pre-service mathematics teachers develop their own conceptual understanding of statistics content (Conference Board of the Mathematical Sciences, 2001), because their conceptual understanding impacts learning opportunities available for their students (Ball, Lubienski, & Mewborn, 2001). In statistics, one concept of critical importance is the sampling distribution (SD). It forms the conceptual basis of much of elementary statistics, including confidence intervals, hypothesis testing, and correlation testing (Lipson, 2003). Thus, if we want students to develop strong conceptual understandings of elementary statistics, it is essential to help our pre-service math teachers develop strong understandings of SDs, as well.

Much research aimed at conceptual understanding of SDs focuses on exploration activities in which students repeatedly sample from a population (Aguinis & Branstetter, 2007; Chance, delMas, & Garfield, 2004; delMas, Garfield, & Chance, 1999; Glencross, 1988; Mills, 2002; Peck, Gould, Miller, & Zbiek, 2013; Watkins, Bargagliotti, & Franklin, 2014). These simulations are meant to show the emerging properties of the SD that: (a) the shape of the distribution is approximately normal, (b) $\mu_{\bar{x}} = \mu$, and (c) $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. However, some have realized that these same simulations might inadvertently be misleading (Watkins et al., 2014). It is beneficial for teacher educators to know what misunderstandings their students might develop from such simulations in order to use them most effectively. Our study seeks to build on this research by describing a previously undocumented misconception seen in pre-service teachers, which gets at the heart of what a SD even is. We also examine how this misconception might be resolved.

Background on the Sampling Distribution

Brief Recap of Sampling Distributions

Many types of statistical studies are based on using a sample to estimate or test certain population parameters, such as the population's mean (μ) or standard deviation (σ). Sampling distributions (SD) are what underlie the statistical methods used to do this. The basic idea of a SD is that, given a fixed sample size n, if all possible samples of size n are taken from the population, then the statistic of interest from all those samples creates a distribution in and of itself (Triola, 2010). For example, a SD for means is constructed by taking the sample means ($\bar{x}$) of all samples of the same size n from the population and putting them together to create a new distribution (see Figure 1). Note that there is

In: Sacristán, A.I., Cortés-Zavala, J.C. & Ruiz-Arias, P.M. (Eds.). (2020). *Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, Mexico. Cinvestav / AMIUTEM / PME-NA. https:/doi.org/10.51272/pmena.42.2020

a different SD for every sample size n that is chosen, because n is the same for all samples within a given SD. The Central Limit Theorem (CLT) then guarantees that a SD for means will always have the same mean as the population, $\mu_{\bar{x}} = \mu$, and a standard deviation given by $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. If n is sufficiently large, often cited as $n > 30$ (e.g., Triola, 2010), the CLT states that the SD will be an approximately normal distribution. Similar properties hold for SDs for proportions. It is important to note that SDs are *theoretical* in nature, in that they do not need to be empirically constructed to be used in statistical analysis. The CLT guarantees those properties that are needed for statistical analysis.

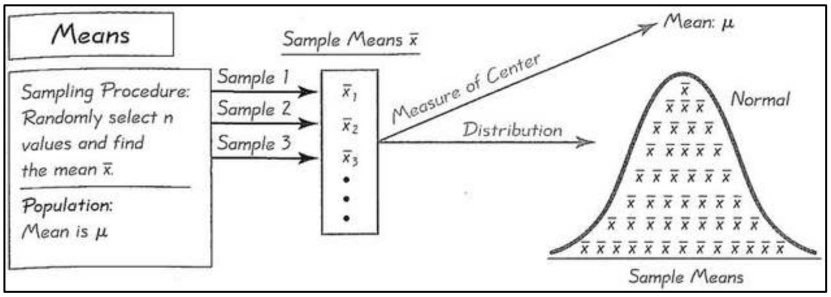

Figure 1: Creation of a Sampling Distribution, taken from Triola (2010, p. 281)

Brief Literature Review on Sampling Distributions

A common tool for teaching SDs and the CLT are simulations, in which physical enactment or computer software is used to create the results of many samples and to partially construct a SD. Using simulations to help students learn about SDs has been recommended as far back as the 1970s (e.g., Committee on the Undergraduate Program in Mathematics, 1972). Some have demonstrated that these simulations can give insight into otherwise theoretically intractable ideas (e.g., Mills, 2002; Simon, 1994). Yet other research has shown that simulations are insufficient by themselves. delMas et al. (1999) found that when students were allowed to experiment with simulations, their understanding did improve by a little bit, but not by as much as expected. They realized that the simulations alone did not force students to notice relevant features, and that activities needed to carefully scaffold student noticing (see also Chance et al., 2004). Lipson (2003) explained that there is a jump between an empirically constructed approximation to a SD using simulations and the actual theoretical SD. However, Lipson's focus was more on the influence that disconnect had on students' understanding of inference. In this paper, we examine how that disconnect directly impacts students' understanding of SDs themselves. Further, because simulated distributions are not perfect representations of the theoretical SD, Watkins et al. (2014) saw that students were sometimes misled by simulations. They observed students who incorrectly believed that the SD's mean got "closer" to the population mean as n increases. In fact, the CLT guarantees that $\mu_{\bar{x}} = \mu$ exactly, regardless of sample size. The misconception we discuss is related to what Watkins et al. observed, and might even be a root cause of it.

Taken together, this literature shows that simulations can be a useful tool in statistics education, so long as they are used carefully. We must be fully aware of potential pitfalls simulations might contain. We should continue to unpack possible issues in understanding SDs with simulations, in order to most effectively use simulations. This study adds a key, previously undocumented misconception, related to what a sampling distribution fundamentally is, that we observed in pre-service teachers who experienced this type of simulation activity.

Theoretical Perspective: Processes versus Objects

We view the concept of SDs as having a close connection to the theoretical notion of *processes* versus *objects* (Sfard, 1991, 1992). This lens gives valuable insight into possible SD misconceptions, and can help produce paths toward their resolution. In short, a process means an activity that can be conceptualized as being carried out, like imagining counting up to 1 million. It does not necessarily need to be enacted to be a process, but imagined. An object, then, is the encapsulation of such a process into a single cognitively entity. The "size" 1 million is a conceptual object, which can emerge out of imagining the process of counting up to 1 million.

SDs inherently deal with the *process* of repeated sampling. That is, one can conceptualize taking a sample and recording its sample mean (or proportion), and then taking another sample and recording its sample mean (or proportion), and so on. However, the full comprehension of SDs is to realize that this process can be encapsulated into a final result: the distribution of *all* sample means (or proportions). The SD is the *object* that results from the *process* of repeated sampling. We view simulations, through this lens, as essentially a representation of the *process* aspect of a SD. It permits the process to be quickly viewed over a large number of samples (as seen in Figure 2). However, in these simulations the *object* aspect of the completed SD is typically not reached. This limitation comes because such simulations usually do not depict when every sample has been represented exactly one time in the simulation, or at least represented in exactly equal proportion to every other possible sample. This matches Lipson's (2003) assertion that there is a jump from the empirically-simulated approximate SD to the actual theoretical SD. In this study, we examine how this issue led pre-service teachers to make incorrect conclusions about the fundamental nature of what a SD is.

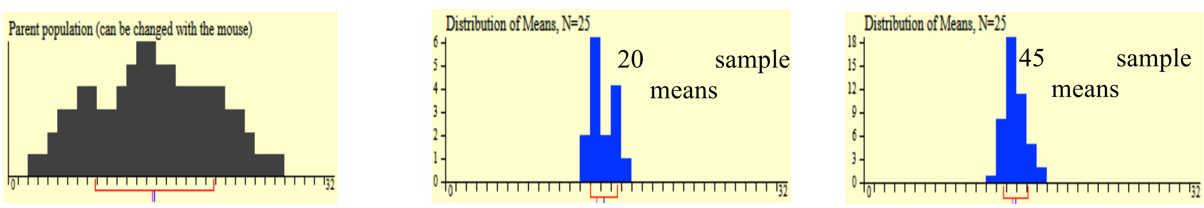

Figure 2: Example of a simulation (http://onlinestatbook.com/stat_sim/sampling_dist/)

Methods

This report emerged from a broader study we were engaged in regarding pre-service mathematics teachers resolving their misconceptions about confidence intervals. Students for the study were recruited from an undergraduate "Teaching Statistics and Probability" course for mathematics education majors, focused on conceptual understanding and on task exploration. The pre-service teachers had all completed a pre-requisite undergraduate statistics course, or AP statistics. In the education class we recruited from, the students used a simulation of repeated sampling, as discussed in the literature review, to develop the ideas of SDs.

The major purpose of the larger study was to understand how students might resolve misconceptions they held and to document the route they took in doing so. To recruit students for interviews, the students were given a quiz in their class regarding misconceptions on SDs and confidence intervals. Because the misconception we report on in this paper had not previously been documented, our specific misconception was not tested for, but emerged during the interviews while the students discussed other aspects of their understanding. From the quiz results, five students with varying levels of misconceptions about confidence intervals were selected to participate in two, hour-long interviews. We give the students the pseudonyms Danielle, Ethan, Corinne, Tiana, and Anna. During the interviews, the students were asked to explain how confidence intervals are constructed, to design their own hypothetical study that would use confidence intervals, and to discuss various aspects of SDs and confidence intervals. While conducting the interviews, the interviewer (Author 1) noticed a trend in terms of how all five students seemed to be talking about SDs. Thus, the interviewer began to follow up on this trend as well, and to make sure each student was asked about it. As the purpose of the larger study was to help students resolve misconceptions, the interviewer also attempted, impromptu, to document instances of students resolving this misconception during the interviews.

To analyze this trend, we went through all of the parts of the interviews where students made statements or gave explanations regarding SDs. In examining all instances of normatively incorrect statement or explanation, we realized they typically dealt with one main misconception about SDs. That is, most incorrect statements or explanations about SDs seemed rooted in the same misunderstanding. Once we singled out this misconception, we went back to the interviews to try to identify where the misconception came from in terms of prior knowledge or in-class activity. In doing so, we saw an important likely connection to the in-class simulation. Finally, as the interviewer had attempted, in the moment, to understand and document these confusions, we tracked the students' evolving conceptions of SDs over the interviews and looked for what was discussed in conjunction with changes in their understanding. This aided us in identifying what might help resolve the underlying misconception.

Limitations

There are some limitations to our analysis of this misconception. First, we had only five students in the sample, which is few. However, this report focuses only on documenting and discussing the misconception, rather than on establishing how common it is. Yet, since all five students in this study did share this same misconception, we posit that it is likely to be more widely shared. Second, this study was not originally designed to uncover this misconception, but it rather emerged from the data. Future work can be done to examine this misconception more systematically and to identify how common it might be among typical pre-service teachers.

Results

The Multi-sample Distribution Misconception

All five students suggested that to do statistics, one essentially uses a "sampling distribution" that contains *some* of the samples of size *n* from the population, as opposed to a completed sampling distribution of *all* possible samples. For example, consider Corinne's explanation.

> Corinne: If you took a bunch of samples and you found their means, you would get a sampling distribution.
> Interviewer: How many samples?
> Corinne: More than 30.

Corinne implied that if one has a certain amount of samples (i.e. "30"), then the distribution of those sample means *is* a SD. In another example, Ethan was describing how his hypothetical statistical study could be done. He seemed to imply that to obtain a SD, one would literally collect several samples in practice and compile them into a SD. The interviewer was initially unsure what he meant, and asked about how he imagined a SD.

> Interviewer: What's a sampling distribution?
> Ethan: The distribution of the means of your samples.
> Interviewer: How many?
> Ethan: Are you talking about in my thing [i.e. hypothetical study], or just in general? … [The SD is] the means of how many samples you take.
> Interviewer: But what if you only take one [sample]?
> Ethan: If you only take one sample then, [pause] I'm lost.

…

Interviewer: [The SD] is the distribution of the means of all your samples. Okay. All the samples that you take, or all the samples that you could take?

Ethan: That you take.

Here, Ethan explained that a SD is created by *empirically* collecting multiple samples during a statistical study. That is, he did not conceptualize a *theoretical* SD with the properties guaranteed by the CLT. In fact, at one point Ethan suggested that "all possible samples" really meant all the samples it was possible for someone to practically take. Ethan was not alone in believing one must collect multiple samples to empirically create a SD to do statistics. Most of these students described that to carry out their hypothetical study, they would need to take many samples to create a SD, as seen in the following statements.

Tiana: I want to take 30 samples of size 100.

Corinne: I would take 100 samples of size 30.

Ethan: I would get at least 200 samples just to be realistic.

Danielle: You take a large number of samples, like say 1,000, to get a sampling distribution of $\bar{x}$.

It is clear that these students were all thinking of a SD as a collection, not of *all* possible samples, but of several literally collected samples. We call the conceptualization of such a distribution of many, but not all, samples the *multi-sample distribution*, denoted *M-SD*. We consider it a misconception when the M-SD is seen as being *the* SD. We claim that the M-SD misconception is closely connected to perceiving only the *process* part of the SD. That is, the process of repeated sampling is understood, but it does not have a theoretically completed end of all possible samples that is the *object* SD. In this way, M-SD is not "wrong," but incomplete in a critical way. The students even seemed to understand that this process could continue, with more samples, to create a "better" M-SD, but they typically did not understand that the process has an end-result *object* that is the theoretical SD.

In conjunction with the M-SD misconception, the students in our study exhibited some misconceptions previously reported on in the research literature. For example, many claimed that $\mu_{\bar{x}} \approx \mu$ rather than $\mu_{\bar{x}} = \mu$ (cf. Watkins et al., 2014), as in the following excerpt from Corinne.

Corinne: [$\mu_{\bar{x}}$] is the mean of the means you sampled… In the real world, we never get to work with the distribution where μ and $\mu_{\bar{x}}$ are equal. We just get closer and closer [with more samples].

In fact, we believe previously reported misconceptions like this regarding $\mu_{\bar{x}}$ may really be a symptom of an underlying M-SD misconception. Note that the mean of the M-SD is technically $\bar{x}_{\bar{x}}$, as opposed to $\mu_{\bar{x}}$ because it is only a *sample* of sample means rather than the *population* of all sample means (where $\bar{x}$ refers to a sample and μ to a population). In this perspective, it is true that $\bar{x}_{\bar{x}} \approx \mu$ as the students claim. It is only in the SD of *all* samples where $\mu_{\bar{x}} = \mu$ exactly.

Possible Origins of the M-SD Misconception

During the interviews, the students described aspects of their thinking that matched with the simulation activity used in their class to discuss SDs. In the classroom activity, each student randomly selected samples from a population and computed the sample mean of them. The activity culminated in the students plotting their sample means together to create a visual representation of what is, essentially, a M-SD. This activity may have fostered M-SD thinking in the students. For example, in her interview, Tiana recounted this simulation activity and explained that she understood the image she saw – of the multiple sample means plotted together – as being *the* SD. Regardless of whether the instructor may have mentioned that the image was not the SD, the strong visual that represented the culmination of the simulation activity seemed powerful enough that she interpreted it as though she were seeing a SD.

By contrast, Corinne did recognize in her interview that there is such a thing as a distribution of *all* sample means. However, she discarded it as anything practically useful in doing statistics, explaining instead, as seen in her excerpt above, that "in the real world, we never get to work" with the actual SD. She explained that you could only use the actual SD "in something like manufacturing where you have data on every item or when you have a small population. But in that case it would be pointless because you could just do a census and know the population parameters." She explained that, practically speaking, in order to do statistics one would need to create the type of distribution seen in their class, that was made up of a collection of multiple sample means rather than all sample means.

Another root of the M-SD misconception may lie in classroom discussion of the CLT. One property of SDs given by the CLT is that if the common sample size for all samples is sufficiently large, often given as $n > 30$, then the SD is approximately normal. However, students may have confused this with believing that they need at least 30 samples for the (M)-SD to be approximately normal. The simulation activity may have inadvertently led them to focus on the wrong thing for "$n > 30$." In the simulation, the students saw that with each new sample mean added, the distribution began to resemble a normal distribution more. For example, when Ethan was explaining his hypothetical statistical study, he settled on wanting to collect 30 samples.

> Interviewer: Why is 30 a magical number to you?
> Ethan: The central limit theorem wants 30 [samples] for the sampling distribution to be normally distributed.

Notice that Ethan is justified in asserting that "30" is connected to the normality property given by the CLT. But, he did not appear to connect $n > 30$ as representing the *sample size* of each of those samples, as opposed to the *number of samples* needed to create a reasonably normal (M)-SD. This result is supported in the students excerpts from the previous section about wanting 30, 100 or 1,000 samples to make a sampling distribution.

Possible Resolutions of the M-SD Misconception

We defined the resolution of this misconception as recognizing that (a) the SD is a theoretical distribution from *all* possible samples and (b) that it does not need to be empirically constructed to be used. From our process-object perspective, the M-SD misconception essentially lacks the object component. Thus, resolution of this misconception is based on extending their process-oriented conception to include an object. Danielle, Ethan, and Corinne each gave some evidence of resolving this misconception. First, consider Danielle. One important part of her resolution of this misconception involved clearly distinguishing between sample size and number of samples.

> Interviewer: How many samples do we need to take before we can use the sampling distributions and assume that they are normal?
> Danielle: I think generally they say it's supposed to be like 35 or 30.
> Interviewer: Samples?
> Danielle: Yeah. That's the size of the sample [pause]. So wait, your question is?
> Interviewer: How many samples?
> Danielle: Oh, how many samples do we need to take. So, usually when we are using these types of things like our equations [refers to a formula sheet] we just take one sample!

Here, Danielle seemed to have realized the mismatch in thinking that multiple samples are needed to literally create an SD versus the fact that the statistical formulas use only one sample. Then, by thinking of just this single sample, she began to create for herself the ideas of a SD.

> Danielle: Any kind of sample you take is going to fall… somewhere. It is possible to get one that is farther away from the population mean… [Draws Figure 3]. If you were to take a sample, just one

sample, then it will fall somewhere along here in this range that is close to the population mean [gestures toward the middle a population distribution in Figure 3].

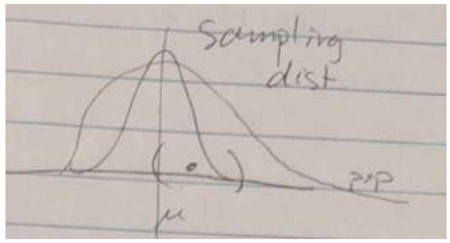

Figure 3: Danielle building on the simulation to now imagine all possible sample means

By reasoning with only a single sample, Danielle began to think more theoretically about where that single sample mean could be. In fact, this theoretical thinking seemed to help her imagine all possible samples, without having to literally collect all of them.

Danielle: So, if you could possibly take every single sample of that certain size and you were to be able to plot that, the sampling distribution would be normal and so we, since that concept is true, then we can just pull one sample point and it will be a point from somewhere on the sampling distribution.

We can see that Danielle had now conceptualized a SD as having all possible samples, and that it was theoretically, not empirically constructed. The single sample used in statistics was a member of this theoretical distribution. In fact, thinking of a single $\bar{x}$ more abstractly appeared helpful for some students in transitioning from the empirically-grounded M-SD to the theoretically-based SD. Corinne also used single $\bar{x}$'s to help make this transition.

Interviewer: What happens if someone only picks one sample? Let's make this the smallest possible $\bar{x}$ and this the largest possible $\bar{x}$. [Here the interviewer writes a number line and marks two points along it.]
Corinne: Without even knowing anything about this, most of them are going to be in the middle. So chances are that this one single [$\bar{x}$], it's here somewhere [gestures to the middle of the number line.]

Here, Corinne made an assertion about where a given $\bar{x}$ might be, without trying to create multiple samples. The interviewer tapped into this by then asking Corinne to imagine where all possible $\bar{x}$'s might fall along this number line. Corinne began to piece together where they might be, including that many $\bar{x}$'s would fall toward the middle. She eventually drew a SD similar to Figure 3. The interviewer asked about some of the specific properties of this new distribution.

Interviewer: Is μ the same as $\mu_{\bar{x}}$? [i.e. assuming all possible samples]
Corinne: I think at this point they are the same.
Interviewer: Why?
Corinne: Because at this point, if we have taken every possible sample, and take their means, and we are finding the mean of all those means, that is mathematically the same as finding the mean of all of those at once, which is finding μ.
…
Interviewer: So you are saying we can use just one sample [to do statistics]?
Corinne: But you're basing it off of information about all possible samples.

By leaving the empirical enactment from the simulation, Danielle and Corinne could begin to reason theoretically about where one $\bar{x}$ might lie, and then to where multiple $\bar{x}$'s might lie, to then where all $\bar{x}$'s might lie. This seemed to help them extend the *process* seen in the simulation to an imagining of a completed SD *object* with all possible $\bar{x}$'s being represented.

Discussion

We agree with the body of research that claims simulations are important for developing students' understanding of SDs (e.g., Mills, 2002; Simon, 1994). We also believe our study helps us better understand why simulation activities might be misleading in some ways, as noted by Watkins et al. (2014). Our process-object perspective suggests that simulations can only account for the process part of the conception of SDs, and cannot adequately portray the object part. We believe this is the reason for the possible disconnect Lipson (2003) described between the empirical simulation and the theoretical SD. If a simulation can only achieve the process component, the M-SD becomes a possible misconception students might develop. To be clear, we see the process component of a conception of SDs as essential, and simulations as a valuable way to develop the process component. That is, if one tried to simply create the *object* SD without first developing the process behind it, one might be left with a pseudostructural conception instead (Sfard, 1992; Sfard & Linchevski, 1994). In other words, the students might conceive of an SD object, but without understanding the underlying process that leads to it. Thus, we promote simulations as a useful way to develop the process, but claim that instruction must, at some point, move past the empirical simulation into a theoretical SD. Of course, simply "telling" students that there is a completed theoretical SD after observing a simulation might be insufficient to bridge the gap between process and object. Rather, it seemed important for some of our students who resolved the misconception to reason more theoretically about the distribution of a single hypothetical $\bar{x}$. This led to where multiple $\bar{x}$'s, and eventually all $\bar{x}$'s, might be distributed.

It was also important for our students to explicitly confront the difference between the sample size and the number of samples, which we believe is related to the misconception that $\mu_{\bar{x}}$ gets closer to μ as sample size increases (Watkins et al., 2014). As sample size increases, it is true that a M-SD will have an $\bar{x}_{\bar{x}}$ closer to μ. However, since a SD deals with all samples, *not* an increasing number of samples, it will always have $\mu_{\bar{x}} = \mu$.

Finally, we wish to emphasize that the M-SD misconception is not "wrong," but simply incomplete. In the spirit of perceiving misconceptions as useful building blocks, rather than faulty thinking that must be removed and replaced (see Smith, diSessa, & Roschelle, 1993/94), we find that resolving this misconception deals with adding on to what is already there, rather than taking away. Viewing the M-SD misconception in this light makes the path toward its resolution clearer, in that we can take the students' understanding as is and help them extend it to a completed *process-object* conception of SDs.

References

Aguinis, H., & Branstetter, S. A. (2007). Teaching the concept of the sampling distribution of the mean. *Journal of Management Education, 31*(4), 467-483.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching (4th ed.)*. New York, NY: Macmillan.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer Academic.

Committee on the Undergraduate Program in Mathematics. (1972). Introductory statistics without calculus (Vol. 2, pp. 472-525). Washington, DC: Mathematical Association of America.

Conference Board of the Mathematical Sciences. (2001). *The mathematical education of teachers*. Providence, RI: American Mathematical Society.

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3), http://jse.amstat.org/secure/v7n3/delmas.cfm.

Glencross, M. J. (1988). A practical approach to the central limit theorem. In R. Davidson & J. Swift (Eds.), *Proceedings of the second international conference on teaching statistics*. Victoria, B.C.: Organizing Committee for the Second International Conference on Teaching Statistics.

Lipson, K. (2003). The role of the sampling distribution in understanding statistical inference. *Mathematics Education Research Journal, 15*(3), 270-287.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1), http://jse.amstat.org/v10n11/mills.html.

Peck, R., Gould, R., Miller, S. J., & Zbiek, R. M. (2013). *Developing essential understanding of statistics for teaching mathematics in grades 9-12*. Reston, VA: National Council of Teachers of Mathematics.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*(1), 1-36.

Sfard, A. (1992). Operational origins of mathematical objects and the quandary of reification--The case of function. In G. Harel & E. Dubinsky (Eds.), *The concept of function: Aspects of epistemology and pedagogy (MAA notes, no. 25)*. Washington, DC: Mathematical Association of America.

Sfard, A., & Linchevski, L. (1994). The gains and the pitfalls of reification: The case of algebra. *Educational Studies in Mathematics, 26*(2-3), 191-228.

Simon, J. L. (1994). What some puzzling problems teach about the theory of simulation and the use of resampling. *The American Statistician, 48*(4), 290-293.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993/94). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Triola, M. F. (2010). *Elementary Statistics* (11th ed.). Boston, MA: Addison-Wesley.

Watkins, A. E., Bargagliotti, A., & Franklin, C. A. (2014). Simulation of the sampling distribution of the mean can mislead. *Journal of Statistics Education, 22*(3), 1-21.