2016 NCTM RESEARCH CONFERENCE
April 11-13 • San Francisco

Paper Title: Elicited Models of Resampling and Bootstrapping
Author(s): Jeffrey A. McLean & Helen M. Doerr
Session Title:
Session Type: Brief Research Report
Presentation Date: Tuesday, April 12, 2016: 11:30 AM-12:00 PM
Presentation Location: San Francisco, California

# Elicited Models of Resampling and Bootstrapping

Jeffrey A. McLean
Sonoma State University
jeffrey.mclean@sonoma.edu

Helen M. Doerr
Syracuse University
hmdoerr@syr.edu

*This study focuses on the construction and development of secondary and tertiary introductory statistics students' (n=68) reasoning about resampling and informal inference. Students engaged a model eliciting activity aimed at eliciting the process of bootstrapping and the drawing of informal inferences. Bootstrapping has become an important tool for statisticians, who assert that it is intuitive to novice statistics students. Some groups of students constructed a resampling process similar to bootstrapping, but fell short of constructing the bootstrapping process by not collecting resamples that were equal in size to the original sample. We will discuss four categories of models of resampling constructed by groups of students and how these resamples were used to make informal inferences.*

## Introduction

Over the past few decades, statistics education has become an integral part of the mathematics curriculum at all levels. Statistics education gained traction in the K-12 classroom with the release of the National Council of Teachers of Mathematics' (NCTM) influential documents *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) and its successor, *Principles and Standards for School Mathematics* (NCTM, 2000). Data analysis and probability was highlighted as one of the five strands of mathematical content that should be emphasized in K-12 education. The *Common Core State Standards for Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010)* also emphasize statistics and probability as a domain of standards for students in the 6$^{th}$ through 12$^{th}$ grades, with statistical topics covered in the domain of measurement and data for grades K-5. This lesser emphasis on topics in statistics at the early grade levels is a noted critique of these standards (Rossman & Shaughnessy, 2013). These documents placed the learning of statistics on equal footing with number sense, algebra, geometry, and measurement in the recommended school curriculum. Prior to these documents, statistics was often “the mere frosting on any mathematics program if there was time at the end of the school year” (Shaughnessy, 2007, p. 957).

A current trend in statistics education is the shift from a focus on theoretical distributions and numerical approximations into an emphasis on data simulation and analysis (Cobb, 2007). New curricula for introductory statistics courses at the secondary and tertiary levels have been created which focus on this trend (Garfield, delMas, & Zieffler, 2012; Pfannkuch, Forbes, Harraway, Budgett, & Wild, 2013; Tintle, VanderStoep, Holmes, Quisenberry, & Swanson, 2011). Bootstrapping is one method of simulating data through resampling that has become an important tool for statisticians, who suggest that it is intuitive to novice statistics students (Lock et al., 2013). Bootstrapping is a process of simulating data beginning by drawing one sample from a population. Resamples are constructed by choosing elements from the original sample, one at a time with replacement, until as many elements are drawn as in the original sample. This new sample is called the bootstrap sample. The process is repeated many times to create a collection of such bootstrap samples. A statistic from each of these bootstrap samples is aggregated to form an empirical bootstrap sampling distribution, which can then be used to make inferential claims about the population from which the original sample was drawn.

Lock et al. (2013) claim that bootstrapping capitalizes on students' visual learning skills and helps to build students' conceptual understanding of key statistics idea. While the research literature has examined students' understandings of the method of bootstrapping, studies have not examined how students construct and develop the bootstrapping method. I suggest that this lack of eliciting the concept of bootstrapping is a gap in the research literature. In this study I examine the models of resampling and bootstrapping that groups of students constructed in order to make informal inferences about a population of data

## Review of Literature

Key to this study was the use of students' informal inferential reasoning to make claims about a population. Informal inferential reasoning is the drawing of conclusions from data that extend beyond the data, from viewing, comparing, and reasoning with distributions of data (Makar & Rubin, 2007; Pfannkuch, 2007). Makar and Rubin (2009) posited three principles that they think are key to informal inferential reasoning: 1) the making of generalizations beyond describing the given data; 2) the use of data as evidence for those generalizations; and 3) the use of probabilistic language to describe the generalizations. Researchers claim that the use of simulation to teach informal inferential reasoning can help students build a deep understanding of the abstract statistical concepts (Maxara & Biehler, 2006) and suggest that students' informal inferential reasoning may support the development of students' understanding of formal inferential reasoning (Bakker & Gravemeijer, 2004; Saldanha & Thompson, 2002). College curricula that use simulations have indicated modest improvement in students' understanding of inference (Garfield, delMas, & Zieffler, 2012; Tintle, Topliff, Vanderstoep, Holmes, & Swanson, 2012).

The focus of analysis for this study was the models of resampling and bootstrapping which participants constructed and developed while engaged in a model eliciting activity (Lesh et al., 2003). Models are "conceptual systems … that are expressed using external notation systems, and that are used to construct, describe, or explain the behaviors of other system(s)" (Lesh & Doerr, 2003, p. 10). Model-eliciting activities encourage students to generate descriptions, explanations, and constructions in order to reveal how they were interpreting situations. Model eliciting activities are designed in order for students to:

- Make sense of the situation drawing on both their school mathematics real-life sense-making abilities;
- Recognize the need to construct a model to complete the activity, rather than produce only an answer.
- Create documentation that shows solution paths, patterns, and irregularities that the students considered while constructing their model;
- Assess when their responses need to be improved, refined, or extended,
- Create models that can be extended to use in a broader range of situations (Lesh et al. 2000).

By using a modeling approach to examine student reasoning, we can view reasoning as dynamic and developing over the course of instruction.

## Design & Methodology

This study is a qualitative case study and part of a larger study that consisted of an eight-class-session teaching experiment that was enacted in four introductory statistics classes at the high school and community college levels (n=68). This study focuses on the student reasoning that developed during one class-session as students participated in a model-eliciting activity that aimed to elicit the method of bootstrapping. During the model-eliciting activity, we collected written classwork from all participants and videotaped full class discussions and presentations. We videotaped four focus groups of students (one from each class) in order to document the groups' model construction and development. We analyzed the videos and written classwork in order to reconstruct the development of the reasoning that were created by the participants.

Groups of students participated in a model eliciting activity, using hands-on manipulatives, where they were given one sample from a population and constructed models of resampling and inference in order to make claims about a population. The activity was designed to elicit a model that could be used to resample from a sample, with replacement, in order to construct a bootstrap sample of the same size as the original sample and use the distribution of these bootstrap samples to determine which are most likely to occur. The model eliciting activity asked groups of students to help a manager in a grocery store predict the percentage of peanuts in a certain brand of mixed nuts. The students were given a sample of mixed nuts in the form of seven craft sticks marked with a 'P' for peanut and 18 not marked to represent other kinds of nuts. The manager planned to buy a large shipment of mixed nuts, but thought that her customers preferred fewer peanuts. From this one sample of mixed nuts, the manager asked the students to determine a likely range for the percentage of peanuts in the entire brand of mixed nuts.

## Results

Groups of students constructed four categories of models for resampling and inference (Table 1). The first category of model (n=14 groups) treated the sample of 25 nuts in a manner similar to a population and collected resamples of the same size (either five or 10 nuts), without replacement, from the 25 nuts. The second category of model (n=2 groups) also sampled without replacement, but varied the size of each resample. The third category of model (n=3) discussed or collected resamples from the sample of mixed nuts by choosing with replacement. The fourth category of model (n=1) created a spinner with areas of a pie chart proportional to the number of peanuts and other nuts in the group's sample. The group then simulated resamples of 15 nuts, one spin at a time. We will briefly discuss the models in the first two categories and provide greater evidence and details for the groups in the final two categories, which developed models of resampling similar to bootstrapping.

Table 1
*Models of Resampling.*

| **Model Category (Groups)** | **Characteristics of Model Category** | **Model Subcategory Characteristics (Groups)** |
|---|---|---|
| Resampled *without* replacement: Resamples of *equal* sizes<br><br>(n=14) | Collected 10-20 resamples without replacement, each containing 5 or 10 nuts.<br><br>Constructed a likely range for the peanuts in the new brand. | Range's end points were *not* equidistance from the mean and were rounded to the tens of a percentage. (n=3) |
| | | Range's end points were equidistance from the mean. (n=2) |
| | | Range contained all values observed in their resamples. (n=5) |
| | | Range contained most values in the resamples. Extreme values that occurred least often were excluded. (n=4) |
| Resampled *without* replacement: Resamples of *varying* sizes<br><br>(n=2) | Collected 10 resamples, varying in size from 9 to 18.<br><br>Constructed a likely range for the peanuts in the new brand. | Range's end points were roughly equidistance from the mean. (n=1) |
| | | Range contained most values in the resamples. Extreme values that occurred least often were excluded. (n=1) |
| Resampled *with* replacement<br><br>(n=3) | Collected (or discussed collecting) resamples of equal sizes, with replacement. | Collected two resamples of 20 nuts, but did not use the data to draw a conclusion. Resampling was viewed as inefficient and did not yield expected results. (n=1) |
| | | Discussed, but the group did not collect 25 samples of 10 nuts. Resampling with or without replacement were viewed as yielding the same results. (n=1) |
| | | Collected one resample of 14 nuts before running out of time. The group discussed that resampling preserved the makeup of the sample when selecting each nut. (n=1) |
| Resampled with spinner<br><br>(n=1) | Constructed a spinner, with two regions equivalent to the proportion of peanuts and other nuts, to simulate samples. Collected 4 resamples of 15 nuts, but did not use the data to draw a conclusion. (n=1) | |

**Resampled without replacement: Resamples of equal size**

Each group that constructed a model in this category collected resamples from the sample of 25 nuts, without replacement, with equal sized resamples of either five or 10 nuts. The variation between these models came from how the groups of students used this data to draw conclusions. Four subcategories of models were used to draw conclusions:

- Groups found the mean number (or percentage) of peanuts in their resamples and constructed an interval around this mean with end points that were *not* equidistant from the mean.
- Groups found the mean number (or percentage) of peanuts in their resamples and constructed an interval around this mean with end points that were equidistant from the mean.
- Groups constructed an interval that contained *all* of the values for the number of peanuts in their resamples.
- Groups constructed an interval that contained *most* of the values for the number of peanuts in their resamples.

**Resampled without replacement: Samples of unequal size**

Two groups of students decided to take samples of varying sizes in asserting that they would have more variation in the samples. One group claimed that by varying the sizes of the samples, they could collect a broader range of samples than those with a constant sample size. The second group of students using this method asserted that varying the resample size was a more realistic scenario. A student stated that when buying bulk food, such as nuts, you don't purchase a certain number of nuts, but rather a scoop taken from a large bulk bin, "we're doing ten trials and we're grabbing a group. When you grab actual bulk food you don't know the certain size you're taking, you can't just grab one [size]." Each group collected 10 resamples, which varied in size from nine to 18 nuts, found the mean percentage of peanuts in their resamples, and constructed an interval around this mean with end points that were equidistant from the mean.

**Resampled with replacement**

Three groups discussed ideas of resampling with replacement. The first group began by choosing sticks one at a time from the bag with replacement, but they chose not to pursue the method because they received results that were unexpected. The group asserted that their results were unexpected since the proportions of peanuts in each sample (2/20 and 5/20) were lower than they had expected. If the proportion of peanuts in this new brand of nuts was 28% (as it was in the group's initial sample) it would be very unlikely to collect a resample of 20 nuts, with replacement, that only has two peanuts. Because the group collected an unlikely sample, they thought that their approach to the task was incorrect. The group also noted that the process was inefficient. The group had planned to collect 20 resamples of 20 nuts, but found resampling with replacement was time consuming. This led to the group viewing the process as inefficient and contributed to the group changing their approach to the activity.

The second group discussed resampling with replacement, but chose also pursue a method of sampling without replacement due to time. A student in this group, William, described the approach that he had wanted to use to go about the activity, but his group decided not to pursue.

Instructor: Talk about what you wanted to do. Your group wouldn't let you do it

William: Yeah, basically, I kind of want to do the same thing, pick out 10, but do it by singles, just like, take out one, and then figure it out by then.

Instructor: Okay.
William: Instead of taking all 10 and counting, just pick out one and say if it's peanut or other and do that, like 25 times. Putting it back in there each time and figuring out the likely range of it being other or peanut.
Instructor: So what about that feels better than, a better method for sampling than doing what your group has done?
William: I wouldn't say that it would be better; I would think that it would be the same. It was just a different method of doing it.

William described a method with resampling with replacement and the collection of resamples that are the same size as the original sample. His reasoning of sampling with versus without replacement was still developing, but his group went in a different direction before he could pursue this reasoning.

The third group constructed a method of resampling similar to bootstrapping, but they had difficulty deciding how to proceed when they first approached the activity. The instructor approached the group of students (Randy, Ted, and Susan) and asked how they were going about the problem.

Susan: I would say from this sample, that a little over a quarter of the peanuts, of the nuts are peanuts, from our random sample.
Randy: But it's only one sample.
Ted: But this is the only sample we have.
Brenda: So if you pick another random sample what's going to happen?
Ted: It's most likely going to change.
Susan: It's going to change, but I feel like it will probably be still about the same.

Susan made an assertion about the percentage of peanuts in the population under the condition that she was basing her assertion on the random sample, which Randy stressed is only one random sample. Susan emphasized that although other samples of nuts will be different from their one sample, they will probably be "still about the same". Susan demonstrated an understanding that the percentage of peanuts in the sample likely represents the percentage of peanuts in the population. This is a key aspect of inferential statistics. When making inferential claims you take for granted that the sample represents the population because as Ted stated, "this is the only sample we have" and as Susan asserted, "it will probably still be about the same" as other samples.

The group initially decided to take a resample of 14 mixed nuts by drawing one at a time from the bag of mixed nuts, without replacement. The group only took one sample with this method, which yielded five peanuts out of 14 nuts. The instructor then came back over to the group to discuss how the group had collected the sample of 14 nuts.

Instructor: So how are you choosing those?
Susan: He [Randy] randomly puts them together, and then I randomly without looking draw them out.
Instructor: Okay. So you're drawing out one at a time?
Susan: Uh-uh.
Instructor: Okay, and you're setting it on the table?
Ted: Yes.
Instructor: Then you're going back in and you're drawing another one?
Susan: Yes.
Instructor: Okay, so …

Ted: Ohh! Wait, you said that you're putting them on the table. Was that like, do you think that we should put them back in the bag after we draw it out? Like for probability simulators?

From this exchange Ted considered how the sampling would change if they resampled with replacement instead of resampling without replacement. By using the term "probability simulators" he was approaching the idea that if you don't replace the stick after choosing each one, the probabilities of choosing a peanut or another nut will change. Ted was combining the idea of the representativeness of the sample to the population, that the group discussed earlier, with the idea of the probabilities staying constant for each choice of nut, the key concepts of bootstrapping. The group continued to discuss how this process of replacement was different than their initial approach without replacement. They asserted that when not using replacement, the nuts that they were drawing from the bag no longer represented the sample after some nuts were drawn and not returned. The group followed this procedure of resampling with replacement to collect a sample of 14 nuts, with five peanuts, and nine other nuts. This was the only sample that they had time to gather before the class was reconvened to discuss each group's approach to determine the likely range of peanuts in this new brand. The group did not use this one sample to draw a conclusion since time ran out for developing their model.

**Resampling with a spinner.**

For the final category for models of resampling, one non-focus group constructed a spinner using a pencil and a hairpin (Figure 1).

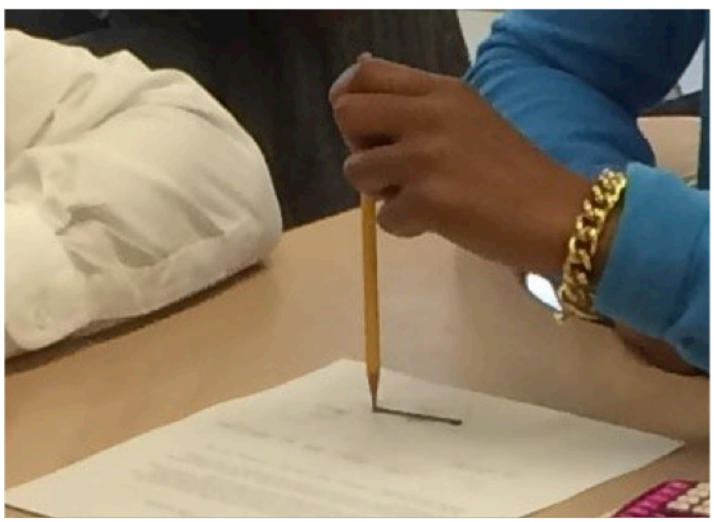

*Figure* 1. Spinner constructed by a group of students to resample the mixed nuts.

A student from the group, Yvette, discussed her group's approach when presenting it to the class (Figure 2):

Yvette: Basically what we did is we made a spinner. So we had 60 trials using the spinner, we did 15 four times [four samples of 15 nuts].

Instructor: I don't think that anyone else thought about a spinner. Can you tell us about, we haven't seen a spinner since…can you tell us about how you came up with this spinner?

Yvette: We counted how many peanuts we had out of the 25 and 7… so that's 28%, and then we constructed that spinner [The instructor shows the spinner to the class from Figure 1]

Allen: Hair pin, pencil, paper spinner, drawn approximately right, right? Slightly bigger than a quarter.

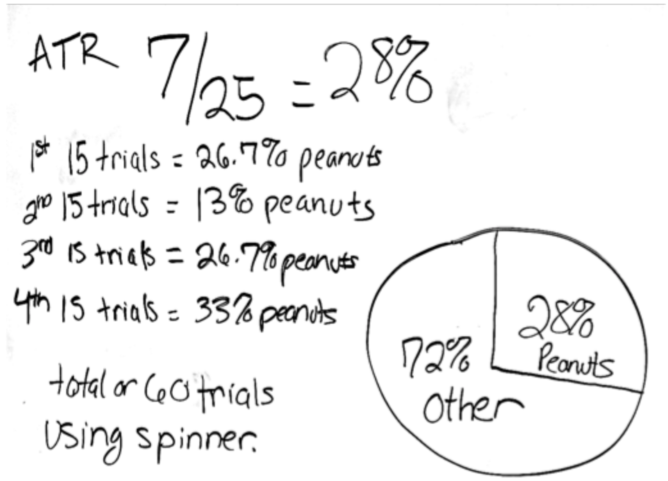


*Figure 2.* Spinner group's presentation of their model of resampling

From the four resamples (Figure 2) the group did not determine a likely range for the percentage of peanuts in the brand of mixed nuts. One of the resamples contained 13% peanuts, while the others contained 26.7%, 26.7%, and 33%. Yvette said that she was "thrown off" by the low value of 13% and wasn't sure what conclusion to draw. The group viewed the sample with 13% peanuts as unlikely sample, which made them question the validity of the method that they used to simulate the resamples.

**Discussion & Conclusions**

The key difference between this study and previous research addressing data simulation and bootstrapping (Garfield, delMas, & Zieffler, 2012; Pfannkuch, Forbes, Harraway, Budgett, & Wild, 2013), is the *elicitation* of bootstrapping methods by groups of students rather than the instruction of students on how to use the method. Four categories of models for resampling and inference were constructed by the groups of students in the in the model eliciting activity (see Table 1). Groups constructing the first two models of resampling applied a partial multiplicative conception of the samples (Saldanha & Thompson, 2002) by treating the samples as representative of the population and resampling from the sample in the same way they had sampled from the population in their previous coursework. What was missing from these models was the quasi-proportional nature of the relationship between the samples and population and how this relationship is affected by collecting samples without replacement. In the model eliciting activity, as each stick was drawn from the bag, without replacement, the proportions of peanuts changed, which meant the sample no longer maintained the representative relationship with the sample. Groups constructing the last two models of resampling took into account how the quasi-proportional relationship between the sample and population was affected by resampling and developed models of resampling which maintained this relationship. These students applied a multiplicative view of the sample (Saldanha & Thompson, 2002) and global view of the sampling process (Pratt & Ross, 2002) to initially construct a resampling process similar to bootstrapping, but fell short of constructing the bootstrapping process by not collecting resamples that were equal in size to the original sample. The students had difficulty drawing inferential claims from simulated data due to the small number of samples gathered from the time consuming process of resampling with replacement.

The findings of this study contribute to the field of statistics education by examining students' reasoning as they constructed and developed bootstrapping methods and investigated the relationship between this reasoning and the drawing of informal inferences. The study demonstrates how a model-eliciting activity can be used in introductory statistics classrooms in order for groups of students to construct models of bootstrapping. This has potential implication for future introductory statistics curricula development.

## References

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield, (Eds.), *The challenge of developing statistical literacy, reasoning*

Cobb, G. W. (2007). The introductory statistics course: A ptolemaic curriculum?. Technology Innovations in Statistics Education, 1(1).

Garfield, J., delMas, R. & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education.* DOI 10.1007/s11858-012-0447-5. Springer.

Lesh, R., Cramer, K., Doerr, H., Post, T. & Zawojewski, J. (2003). Model development sequences. In R. Lesh, & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on problem solving, learning, and teaching* (pp. 35–58). Hillsdale, NJ: Lawrence Erlbaum and Associates.

Lesh, R., & Doerr, H.M. (2003). Foundations of models and modeling perspective on mathematics teaching, learning, and problem solving. In R. Lesh & H.M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 3–33). Mahwah, NJ: Erlbaum.

Lesh, R., Hoover, M., Hole, B., Kelly, A. & Post, T. (2000). Principles for developing thought revealing activities for students and teachers. In Kelly, A., & R. Lesh (Eds.), *Handbook of research design in mathematics and science education*. (pp. 591-645) Mahwah, NJ: Lawrence Erlbaum Associates.

Lock, R., Lock, P.F., Lock-Morgan, K., Lock, E.F., & Lock, D.F. (2013). *Statistics: Unlocking the Power of Data.* Wiley.

Makar, K., & Rubin, A. (2007). Beyond the bar graph: Teaching informal statistical inference in primary school. In J. Ainley & D. Pratt (Eds.), *Reasoning about Statistical Inference: Innovative Ways of Connecting Chance and Data. Proceedings of the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy* (pp. 1- 29). The University of Warwick.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.

Maxara, C., & Biehler, R. (2006). Students' probabilistic simulation and modeling competence after a computer intensive elementary course in statistics and probability. Paper presented at the Seventh International Conference on Teaching Statistics. Salvador da Bahia, Brazil

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for K–12 mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Washington, DC: Authors. http://www.corestandards.org/Math/

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education, 2(3)*.

Pfannkuch, M., Forbes, S., Harraway, J., Budgett, S., & Wild, C. (2013). *Bootstrapping students' understanding of statistical inference.* Summary research report for the Teaching and Learning Research Initiative, www.tlri.org.nz

Pratt, D., & Noss, R. (2002). The micro-evolution of mathematical knowledge: The case of randomness. *Journal of the Learning Sciences, 11*(4), 453-488.

Rossman, A., & Shaughnessy, M. (2013). Interview with Henry Shaughnessy. *Journal of Statistics Education, 21*(1). Retrieved from http://www.amstat.org/publications/jse/v21n1/rossmanint.pdf

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257-270.

Shaughnessy J. M. (2007). Research on statistics learning and reasoning. In F.K. Lester (Ed.), *The second handbook of research on mathematics* (pp. 957–1010). Reston, VA: National Council of Teachers of Mathematics (NCTM).

Tintle, N., Topliff, K., Vanderstoep, J., Holmes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. Statistics Education Research Journal, 11(1), 21-40

Tintle, N., VanderStoep, J., Holmes, V-L., Quisenberry, B. & Swanson, T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education, 19*(1).