Teachers' Use of Informal Conceptions of Variability to Make Sense of Representativeness of Samples

Gabriel Tarr
Arizona State University

April Strom
Scottsdale Community College

A key factor in statistical thinking is reasoning about variability. This paper contains data on how in-service middle school teachers and a community college faculty member reasoned through two statistical tasks. The researcher presents his analysis of the data through the lens of how teachers reasoned about variability as they worked through the two statistical tasks.

Keywords: Statistics, Professional Development, Teacher Education

As data and statistical thinking have become more important in the information age, middle school mathematics teachers have been tasked with placing greater emphasis on statistical concepts than may have been previously required before the introduction of the Common Core State Standards for Mathematics [CCSSM] (National Governors Association, 2010; Tran, Reys, Teuscher, Dingman, & Kasmer, 2016). To reason about and teach statistical concepts in a productive manner, it is critical for teachers to attend to the notion of variability in their personal reasoning and in teaching statistics (Franklin et al., 2015; Garfield & Ben-Zvi, 2005; Moore, 1990; Shaughnessy, Watson, Moritz, & Reading, 1999; Wild & Pfannkuch, 1999). Specifically, Garfield and Ben-Zvi present a framework from which teachers can develop tasks to support and assess strong conceptions of variability in their students. Developing intuitive ideas of variability, using variability to make comparisons, and using variability to predict random samples or outcomes are key ideas in this framework.

In this paper, we compare the following: (1) how middle school math teachers used intuitive notions of variability to make comparisons between sets of outcomes of random processes for two statistical tasks and; (2) how a community college instructor answered the same tasks using more formal notions of variability. We attempt to answer the following question: *How do teachers' informal ways of reasoning using variability compare to an expert's more formal ways of reasoning about variability while working through statistical tasks?*

Methods

The data collected for this study were gathered in a large-scale professional development and research program, focusing on middle school teachers in a Southwestern state in the United States. Each teacher in the program was asked to participate in professional development activities for two years. The project focused on increasing teachers' mathematical and pedagogical content knowledge.

In the second year of the project, participants were involved in nine full-day workshops focused on both functions and statistics content. Upon the teachers' completion of their second year in the program, the researcher conducted four videotaped, individual, task-based, clinical interviews (Clement, 2000). The subjects for this study were three middle school teachers (Joy[1], Leia, and Nina) and one community college faculty member (Kory) responsible for co-leading the statistics content for the workshops.

---

[1] All subject names are pseudonyms.

Prior to their second year on the project, the three teachers responded to a free-response survey question that asked them to describe their own personal background in statistics. Joy responded that she was "very limited" in statistics with her background being one college course in educational statistics, the material that she taught to her students, and the statistics-related material learned in her collaborative community of learners[2] (CCOL) facilitated by the other co-leader of the statistics content at the workshop. Leia described her background as limited in the variety of statistics that she used and uncomfortable in justifying her methods for doing statistics, but comfortable in analyzing data and displays. Leia also participated in the same CCOL as Joy. Nina described herself as having "very little background" in statistics other than working with spreadsheets. Kory taught statistics at the secondary and post-secondary levels for over 20 years.

The researcher analyzed each subject's raw and transcribed video data with careful attention to habits or inclinations that the subjects may have shown while reasoning through the tasks. The researcher then created themes based on these habits or inclinations before refining said themes through subsequent passes through the data using open coding principles developed by Strauss and Corbin (1998). Once the researcher felt these themes were sufficiently well-defined, his attention shifted to themes that related to how the subjects utilized concepts pertaining to variability to reason through the statistical tasks. The researcher then analyzed these variability-related themes for each subject individually before coordinating common themes across subjects.

Task Description

Participants engaged in two tasks: The Coin Flip Task[3] (Figure 1) and The Orange Bin Task (Figure 2). The Coin Flip Task was chosen to determine how participants would reason about the probabilities of two distinct events when the proportion of outcomes for each event was the same. From prior data collection efforts, the researcher suspected that the teachers would have limited mathematical background, thus making this The Coin Flip Task non-trivial. Thus, the teachers would need to reason about the similarities or differences between the two events in order to provide an answer they deemed to be reasonable.

Event A: A machine flips a fair coin 10 times with the outcome of 7 heads.
Event B: A machine flips a fair coin 1000 times with the outcome of 700 heads.

Which one of the following is true?

1) Event A and Event B are equally probable.
2) Event A is more probable than Event B.
3) Event A is less probable than Event B.
4) Unable to determine given the information.

Figure 1: The Coin Flip Task

The Orange Bin Task was chosen to determine if subjects would reason about the role of sample size in the variability of outcomes and use this reasoning to support the grocer's choice of whose method to choose. As an explicit verification of how subjects made a connection between sample sizes and the variability in the two sets of potential mean weights, the following question was asked to any teacher who did not give an example of two mean weight lists[4]: *Suppose that both*

---

[2] The project's version of a Professional Learning Community (PLC).
[3] Modified task from Schrage (1983) p. 353
[4] Each mean weight was determined by oranges randomly generated from a normal distribution with a mean of 131 grams and a standard deviation of 15 grams.

of these people had repeated their method only five times. One person yielded the following five mean weights. 133 grams, 124 grams, 129 grams, 129 grams, 140 grams. The other person yielded the following five mean weights. 126 grams, 130 grams, 128 grams, 135 grams, 133 grams. Which set of mean weights belong to which person, and why?

A grocer wanted to determine the typical weight for oranges in his store. One employee, Jeff, suggested finding the mean weight of 5 randomly-selected oranges, placing the oranges back into the bin, and repeating this process several times. Another employee, Krystal suggested finding the mean weight of 15 randomly-selected oranges, placing the oranges back into the bin, and repeating this process several times.

1) Generate what you believe could be the outcomes for Jeff's process.
2) Generate what you believe could be the outcomes for Krystal's process.
3) The grocer decided to go with Krystal's suggestion of using 15 randomly-selected oranges instead of Jeff's 5 randomly-selected oranges. Give specific reasons for why you believe the grocer decided to go with Krystal's suggestion?

Figure 2: The Orange Bin Task

Preliminary Results and Discussion

The Coin Flip Task

Kory read the problem and immediately determined that Event A was more probable than Event B due to the difference in the number of trials for each event. To support his reasoning, Kory drew upon his prior statistical knowledge and saw the context of the coin flip problem as a binomial situation. Kory calculated the expected value and standard deviation for each event to determine how far each outcome deviated from the expectation. Kory explained for Event A that "seven is a little bit more than a standard deviation away from the center, it's fairly likely to occur. It's not incredibly likely, but it's still within the realm of possibilities." However, Kory voiced the opposite stance when considering the probability of Event B: "If you think about where 500 and 700 is, that's over 10 standard deviations away, that more like 12 or 13 standard deviations….Yeah not gonna [sic] happen."

$$\mu = 10(.5) = 5$$
$$\sigma = \sqrt{10(.5)(.5)} = 1.58$$

$$\mu = 1000(.5) = 500$$
$$\sigma = \sqrt{1000(.5)(.5)} = \underline{15.8}$$
$$500 \rightarrow 700$$

Figure 3: Kory's Calculations of Mean and Standard Deviation

Leia and Nina anticipated that the outcomes of the coin flips should be equally represented. For both teachers, the number of trials in Event A made it seem possible that seven heads could occur in 10 coin flips because this event didn't deviate too much from their initial anticipation. However, when reasoning about Event B, both teachers saw this as a major aberration from what they had anticipated. In fact, when first engaging with Event B, Nina immediately said "Wow! … That seems crazy. Which one of the following is true? The coin is not fair." Nina elaborated on her reasoning about why Event A was more probable than Event B by stating the following:

"It's a fair coin, … the more times I do it (trials), it should be, it (distribution of outcomes) should approach tails and heads should be appearing in an equal frequency, 50% 50%." Leia provided similar explication for her stance on why Event A was more probable than Event B by invoking her image of the law of large numbers "The higher the number of trials is, the less likely it is that these numbers are going to be far away from that theoretical probability….When you only do ten, there's a lot of chance for it to be different." In both instances, the Leia and Nina reasoned that a larger number of trials would decrease the amount of variation that between what they would anticipate for the outcomes of flipping a fair coin, and what the observed outcomes would be.

The only subject to answer The Coin Flip Task incorrectly was Joy. Joy gave the response that Event A and Event B were equally probable because "seven tenths is the same as $\frac{700}{1000}$." When asked by the researcher to create a new event in the same context that would have the same probability of Event A and Event B, Joy created Event C and Event D where the outcomes were 100 coin flips with 70 heads and 50 coin flips with 35 heads, respectively. As Joy reasoned about the differences between Event A and Event B, the only difference that she verbalized was the fact that Event B had 100 times as many trials.

When Joy was presented The Coin Flip Task, she immediately made mention to a prior teaching experience where she presented her students with a probability lesson related to coin flipping. She jokingly lamented about a troublesome student who had challenged her claims to a coin flip outcome because "the coin wasn't fair because one of the sides was heavier." She stated that "he likes to get nitpicky about stuff like that." Initially, the researcher considered this to be a throwaway comment. However, shortly after the Joy recounted her classroom experience, reading the phrase *fair coin* seemed to trigger the response "It's a fair coin, that's what I should have said, one of the sides is not heavier than the other," as if she had just realized a quelling to her troublesome student's refutation. These utterances seem to provide evidence for why Joy had not analyzed the situation in an analogous way to the other subjects. By not considering fairness, it is possible Joy was not anticipating that the outcomes of the coin flips should be equally represented given a large enough size of observations. By not anticipating this equal share of outcomes, she was not perturbed by 70% of the outcomes being heads as deviating far from 50% of the outcomes being heads in either the 10 or 1000 flip case.

The three subjects who were able to correctly respond to The Coin Flip Task shared a common focus that seemed to be paramount in their thinking. Establishing a reference for what they had *anticipated* would happen allowed the subjects to compare the degrees of variability for each event. Kory compared expectations with observations using a measurement tool of standard deviations. Leia and Nina showed more informal reasoning that a small number of trials would result in a greater chance for aberration from expected outcomes than would a greater number of trials. In reasoning through this task, all three successful subjects reasoned using some aspect of variability to fuel their thought processes.

The Orange Bin Task

Using some aspect of variability to reason about a statistical situation was present again for participants for The Orange Bin Task. Knowing that there would be variability in the collection of sample means, Kory set off to make the variability explicit. Kory calculated the standard deviation for each sampling mean distribution under the assumption that orange weights would be normally[5] distributed. Comparing the spread of the distribution of sampling means for both Jeff and Krystal's methods allowed Kory to reason about how each sample would be distributed around the true mean weight for the population of oranges (Figure 2). Kory reasoned that: "… she's going to have much less variation…. The grocer will probably go with, he goes with Krystal's suggestion because she's probably closer to the truth than Jeff is. She has got much less variability in her distribution of sample means. Larger samples give better results typically."

Leia, Nina, and Joy also utilized the fact that sample size was the factor in determining why the grocer chose Krystal's method. A common theme in their approach was how extreme values would potentially influence the mean weights for each method. Joy provided two lists of mean weights (Figure 3) where the underlying reasoning was that larger samples would produce more consistent means.

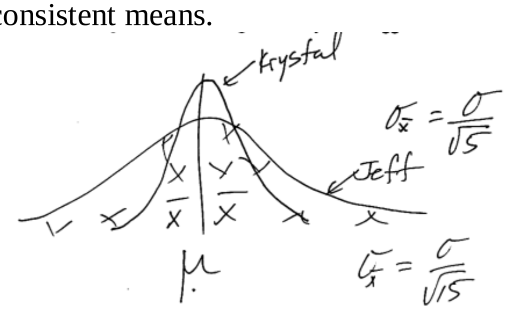

Figure 4: Kory's Representation of Distributions of Sample Means

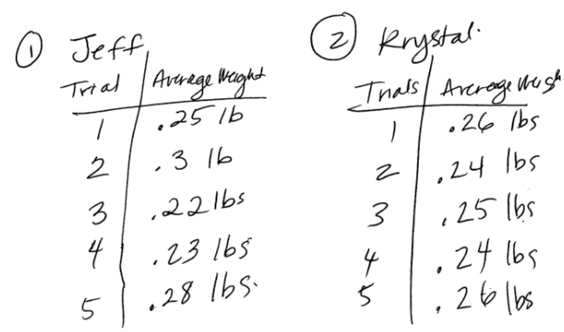

Figure 5: Joy's Constructed Lists of Mean Orange Weights

Thus, Joy felt the variation for Jeff's means would be larger than the variation for Krystal's means. While, Leia and Nina did not explicitly produce lists of mean weights, when given the follow-up question, both Leia and Nina reasoned that since the second list of numbers showed less variation, the list had to belong to Krystal.

The teachers were able to use informal ways of reasoning about variability in mean weights to support their arguments about how larger sample sizes would produce more representative sampling results. The teachers were able to reason that a larger sample size would allow for fewer relative aberrations, which in turn would make the collection of means for this larger sample size more consistent with each other. Though not synonymous with Kory's way of reasoning, the three teachers definitely displayed reasoning that can potentially precede thinking about the standard deviation of the sample mean.

Conclusion

While the ways of statistical thinking presented by the community college faculty member are beyond what most middle school mathematics teachers would teach, these underlying ways can be preceded by the ways of thinking that the teachers displayed. Developing these intuitive ways to use variability to reason about statistical situations can lead teachers to develop more normative, robust ways of reasoning. This in turn may allow middle school teachers to better understand where their students are headed in terms of concepts which will enable them to prepare a better statistical foundation for their students.

Acknowledgements

This work was supported by MSP grant #1103080 through the National Science Foundation.

[5] Fruit weights are typically lognormal distributed.

References

Clement, J. (2000). Analysis of clinical interviews : Foundations and model viability. In R. Lesh & A. Kelly (Eds.), *Handbook of Research Methodologies for Science and Mathematics Education* (pp. 341–385). Hillsdale, NJ: Lawrence Erlbaum.

Franklin, C. A., Bargagliotti, A. E., Case, C. A., Kader, G. D., Scheaffer, R. L., & Spangler, D. A. (2015). Statistical education of teachers (SET).

Garfield, J. B., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability 10. *Statistics Education Research Journal*, *4*(1), 92–99.

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New Approaches to Numeracy* (pp. 95–137). Washington, D.c.: National Academy Press.

National Governors Association. (2010). *Reaching Higher: The Common Core State Standards Validation Committee*.

Schrage, G. (1983). (Mis-) Interpretation of Stochastic Models. In R. W. Scholz (Ed.), *Decision Making Under Uncertainty* (pp. 351–361). North-Holland: Elsevier Science Publishers B.V.

Shaughnessy, J. M., Watson, J., Moritz, J., & Reading, C. (1999). School Mathematics Students' Acknowledgment of Statistical Variation. In *Paper presented at the NCTM Research Presession Symposium: There's More to Life than Centers* (pp. 1–35).

Strauss, A. L., & Corbin, J. M. (1998). *Basics of Qualitative Research* (2nd ed.). Thousand Oaks, California: Sage Publications, Inc.

Tran, D., Reys, B. J., Teuscher, D., Dingman, S., & Kasmer, L. (2016). Analysis of Curriculum Standards: An Important Research Area. *Journal for Research in Mathematics Education*, *47*(2), 118–133.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review / Revue Internationale de Statistique*, *67*(3), 223–256.