ENHANCING ELEMENTARY PRESERVICE TEACHERS' CONCEPTIONS OF VARIATION IN A PROBABILITY CONTEXT

Dan Canada
Eastern Washington University, United States
dcanada@mail.ewu.edu

While recent and ongoing research has begun to reveal ways that precollege students think about variation, little research has been done with the preservice teachers who will eventually serve such students. Specifically, more research is needed to understand what are the conceptions of variation held by elementary preservice teachers (EPSTs), and also how to shape the university courses where those preservice teachers learn. This paper, sharing an excerpt from an exploratory study aimed at EPSTs, describes changes in class responses to a probability task where variation is a key component. Overall, going from before to after a series of instructional interventions, responses reflected a more appropriate sensitivity to the presence of variation.

INTRODUCTION

The purpose of this paper is to report on research aimed at elementary preservice teachers' conceptions of variation. Other research has already begun to illuminate precollege student thinking about variation in several contexts, such as sampling, data and graphs, and probability situations (e.g., Reading and Shaughnessy, 2004; Watson and Moritz, 1999; Shaughnessy and Ciancetta, 2002). However, as the picture begins to get painted about how precollege students reason statistically, the research on how teachers reason about variation remains thin. In particular, there is a paucity of research about how preservice teachers think about variation, or variability in data. Therefore, doctoral research was undertaken to explore what are the components of a conceptual framework that help characterize elementary preservice teachers' (EPSTs') thinking about variation, how EPSTs' conceptions of variation before an instructional intervention compared to those conceptions after the intervention, and what tasks were useful for examining EPSTs' conceptions of variation in the contexts of sampling, data and graphs, and probability. The specific research question derived from the larger study and addressed in this paper is: How do EPSTs' responses concerning variation in a probability context compare from before to after an instructional intervention? After describing the conceptual framework and methodology for the study, the results will next be presented, followed by further discussion.

CONCEPTUAL FRAMEWORK

Three key aspects of understanding variation that governed the overall study focused on how students were *expecting*, *displaying*, and *interpreting* variation. In dealing with expectations, students need an opportunity prior to conducting statistical investigations to express both what they expect and why. After gathering or analysing data, they can then go back and discuss their a priori expectations in light of their emergent understanding, with particular attention paid to the extent to which their intuitive expectations had emphasized centers to the neglect of spread (Reading and Shaughnessy, 2004). With displays of data, students need to create their own graphs to either highlight or disguise variation, depending on the context of the situation. They also need to evaluate displays and compare distributions in ways that take an aggregate view of data, considering shape and spread in addition to centers (Shaughnessy, Ciancetta, Best, and Canada, 2004). From discussions about probabilistic and statistical situations, students' interpretations of variation emerge as they speculate on both causes and effects of variation and also on ways of influencing variation and expectations. Intuitively, physical causes are often the easiest for students to imagine, but an understanding of randomness as an inherent cause of variation can be a more tenuous concept. Effects of variation were seen as a part of the conceptual framework in terms of the effects on students' perceptions and decisions. For example, if students perceive that the presence of variation precludes any kind of confidence in making inferences, they may likewise be reluctant to make decisions based on data. Influences on expectation and variation generally reflect the sizes of samples used in conducting investigations, or the numbers of trials performed in simulations.

METHODOLOGY

The thirty subjects in the study of EPSTs (24 women, 6 men) were enrolled in a ten-week preservice course at a university in the northwestern United States designed to give prospective teachers a hands-on, activity-based mathematics foundation in geometry and probability and statistics. During the first week of the course, prior to instruction in probability and statistics, subjects took an in-class survey (called a PreSurvey) designed to elicit their understanding on a range of questions about sampling, data and graphs, and probability. The probability question (PreSurvey Q7c) that relates to the current paper concerned six sets of fifty flips of a fair coin. For each of the six sets, students were asked how many times out of the fifty flips the coin might land heads-up. They were also asked why they had chosen the numbers they did. Following the PreSurveys but prior to the class instruction on probability and statistics, individual interviews were conducted with ten subjects to allow further probing of their thinking. After instructional interventions took place in class, a similar PostSurvey question (PostSurvey Q1c) was asked concerning six sets of fifty spins of a fair half-black and half-white spinner. For each of the six sets, students were asked how many times out of the fifty spins the pointer might land on black, and also why they had made the choices they did. Finally, after the PostSurveys the same students who had been earlier interviewed were interviewed once again.

The class interventions were a series of small-group and whole-class activities and simulations that engaged all three contexts of data and graphs, sampling, and probability situations, and were designed to elicit discussion about variation. The two activities comprising the Class Intervention for the context of data and graphs were called “Four Questions” and “Body Measurements.” The “Four Questions” activity offered a good opportunity to discuss both average and spread in data sets, and Sam started the class exploration of statistics in the fifth week by having the entire class gather data from one another in response to four questions:

> *Four Questions Activity Prompt*
> How many pets do you have?
> How many years have you lived in this city (to nearest half-year) ?
> How many people are in your household?
> How much change (in coins) do you have today?

After graphing the data in different ways, the class had a discussion about levels of detail provided by each type of graph and about what were “typical” values for an individual student or for the whole class. The tension between centers and spread of data was one theme to emerge from the discussion over graphs from the “Four Questions” activity. For the second activity in the class intervention focusing on the context of data and graphs, “Body Measurements,” everyone’s own armspan, height, handspan, head circumference, and pulse rate per minute were recorded. Also, all students in class measured a designated person’s armspan, to gather data from a repeated-measurements experiment. Again, we had a class discussion about the data and graphs for the body measurements, this time focusing more on causes of variation.

In the seventh week of class, the two activities “Known Mixture” and “Unknown Mixture” were done with Sam’s students. Prior to the “Known Mixture,” we started with a general discussion of what samples were, who uses samples, and what samples were good for. Then the following scenario for the Known Mixture Activity was given as a part of a handout:

> *Known Mixture Activity Prompt*
> The band at Johnson Middle School has 100 members, 70 females and 30 males.
> To plan this year’s field trip, the band wants to put together a committee of 10 band members. To be fair, they decide to choose the committee members by putting the names of all the band members in a hat and then they randomly draw out 10 names.

The class discussed initial expectations for this scenario, especially focusing on what would happen if the random draw of 10 names were to be repeated thirty times. After students talked about predictions for drawing thirty samples each of size ten, we simulated this activity using chips in a jar. Actual data was gathered and graphed. Then we had a discussion about how the graphs of the predicted data compared to one another, how the graphs of the actual data compared to one another, and also how the predicted graphs compared to the actual graphs. We then made a transition into the second activity in this intervention, the Unknown Mixture. Now we had larger jars, each containing 1000 chips of yellow and green with the same mixture. However, the exact mixture was not known to the class (it was actually 550 yellows and 450 greens). The students were asked to decide in their groups what sample size they wanted to use and how many samples they wanted to draw. Then they were to carry out their plans, do the sampling, graph the results, and make some conjectures about the true mixture in the jar. After the simulation was carried out, we had a class discussion about the different choices made in sampling, the class results, and we tried to forge a class consensus about what the true mixture was.

There were two activities that made up this intervention, "Cereal Boxes" and "The River Crossing Game." These were chosen specifically because of the probability aspects involved in the activities. Cereal Boxes relies on the use of spinners and River Crossing on the use of dice as random generators, and these two activities were the main ones done in class involving random devices. The first activity for this intervention (Cereal Boxes) actually took place in the first class session of week 2, just before we gathered data for Body Measurements. As explained earlier, there was considerable overlap in the three contexts, and Cereal Boxes is a good example of this overlap. Cereal Boxes is sample-until scenario, assuming that any of five different stickers can be obtained within each box of cereal, and that the five stickers have equal chances of being obtained. The question is about how many boxes would need to be opened to obtain all five stickers, and the situation was simulated by using an equal-area five-region spinner. Cereal Boxes brings together probability, sampling, and data and graphs in a way that highlights variation. The second activity for this intervention (River Crossing Game) involved finding the sum of two dice. Both the Cereal Boxes activity and River Crossing Game are part of the *Math and the Mind's Eye* curriculum (Shaughnessy and Arcidiacono, 1993). Using two players, each player receives 12 chips to place on their side of a "river," along spaces marked 1 through 12. After configuring their chips in an initial arrangement along the spaces, players take turns tossing a pair of dice. If either player has any chips on the space showing the total for the dice, one chip can "cross the river" and be removed from the board. The winning player was the first one to remove all the chips on his or her side. As with Cereal Boxes, in the River Crossing Game we made predictions, gathered and graphed data, and discussed results.

The activities in the all the interventions were designed to elicit discussion about variation. For instance, the intervention on data and graphs included different types of graphs and the amounts of variation they showed. Body Measurements got at the ideas behind repeated measurements, as did the muffin weight questions on the PostInterview. The Known and Unknown mixtures had students actually draw chips from a container to experience drawing candies from Large and Small Jars. Cereal Boxes and the River Crossing Game had students use traditional random generators such as spinners and dice to get a sense of what was likely in a probability context. The *Fathom* software (Finzer, 2001) was used to aid graphical representations and to extend the simulations that the class had already participated in manually.

RESULTS

Both parts of the probability question (*what* students expected and *why*) were taken into consideration for coding purposes, primarily to retain consistency with an analogous rubric derived for a similar question asked in a sampling context (Shaughnessy *et al*., 2004). The rubric places a higher value on responses that integrate proportional reasoning as well as variation. The codes and class results for this subquestion are presented in Table 1. Only inappropriate choices for listing *what* was expected (or blank answers) were coded at Level 0. Deciding what would constitute an appropriate choice for the results on six sets of flips or spins involves making a judgment call, and the subcodes used for this subquestion question help identify inappropriate choices as (W)ide, (N)arrow, (H)igh or (L)ow. Of the 30 students enrolled in the class, 27 were in attendance to complete the PreSurvey and 29 completed the PostSurvey.

Table 1: Results for PreSurvey Q7c and PostSurvey Q1c

Code Level	Description of Category	Number of Students (Pre)	Number of Students (Post)
L3	Appropriate choice and Explanation explicitly involves proportional reasoning as well as variation	2 (7.4%)	9 (31.0%)
L2	Appropriate choice and Explanation reflects proportional reasoning or notions of spread	10 (37.0%)	15 (51.7%)
L1	Appropriate choice and Explanation left blank or lacks any specific reasons relating to details of the distribution	4 (14.8%)	3 (10.3%)
L0	Inappropriate choice (Regardless of Explanation) W(ide) = Range > 19, N(arrow) = Range < 2, H(igh) = Choices > 24, L(ow) = Choices < 26	11 (40.7%)	2 (6.9%)

Of the eleven inappropriate PreSurvey responses, one was narrow, one was high, one was low, and four were wide (the remaining were left blank). Of the two inappropriate PostSurvey responses, both were wide. A few of the Level 0 exemplars are:

[L0] Alice (Q7c) {25, 25, 25, 25, 25, 25} I don't see how the chances of getting heads will change if he does more sets of 50 flips.
Brita (Q7c) {7, 21, 23, 25, 29, 31} I chose numbers close to 25 because I think with a 50% probability, the results would come out pretty close to 25. I put the oddball 7 in for fun, because there is always that element of chance.
Susie (Q1c) {5, 15, 30, 40, 45, 50} It is chance.

Alice narrow response is obviously over-influenced by the expected value, but it seems surprising that more subjects did *not* put all 25s for their choices in the PreSurvey, given results discussed by other researchers (e.g., Shaughnessy *et al*., 1999). Brita's choice of 7 is extremely unlikely and makes her overall range too wide, although her upper bound of 31 is plausible. Susie's choices are too extreme at both the upper and lower ends.

Level 1 responses had appropriate choices for *what* was expected but the reasons *why* did not specifically reflect distributional thinking:

[L1] Carrie (Q7c) {22, 23, 24, 25, 26, 27} It's usually not the same.
Maria (Q1c) {20, 23, 25, 25, 26, 30} I think he will hit 25/50 one time. The rest of the times, he will be close, but not exactly on. Also I think he will be controlling the way he hits the spinner more on the second day, which accounts for no 23 or 28.

Maria points to *causes* of variation in noting the physical manipulation of the spinner, and other subjects also seemed to indicate that spinners are not viewed as true random devices because the user can ostensibly control outcomes by altering the way the pointer is spun.

The Level 2 responses included an indication of reasoning using an average, proportion, or a measure of spread:

[L2] Sofia (Q7c) {20, 20, 24, 25, 26, 27} Because they average to about 25.
Sally (Q7c) {22, 23, 24, 25, 26, 27} They are all close to 25, ½ of 50.
Leila (Q1c) {23, 24, 24, 25, 25, 26} The numbers are pretty close to half or 50%.
Rocky (Q1c) {20, 22, 23, 27, 28, 30} These numbers represent a distribution across a range of likely results.

Note how Sally's Level 2 response includes the same choices as in Carrie's Level 1 response shown earlier. However, Sally gives more specificity than Carrie in describing her reasoning, which is proportional in Sally's case. Rocky doesn't include the expected value in his choices, but feels he has given a likely range, and his sophisticated language borders on a Level 3 response.

What distinguished the Level 3 responses was an indication of reasoning using *both* centers and spread:

[L3] Maya (Q7c) {23, 24, 25, 25, 25, 26} Because there should be variation around the mean. The average should be 25.

Ross (Q7c) {22, 23, 24, 26, 27, 28} While 25 flips are likely to be heads, in reality some variation is likely. My numbers represent a range that averages 25.

Sally (Q1c) {21, 24, 25, 26, 28, 29} All numbers are 25 or close to 25 (1/2 of # of spins). Not all are 25 in order to account for variation.

Daisy (Q1c) {18, 21, 24, 26, 28, 31} Because they are close to the 50% chance to get 25 hits of black allowing for variation due to random spinning hits. But none of the #'s are too high or too low (far from the 25) which would be hard to hit based on the 50% odds.

There were more Level 3 responses in the PostSurvey than in the PreSurvey, and the relative sophistication is apparent as subjects reconcile the tension of having results be close to an average value while also acknowledging the presence of variation. Also, there were more subjects in the PostSurvey than in the PreSurvey whose choices did not include the expected value of 25 (such as Susie, Rocky, and Daisy), suggesting that the class experiences helped counter the natural tendency to pin expectations solely to a theoretical average without an appreciation of the variation in repeated trials. Average class performance on the task also increased, from a mean of 1.11 in the PreSurvey to 2.07 in the PostSurvey.

DISCUSSION

It was clear from subsequent discussions in class that students initially felt uncomfortable venturing a guess for six results, often expressing their perception that it was difficult to guess correctly, or that "anything could happen." One interesting feature in the results was the low incidence of narrow responses from the EPSTs, which contrasts with research involving 93 high schoolers and a sampling task, whereby almost 26% of responses were narrow (Shaughnessy *et al*., 2004). Another interesting feature in responses was the tendency avoid repeating choices when making predictions for multiple trials in the PostSurvey. For example, in giving choices for on the PreSurvey, most students gave some repeated values for their choices, such as James' (20, 22, 25, 25, 26, 27) or Daisy's and Emma's (23, 24, 25, 25, 26, 27). There is nothing wrong per se with having repeated values in six conjectured results, but most of the PostSurvey choices contained no repeated values. Ross's choices on the PostSurvey were (22, 23, 24, 26, 27, 28), and he commented how his choices still were "similar, but not identical" and how "there's no repeats." Sandy said of her PostSurvey choices (20, 23, 24, 26, 28, 29) that "they could repeat, but I just did a range – from 20 to 30, just to choose... different numbers, but still somewhere in that range." Ross and Sandy, like many others, also seemed to deliberately avoid including the expected value among their choices in the PostSurvey.

One reasonable hypothesis for why the class as a whole seemed to shift to a greater awareness of variation in results stemming from a probability experiment is that their collective engagement in the activities, simulations, and subsequent class discussions made them more expectant of variability in data. More than half of the students referred directly to class activities or simulations in explaining their thinking, and comments like these are representative:

Dixie: In our class experiments, when I repeated an experiment you'd often have some new variations pop into the picture but the central probability remains the same.

Rosie: Because we had the same activity in class, the same concept: The more chances or tries you have more different answers you can get.

Frida: I based it on the activities we have done in class w/ computer program as well as hands-on activities.

Sheila: I know this because we saw it on the computer program in class.

IMPLICATIONS

Implications for teaching EPSTs include the suggestion that having hands-on activities, bolstered by small-group and whole-class discussion focused specifically on variation, can be a powerful way to move EPSTs toward a better appreciation of how variation plays a role in statistical thinking. The class interventions involved all three main aspects of understanding variation (expecting, displaying, and interpreting) in the contexts of sampling, data and graphs, and probability situations, all of which are important for elementary schoolchildren to address. If school teachers are to shape their lessons so as to encourage statistical thinking in their own students, then university teacher training programs need to provide an environment where preservice teachers can learn in a similar way that they themselves will aim to teach. In the environment where this research took place, the teaching philosophy of the course encouraged a great deal of discourse among students, which served to naturally provide springboards from which class discussions of variation could emerge. Also, a key design component of the surveys, interviews, and class interventions was that EPSTs were expected to make conjectures and discuss their reasoning before actually doing any activities. By laying out ahead of time what everyone in class thinks, groundwork can be established for making comparisons after actual data has been collected by doing the probability experiments. The computer simulations, brought out only after students have physically run simulations themselves, also seem to hold much promise for getting EPSTs to understand long term trends. EPSTs, like the children they will one day teach, need to investigate variation in probability settings by conjecturing, reasoning together, doing experiments, and discussing findings. The task for teacher educators is to continue to develop ways to structure their classes in ways that support EPSTs reasoning about variation.

More research is warranted to discern the best ways to use software in connection with hands-on activities in ways that promote an understanding of variation and also reflect what the preservice teachers themselves will aspire to do with their own students in the schools. Since university teacher preparation programs are concerned with the knowledge of teacher candidates, it makes sense to wonder about conceptions of variation held by preservice teachers. If a goal is for teachers to provide students with authentic, inquiry-based tasks meant to develop children's reasoning about variation, then a natural step in achieving this goal is to improve teacher training courses. Thus, by discerning components of preservice teachers' reasoning, teacher educators can better design university experiences that promote an understanding of variation for preservice teachers, as well as an understanding on how precollege students come to learn this topic.

As research in the field of statistics education advances, one goal is that teacher education can improve not only the subject matter knowledge of EPSTs, but also the pedagogical content knowledge of teaching about variation. Steps toward improved pedagogical content knowledge can certainly be informed by recent research about how precollege students learn. Meanwhile, steps toward improved subject matter knowledge can be informed by a consideration of what are the conceptions of variation held by preservice teachers as they enter university programs. Collective discourse in the class, bolstered by activities and simulations targeted at eliciting conceptions of variation and developing these concepts, hold promise as ways of building EPSTs knowledge while also reflecting the kinds of practice they themselves will want to demonstrate in their own classrooms.

REFERENCES

Finzer, W. (2001). *Fathom*! (Version 1.16). Emeryville, CA: Key Curriculum Press.

Reading, C. and Shaughnessy, J.M. (2004). Reasoning about variation. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shaughnessy, M. and Arcidiacono, M. (1993). *Visual Encounters with Chance (Unit VIII, Math and the Mind's Eye)*. Salem, OR: The Math Learning Center.

Shaughnessy, J.M. and Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society,* Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Shaughnessy, J.M., Ciancetta, M., Best, K., and Canada, D. (2004). *Students' Attention to Variability when Comparing Distributions.* Paper Presented at the Research Presession of the 82nd Annual Meeting of the National Council of Teachers of Mathematics, Philadelphia, PA.

Watson, J. and Moritz, J. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics,* 37(2), 145-168.