MIDDLE SCHOOL STUDENTS' AND PRESERVICE TEACHERS' INFORMAL INFERENCES ABOUT RANDOM WALKS

Dr. Dan Canada
Department of Mathematics, Eastern Washington University
Cheney, WA 99004 (USA)
dcanada@ewu.edu

The purpose of this paper is to report on the emerging results of a project that used an instructional intervention designed to improve middle school students' and preservice teachers' informal expectations of variability in a two-dimensional context based on taking a random walk. Specifically, one aim of the project was to compare how participants reasoned about variability to make informal inferences both before and after modelling a task physically. A simultaneous goal was to have participants pursue their own additional questions, beyond the initial prompts given, inspired by an analysis of the data they had gathered. In the work with preservice teachers, computer simulations were also used to generate more data, which in turn gave rise to more questions about variability and probability.

INTRODUCTION & BACKGROUND

To properly situate the work described in this paper, considering the variability in taking a "Random Walk", it may be helpful to know the contextual backdrop. While serving as a Fulbright Scholar to Tanzania for ten months, I had previously worked with some Tanzanian middle school students on an instructional intervention about "Falling Raindrops", which asked where sixteen drops might fall across a 4x4 grid (Canada, 2019). The underlying task for the Falling Raindrops intervention was based on the work of others (e.g. Engel & Sedlmeier, 2005; Green, 1982; Piaget & Inhelder, 1975), and this motivated the current project, which again aimed to look at using a two-dimensional probability phenomenon to motivate thinking about and questions related to variability in data. Thus, the project for Random Walks, similar to that of Falling Raindrops (Canada, 2019), reflected the first four aspects of Engel's (2002) five-step procedure: Making initial observations and conjectures regarding a specific probabilistic phenomenon, developing a model to simulate the phenomena, gathering data from that simulation, and comparing subsequent results to initial predictions. The fifth step (formal mathematical computation and analysis) was deemed beyond the skillset of the project's subjects.

Although the idea for the Random Walk intervention began informally in rural Tanzania with middle school students, it continued with a more formal structure at a university in the United States, with preservice teachers (PSTs) training for licensure in the middle grades. Whereas in Tanzania the intervention did not use technology, a subsequent extension was added to the work with the PSTs in the United States, which did include computer simulations. Therefore, this paper will first expand on the design of the intervention by situating the task as it was done in Tanzania, without technology. This design carried over analogously to the work with PSTs, in the sense that the physical simulations done in the United States did not differ substantively from what was done in Tanzania. Next, the paper shares some informal results from the students in Tanzania, and how those responses compared with emerging results from the PSTs. The impact of using computer simulations with the PSTs is also discussed, while the conclusion points to the robust statistical inquiry that arose from getting people involved in stochastic processes even when technology is absent.

DESIGN

The instigation for this project involving Random Walks occurred with eleven middle grades students in a rural part of Tanzania. The students' background in probability and statistics seemed limited to describing relative likelihoods, calculating probabilities of simple (single) events, and creating and interpreting basic graphs. They also knew how to find some summary statistics such as range, median, mode, and mean. To set up the premise for the task, we went outside of the classroom, where a long segment was dug down the middle of the straight dirt path alongside the building. At one end of the segment (the Starting Point), a student stood facing forward toward the remaining segment on the path ahead. A coin was produced, with the premise that a Heads (H) would result in taking a step

In: Kaplan, J. & Luebke, K. (Ed.) (2024). *Connecting data and people for inclusive statistics and data science education. Proceedings of the Roundtable conference of the International Association for Statistics Education (IASE), July 2024, Auckland, New Zealand.*

diagonally in one direction (i.e. forward and to the left), while a Tails (T) would result in taking a step diagonally in the other direction (i.e. forward and to the right). The question posed was simply: How many steps do you think it will take to get to one edge of the path or the other?

Whereas the path had relatively parallel edges, with the building on one edge and grass on the other, it became clear immediately that some students were taking longer steps than others, and so one of the first clarifications that emerged was to standardize the step lengths. We did this by creating three lanes on each side of the primary segment we had dug down the middle of the path using stones. The net effect was that we had the primary (middle) segment, the two outer edges of the path, and then two lines of rocks on either side of the middle segment. Putting a student at the starting point, two questions were asked: “Which direction do you think the student will step next?” and “How many steps do think it will take to reach an outer edge?” Students also provided rationale for their initial expectations. Then, using lined paper and pencil, the entire walk was recorded, which students called a trace. Figure 1, while not a photo reproduction of the trace, gives the essential information of this first random walk:

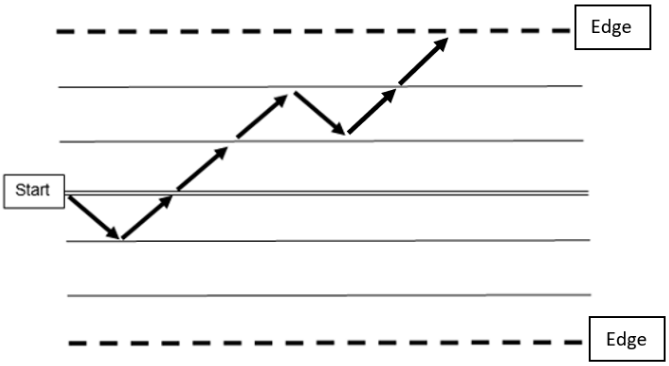

Figure 1. The trace of a single trial (of a Random Walk)

The immediate number to emerge from Figure 1 was that it took 7 steps to end that random walk, which became known as the first trial of what eventually was called the *Out of Bounds in Three* game. We subsequently talked about doing a batch of multiple trials.

Having seen how to conduct and record a single trial, students got into four groups of two each, and one group of three. As each group approached the path, one student took the actual steps, while the other student (or other two) flipped the coin and recorded the walk (trace). After each group had done at least a couple of trials, taking turns either doing the steps or making the recordings, we went indoors and posted the results from all the trials done so far across the classroom walls. Including the first trial, done together as a large group, as well the trials from the smaller groups, we had what we called a batch of fifteen trial results. We then reflected on this batch, prompted by what was noticed and what was wondered, and students were asked how the resultant data from the physical simulations compared with what they thought might happen beforehand.

One thread of inquiry led students to wonder how another batch (of fifteen trials) would compare to what they had just done. For example, the average number of steps needed for the fifteen trials in our first batch was about 10: Would another batch of fifteen trials yield a similar average? By simply using recording paper and coins, we were indeed able to produce a second batch of fifteen trial results, and then a third batch. In looking at the aggregate (all forty-five trials), as well as comparing the first batch to the second to the third, a lively conversation about statistics and probability ensued, some of which is elucidated further in the next section. Of key interest were responses from students when they were asked about what was surprising to them, and also about how likely certain trial results were adjudged: For example, how likely would they consider a trial that took only 3 steps? How about a trial that took longer than 20 steps? At the end of the intervention, students predicted again how long a single trial might take.

In continuing the use of the Random Walks intervention with preservice teachers (PSTs) in the United States, the preceding design worked in a structurally similar manner. Most recently, in a class of 28 PSTs taking a course in mathematics for elementary and middle grade teachers, the prerequisite knowledge of probability and statistics was minimal, with the overall effect being that their incoming skills were similar to those of the middle school students I had worked with previously. Instead of going outdoors, we filed into the adjacent hallways of our building, which conveniently had square tiles parallel to a lengthy wall. We used masking tape to mark the edges and the middle of our lanes along the floor, with students taking diagonal steps from the corners of one square tile to another. The results of their random walks were then recorded on graph paper.

In accordance with the research aim of seeing how direct simulation impacted PST predictions, the responses of PSTs were recorded both before and after participation in physically doing multiple random walks. As far as the initial questions and subsequent discussion of results with the PSTs, the flow was very similar to what was done with the middle school students. For example, PSTs were asked prior to gathering results what they expected and why. When the first batch of results were being discussed as a whole group, PSTs were prompted about what they noticed, what they wondered about, and what was surprising to them. One difference in doing Random Walks with the PSTs is that, because we had access to computer simulations, we were able to expand our investigation easily into an Out of Bounds in Four scenario. This scenario takes longer to enact physically since the average number of steps is greater than in the Out of Bounds in Three version but can be enacted rapidly on the computer. Finally, by using computer simulations we were able to pursue a more thorough investigation into some of the questions the PSTs (and middle school students) had asked. Specifically, the greater amounts of data generated by the computer led to an increase in confidence in what the class expected.

RESULTS

Middle School Students

Initially, from the conversation with the middle school students in Tanzania, a mix of predictions about their first trial ranged from overly deterministic ("He'll get it in 3 steps because he wants to go that way") to overly fatalistic ("He'll get to the edge one way or another, it's in God's hands when") to a resignation of not knowing ("Who can tell, he might just go back and forth forever"). Whereas this was an informal activity that spawned further work with the PSTs later on, I simply annotated what the students had to say and the kinds of questions they asked.

Although predictions before doing the trials ranged from 3 steps to infinity, most were willing to suggest that the number of steps should be large, such as in the teens or twenties. Once we looked at our first batch of 15 trials, we took the number of steps needed in each trial and made a dotplot on a large piece of paper of that data. Although Figure 2 is not an actual photo of the dotplot made, it faithfully captures the results for how many steps each of 15 trials took.

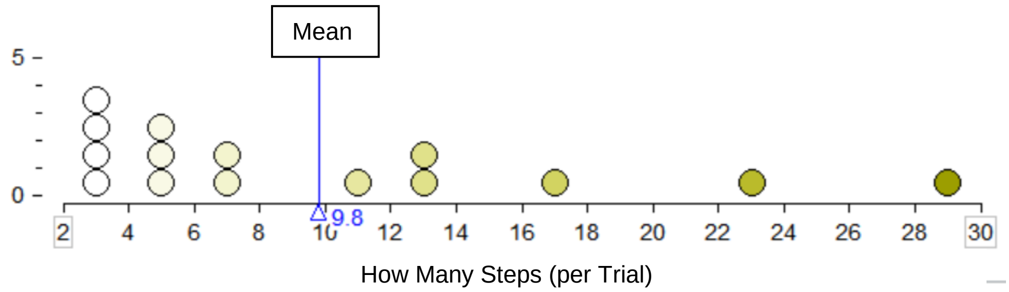

Figure 2. Students' first batch of 15 trials

Indeed, we saw that 4 trials ended in 3 steps, and 3 trials ended in the teens, with 2 trials ending in the twenties (the other trials were between 5 and 11 steps, inclusive). The mean number of steps was 9.8 steps, which seemed surprisingly low to some students, but not as low as the median of 7 steps. Other students felt vindicated in their prediction of 3 steps, since "That happened the most". Some students thought the maximum (29 steps for our first batch shown in Figure 2) was lower than they expected.

After the second batch of 15 trials was completed, again with a dotplot showing the number of steps taken for each trial, students commented on similarities among and differences between the two batches: The mode was at 5 steps, the mean was 7.5 steps, the median was again 7 steps, and the maximum was only 19 steps. Upon comparing the second batch (of 15 trials) with the first batch, there was considerable excitement and debate amongst the students. For example, maybe the best guess for how long an arbitrary trial might take was less than 9.8 steps (the mean from the first batch). Perhaps the maximum of 29 steps from the first batch was highly unusual, and even the minimal amount of 3 steps might not be the most likely. Students sought to reconcile a focus on measures of center with a consideration of measures of spread when making inferences about what to expect, reflecting aspects about comparing distributions mentioned in other research (e.g. Konold & Higgins, 2002; Leavy, 2006; Makar & Canada, 2005; Makar & Confrey 2007; Shaughnessy, et al., 2004;).

My time was limited with this group of students, so I focused instead on the other questions that arose, which were not inspired by the distributions such as those shown in Figure 2. In particular, students were looking at the actual trials themselves (the traces of the actual paths), for which we had dozens displayed across the walls: Fifteen trials for the first batch, fifteen for the second batch, etc. What captured some students' attention was not simply how many steps a trial took, but a specific trait about the traces that they called a "turn", meaning a simple change in direction. For instance, in the portrayal of a trace for a trial taking 17 steps, as depicted in Figure 3, students counted exactly seven turns, where they were headed in one direction, but the next step was in a different direction. They could easily identify these turns as obvious vertices or corners on the traces. Clearly a minimal 3-step trial would have no turns, and so a trial longer than the minimum would involve at least one turn. What the students wondered, to reformulate their question after it got distilled through conversation, was if there was a relation between the total number of steps a trial might take and how many turns it had?

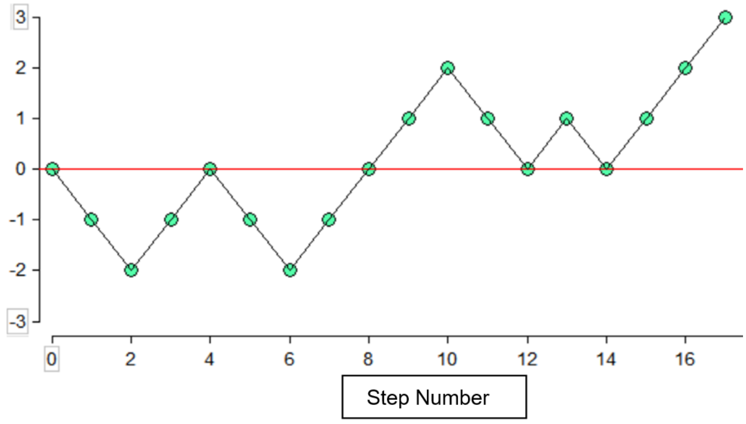

Figure 3. The trace for trial of 17 steps, showing seven "turns"

Other questions that arose from inspecting the many traces we had on display included: How many times did a trace cross the center line? How could we guess the number of times crossing the center line if we knew the total number of steps? Does a trace usually spend more time on one side of the center line than on the other? With the limited data we had, confidence in the conjectures for those questions was low.

At the end of our discussion of all the 45 trials we had done (across three batches of 15 trials per batch), we imagined the following scenario: A student who was not with us today will show up tomorrow and do a random walk as we did. How many steps do think it will take? In revisiting the same question asked at the outset, there seemed to be more attention to variability in responses, such as from these three students -

Patrice: It probably would be between 7 and 11 steps
Neema: Wouldn't surprise me if it was less than 8 steps
Godwin: Maybe 8 or 9 steps

I selected those simply to illustrate moving beyond a point estimate and giving a single number, to more a likely range based on what they had seen. Students seemed less likely to think "No one can tell" and more confident in giving a response based on the data available.

To wrap up the time we had together, I posited one more question based on our imaginary student who was not with us, asking: "Suppose we told the student to go home and do the random walk, record the trace, and bring it back here. The student produced the following (Figure 4)." The question posed was: "Do you think the student just made up those results without actually doing the walk, or are those real results, or can you not tell?" Of the eleven middle school students, seven felt strongly that Figure 4 showed fake results, with the consensus reasoning being that none of the trials we had done exhibited that kind of pattern "where it's one above, then one below, over and over again". Three students suggested it seemed like real results, in the sense that "it could happen", and one student argued that it was impossible to tell "for sure".

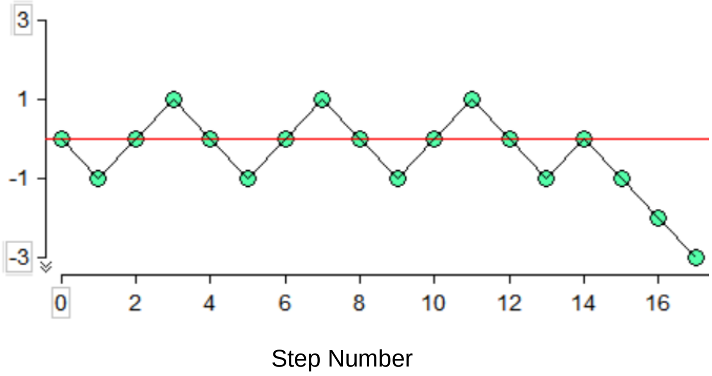

Figure 4. Real or Fake? A purported trial of 17 steps.

My lasting thought with this group of students was a wish to have more data so that, rather than being able to tell "for sure", we could at least speak with more confidence about what was likely.

Preservice Teachers

The experience with the middle school students was largely informal due to its somewhat spontaneous nature, but my main takeaways were twofold: One, I was impressed at how much interest was generated, the questions that emerged and the discussion that ensued, all without the use of technology. The figures presented in this paper were done on a computer, but in the field it was all done by hand, including the simulations themselves. Two, I was keen on trying this with my PSTs, hewing as closely as I could to what had unfolded in Tanzania. That is, even knowing the technology was available for the PSTs to use, I first wanted to do as much as possible without the use of computer simulations. What follows is a summary of how PSTs responded, not only to the same kinds of prompts that were used with the middle school students, but also regarding some new avenues for investigation that arose after some computer simulations were introduced.

Prior to doing any physical simulations, the PSTs' predictions were similar to those of the middle school students qualitatively. Only three had predictions beyond a single point estimate, such as "*Around 10 steps*" or "*Maybe 12 steps, give or take*". Ten predictions were just single numerical estimates, which were all less than nine steps (and four of those ten students put "*Three steps*"). Another eleven predictions were point estimates that were greater than or equal to nine steps, with the four remaining responses being that it either could go on forever or that there was no way to predict.

After doing the physical simulations, we ended up with three batches of thirty trials each, and this was due to having more people working together. By using the space along the hallway walls, we displayed all 90 trials. I asked PSTs in their small groups to write down as many observations and questions as they could think of, with an added prompt to consider "What do you think your students would notice or wonder about?" In subsequent whole-class discussion, it became apparent that many of the same characteristics of the traces of individual trials were of interest to the PSTs as to the middle school students. For instance, there were several comments about how many times a trace might cross the center line, or if a trace spent more time on one side of the center line than the other, and also about the number of turns (which the PSTs called "switchbacks"). We also invoked the scenario of the imagined student who shows up to do one more trial, asking predictions for the total number of steps needed. There was a clear majority (19 of the 28 PSTs) whose responses showed a more appropriate sensitivity to the variability of the situation, such as: "*Around 8 to 10*", and "*Pretty close to 9*", and "*Maybe 10, give or take a few steps*."

Lastly, PSTs also responded to the Real or Fake scenario, offering similar comments as those made by the middle school students. One noteworthy difference was in the way that PSTs talked about the uniform size of the switchbacks in the Real or Fake scenario. As seen in Figure 4, each of the seven switchbacks is only size 1 (meaning only one step above or below the center line). The idea of a size of a switchback, at first a reference to how many steps away a trace was from the center line, became reinterpreted by the PSTs to reflect how close a trace was to ending (that is, to the edge of the path as from the physical simulations). What PSTs attended to and wondered about was what they called a "major switchback": By this they referred to when a trace got within 1 step of finishing but ended up on the other side of the middle line. Looking at Figure 3, notice how by the 2nd step the person is just one step away from being finished – Just one more step "*in the same direction*" would end the trial. The end of the trial, however, is not on the same side of the middle line as the 2nd step: instead the trace goes clear across the middle line to finish on the other side. In fact, Figure 3 depicts two major switchbacks: at the 2nd step and again at the 6th step. It seemed to most PSTs that "*if you got that close, you'd end up on that side*". As they saw in Figure 3, by the 10th step the person was again close to finishing (+2), and even though there were still three other switchbacks the trial did indeed end on that side of the path (+3). Thus, in the language of the PSTs, Figure 3 exhibits seven total switchbacks, two of which are major switchbacks. Examples of the questions they brought up include: How likely would it be that a trial lasting 9 steps would have at least 2 switchbacks? At least 1 major switchback? We used computer simulations to generate more data, which allowed PSTs to speak with greater confidence about empirical probabilities of these questions that were raised. The use of empirical probabilities in investigating the likelihood of major switchbacks was particularly relevant to these PSTs, since they did not have the mathematical prerequisites to compute a theoretical probability.

Most importantly, we were able to produce more batches of the same number of trials. That is, instead of only three batches (each of thirty trials) that we had produced via physical simulation, we used the Fathom software (Finzer, 2001) to get dozens of such batches. Looking only at the mean number of steps needed per batch, collecting and plotting that statistic laid the foundation for the Central Limit Theorem. For each of the fifty batches, the mean number of steps for the constituent thirty trials was collected and put into a histogram. The overall mean for the data was 9.01 steps. We are planning on using the data from the fifty batches when we focus more on confidence intervals. For now, the PSTs were satisfied that even though individual trials vary in terms of steps needed, and even though batches of thirty trials vary in their mean number of steps needed, over the course of looking at many batches we're still coming in at an average of about 9 steps for the Out of Bounds in Three scenario.

DISCUSSION AND CONCLUSION

In the initial work with this scenario, occurring with middle school students in Tanzania, I had only anticipated a discussion on the primary question of "How many steps do you think it will take?", which in itself can be a powerful generator for balancing the idea of an average versus variability over many trials. By only using hand-generated graphs to record data from multiple trials done via physical simulation, students reflected on the variability apparent across trials, and by comparing several batches of 15 trials a better sense of what to expect emerged. More importantly to the endeavor of fostering statistical inquiry, student attention to the traces of the walks represented in each trial became a line of questioning that deepened the need for more data. We were limited in our estimates of empirical probability to the total number of trials we had collected, but students still reasoned about the estimated likelihood of things like how many "turns" a trace might have for trials lasting a certain number of steps. Overall, the Out of Bounds in Three proved to be a useful scenario for doing physical simulations, in the sense that the outliers of walks taking a long number of steps were relatively infrequent, so that many trials could be completed in a short amount of time.

With the PSTs, many of their conjectures and observations were reflective of those given by the middle school students, particularly in the initial part of the project which did not rely on technology. That is, by restricting our simulations to only those done physically, and using graphs made by hand, the PSTs' dialogue regarding the expected number of steps needed included more attention to and consideration of the variability inherent in the scenario. Moreover, the PSTs came up with additional questions about the characteristics of the trace for single trials, going beyond just the number of switchbacks to wonder about the likelihood of a major switchback. They also relied on their insight into variability for trials to gain confidence in deciding if a purported trial was real or fake.

The next phase of this project will be to develop a rubric based on an appropriate conceptual framework for a more systematic way of qualitatively characterizing the responses, and to then use that rubric with more PSTs in developing a fuller picture of their understanding of variability in this context of random walks. Meanwhile, from the emerging responses by both middle grade students and PSTs, the importance of the Random Walk scenario to statistics education is in how it provides an entry point for people to engage in authentic inquiry about a two-dimensional probability context rife with variability, and to do so in a way that initially requires few technological resources.

REFERENCES

Canada, D. (2019). Middle school preservice teachers' informal inferences about falling raindrops. *Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE)*. International Association for Statistical Education.

Engel, J. (2002). Activity-based statistics, computer simulation, and formal mathematics. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. International Statistical Institute.

Engel, J. & Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *Zentralblatt füur Didaktik der Mathematik*, *37*(16)

Finzer, W. (2001). Fathom! (Version 1.16) [Computer Software]. Emeryville, CA: Key Curriculum Press.

Green, D. R. (1982). A survey of probability concepts in 3000 students aged 11–16 years. In D. R. Grey (ed.), *Proceedings of the First International Conference on Teaching Statistics*, Teaching Statistics Trust, 766–783.

Konold, C., & Higgins, T. (2002). Highlights of related research. In S.J. Russell, D. Schifter, & V. Bastable (Eds.), *Developing mathematical ideas: Working with data*, (pp. 165-201). Dale Seymour Publications.

Leavy, A. (2006). Using data comparisons to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal* 5(2), 27-45.

Makar, K., & Canada, D. (2005). Preservice teachers' conceptions of variation. In Chick, H. L., & Vincent, J. L. (Eds.) *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education.* International Group for the Psychology of Mathematics Education.

Makar, K., & Confrey, J. (2007). Beyond the bar graph: Teaching informal statistical inference in primary school. Paper presented at the *5th Conference on Statistical Reasoning, Literacy, and Thinking (SRTL 5).* University of Warwick, UK.

Piaget J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. Routledge & Kegan Paul.

Shaughnessy, J., Ciancetta, M., Best, K. & 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.