TEACHER ASSISTANCE IN MODELING CHANCE PROCESSES

Panchompoo Wisittanawat and Richard Lehrer
Vanderbilt University
panchompoo.wisittanawat@vanderbilt.edu

This research characterizes recurrent forms of teacher assistance in a sixth-grade class as students invented and revised models. The primary focus is on how the teacher supported students to construct an image of repeated process and to subsequently employ this understanding as a guide to considering model fit and criteria for good models. Sources of data include field notes and video recordings of whole-class and small-group conversations. The analytic approach employs methods of constant comparison but is also informed by findings from end-of-year interviews with students.

BACKGROUND

One goal of statistics education is to support students to understand how inferences can be made in light of variability. However, formal approaches to statistical inference require understandings of probability density functions that are typically outside the scope of K-12 mathematics education. Instead of relying on these formal approaches, we take a model-based, informal inference approach in which we engage students in constructing and revising models of chance to guide inferences about various phenomena.

This research builds on a data modeling perspective that positions students as participants in approximations of professional practices, especially those of visualizing, measuring, and modeling variability (Lehrer, Kim, & Schauble, 2007). These professional practices steer inference under conditions of uncertainty. In this research, prior to the focal lessons that we analyze in this paper, a class of sixth grade students (age 11) measured the perimeter of the same table with two different measuring instruments (a 15 cm ruler and a cm tape measure). Then they invented data representations of their measurement data. Comparing and critiquing a variety of invented displays, students came to appreciate how inventors' representational choices revealed and obscured shape and pattern in the data. Students went on to invent measures of these data's center (i.e., best guess of true table perimeter) and variability (i.e., precision or tendency of the measures to agree). Students compared data representations and measures of data for each measurement tool with an eye toward considering how change in the measurement process (i.e., the change in the measurement tool) produced changes in the shape and measures of the data. They also considered whether their inventions were robust to other changes in data that they could imagine (e.g., an occasional outlier, a change in sample size). With practices of visualizing and measuring characteristics of data in hand, students examined behaviors of simple random devices (e.g., a hand-held spinner), investigating how varying either the structure of a device or the sample size influenced the distribution of repeated outcomes. They also investigated the sampling distribution of sample statistics (e.g., percent red of a two-color, red-blue spinner).

With these preambles in mind, this paper focuses on the next phase of the instructional sequence, where students used TinkerPlots, a digital data visualization and modeling software (Konold & Miller, 2005), to construct models of chance processes, and used these models to answer a question or make inferences about various phenomena, other than the random devices themselves. The modeling sequence was as follows: First, students constructed models to estimate compound probability in standard textbook probability problems (e.g., basketball free throws, games of chance). Next, students constructed models to account for variability in an observed sample in contexts of signal and noise (e.g., their class's table measurement data or the production of a batch of cookies). Students judged model fit based on a sampling distribution of a model statistics and made inferences about claims of changes to processes of measuring (e.g., same or different object measured) and of producing (e.g., change in production method). The instruction concluded with student invention of models of natural variability in contexts of a psychophysics experiment about visual illusion and of plant growth at different initial days of planting. Here students used their models to warrant claims about whether differences between conditions could be ascribed to chance. Although we do not focus on it here, the contexts varied in the visibility of the variability-generating processes.

Although students had many opportunities to invent and revise models of chance during this

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

instruction, we observed that these opportunities were significantly enhanced by teacher practices. Following the view that teaching consists in assisting performance (Tharp & Gallimore, 1988), we describe some of the conceptual challenges faced by students, and how the teacher helped them respond to these challenges during the course of instruction. Sources of data include field notes and video records of whole-class and small-group interactions involving the teacher. The analytic approach employs methods of constant comparison but is also informed by findings from end-of-year interviews with students (Lehrer, 2017). These interviews revealed that most students had developed a hierarchical image of sample—seeing samples as constructed and as constituted by individual cases that vary, while simultaneously seeing samples as members of an imagined collection of samples that also vary (Saldanha & Thompson, 2014). Hence, here we focus on teacher practices that appeared to support student construction of this statistical image of sample. We examined video of classroom activity during all 16 lessons and developed a comprehensive description of how the teacher helped induct students into the practice of modeling variability, but in this presentation, we targeted initial and later forms of teacher support for students' conceiving of stochastic process as an explanation of sample variability and of sampling variability. Although such support recurred throughout all 16 lessons, in this paper we illustrate initial forms of teacher assistance that focused student attention on the role of repeated process in estimating probability and then illustrate later forms of assistance that expanded students' reach to encompass stochastic process as a conceptual tool for considering model fit and for developing model aesthetics (e.g., the grounds for deciding whether a particular model was "good").

FINDINGS

Supporting construction of an image of repeated process

Students first (Lesson 1) constructed models to estimate the probability that an opponent basketball player with a history of free-throw percentage of 50 would miss all three shots when her team was one point behind at the very end of the game and lose the game. Students' initial modeling attempts often followed the story line of the problem context (Noll, Clement, Dolor, & Peterson, 2017), so they constructed models that would simulate one game:

Teacher: So tell me what you did.

Sean: We made 3 spinners for the 3 shots. And then we halve them, because she makes half the shots. And then miss and hit on all of them. And then we repeated it once, so draw 3. And she won the game because she got one hit.

In this excerpt, Sean followed the story line and made explicit the mapping between model components (e.g., "we halve them") and corresponding world entities (e.g., "she makes half the shots" [over the year]). Sean also interpreted the outcome of this simulated game (i.e., "she got one hit") in the context of the game situation (i.e., "she won"). At this point, many students seemed to have taken the modeling task to be configuring random devices to simulate one game of basketball. None of the pairs that the teacher talked with (about half of the class) as she roved the classroom spontaneously used their model to generate more than one game at a time. In response, the teacher gently prompted them to model a collection of games rather than a single game:

Teacher: So that one was a miss-make-make, so you didn't win the game. But that's just totally one time it happened, we want to know the probability of it happening anytime, so what could we do?

Carson: You could turn the repeat higher.

By asking students to think beyond what happened "just totally one time" to what happened "anytime," the teacher was linking one individual outcome to the long run, and was establishing a shared goal that what "we want to know" was not what happened each time, but what happened over time. In doing so, the teacher also extended the role of students' model, from a representation of the situation, to a data-generating tool that students could use to generate data that helped them know something they did not know before.

Even with this prompting, students tended to simulate a relatively small number of games, and the teacher seized this opportunity to highlight sample-to-sample variation:

Teacher: Okay, so what do we want to do?
McKinley: Maybe [repeat] 20?
Teacher: Okay, so let's see what happens there? … Oh, you got one [miss-miss-miss]. You see it? Okay, so what percent of the time did that happen?
Students: 1 out of 20, 5%
Teacher: 5? What do you think would happen if you change or you do it again?
Carson: You might have more or you might have less. Probably more.
Teacher: Let's try it.
Carson: This time it has 1, 2, 3. So that's 15.
Teacher: Okay. So 5 to 15%, that's a big jump, huh? So how can you narrow in on the ((gesturing two fingers closing in)), closer to the theoretical?
McKinley: Maybe do more than 20?

The teacher routinely asked students to rerun their model, and here, in doing so, the teacher treated 20 games as a collection, whose collection process could be repeated, and now the repeated process was extended to include repeated samples (here, of 20 simulated games). By asking students to rerun their model, the teacher also produced sample-to-sample variation for students to see, and in the above episode, she highlighted it as "a big jump." At this early point, another approach to coping with sample-to-sample variability was to ignore it. Students wanted to only attend to the sample outcome that matched their expectations, rooting for the model to produce such sample outcome (around 12-13% miss-miss-miss in this case). The teacher pressed very hard for students to pay attention to the variation and to justify how we could know the answer given this variation between samples (of 100 simulated games):

Students: Oooh, 12, exactly
Teacher: So now they got 12% [miss-miss-miss], how do they know that's what they're really looking for? … What could they do?
Model: The model was rerun, and in the new sample, miss-miss-miss went up to 19%.
Student: Oh man.
Teacher: Oh woah, did they just?
RL: Just ran it again?
Teacher: Now it's on 19%. So which one's right?
Students: 12.5!
Teacher: You don't know it's 12 … How do you know?
Sean: Run it once more.
Teacher: Once more?
Sean: Yes because it will see which is closer.
Model: The model was rerun, and in the new sample, miss-miss-miss was 13%.
Students: 13!
Student: Do it again.
Teacher: So Sean said run it again, and the third time will tell us. Is it going to be the third time that tells us every time?
Sean: No, not every time, but if you want to do it quickly.

Similar to focusing on one game at a time earlier, students now focused on one sample at a time. They were excited when an individual sample outcome matched their expectation (e.g., "13!"), and disappointed when it did not (e.g., "oh man"). Faced with sample-to-sample variation, while having to make a decision about a good estimate for empirical probability, one student suggested that "the third time" was as good as any other time. To counter this tendency to ignore the long run, the teacher often suggested collecting a large number of samples:

Teacher: Well, Elijah, how many times do you want to do it, just one more time, again, and again, and again, and again?
Elijah: No, probably like 10.
Teacher: So could we just maybe collect it?
Sean: Maybe we could just collect it 10 times.
Teacher: Okay before you collect a bunch of them, so right now, we are just collecting this

((pointing at the miss-miss-miss stack)), and we had a 12, a 19, and a 13. What shape when we collect, we're gonna collect 300 of them, I want you to talk to your neighbor, what do you think the shape of that graph will be, where do you think the middle, or the center, or the clump, all those things, where do you think those will be?

The teacher revoiced what students had been saying—"do it, just one more time"—with an added emphasis on an image of long-term repeated process—"again and again, and again, and again." By asking "how many times" to run again, the teacher prompted students to think about a collection of samples. Students suggested a collection of 10 samples, which the teacher quickly extended to 300, a relatively large number of samples. The teacher again highlighted the sample-to-sample variation in outcomes that the class just witnessed together: "a 12, a 19, and a 13," and then asked students to anticipate the aggregate structure of a collection of sample outcomes. While in this first lesson, the teacher suggested collecting many samples in order to emphasize variability, in later lessons, she often asked students to justify why they might want to collect a large number of samples:

Teacher: Okay. So, thumbs up, thumbs down, do you want me to collect 100 [samples]?
Students: Yes
Teacher: What's that gonna tell me? I'm not going to run unless you can tell me what it's gonna help me know.
Gideon: You can see what happens at many different times.
Sean: If she did this over and over again, what would be the results of that.

By pressing for student justification for "what it's gonna help me know," the teacher positioned collecting sample statistics as an epistemic tool, as it became a means to know something they did not know at the moment.

We have described how the teacher initially supported attention to repeated process, a process that had to be imagined by students in the textbook-like problems on compound probability they initially modeled. Now we describe how in later lessons the teacher and students constructed and appealed to repeated stochastic process to consider model fit and aesthetics of "good" models.

Evaluating model fit

Model evaluation and revision require criteria for model fit. As we noted, students initially evaluated model fit based on a few instances of correspondence between empirical and simulated sample statistics or between empirical and simulated case values. But the teacher continued to emphasize a repeated process, often pressing students to consider what would happen if they ran their model again, as a way to allude to sample-to-sample variation, and then to motivate using a model-generated sampling distribution for judging model fit. For example, in Lessons 11-12, students were challenged to construct a model of a psychophysics experiment. In this experiment, participants estimated the location of the midpoint of a horizontal line under three different conditions, two of which were designed to create illusions that would bias student estimates. Students first explored their class's data for the unbiased no-illusion condition, locating the center (median) and measuring the variability (IQR) of their estimates. Isaac and Dean (below) had spent some time constructing and revising a model to account for the center and variability of the sample, and when the teacher started talking with them, they were at a point where their model had generated a sample "with a median that agrees with [the observed] median," which was 100, and with an IQR "only 2 off," and thus, was "close enough." The teacher simply asked:

Teacher: What'd happen when we do it again? ((Model rerun, new sample median was 96.))
Isaac: What the?!
Teacher: Still like your model?
Isaac: No.
Teacher: Why not?
Isaac: I don't know.
Teacher: I couldn't have planned that better if I wanted to. That makes me so happy that that just happened.
Isaac: Well, ah. ((all laughing)) I don't know what happened.
Teacher: I just ran it again.

Isaac: Run it again. ((Model rerun; new median was about 98.)) There we go, a little bit better.
Dean: ((pointing to the error device)) It isn't the same to get +3 and -3 …
Teacher: ((keeps rerunning the model)) Look at that one.
Isaac: Go 100.
Teacher: If we want to see what happens over lots of time, can we collect the statistics of that, collect our statistics and see if we still like our model, or our model is a little too cranky?

The teacher asked students to anticipate another simulated sample. Again, the teacher appealed to an image of repeated process. The students seemed astonished by the shift in the new sample median generated by their model. Isaac was rooting for his model to do better. Dean pointed out a potential defect in the random device and started revising it. The teacher reran the model a few times, before she suggested collecting the model statistics to "see what happens over lots of time," which would become grounds for "see[ing] if we still like our model." In doing so, the teacher began to establish that model fit required attending to the long run, to avoid falling prey to a "cranky" model that could produce a few samples of corresponding statistics just by chance.

With sampling distributions of model statistics students could extend their evaluation of model fit to include considering how often samples like this occurred:

Dean: Do you like it like that?
Isaac: Yeah, I say it's a pretty good model. You know, 55% of the time our target, our median was (100).
Dean: Over half
Isaac: Oh, but the width of crown [IQR] needed to be like 5. So, let's see how much of the time we got it to 5. 26% and that's a fourth. I mean, is it really? A fourth of the time we got 5.

The following day, Isaac and Dean shared their results with the teacher, and asked,

Isaac: Would you say this is a good representation?
Teacher: I don't know. Would you?
Isaac: Yes

The students seemed satisfied that their simulated medians were 100 over half of the time, but unsatisfied and a bit skeptical that their simulated IQR's were 5 only a fourth of the time. Implicit in this reasoning about the likelihoods of model statistics—how often do samples like this occur?—was an image of long-term repeated process. By appealing to an image of repeated process, the teacher motivated employing sampling distributions of model statistics to evaluate model fit.

Considering criteria for "good" models

The end-of-year interviews revealed that students came to think that good models explain or represent a process. They viewed models as approximations of processes (rather than as exact copies). Interpreting models as approximations entails imagining possible outcomes even when they are not present in an empirical sample. In the episode below (Lesson 12), the class contrasted two positions on how to construct a model of the psychophysics experiment described above. The teacher animated a student's position that a value absent from their class's empirical sample (an error of -5 or an estimate recorded at 95) should not be generated by the model. Appealing to an image of a repeated random process, many students disagreed with the teacher's position, arguing that values missing from just this one sample should be included in the model as "possible" values:

Teacher: Let's vote. Who think they should not have -5 up there? Who's with me? ((A few students raised their hands.)) Because it didn't happen?
Student: So?
Students: It's possible. / It's still possible. / It's plausible.
Student: Just because it didn't happen-
Teacher: Based on the data, it's impossible. It didn't happen. ((pointing at the empty spot at 95 in the observed sample))
Boston: Well, it's between the two numbers that did happen, so it's possible.
Bryant: That's just one data set.
Teacher: Oh, so Bryant said this is just one data set.
Students: Yes

Teacher: So, Bryant, why do I care that that's just one data set? What does that tell me?
Bryant: Because not everyone is gonna get every single number. If we run it more times, there're more possibilities.
Teacher: What do you mean run it more times? Run the model?
Bryant: If we were to do the line more times, again and again, if [another class] do it, then it'll be-
Charlie: Like if we do the line again.
Teacher: Oh, you mean if I collect the real data again it's possible to get 95, so since it's possible for it to really happen in real life, my model might want to represent that happening?
Students: Yes.

Even though a value was missing from an empirical sample, students could imagine it occurring in another hypothetical repeated experiment. The teacher made this explicit by asking a clarifying question, so that students were imagining the model as repeating the actual experiment itself more times. The aesthetics of a "good" model of a process extended beyond mere fit with the data to include how well the chance model approximated the process being modeled.

CONCLUSION

We described some of the ways in which a teacher encouraged students to construct an image of a sample as a collection of outcomes from a repeated stochastic process and how she challenged students to elaborate and extend this image as they considered model-fit and criteria for "good" models as reflecting these repeated processes. Students initially tended to think of models as capturing data, so that when a model's simulation replicated an aspect of the empirical data or an aspect of the problem situation, they were satisfied. With teacher questions, such as what would happen in the next simulated sample, or what would happen if the process were repeated "again and again," and with teacher positioning of students as debating the grounds of good models, students' stances gradually shifted to view particular samples as members of an imagined collection of samples that varied. This image was appropriated by students to justify model fit: Students often challenged models offered by peers for not generating "possible" values consistent with a process, but by chance not present in an empirical sample. The enterprise of construing a sample as hierarchical was not simply a matter of insight—students often "fell back" to capture criteria, albeit in more sophisticated ways, such as a model that simulated sample medians corresponding to the empirical sample, while failing to notice that the model did not represent the variability of the process as adequately. But the teacher addressed these mishaps in modeling by continually emphasizing chance as describing long-term process, and this stance was appropriated by students who often used it as a basis for model critique—of their model and those of peers. We have focused here on a relatively narrow yet important form of teacher assistance. However important this form of assistance, it was complemented by others that deserve a fuller description.

REFERENCES

Konold, C., & Miller, C. D. (2005). TinkerPlots: Dynamic Data Exploration. Emeryville, CA: Key Curriculum Press.

Lehrer, R. (2017). Modeling Signal-Noise Processes Supports Student Construction of a Hierarchical Image of Sample. *Statistics Education Research Journal, 16*(2), 64-85.

Lehrer, R., Kim, M.-j., & Schauble, L. (2007). Supporting the Development of Conceptions of Statistics by Engaging Students in Measuring and Modeling Variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195–216.

Noll, J., Clement, K., Dolor, J., & Peterson, M. (2017). *Students' use of narrative when constructing statistical models in TinkerPlots™.* Paper presented at the The Tenth International Research Forum on Statistical Reasoning, Thinking and Literacy.

Saldanha, L. A., & Thompson, P. W. (2014). Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments. *Journal of Mathematical Behavior, 35*, 1-30.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge: Cambridge University Press.