Structuring Variability by Negotiating its Measure

Richard Lehrer & Min-Joung Kim
Vanderbilt University

Although variability and structure are often considered as antonyms in many everyday settings, a mathematically disciplined view contradicts this opposition. To initiate fifth- (10 years old) and sixth-grade (11 years old) students in this disciplinary view, we engaged students in practices of modeling data. These practices included inventing and revising data displays, inventing and revising measures of centre and variability, and inventing and revising models of chance to account for variability. Here we focus on prospective correspondences between students' invented measures (statistics) of variability and those favoured by the discipline. We suggest that inventing measures positions students to transform their vision of variability from mere difference to more structured forms, some of which coordinate centre and spread. By tracing interactions among an inventor, her classmates, and the teacher, we trace how structuring variability and constituting its measure co-originated during the course of negotiations about the meaning of the measure. Consideration of the coherency, transparency and generalisability of a statistic, all of which are valued by the discipline of statistics, emerged during the course of invention.

Variability is foundational to statistics. G. W. Cobb and Moore (1997) suggest that the very need for the discipline of statistics arises from the omnipresence of variability, and Wild and Pfannkuch (1999, p. 226) suggest "noticing and acknowledging variation" as a critical initial step for initiating statistical reasoning. Yet in everyday discourse, variability is often associated with a lack of structure or pattern, as mere difference among data. The disciplined view is very different: Variability is structured as distribution, and the nature of the distribution reflects the operation of a repeated random process (deGroot, 1975; Thompson, Liu, & Saldanha, 2007). Random is not a synonym for haphazard, but is instead a description of phenomena having uncertain individual outcomes and a predictable pattern, given sufficient repetition (Moore, 1990).

Moore (1990) suggests that recognition and coordination of uncertainty and pattern is foundational for reasoning statistically about variability, but many studies suggest that integrating chance with variability is often challenging. For example, Metz (1998a, 1998b) investigated children's and adults' conceptions of the operation of three random devices (a marble tilt box, spinners, and urns). Although there was an increasing tendency with age to conceive of the operation of these devices as involving chance, university students often could not reconcile uncertain individual outcomes with predictable aggregate structure.

Konold (1989) found that for many phenomena, there is a pronounced tendency to treat a single outcome or event (e.g., rain tomorrow) as isolated and not as participating in a long-term process involving similar conditions (i.e., other occasions when the atmospheric conditions were similar). An implication of the outcome view is that coming to see phenomena as part of

a stochastic process is often challenging. Even when the repetition of a repeated random process is made explicit, and the sample space is relatively simple, students at all ages do not readily perceive relations between the mathematical structure of the random process and the distribution of variability that results from this repeated process (Shaughnessy, Canada, & Ciancetta, 2003; Shaughnessy, Watson, Moritz, & Reading, 1999; Torok & Watson, 2000; Watson, 2006).

The challenges to reasoning statistically about variability are evidently formidable, suggesting the need for the design of learning environments that support the growth and development of this specialised form of reasoning. The central challenge for instructional design is to generate prospective pathways that potentially help students transform initial understandings of variability as unstructured, and chance as haphazard (or even as personally controlled), to forms of reasoning that coordinate chance with variability. To support development of this form of reasoning, we engage students in an approach that we term *data modeling* (Lehrer & Romberg, 1996; Horvath & Lehrer, 1998). As the name suggests, data modeling positions students to invent and revise models of chance as accounts of observed variability. As we describe more completely in the next section, data modeling is grounded in related forms of activity that include decisions about which aspects of the world are relevant to a particular question, how best to measure them, and how to structure and represent the resulting measures so that pattern (reproduceable features) in variability can be manifested. Without structuring variability as pattern, modeling founders, because model fit is determined by approximating the structure of what one is attempting to model.

In this paper, we focus on the role of inventing measures of variability as a means for structuring variability. Inventing measures (statistics) of variability affords opportunities for coming to see differences among cases in new ways. As in the discipline of statistics (e.g., Hall, Wright & Wieckert, 2007), inventing statistics is not a solitary act: The meaning of a statistic is negotiated in a classroom community. In the next section, we situate the role of inventing statistics within the larger framework of an instructional design intended to leverage historic relations between repeated measure and statistical reasoning.

Designing Instruction to Link Variability and Chance

We approach the problem of coordinating variability and chance by engaging students in a series of design challenges, each of which is intended to promote conceptual change through an interaction of tasks (e.g., the explicit problem posed), material means (e.g., paper-and-pencil, computer tools), modes and means of argument (e.g., justifying a particular design solution by appealing to its generality), classroom norms (e.g., student justifications need to be rendered in ways that are sensible for classmates), and activity structures (e.g., producing displays, methods, critiquing displays, methods). The design challenges are aimed at students in the late elementary and early middle school years (Grades 5 to 7 in the USA). Our focus on design challenges reflects an explicit commitment to student authoring (invention), because we intend to support an epistemology of

mathematics as a productive and generative enterprise (Boaler, 2002; Greeno & MMAP, 1998). Invention positions students as authors and provides a pathway for understanding conventions that are used in the discipline. After invention, students have a clearer idea about the problems that conventions solve. However, careful pedagogical design is needed to create opportunities for students to consider multiple aspects of a problem and to consider particular solutions as representing trade-offs. The context of these design challenges is one of repeated measure: See Konold & Lehrer (2008) for a discussion of the pedagogical affordances and constraints of different contexts for the study of variability.

Repeated Measurement

To begin, all students *measure* the same attribute, typically a length such as the circumference of a person's head or the distance of their outstretched arms, fingertip to fingertip. We ensure that students have had opportunities to explore qualities of spatial measure as a preamble, because we wish to draw upon measurement as a resource for learning (Lehrer, 2003). While measuring, students directly experience processes involved in generating a measure, such as the repeated iteration of a ruler, which we later draw upon to make sense of the resulting collection of measurements; and because they are measure-agents, students are in a position to think simultaneously about individual (i.e., a particular value) and collective levels (i.e., the batch of measurements). Moreover, students measure with different tools, a change in process that affects the variability of the resulting collection. The repeated measurement context affords ready interpretation of statistics as reflecting signal and error (Konold & Pollatsek, 2002; Petrosino, Lehrer, & Schauble, 2003).

Designing and Comparing Displays

Individual measurements are recorded in some fashion and often literally stuck on a whiteboard or other surface. Students work in small group to design paper-and-pencil displays that present some "pattern" or "trend" that they notice, although occasionally students will insist that the values are only different and are without any structure "unless someone is imagining things," as one sixth-grade team acerbically suggested. Students produce a variety of displays, and we use this variety to highlight issues of structure, such as order, count, and interval. These collectively contribute to the shape of the data. Computer tools typically construct intervals for students. Because we wish to problematise interval, we have students use paper-and-pencil, so that the choice of whether to construct or even consider intervals is left to them.

By comparing displays, suggesting what they show and hide about the measurements, students consider how the choices of designers affect the resulting shape of the data. For example, some students order data, resulting most often in lists, while others create intervals and counts. As students account for how attention (or lack thereof) to order, interval, and count give rise to a particular representational form, they are also developing meta-representational competence (diSessa, 2004). We select a representation

based on frequency (typically invented by students, albeit not always in a conventional form), and students consider how the processes of measurement could account for the shape of the data. Why, for example, is the data clumped in the middle? Our intention is to link process and a now more structured variability—one where there are regions of values and not mere differences.

Inventing Measures of Distribution

The second design challenge posed to students invites closer inspection of the qualities of the aggregate now literally visible in displays. Students invent measures of the "best guess of the real measurement", and they invent measures of "precision." The former effort capitalises on the signal interpretation of the data: Children do not believe that the actual length of the object has changed, but they are confronted with the problem that the true score does not announce itself and hence must be estimated (Lehrer & Schauble, 2002). Measuring precision entails further challenge: What might precision mean? How might a measure be designed to support the intended meaning? We suggest precision rather than spread, because "spread" often signals range and thus disrupts tendencies to scrutinize variability in relation to centre (delMas & Liu, 2007; Garfield, delMas & Chance, 2007; Petrosino et al., 2003). We also introduce students to the Tinkerplots software at this junction, because this tool affords a variety of means for partitioning and organizing data (Konold, 2007).

As children invent measures, they propose their inventions to their classmates in whole-group settings. The variety of invented statistics sparks classroom conversation about which aspects of the collection of values are attended to by particular methods and engages students in productive comparison among these invented statistics. Students realise that multiple solutions to the problem are possible and perhaps even desirable. During classroom conversations, we focus especially on those invented statistics that coordinate conceptions of centre and spread. As we will elaborate later, it is during whole-class conversation that the meaning of a measure is negotiated. It is rare that an invented measure goes unscathed: Students typically do not uncritically accept initial proposals. The variety of invented statistics provides a secure foundation for better understanding of practices within the discipline of statistics, where multiple measures of centre and variability are used. At this point in children's development, we are content with securing this starting point, although one criterion for choosing among measures that we support explicitly is that of generalisation. To encourage this disciplinary practice, after children invent measures (statistics) with the sample generated with one measurement tool (e.g., a 15 cm. ruler), they explore the utility of their measures (statistics) with a second sample generated with the tool that allows more precise measurement (e.g., a metre stick). Children justify whether or not their measures generate values that make sense in light of the change in distribution. For example, the centres of the two distributions are usually similar but their variability is distinct. Do the measures adequately reflect these properties?

Modeling Chance

We next support coordination between variability and chance by challenging students to invent models of the measurement process that take into account the tendency toward true score as well as the variability attributable to "mistakes." This is a challenging step, because it involves the supposition that the syntax of a chance process is akin to the process employed by the agent-measures. As in most situations, the role of uncertainty is not apparent in the representation of the data. We approach this difficult threshold by engaging students in the analysis of error: Students make decisions about likely sources of error. After enumerating various sources of mis-measurement, they design a chance device for each source of error in which they specify magnitudes of error (e.g., + or – 2 cm.) and likelihood of error (e.g., a probability of ½). Models of the measurement process are represented as a combination of these chance devices and a constant that designates the true score (Lehrer, Kim & Schauble, 2007). In measurement theory, this is the familiar additive model of true score and error (Lord & Novick, 1968).

Design Studies

This prospective pathway for learning encompasses resources for reasoning and explicit attention to the conditions under which learning can best be supported. Thus, investigation takes the general form of a design study (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). One understands learning by designing the elements of a learning system and then studying the functions of these inter-related elements as the design is put into play. The purpose is to contribute to greater understanding of domain-specific processes of learning, not to evaluate one design against another. Instead, the intent is to construct and revise what diSessa (Cobb et al., 2003) calls a "humble theory," that is, an account of knowledge development tightly tied to the particulars, even if the particular topic in question is as omnipresent and important as variation. Design studies typically entail multiple iterations of refinement and ours is no exception: We have conducted four iterations of this prospective pathway for learning.

Exploring Measurement of Variability

Having situated student invention of measures (statistics) of actual value and variability within the framework of the trajectory of learning envisioned by the design, we turn now to focus on the role that inventing measures of variability plays in students' efforts to structure variability. Our focus on measure follows from its ubiquitous properties: Developing measures of a phenomenon typically involves analysis of qualities and transformation of these qualities into a metric. In some views, the progress of science is attributed to efforts to quantify natural systems —to develop measures and establish relations among these measures for phenomena ranging from falling objects to magnetic fields (Crosby, 1997). The historic development of statistics fits squarely in this realm (Porter, 1986). Hence, inventing measures reflects the history of the development of the field,

although the pedagogical utility of positioning students to invent measures remains open to question.

As in professional or discipline-based practice, invention does not occur in a vacuum or solely in the heads of interested individuals. Once drafted, measures are subject to further rigors, including the perception of their utility and intelligibility to a wider community of prospective users. As we described previously, to simulate this aspect of discipline-based practice, students participate in whole-class showcases, during which the meaning of a particular measure, or component of a measure, is often contested, so that the resulting measure emerges as a negotiated outcome. Local contests reveal what students understand as a legitimate measure, and we infer from these interactions the nature of the criteria that appear to guide negotiation. These negotiations are also windows to the qualities of the distribution that the author(s) of the measure chose as worthy of measure, as are the forms of measure ultimately developed.

To illustrate the prospective contributions of inventing statistical measures to structuring variability, we first present representative samples of students' inventions. We suggest that these student inventions provide a window on to how students sought to structure variability. Then, we turn to an extended episode of interaction with one student, Shakira, because she represents a difficult and protracted negotiation, one that threatened to break down at any moment. Yet despite its tenuous character, the case illuminates how measurement of variability serves to create structure, and how this structuring was interactively constituted.

Method

Participants

Participants (10 males and 8 females) attended an urban school serving primarily underrepresented youth in the southeastern region of the United States. The proportion of children attending the school who qualify for free or reduced-price lunch ranges from 60 to 80 per cent from year-to-year, suggesting comparatively lower SES status. Students were 11-12 years of age and came from a wide range of ethnic backgrounds. Although this class is the primary source of data, we also occasionally refer to student work collected during the previous iteration of the design study, conducted with students attending the same school.

Procedure

One of us (RL) served as the primary classroom teacher for mathematics during the school year. Mathematics classes were conducted twice each week for 1.5 hours each. During most weeks, students responded to a weekly assessment for an additional 45 minutes per week. Each classroom lesson was videotaped and digitally rendered for further analysis. One small group of students was videotaped with a second, wall-mounted camera throughout the year. Field notes contextualized the video recordings and served as a platform for further reflection and planning. Within this broader

context, we conducted a design study over several months where students were introduced to distribution, statistics, data and chance. As we indicated previously, the design culminated in students' efforts to invent and revise models.

For the extended analysis described below, we transcribed portions of the video of one lesson that included the whole-class conversation about Shakira's method for measuring precision (variability). The transcript conventions that we employed included a ? to indicate a rising intonation, a ?? to indicate an interrogative accompanied by a rising intonation, a period (.) to indicate a falling intonation, ### to indicate unintelligible speech, < > to indicate overlapping speech, ( ) to indicate specific behaviors accompanying speech, and = to indicate latched talk (talk immediately following an utterance, without pause).

Results

In the first section, we present a representative sample of some of the methods and measures of variability invented by children during the course of the last two iterations of the design that contain the seeds of statistics employed by the disciplinary community. Recall that the challenge posed to students was one of inventing a measure of the precision of their measurements.

Figure 1 displays a measurement method for finding the precision of the observed measures of the length of a teacher's arm-span. Note that this invented measure focuses on the sum of pair-wise differences among the ordered set of measures. The student, Renee (all student names are pseudonyms), justified the validity of the measure by demonstrating that the sum decreases as the precision increases (and the variability decreases). Other students noticed that this measure bore some resemblance to the range, invented by a classmate, in that it focused on difference, but Renee's measure included all the data, instead of just two cases. Students thought that this might be a good idea, because the range presented the whole class with the two "worst measurers."

Wing Span Crude Tool   Wing Span Precise Tool

```
118                              141
127 > 9cm    Procedure!         145 >
136 > 8cm    my procedure       146 >
137 > 1cm    is that take       148 >
143 > 6cm    all the           150 >
147 > 4cm    measurements       152 >
148 > 1cm    and write          155 >
152 > 4cm    them down          155 > 0cm./0/
152 > 0cm./0/ on the paper.     156 > 1cm.
154 > 2cm    then count         157 >
155 > 1cm    how far            157 > 0cm./0/
156 > 1cm    apart each         157 > 0cm./0/
158 > 2cm    measurements.      157 > 2cm.
158 > 0cm./0/ Then add          159 > 0cm./0/
160 > 2cm    them all           159 > 0cm./0/
161 > 1cm    together           159 > 2cm.
163 > 2cm    (separately)       161 > 0cm./0/
168 > 5cm    you will           161 > 0cm./0/
169 > 1cm    find 2 different   161 > 1cm.
172 > 3cm    numbers.           162 > 0cm./0/
177 > 5cm    They are           162 > 1cm.
181 > 4cm    39 numbers         163 > 2cm.
184 > 3cm    apart.             165 > 4cm.
193 > 9cm                       169
    (66)                            (27)
```

Figure 1. Renee's pair-wise distance method.

Figure 2 displays a measure "piggybacked" on that invented by Renee by two other classmates. It also focuses on differences among cases as the basis of measure, but these students suggested an improvement based on a common point of reference – the sample median – perhaps because they had previously invented the median as a best estimate of the true value. They suggested that precision could be thought about as distance from the real length (the median). They were also confident that the sums of these (absolute value) differences were valid measures because the sums decreased as the variability decreased (317 vs. 112), suggesting greater "closeness" among the measurements of the teacher's arm-span.

Crude Tool		Precise Tool	
Measurement	- Median	Measurement	- Median
118 -157 =	|-39|	141 -157 =	|-16|
127 -157 =	|-31|	145 -157 =	|-12|
136 -157 =	|-21|	146 -157 =	|-11|
137 -157 =	|-20|	148 -157 =	|-9|
143 -157 =	|-14|	150 -157 =	|-7|
147 -157 =	|-10|	152 -157 =	|-5|
148 -157 =	|-9|	155 -157 =	|-2|
152 -157 =	|-5|	155 -157 =	|-2|
152 -157 =	|-5|	156 -157 =	|-1|
154 -157 =	|-3|	157 -157 =	0
155 -157 =	|-2|	157 -157 =	0
156 -157 =	|-1|	157 -157 =	0
158 -157 =	1	159 -157 =	2
158 -157 =	1	159 -157 =	2
160 -157 =	3	159 -157 =	2
161 -157 =	4	161 -157 =	4
163 -157 =	6	161 -157 =	4
168 -157 =	11	161 -157 =	4
169 -157 =	12	162 -157 =	5
172 -157 =	15	162 -157 =	5
177 -157 =	20	163 -157 =	6
181 -157 =	24	165 -157 =	8
184 -157 =	27	169 -157 =	12
193 -157 =	36		
	317		**112**

Figure 2. Mikela's method, using the differences between the measurements and the median

Figure 3 displays an alternative approach to coordinating centre with variability drawn from a fifth-grade class during a previous iteration of the design study. The measurements are of the height of the school's flagpole, estimated by use of similar triangles. Henry focused on the neighborhood of values surrounding the median and used TinkerPlots to create a "centre clump region", with precision defined by the percentage of cases in the centre clump. As precision increased, so too did the percentage of cases in the centre clump. However, he did not have a method for determining the boundaries of the centre clump.
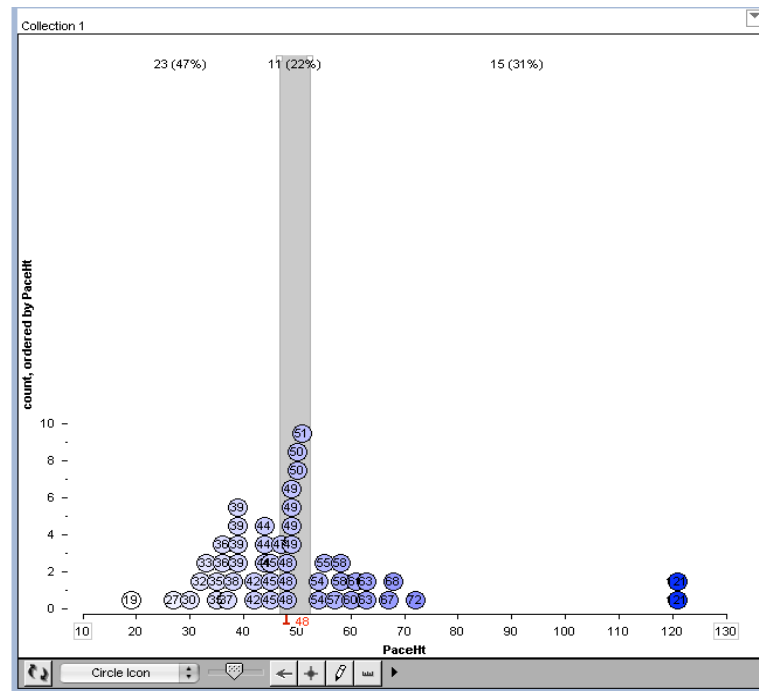
Figure 3. Henry's method, using the percentage of mesurements in a "centre clump".

Although none of these inventions are conventional statistics, they contain the seeds of the structuring of variability upon which more formal statistical conventions rely. The solutions represented in Figures 2 and 3 especially illustrate coordination of centre with variability and hence provide the grounds for further productive refinement, both for the individuals involved and for the class collectively.

Although we have focused on particular products, each of these measures was constituted interactively—with peers and with the teacher—and were often refined, revised, or outright abandoned. In the next section, we illuminate one of these episodes of negotiation to clarify how structuring variability emerged during interactions with peers and the teacher.

Negotiating Variability

Shakira's method for measuring precision relied on her intuition that measurements ought to cluster around the median, which was previously established in this class as an indicator of true measure. She also believed that precision should take the modal clustering of the data into account. She used TinkerPlots to first partition the data into two large bins. This rough partition (a default setting of TinkerPlots) effectively eliminated outliers from consideration, on the grounds that these outliers were clear instances

of disagreement and hence not worth including. Shakira used the drawing tools provided by TinkerPlots to graphically structure her sense of a neighborhood of values surrounding the modes and median of the data, displayed in Figure 4. Shakira's initial description was figurative, in the sense that she said: "What I did was I circled the median and mode, our numbers that were close to it."
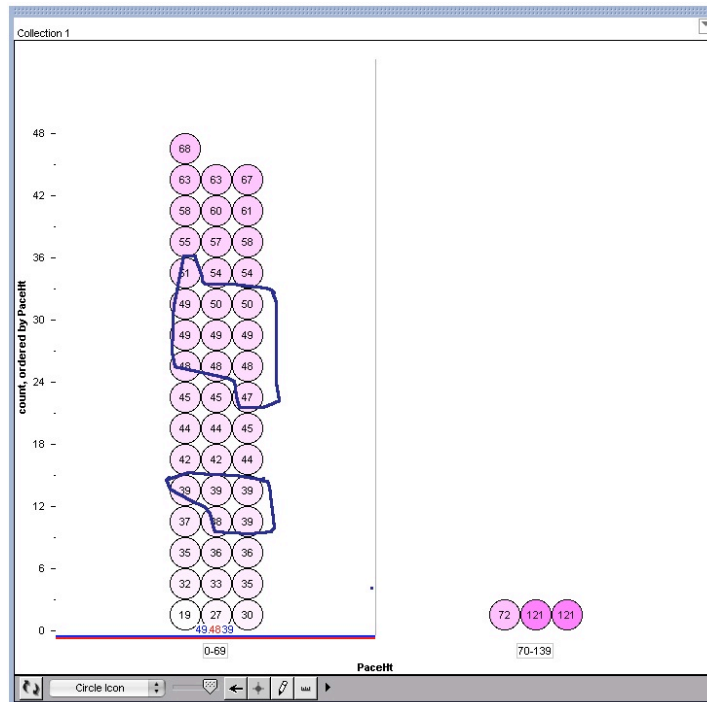

Figure 4. Shakira's "wrap-around" method for measuring precision.

She immediately proceeded to justify her measure by relating it to previous conversations in the classroom about what the data might look like if one had more or less agreement (e.g., if everyone agreed, there would be no variability):

> All the numbers will figure out how do difference, all the numbers that are outside of the circle, if we disagree on them, they will get bigger [pointing to measurements outside of the regions depicted in Figure 4]. All the numbers inside here [pointing to outlined region in Figure 4]. See how small they are? If we agree on them, they will get bigger. But if we disagree on them, they will get smaller [pointing again to the interior of the outlined region in Figure 4].

Although Shakira elaborated further, her references were indistinct in the sense that neither the teacher nor her classmates were certain about what she meant by "bigger" or "smaller." For example, Jamir asked: "What do you mean? Like 'we, we disagree on that it would be smaller.' What do you

mean by that?"

The conversation was redirected by several students to request justification of values included in the region, as exemplified in the following exchange between Lena and Shakira in which Shakira proposes criteria of "close" and "surround:"

Lena:	Why did you add 50 to that and along with 51?? Why did you add that??
Shakira:	Add 58??
Lena:	No. You added 50.
Teacher:	Fifty and 51, Lena is pointing out.
Lena:	=Ya.
Teacher:	=She is asking you why those, and not others.
Lena:	<Because###>
Shakira:	Forty. Because 49 is the mode, and these are two closest numbers (Shakira points at 50 and 51) that surround 49. And others 48. And 48 is the median. And 47 and 49 both surround 48.
Teacher:	<Okay>.
Lena:	<I still> don't get why you put 50 and 51.

In response to another series of queries initiated by Mikela (Line 1, below), Shakira provided further elaboration of "numbers that are close", aligning closeness with agreement, in response to another series of queries initiated by Mikela (Line 1, below). Renee attempted to prompt Shakira to elaborate her criteria for closeness by asking whether or not the same standard should be applied to the region surrounding the second mode as well (Line 1). Lena suggested not, referring to "important" numbers (Line 21), but Renee explicitly rejected Lena's proposal that an exception to the standard of closeness be made (Line 22). Shakira's response appeared to rely on her authority as the drawer of the boundary, and the exchange ended with Renee challenging this assertion (Line 28).

1	Mikela:	How do you get of the agree, if 47, 48, 49, 50 and 51?
2	Shakira:	=Okay.
3	Mikela:	How are those agreeing?
4	Shakira:	First you get the median and mode. Okay. For the median and mode here 49 and 48. So? those two numbers you keep automatically. Okay. What are the numbers that closest to 49??
5	Amanda:	47.
6	Lena:	And 48.
7	Shakira:	What??
8	Mikela:	47…, 48…, 50…, 51...
9	Shakira:	No. They are close to 49 up here. 50 is one right after 49, and 51 is going after 50. So those two are closest to the 49. They agree the most.
10	Mikela:	Why did you? [Hand gesture, open palms indicating confusion]

11   Amanda: Why ###
12   Mikela: <Why did she?> [Directing her gaze toward to her table partner, Amanda and hand gesturing again to indicate her confusion]
13   Shakira: Renee.
14   Renee: Okay. How can you put 47 and 49? But you can't put 37 and 39??
15   Shakira: Huh??
16   Renee: In the two little circle things you have up there, in the little groups how can you put 47 and 49 but not 37 and 39??
17   Shakira: What did you say?
18   Renee: How can you put 47 and 49, but you can't put 37 and 39??
19   Shakira: Okay, you said put 47 and 49, but I can't put 37 and 39??
20   Renee: No. Why can't put 47 and 49 together, no, no, no, why can't you put 37 and 39 together when you can? put 47 and 49 together?
21   Lena: Because those are not the important numbers. <47>
22   Renee: <No!> (refers to Lena) But she said the mode is 37 no, the mode is 39 and 49.
23   Shakira: <The mode is 39.>
24   Lena: <I know. >
25   Renee: <So> why she can't put 37 together 39? 39 and 49 are the important numbers.
26   Shakira: Okay. Okay. 37 isn't the one closest numbers to 39.
27   Teacher: Okay, So you're. <If I understand correct>
28   Renee: <But 47 is??> [Tone indicates disbelief]

In the next conversational exchange, the teacher revoiced Shakira's approach, pointing out its appeal to trust in repeated measures and aligning her approach with Jamir, who had presented his measure and rationale the day before. The teacher then attempted to negotiate a boundary for the Shakira's regions of agreement that might extend beyond her personal writ: "And what's your rule, how come 50 is OK [Teacher gestures toward upper clump], but 54 isn't? How come 51 is OK [pauses]? That's, I think, what we are asking."

Shakira responded by seeking clarification: "How close is 49 to 54?" The teacher followed up by asking: "Are you asking us? How would we find out? How close is it?" Lena immediately suggested subtracting 49 from 54, with the result of 5. In the next exchange, a boundary condition was negotiated:

Shakir   How close is 49 and 51??
Teacher: Okay. It's two away??
Shakir   Right. Those are the numbers, 49 and 50 and 51 are closest numbers in the clump to the 49.
Teacher: So, are you defining (gestures toward region)?? Are

	you suggesting that if the number is within two??
Shakir	Right.
Teacher:	Okay. But see we didn't understand that. This makes it clear. If the number is within two, then you would include it in the clump. Does it matter the direction?
Shakir	The direction? No.

Although the teacher prompted this exchange, in the subsequent exchange two other classmates held Shakira's (perhaps reluctant) acceptance of a defining difference to account. Mikela repeated Renee's earlier inquiry: "Why didn't you include 37?" Renee added that "37 is 2 away from 39," and Shakira signaled her agreement by altering the display depicted in Figure 4 to enlarge the lower region to include the value of 37. Shakira then reminded the class again about how the regions depicted related to agreement, asking the class to consider "the tightness of the clump", "how the measurements agree", and "think about what would happen if everybody agree on those numbers". After a few exchanges which established that the class considered the size of the region as a legitimate indicator of agreement, the teacher again prompted further consideration of the measure: "I think we understand now how to count whether or not something belongs or not. So what do we do now? If this is giving us a precision number, what would that number be?"

Shakira suggested: "You see all the numbers that wrapped around. All the numbers wrapped around 48. Become wrap-around-number. That would be your precision number." The teacher prompted: "How do I know how much wrapping around they are doing?" The teacher appealed to the class for help in generating a precision number.
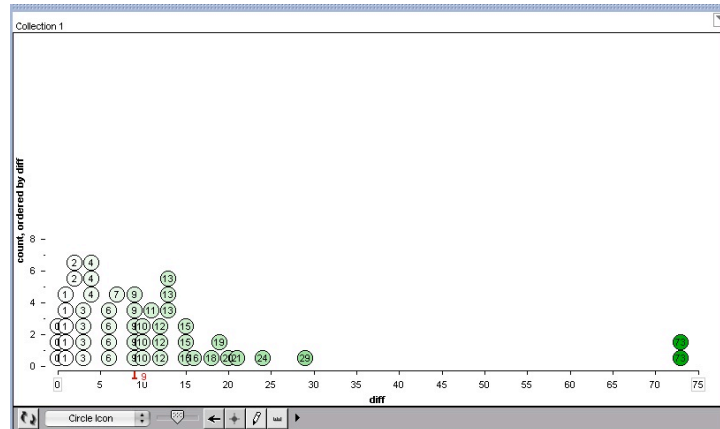

Figure 5. Mikela's measure of precision, the median (9) of differences among each observed value and the sample median.

This appeal was responded to by Mikela, who had previously invented a measure of precision based on the median of the absolute values of the differences between each measurement and the sample median,

demonstrating that when crude tools were used to measure the height of the flagpole, its value was high, indicating lack of agreement, and when the tools used were better, its value was low, indicating more agreement. Mikela's measure is displayed in Figure 5. (It was a revision of the measure displayed in Figure 2, prompted by considering a hypothetical instance proposed by the teacher involving a large number of very small deviations.) Mikela re-deployed her focus on difference in her invented measure to suggest using each mode as a reference point, and then finding differences between each value and the corresponding mode. She recommended that Shakira then consider using the median of the absolute value of these differences as the measure. Because the class was nearly over, the teacher asked for some clarifications of Mikela's proposal and then endorsed distance as a good way of looking at the "tightness" of a region. The teacher also suggested that Shakira consider counting the number of values within a cluster, asking if that might also indicate agreement. The class concluded with the teacher situating Shakira's measure within the emerging collection of measurers being developed in the class: "Could you do some thinking about that and get back to us about different ways that we can take advantage of your method? And, try to help us understand how we know where to draw the clumps, because we have to be able to tell other people."

Discussion

Measurement is often conceived as a mundane activity, and in school it often arrives pre-formed. Students may learn "skills" of measuring, but rarely are they asked to grapple with the foundational problematic of the relation between a measure and a particular phenomenon. We suggest an alternative tactic, one in which students are positioned to invent and revise measures, because inventing measures inherently involves structuring phenomena. In this instance, we positioned students to invent and revise measures of variability. Many of these invented solutions coordinated centre and spread in ways that anticipate the kinds of solutions that are used in the discipline of statistics, such as inter-quartile range and deviation-based metrics such as average deviation and standard deviation. This form of structuring variability often escapes much older students. For example, Garfield, delMas and Chance (2007) indicated that despite "multiple lessons on measures of variability" (p. 141), college students did not understand the grounds of these measures of variability until nearly the end of a course of conceptually oriented instruction.

Although invention is often associated with individual acts, the case study of Shakira suggests the importance of both collective and individual views, what Cobb (1999) referred to as an *emergent perspective*. Shakira initially structured variability by literally drawing regions surrounding the modes and median of the sample. The mode corresponded to literal agreement among measurers and the median as the best guess of the true score. But this first approximation, although intuitively sound, relied primarily on Shakira's authority. Her classmates pressed her to make her decisions more transparent and consistent by seeking definition of the bounded regions and by suggesting that she re-conceptualise the regions as deviation scores. Other students' experiences and invented measures came

into contact with Shakira's, resulting in some transformation of her initial intuitions. As the episode drew to a close, Sharika began to consider how to develop a quantity, a work in progress which was made possible by the emergence of a more explicit structure of variability.

The students in the classroom, although clearly not practicing statistical scientists, were embarked on the development of what Goodwin (1994) called *professional vision*. That term refers to practices that uniquely allow a practitioner to experience the profession in ways that align one practitioner's activity with another's. Although clearly we are not involved here in developing professional statisticians, nonetheless students' participation showcases a forum for developing practices of measure, and as a result, coming to see variability in ways that were more like those of the discipline. Signals of participation in a practice, rather than a string of activity, were suggested by Shakira's initial attempt to justify her measure by appeal to grounds of sensibility (e.g., how the measure reflected a quality of agreement among measurers), and by her classmates' insistence, and her acceptance, that a measure be consistent and coherent. These negotiations suggested development of an aesthetic consistent with disciplinary values of clarity, consistency, and prospective generalisation. The nature of the interactions also indicates that mathematical identities reflecting this aesthetic were being constituted through participation in the practice of measure critique (Boaler, 2002). The teacher's role, in addition to orchestrating the ongoing negotiation about the nature of the measure, also was to promote the notion of quantity by suggesting that the measure of variability should correspond to changes in state (e.g., amount) of variability. In our instructional design, changes in variability arose from use of two tools, one crude and one more refined. In subsequent activity, the structuring of variability generated by inventing displays and statistics set the stage for re-interpreting these structures as reflecting the operation of chance processes, as students invented models of the original batch of measurements that included random error.

In summary, although variability is at first glance an antonym of structure, a view from the mathematics of chance (Stevens & Hall, 1998) suggests instead that variability reflects a structure, however initially obscure. For both practitioners of the discipline and students, variability is merely difference until it has been cognitively and materially transformed. Since structure, as we have seen, has to be constructed, the utility of particular measures (statistics), and their potential to capture and represent such a structure, always remain open to contest. The students in this class were embarked on this realization, in a manner that we anticipated would promote their broader participation in the mathematics of data modeling.

References

Boaler, J. (2002). Experiencing school mathematics. *Traditional and reform approaches to teaching and their impact on student learning.* Mahwah, NJ: Lawrence Erlbaum.

Brown, A. L. (2002). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141-178.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*(9), 801-823.

Cobb, P. (1999). Individual and collective mathematical learning: The case of statistical data analysis. *Mathematical Thinking and Learning, 1,* 5-44.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in education research. *Educational Researcher, 32*(1), 9-13.

Crosby, A. W. (1997). *The measure of reality.* Cambridge MA: Cambridge University Press.

DeGroot, M. H. (1975). *Probability and statistics.* Sydney: Addison-Wesley.

DelMas, R. C., & Liu, Y. (2007). Students' conceptual understanding of the standard deviation. In M. C. Lovett & P. Shah (eds.), *Thinking with data (*pp. 87-116). New York: Lawrence Erlbaum.

DiSessa, A. (2004). Metarepresentation: Native competencies and targets for instruction. *Cognition and Instruction, 22*(3), 293-331.

Garfield, J., DelMas, R. C., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. C. Lovett & P. Shah (eds.), *Thinking with data (*pp. 117-147). New York: Lawrence Erlbaum.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96,* 606-633.

Greeno, J. G., & MMAP (1998). The situativity of knowing, learning and research. *American Psychologist, 53,* 5-26.

Hall, R., Wright, K., & Wieckert, K. (2007). Interactive and historical processes of distributing statistical concepts through work organization. *Mind, Culture, and Activity, 14*(1&2), 103-127.

Horvath, J., & Lehrer, R. (1998). A model-based perspective on the development of children's understandings of chance and uncertainty. In S. P. Lajoie (Ed.), *Reflections on statistics* (pp 121-148). Mahwah, NJ: Lawrence Erlbaum.

Konold, C. (2007). Designing a data analysis tool for learning. In M. C. Lovett & P. Shah (Eds.), *Thinking with data (*pp. 267-291). New York: Lawrence Erlbaum.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6,* 59-98.

Konold, C., & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L. D. English (Ed.), *Handbook of International Research in Mathematics Education* (2nd ed). Philadelphia, PA: Taylor & Francis.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration.* [Computer software]. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33,* 259-289.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69-108.

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin, & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 179-192). Reston, VA: National Council of Teachers of Mathematics.

Lehrer, R., & Schauble, L. (2002). Symbolic communication in mathematics and science: Co-constituting inscription and thought. In E. D. Amsel & J. Byrnes (Eds.), *Language, literacy, and cognitive development*: *The development and consequences of symbolic communication.* (pp. 167-192). Mahwah, NJ: Lawrence Erlbaum.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. C. Lovett & P. Shah (Eds.), *Thinking with data.* (pp. 149-176). New York: Lawrence Erlbaum.

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.

Lehrer, R., Konold, C., & Kim, M.J. (2006April). *Constructing data, modeling chance in*

the middle school. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental tests. Reading, MA: Addison-Wesley.

Metz, K. E. (1998a). Emergent understanding and attribution of randomness: Comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction*, *16*(3), 285-365.

Metz, K. E. (1998b). Emergent ideas of chance and probability in primary-grade children. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 149-174). Mahwah, NJ: Lawrence Erlbaum.

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*, 131-156.

Porter, T. M. (1986). *The rise of statistical thinking 1820-1900.* Princeton: Princeton University Press.

Shaughnessy, J. M., Canada, D., & Ciancetta, M. (2003, July). *Middle school students' thinking about variability in repeated trials: A cross-task comparison.* In N. A. Pateman, B. Dougherty & J. Zilloux (Eds.), *Proceedings of Joint Meeting of PME27 and PMENAXXV* (Vol. 4, pp. 159-166). Honolulu, HI: Program Committee..

Shaughnessy, J. M., Watson, J. M., Moritz, J., & Reading, C. (1999, April). School mathematics students' acknowledgement of statistical variation. In C. Maher (Chair), *There's more to life than centers.* Pre-session Research Symposium, 77th annual meeting of the National Council of Teachers of Mathematics, San Francisco, CA.

Stevens, R., & Hall, R. (1998). Disciplined perception: Learning to see in technoscience. In M. Lampert & M. L. Blunk (Eds.), *Talking mathematics.* Cambridge, MA: Cambridge University Press.

Torok, R., & Watson, J. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal, 12*(2*)*, 147-169.

Thompson, P. W., Liu, Y., & Saldanha, L. A. (2007). Intricacies of statistical inference and teachers' understanding of them. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 207-231). New York: Lawrence Erlbaum.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals.* Mahwah, NJ: Lawrence Erlbaum.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223-248.

Authors

Richard Lehrer, Vanderbilt University, USA. Email: <rich.lehrer@vanderbilt.edu>

Min-joung Kim, Vanderbilt University, USA. Email: <Min-joung.kim@vanderbilt.edu>