

EXPLORING STUDENTS' CONCEPTIONS OF THE STANDARD DEVIATION

ROBERT DELMAS
University of Minnesota
delma001@umn.edu

YAN LIU
Vanderbilt University
yan.liu@vanderbilt.edu

SUMMARY

This study investigated introductory statistics students' conceptual understanding of the standard deviation. A computer environment was designed to promote students' ability to coordinate characteristics of variation of values about the mean with the size of the standard deviation as a measure of that variation. Twelve students participated in an interview divided into two primary phases, an exploration phase where students rearranged histogram bars to produce the largest and smallest standard deviation, and a testing phase where students compared the sizes of the standard deviation of two distributions. Analysis of data revealed conceptions and strategies that students used to construct their arrangements and make comparisons. In general, students moved from simple, one-dimensional understandings of the standard deviation that did not consider variation about the mean to more mean-centered conceptualizations that coordinated the effects of frequency (density) and deviation from the mean. Discussions of the results and implications for instruction and further research are presented.

Keywords: Standard deviation; Variability; Conceptions; Strategies; Interviews

1. INTRODUCTION

1.1. STUDENTS' UNDERSTANDING OF VARIABILITY

The current research study is partially motivated by the first author's collaborative research with Joan Garfield and Beth Chance on students' understanding of sampling distributions. One of the main findings from a decade of research is that this is a very difficult concept for students to develop (see Chance, delMas, & Garfield, 2004; delMas, Garfield, & Chance, 2004). DelMas, Garfield, and Chance (2004; Chance, delMas, & Garfield, 2004) speculate that the difficulty is partially due to students not having firm understandings of prerequisite statistical concepts such as center, distribution, and variability. With respect to variability, part of the difficulty may also stem from students' misunderstanding of how variability can be represented graphically. For example, when presented with a histogram, some students judge the variability of the distribution on the basis of variation in the heights of bars, or the perceived "bumpiness" of the graph, rather than the relative density of the data around the mean (Garfield, delMas, & Chance, 1999). Helping students develop a better understanding of variability and its representation may be one way to support a better understanding of sampling distributions.

Very little research has been conducted on students' understanding of variability (Reading & Shaughnessy, 2004; Shaughnessy, 1997), despite the central role the concept plays in statistics (Hoerl & Snee, 2001; Moore 1990; Snee, 1990) and an apparent conceptual gap in students' understanding of

variability (Reading & Shaughnessy, 2004; Shaughnessy, 1997). A few investigations have been conducted into students' understanding of sampling variability and instructional approaches that affect this understanding. Reading and Shaughnessy (2004) present evidence of different levels of sophistication in elementary and secondary students' reasoning about sample variation. Meletiou-Mavrotheris and Lee (2002) found that an instructional design that emphasized statistical variation and statistical process produced a better understanding of the standard deviation, among other concepts, in a group of undergraduates. Students in the study saw the standard deviation as a measure of spread that represented a type of average deviation from the mean. They were also better at taking both center and spread into account when reasoning about sampling variation in comparison to findings from earlier studies (e.g., Shaughnessy, Watson, Moritz, & Reading, 1999).

Little is known, however, about students' understanding of *measures* of variation, how this understanding develops, or how students might apply their understanding to make comparisons of variation between two or more distributions. The latter ability represents an important aspect of statistical literacy that is needed both for interpreting research results and for everyday decision making. An understanding of statistical variation and of measures of variation is also needed to understand conceptually complex concepts such as sampling distributions, inference, and p-values (Chance, delMas, & Garfield, 2004; delMas, Garfield, & Chance, 1999; Saldahna & Thompson, 2002; Thompson, Saldahna, & Liu, 2004). An incomplete understanding of the standard deviation may limit students' understanding of these more advanced topics.

1.2. A CONCEPTUAL ANALYSIS OF STANDARD DEVIATION

Shaughnessy (1997; Reading & Shaughnessy, 2004) noted that the standard deviation is both computationally complex and difficult to motivate as a measure of variability. Part of this difficulty may stem from a lack of accessible models and metaphors for students' conceptions of the standard deviation (Reading & Shaughnessy, 2004). Most instruction on the standard deviation tends to emphasize teaching a formula, practice with performing calculations, and tying the standard deviation to the empirical rule of the normal distribution. This emphasis on calculations and procedures does not necessarily promote a conceptual understanding of standard deviation. A conceptual model of the standard deviation is needed to develop instruction that promotes the concept. We conjecture that such a model involves the coordination of several underlying statistical concepts from which the concept of standard deviation is constructed.

One of these fundamental concepts is distribution. The students in the current study worked with distributions of discrete variables, so the concept of distribution in this paper will be described in those terms. Essentially, an understanding of distribution requires a conception of a variable and the accumulating frequency of its possible values. Therefore, a visual or graphical understanding of distribution involves the coordination of values and density.

A second fundamental concept is that of the arithmetic mean. A conceptual understanding of the standard deviation requires more than the knowledge of a procedure for calculating the mean, either in procedural or symbolic form (e.g., $\Sigma x/n$). Imagery that metaphorically considers the mean to behave like a self-adjusting fulcrum on a balance comes close to the necessary conception. Such imagery supports the development of the third foundational concept, deviation from the mean. It is through the coordination of distribution (as represented by the coordination of value and frequency) and deviation (as distance from the mean) that a dynamic conception of the standard deviation is derived as the relative density of values about the mean. A student who possesses this understanding can anticipate how the possible values of a variable and their respective frequencies will, independently and jointly, affect the standard deviation. An illustration from the computer program used in the present study may help to clarify this conception.

To clarify the above ideas, Figure 1a presents screen displays from a computer program developed by the first author to study and promote students' understanding of the standard deviation. The program represents the distribution of a variable along the number line as a histogram made up of bars composed of a certain number of rectangles, each of which represents one data point (or observation). The location of a bar indicates the value of all the data represented by the bar (e.g., the

tallest bar has eight data points each with a value of 4). The point location of the mean is indicated by an arrow along the number line. The standard deviation is reported as a numerical value below the arrow, and its size is represented by the length of a horizontal bar that extends below and above the mean. The deviation from the mean of each data point is printed within each rectangle. Overall, the program displays information that has the potential to enable students to mentally coordinate how changes in data points simultaneously affect the mean, deviations from the mean, and the standard deviation.

For example, a student can be asked to anticipate how the mean, deviations, and standard deviation shown in Fig. 1a are affected if the two lowest data points had a value of 1 instead of 2, resulting in moving the lowest bar on the left from 2 to 1 (see Fig. 1b). A student with a fully coordinated conception of standard deviation should anticipate that moving the lowest bar to a value of 1 shifts the mean to a slightly lower value and that all deviations from the mean would change simultaneously, i.e., the deviations of the two tallest bars increase, the deviations of the third tallest bar decrease, and possibly the deviations of the shortest bar increase. A student who is able to coordinate density (or frequency) with deviation should realize that the larger frequencies of the two tallest bars coupled with increases in deviation are likely to outweigh the few values in the third tallest bar that had a slight decrease in deviation. This would result in a larger density of values farther away from the mean and, therefore, an increase in the standard deviation.

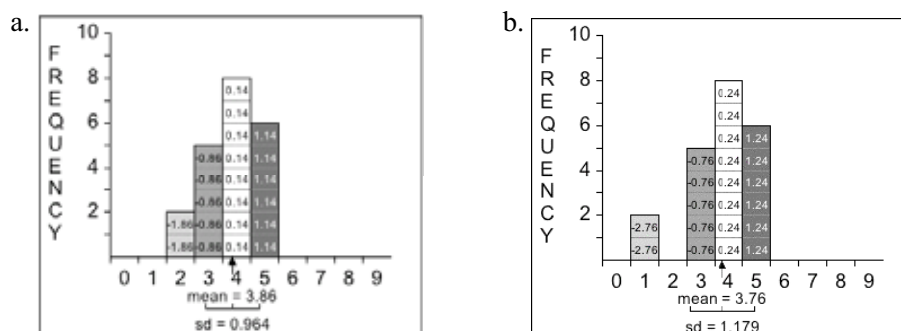


Figure 1. A graphic representation of standard deviation and its related concepts.

A student who understands these relationships should be able to reliably compare two distributions with respect to their standard deviations. Consider the two pairs of graphs presented in Figure 2a and 2b. Knowing only the location of the mean in both distributions, a student might reason that the graph on the left in Figure 2a has a larger standard deviation because there is a lower density of values around the mean. The same conception would lead to the conclusion that the graph on the right in Figure 2b has the larger standard deviation.

1.3. GOALS AND APPROACH

The main research goal of the current study was to gain a better understanding of how students' understandings of the standard deviation develop as they interact with a designed computer environment and an interviewer. The conceptual analysis provided in the previous section served as a framework for describing and interpreting students' understanding. The investigation was exploratory in nature and did not attempt to fully control factors that may or may not contribute to students' understanding of the standard deviation. Nonetheless, careful planning and consideration went into the design of the interactive experience so that it had the potential to promote the goal of a fully coordinated understanding of the standard deviation.

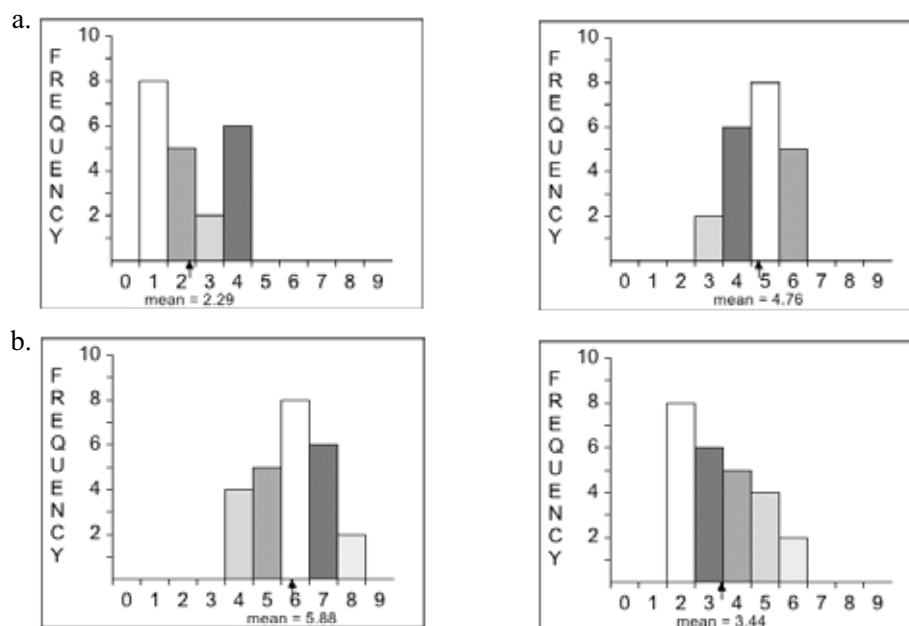


Figure 2. Comparing standard deviations in pairs of graphs.

In order to capture students' understanding, a computer application was written in Java and compiled for the Macintosh computer. The program was based on design principles for creating conceptually enhanced software (delMas, 1997; Nickerson, 1995; Snir, Smith, & Grosslight, 1995). Such software facilitates the development of conceptual understanding by accommodating students' current levels of understanding, providing familiar models and representations, supporting exploration by promoting the frequent generation and testing of predictions, providing clear feedback, and drawing students' attention to aspects of a situation or problem that can be easily dismissed or not observed under normal conditions. Conceptually enhanced software provides some of the conditions that foster conceptual change in science education (Chin & Brewer, 1993; Posner, Strike, Hewson, & Gerzog, 1982; Roth & Anderson, 1988). Among these conditions, the most pertinent for the present study is the engagement of students in a task that has the potential to repeatedly produce credible data that is inconsistent with students' current understanding in order to support reflective change of the underlying misconceptions.

Based on a pilot study, the computer program and interview protocol were designed with several instructional goals in mind. The primary instructional goal was to help students develop an understanding of how deviation from the mean and frequency combined to determine the value of the standard deviation. This entails that students distinguish value from frequency, recognize each value's deviation from the mean, understand that a distribution and its mirror image have the same standard deviation, and understand that the value of the standard deviation is independent of where the distribution is centered. A second instructional goal was to promote an understanding of how the shape of a distribution is related to the size of the standard deviation (e.g., given the same set of bars, a unimodal, symmetric distribution tends to have a smaller standard deviation than a skewed distribution).

2. METHODS

2.1. PARTICIPANTS

Students were recruited from four sections of an introductory statistics course at a large Midwest research university in the United States at the beginning of the spring 2003 term. Three of the course sections were taught by an instructor with a masters degree in mathematics education and four years

of experience teaching the course. The other section was taught by a Ph.D. student in mathematics education with an emphasis in statistics education who had one year of experience teaching the course. The first author made a brief presentation about the nature of the research study to each of the four course sections during the second week of the term. Students were offered a \$20 gift certificate to the university bookstore as an incentive. Of the 129 registered students, 27 initially volunteered to participate in the study. However, only thirteen of the students scheduled and participated in an interview. One of these students was dropped from the analysis because his responses indicated that he did not understand the nature of the research task. The final set of twelve students consisted of five males and seven females. As illustrated in Table 1, the group of participants did not differ noticeably from the non-participants, with the possible exception that a higher percentage of the participants received a final grade of A in the introductory statistics course when compared to the non-participants, although the difference is not statistically significant.

Table 1. Comparison of statistics students who did and did not participate in the study.

Variable	Participants		Non-Participants		Test Statistic	p-Value
	n	Mean	n	Mean		
Mean ACT Composite Score	11	19.8	108	19.2	$t(117) = 0.70$	0.486
Mean ACT Mathematics Score	11	19.0	108	18.2	$t(117) = 0.95$	0.342
Mean High School %-tile Rank	9	55.3	101	53.5	$t(108) = 0.36$	0.719
Cumulative College GPA	12	2.82	115	2.90	$t(125) = 0.43$	0.668
Cumulative College Credits	12	31.5	117	39.2	$t(127) = 1.00$	0.317
Percent Female	12	58.3	117	68.4	$\chi^2(1) = 0.15$	0.701
Percent Caucasian	12	58.3	117	58.1	$\chi^2(1) = 0.00$	0.999
Percent Course Grade of A	12	58.3	117	30.8	$\chi^2(1) = 2.58$	0.108
Percent Course Grade A or B	12	64.6	117	61.6	$\chi^2(1) = 0.001$	0.971

2.2. PRIOR INSTRUCTION

By the time this study started during the fourth week of instruction, the course had covered distributions, methods for constructing graphs (stem-and-leaf plots, dot plots, histograms, and box plots), measures of center (mode, median, and mean), and measures of variability (range, interquartile range, and the standard deviation). With respect to the standard deviation, students had participated in an activity exploring factors that affect the size of the standard deviation. During this activity, students compared nine pairs of graphs to determine which one had a larger standard deviation. Students worked in groups of two or three, identified characteristics of the graphs thought to affect the size of the standard deviation, recorded their predictions, and received feedback. The goal of the activity was to help students see that the size of the standard deviation is related to how values are spread out and away from the mean, the only factor associated with the correct choice across all nine pairs. Therefore, before the interviews, all of the students had received considerable exposure to the standard deviation as a measure of variability.

2.3. RESEARCH MATERIALS AND PROCEDURES

Students interacted with the computer program during interviews conducted by the first author. A digital video camera was used to capture a student's utterances and actions. The computer program wrote data to a separate file to capture students' actions and choices. Each movement or choice was time stamped so that it could be coordinated with the digital video recording.

Each interview was conducted in three phases: introduction, exploration, and testing. The interface for the introduction phase was designed so that students could become familiar with the controls and information displayed in the window. The interface for the exploration phase was similar

to that of the introduction and presented students with tasks that required them to meet specified criteria. The interface for the testing phase presented ten test items designed to assess students' ability to compare the sizes of the standard deviations of two distributions after completing the exploration phase. Some of the same information presented during the exploration phase was available to support students' decisions. The following sections present more detail about each phase.

Introduction Phase

The introduction phase was relatively short and took only a few minutes. Students were introduced to the computer environment by moving two bars respectively representing frequencies of 5 and 8 within the graphing area. The program instructed the student to turn specified buttons on and off to display different information (mean, standard deviation, deviations, squared deviations, and the standard deviation formula). The interviewer described each type of information. Students were also asked to move the two bars within the graphing area while observing how the various values changed. When students indicated they were ready, they clicked a button labeled NEXT to start the exploration phase.

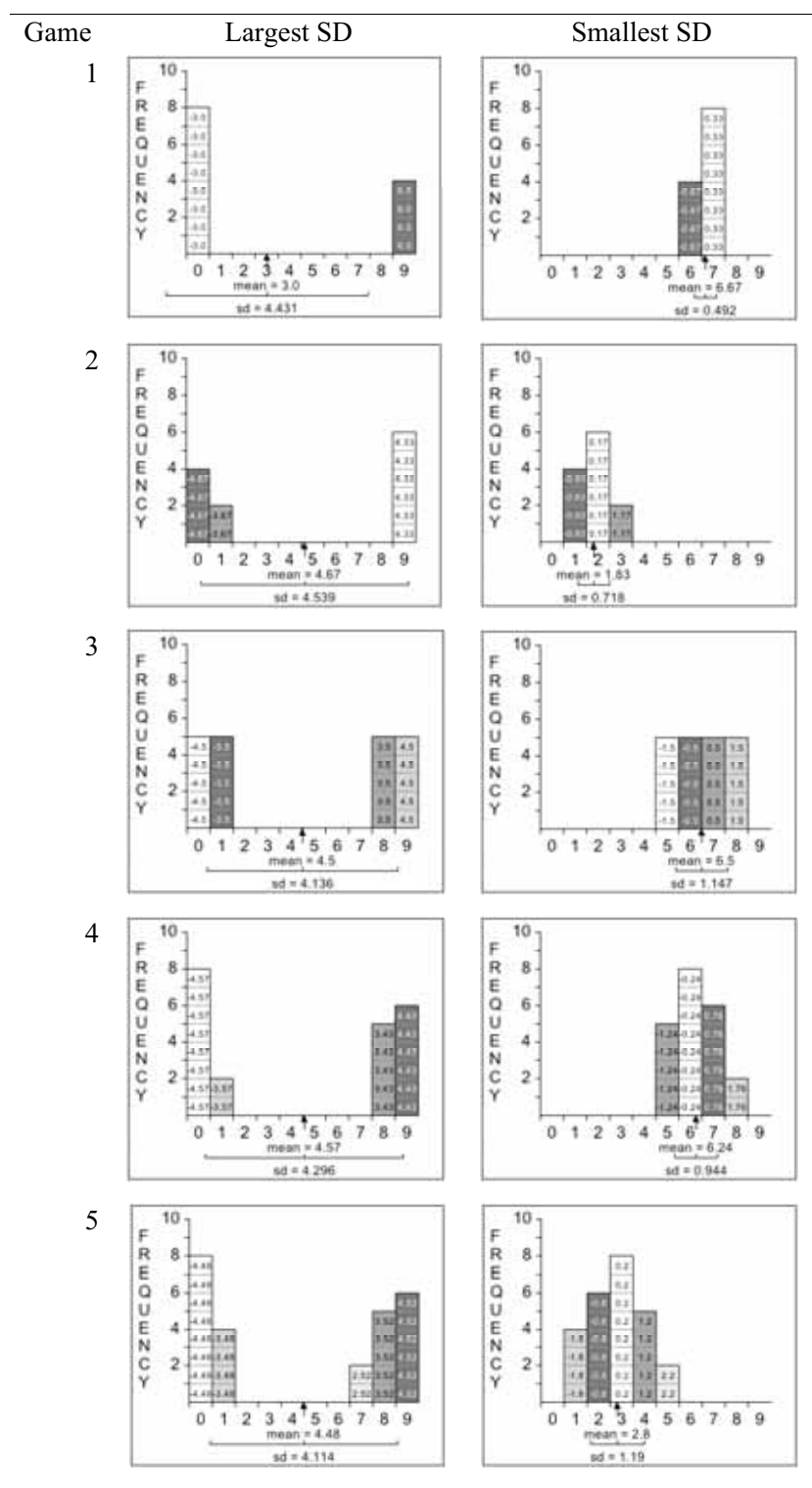
Exploration Phase

The exploration phase was designed to help students learn about factors that affect the standard deviation. Students could move frequency bars within the graphing area and observe simultaneous changes in the values of the mean, standard deviation, individual and summed deviations and squared deviations, and the standard deviation equation. Each student was presented with five different sets of bars (see Table 2), one set at a time. The number of bars ranged from two in the first set to five in the last. For each set of bars, the first task asked a student to find an arrangement that produced the largest possible value for the standard deviation, followed by a second task of finding another arrangement that produced the same value. The third through fifth tasks required three different arrangements that each produced the smallest value for the standard deviation. Each set of bars along with the five tasks comprised a "game".

Students used a CHECK button to determine if an arrangement met the stated criterion for a task. Students were asked to describe what they were thinking, doing, or attending to before checking. In addition, the interviewer asked students to state why they thought an arrangement did or did not meet a criterion once the CHECK button was selected and feedback received. The interviewer also posed questions or created additional bar arrangements to explore a misunderstanding or the stability of a student's reasoning.

Conceptual change can be facilitated through a combination of discovery learning and direct instruction (e.g., Burbules & Linn, 1988) and by drawing students' attention to relevant aspects that might be neglected (e.g., Hardiman, Pollatsek, & Well, 1986). When students had difficulty finding an arrangement that met a task criterion, or when they appeared to be moving in the direction of one of the goals, the interviewer used several approaches to support the student. A mean-centered conception of the standard deviation was promoted by drawing attention to the values of the deviations, by asking students how hypothetical movements of the bars would potentially affect the mean, and by the interviewer modeling reasoning of how the distribution of deviation densities affected the values of the mean and standard deviation. The second goal of promoting an understanding of how distribution shape is related to the size of the standard deviation was also supported. Students were asked to predict how the mean and standard deviation would be affected by bell-shaped and skewed arrangements of the same bar sets. Students were also asked to judge the extent to which a distribution was bell-shaped or whether the bars could be arranged to produce a more bell-shaped (or symmetric) distribution. If the student made a change to a distribution in response to a conjecture, the interviewer drew the students' attention to the value of the standard deviation and the direction of the change.

Table 2. Bar sets with possible solutions for the five games in the Exploration phase.

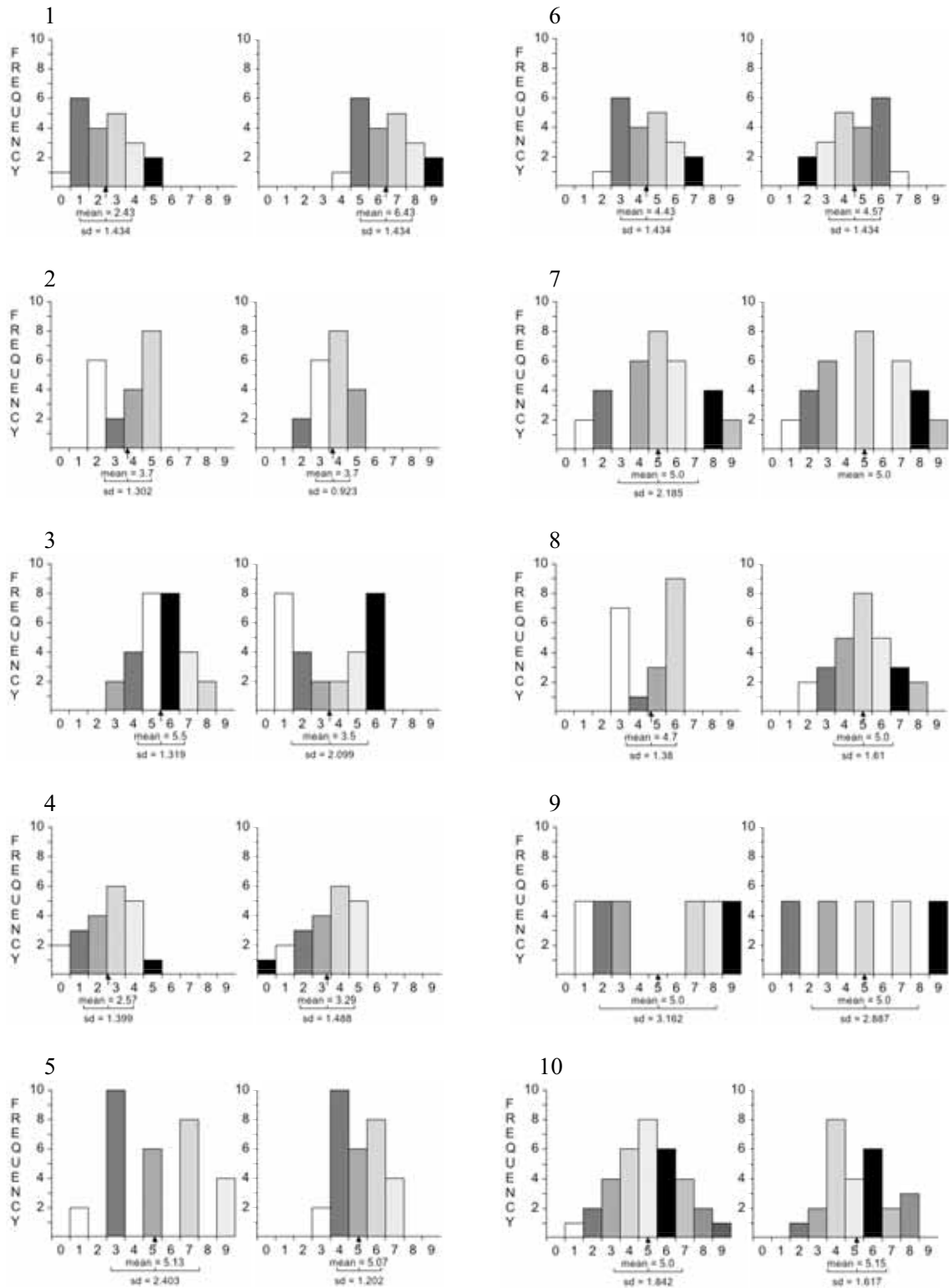


Testing Phase

The testing phase was designed to assess students' understanding of factors that influence the standard deviation. Students were asked to answer 10 test items where each test item presented a pair of histograms (see Table 3). For each pair of histograms, the average and standard deviation were displayed for the graph on the left, but only the average was displayed for the graph on the right. The students were asked to make judgments on whether the standard deviation for the graph on the right

was smaller than, larger than, or the same as the graph on the left. Once a student made a judgment, the investigator asked the student to explain his justification and reasoning. The student then clicked the CHECK button to check the answer, at which time the program displayed the standard deviation for the graph on the right and stated whether or not the answer was correct.

Table 3. Ten test items presented in the Testing phase.



The test items were based on the two stated goals of the study. None of the bar sets were identical to the sets of bars presented during the exploration phase. Each pair of graphs in test items 1 through 7 had identical bars placed in different arrangements. Test item 1 required students to recognize that both graphs had the same bar arrangement in different locations, whereas students had to recognize that one distribution was the mirror image of the other in test item 6. Test item 4 was specifically designed to see if students understood that given the same frequencies and range, a distribution with a stronger skew tended to have a larger standard deviation. Test items 2 and 3 tested students' sensitivity to the density of values around the mean. Test items 5 and 7 were similar in this respect, however the bars in both graphs were in the same order. These latter two items were designed to identify students who might attend only to the order of the bars and not to the relative density of deviations from the mean. The pair of graphs in test items 8, 9, and 10 did not have identical bars and also had a different number of values represented in each graph. In a pilot study, some students initially thought that evenly spacing the bars across the full range of the number scale would produce the largest value for the standard deviation. This misunderstanding was tested directly by test item 9, although this thinking could also be used to answer test item 5.

Test items 8 and 10 were designed to challenge the belief that a perfectly symmetric and bell-shaped distribution will always have a smaller standard deviation. Students were expected to find these items more difficult than the others. For both test items, each graph had characteristics that indicated it could have either the smaller or larger standard deviation of the pair. For example, in test item 8, the U-shaped graph on the left appeared to have less density about the mean while the graph on the right was perfectly symmetric with apparently a large portion of the density centered around the mean. Based on this, a student might have reasoned that the graph on the right had the smaller standard deviation. However, the graph on the left had a smaller range and represented a smaller number of values. Both of these characteristics made it possible for the graph on the left to have relatively more density around the mean than the graph on the right (which was the case). A reasonable response to test items 8 and 10 would have been that both standard deviations (or the variances) needed to be calculated to determine how they differ. The program provided the sum of squared deviations for both graphs to support these calculations.

3. RESULTS

3.1. EXPLORATION PHASE

The transcript analysis of the exploration phase focused on categorizing and describing the justifications students presented when asked why they expected an arrangement to satisfy a task criterion (e.g., the largest possible standard deviation). These justifications were taken to represent the students' conceptual understandings of the standard deviation as they developed during the interview. While some students appeared to start the interview with a fairly sophisticated understanding of factors that affect the standard deviation, most students started with a very simple, rule oriented approach.

Eleven broad categories of justifications were identified with the following labels (in alphabetical order): Balance, Bell-Shaped, Big Mean, Contiguous, Equally Spread Out, Far Away, Guess and Check, Location, Mean in the Middle, Mirror Image, and More Bars in the Middle. Each justification is described and illustrated in more detail in the subsequent sections. This section is organized in five subsections: Justifications for arrangements intended to produce the largest standard deviation (task 1) are presented first, followed by justifications for arrangements intended to produce the smallest standard deviation (task 3), and then justifications for arrangements intended to produce the same standard deviation (tasks 2, 4, and 5). These subsections are followed by a general discussion of students who began to coordinate conceptions about the standard deviation, and a final section that summarizes the findings from the exploration phase.

Graphs of students' bar arrangements are presented to illustrate excerpts from student interviews. Each bar in the software program was presented in a different color, however, the bars presented in the figures are in gray tones. Table 4 presents a legend for the bar colors.

Table 4. Legend of Bar Colors

□ = red	■ = blue	■ = orange	■ = green	□ = yellow
---------	----------	------------	-----------	------------

Largest Standard Deviation

For the first game, most students placed one bar at each of the extreme positions of the number line to produce the largest standard deviation. The subsequent games with more than two bars revealed more details of the students' thinking and strategies. In general, the arrangements represented placing the bars far apart or having them spread out. The typical order was to place the tallest bars at the extremes with subsequent bars placed so that heights decreased toward the center of the distribution. Three general strategies were identified based on students' justifications for their arrangements.

Far Away – Values

Some students stated that an arrangement should produce the largest possible standard deviation if the values (or the bars) are placed as far away from each other as possible. No mention was made of distance (or deviation) from the mean in these statements. This type of justification was prevalent across all the games.

Equally Spread Out

In the Equally Spread Out justification, the student believed that the largest standard deviation occurred when the bars were spread out across the entire range of the number line with equal spacing between the bars. This belief may have resulted from the in-class activity described earlier. Students may have translated “spread out away from the mean” to mean “equally spread out”. For example, with three bars, the student places one bar above a value of 0, one above 9, and then places the third bar above 4 or 5. Some students using this approach realized that the third bar could not be placed with equal distance between the lowest and highest positions, so they would move the third bar back and forth between values of 4 and 5 to see which placement produced the larger standard deviation.

Nancy provided an example of this expectation. She first created the arrangement in Figure 3a when asked to produce the largest standard deviation in Game 2, followed by the arrangement in Figure 3b. The interviewer then asked her about the arrangements.

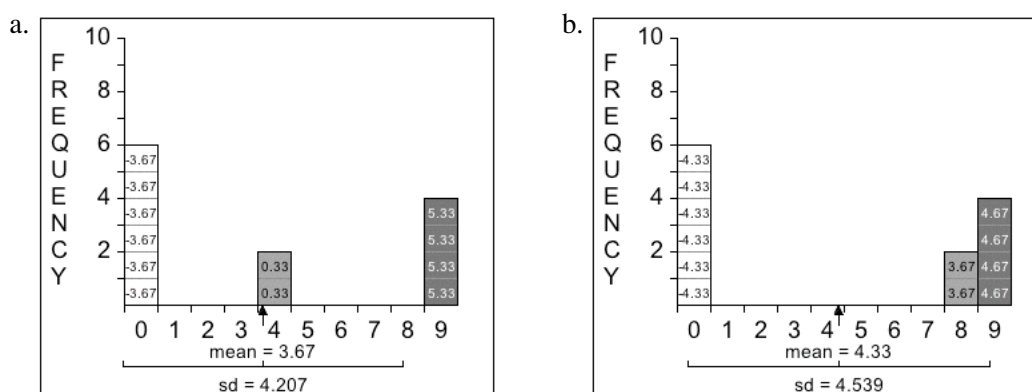


Figure 3. Nancy's arrangements for the largest standard deviation in Game 2.

Intv: And can you tell me what you were thinking, or trying out there.

Nancy: Well, at first I had the biggest bar...I just put the biggest bar because it was the first one on there. And I just put it somewhere. And then, at first I was just going to put the blue

bar and then the orange bar, um, as far as, you know, just have them kind of equally spread out. But then I thought, maybe if I got the orange bar closer to the mean, there's less data on the orange bar, so, er, um, yeah, there's less data there.

Intv: There's fewer values.

Nancy: Fewer values, yeah. So, I wanted to put that closer to the mean than the blue bar because the blue bar has more, more data.

Intv: Mm hmm.

Nancy: So, and then, I didn't know if I put the orange bar right on top of the mean, or if I put it over here by the blue bar, how much of a difference that would make. But I figured the more spread out it would be the bigger the deviation would be, so.

Intv: Okay, so by having it kind of evenly spaced here, is that what you're thinking as being more spread out?

Nancy: Yeah. Yeah. Having, like, as much space as I can get in between the values.

Far Away – Mean

This rule is similar to Far Away-Values, except that the largest standard deviation is obtained by placing the values as far away from the mean as possible. Carl provided a clear expression of this rule in the first game, although he was not sure that it was correct.

Intv: Now, before you check, can you tell a little about what you were thinking as you were trying things out there?

Carl: Well...um...what I know about standard deviation is, um, given a mean, the more numbers away, the more numbers as far away as possible from the mean, is that increases the standard deviation. So, I tried to make the mean, um, in between both of them, and then have them as far away from the mean as possible. But I don't, yeah, I don't know if that's right or not, but, yeah.

Later, in Game 2, Carl was trying to arrange three bars to produce the largest possible standard deviation. He first used an Equally Spread Out approach (see Figure 4a) and then appeared frustrated that it did not produce the intended result. The interviewer intervened, trying to help him extend the Far Away – Mean rule he had used earlier to the new situation. With this support, Carl successfully produced the distribution in Figure 4b. The interviewer then asked Carl to explain why he thought the arrangement resulted in the largest standard deviation.

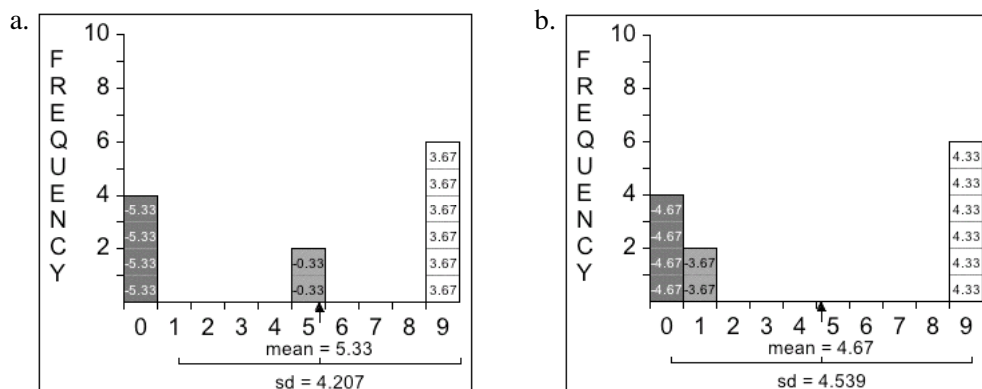


Figure 4. Carl's arrangements for the largest standard deviation in Game 2.

Carl: Because...basically because I have the most amount of numbers possible away from the

mean. Um, because if I were...say I have two bars that are bigger than the one smaller bar,

Intv: Mm hmm.

Carl: If I were to switch these two around.

Intv: The blue and the orange one?

Carl: The blue and the orange. The standard deviation would be smaller then because there's more of the numbers closer to the mean. So, yeah.

Balance and Mean in the Middle

With the Balance strategy, the bars are arranged so that near equal amounts of values are placed above and below the mean. Typically, the bars were arranged in a recognizable order; descending inward from the opposite extremes of the number scale to produce the largest standard deviation. This justification started to emerge during Game 2 with three bars of unequal frequency. Mary provided an example of a Balance justification during Game 2.

Mary: Because, and just with a frequency bar of six over at the high-end, um, it makes it so that the mean can stay more towards the middle. Because if you put too much frequency at one end it kind of makes the mean, the mean kind of follows it. Or, I, so I think that this kind of makes it, evens it out so that it, it can be, um, more frequency away from the mean to get the highest standard deviation.

A Balance justification was often combined with another justification, such as Far Away-Value or Far Away-Mean, during the later games. Another justification that tended to occur with a Balance statement was the Mean in the Middle justification. Alice provided an example where both Mean in the Middle and Balance were used to justify an arrangement (see Figure 5) for producing the largest standard deviation in Game 4.

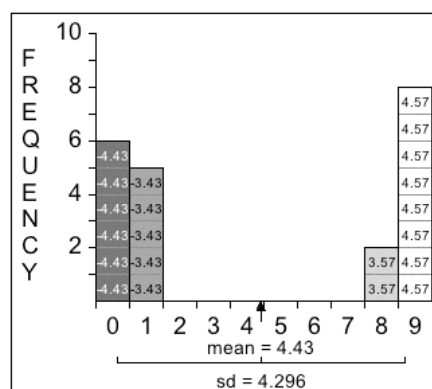


Figure 5. Alice's arrangement for the largest standard deviation in Game 4.

Alice: OK. I think this one will.

Intv: And can you tell me why?

Alice: Because, again I wanted to kind of keep the same values on both ends to kind of get the mean to fall in the center.

Intv: Mm hmm.

Alice: And then, that way they'd both be as far away from the mean as possible. And I think if you put the two middle bars together, then the two, like the largest and the smallest one, you'll get about the same value.

Intv: About the same number of values at each end?

Alice: Yeah, the same number of frequencies.

Big Mean

Troy was the only student who initially believed that a distribution with a higher mean would have a larger standard deviation. The first statement of this belief came in Game 1 after finding an arrangement that produced the largest standard deviation. He immediately posed the question “What would it be if I switched the bars around?” When asked what he thought would happen, he stated “I was thinking that if the mean was, I knew the mean was going to be smaller. I thought that would make the standard deviation smaller.”

Troy’s reasoning seemed to be primarily one of association: the mean gets lower, so the standard deviation will become lower. While he appeared to have an understanding of how value and frequency affect the mean, he had not coordinated these changes with changes in the standard deviation. Even though Troy witnessed that the standard deviation did not change when he produced a mirror image of the distribution, his belief that the size of the mean affects the standard deviation persisted. The following excerpt is from Game 1 after Troy found a distribution with the smallest standard deviation.

Troy: So I suppose it wouldn’t, again, it wouldn't matter if I, like the other graph, if I moved them both together to a different, uh, number on the x-axis, right?

Intv: Yep, well, you’re going to get a chance to find out.

Troy: OK. Again. Move at least one bar.

Troy: Check?

Intv: Mm hmm.

Troy: Yeah, it’s the same.

Intv: Mm hmm. So does that surprise you, or, that the standard deviation is the same?

Troy: I guess a little bit. I, I don’t know. I just, I guess, again, I thought if you. Again I guess it’s thinking that if the mean goes up that the standard deviation goes up. But obviously that's not true a case.

Troy seemed to have learned that the location of an arrangement did not affect the standard deviation, although he did not appear to understand why at this point. He did not, however, present a Big Mean justification during any of the remaining games.

Smallest Standard Deviation

Nearly all of the students placed the bars next to each other in some arrangement to produce the smallest possible standard deviation. Several distinct variations were identified.

Contiguous

Students using this approach stated that they placed the bars next to each other, or as close together as possible, to obtain the smallest standard deviation. In addition to the statement of contiguity, three-fourths of the students placed the bars in an ascending or descending order based on bar height. Alice used a Contiguous – Ascending Order (see Figure 6) in her first attempt to come up with the smallest standard deviation for the three bar situation in Game 2.

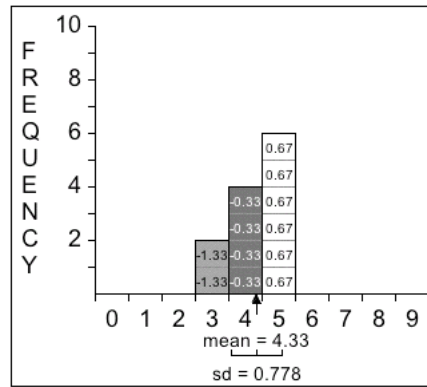


Figure 6. Alice's arrangement for the smallest standard deviation in Game 2.

- Intv: Can you tell me about, uh, what you were doing there?
- Alice: Um, well I knew I wanted the bars to be right next to each other.
- Intv: Mm hmm.
- Alice: Um, and I wanted the mean to kind of fall somewhere in the blue part right here, the middle one.
- Intv: OK
- Alice: And, so I was just going to check to make sure that if I put the smaller one in the middle that I didn't change, like, the deviation would be smaller or larger.
- Intv: OK. And I noticed that the bars are going from the shortest to the next tallest to the tallest.
- Alice: Yep.
- Intv: And any reason why you have put them in that order?
- Alice: Um, well I was thinking, like I think of, um, numbers. And if I count like the boxes there is one, two, three, four, five here, four there, and two there.
- Intv: Mm hmm.
- Alice: I kind of calculate where the median would be and then make sure that the numbers kind of fall somewhere in the middle. So.
- Intv: OK. You want to check this one out?
- Alice: OK
- Intv: So it says it can be even a little smaller.

Alice's reference to the median is not clear, but she may have identified the second tallest bar as the bar with the "median" height, and this was the motivation for placing it in the middle. Alice found out that this arrangement was not optimal and quickly produced an arrangement that did result in the smallest standard deviation.

All of the students who produced an ascending or descending order in Game 2 appeared to abandon this approach for the later games, with one exception. Mona was the only student to use a Contiguous-Descending Order in Game 4 (see Figure 7), even though she produced a bell-shaped distribution in Game 2 for the smallest standard deviation. The following excerpt illustrates her thinking about why the arrangement will produce the smallest standard deviation.

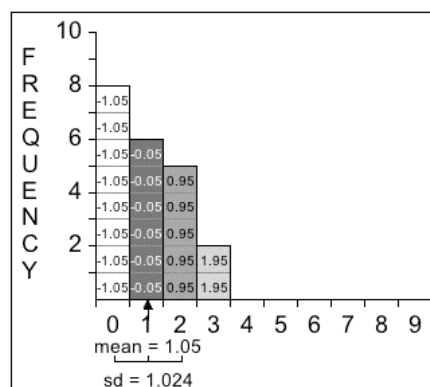


Figure 7. Mona's arrangement for smallest standard deviation in Game 4.

Mona: Yeah, I think this one is right.

Intv: OK. Do you want to check?

Mona: Wait, no, it's not right.

Intv: Now why were you thinking it was right, and why are you thinking it's not right?

Mona: Wait, OK, maybe it is right. OK, because um, I know to make it like, to make it, the standard deviation lower, it has to be closer to the mean. And a lot of them have to be closer to the mean. So, this one is right by the mean almost.

Intv: The red [tallest] bar?

Mona: Yeah.

Intv: Mm hmm

Mona: So it's like, the more of these I have closer to it the better it is. And, and since I only have two that's not, then...

Intv: Yeah. I see what you're saying. You want to check it out?

Mona: OK

Intv: So now it says you can make it smaller.

Mona's statements indicate an understanding that most of the values need to be close to the mean in order for the standard deviation to be small. Her statements indicate that she was considering the placement of values in relationship to the mean, but not necessarily considering deviations from the mean, or the relative density of the deviations. Her thinking ("a lot of them have to be closer to the mean") may have been too general to guide an optimal solution. Mona did eventually produce a bell-shaped distribution with the smallest standard deviation, and she did not use an ascending or descending order in Game 5.

More Bars in the Middle

Some students stated that one of the reasons the standard deviation would be the smallest was because more values or the tallest bars were placed in the middle of the distribution or close to the mean. This justification started to appear in Game 2. A More Bars in the Middle statement, coupled with a statement of contiguity, was the predominant justification offered in Game 4. Mona provided a More Bars in the Middle statement in the previous excerpt when she stated, "a lot of them have to be closer to the mean." Carl used this justification for his first arrangement in Game 2 (see Figure 8).

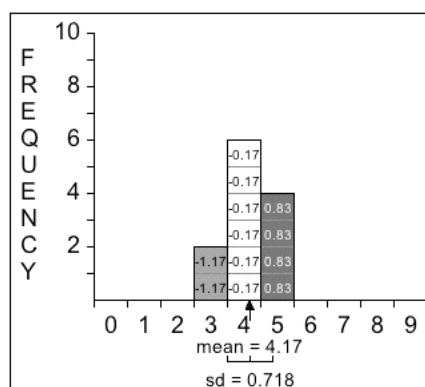


Figure 8. Carl's first arrangement for the smallest standard deviation in Game 2.

Intv: So what were you doing here?

Carl: I put...I tried to scrunch up all of the numbers as close as I could to the mean, and tried to make a new mean in this case. And, uh, it wouldn't have worked if I were to have switch, if I were to have it in progressive order of small orange, medium blue, to big red.

Intv: Mm hmm.

Carl: Because the red is the biggest. You want it to be that in between so the mean is going to, so that you'll have the mean more with the more amount of numbers, basically.

Nancy also used a More in The Middle justification for her first smallest standard deviation arrangement in Game 4. She started off with an arrangement that was near optimal (see Figure 9a) and then decided to switch the positions of the two shortest bars to produce the graph in Figure 9b. Although the result was fairly symmetric and bell-shaped, she talked only about placing lots of values close to the mean.

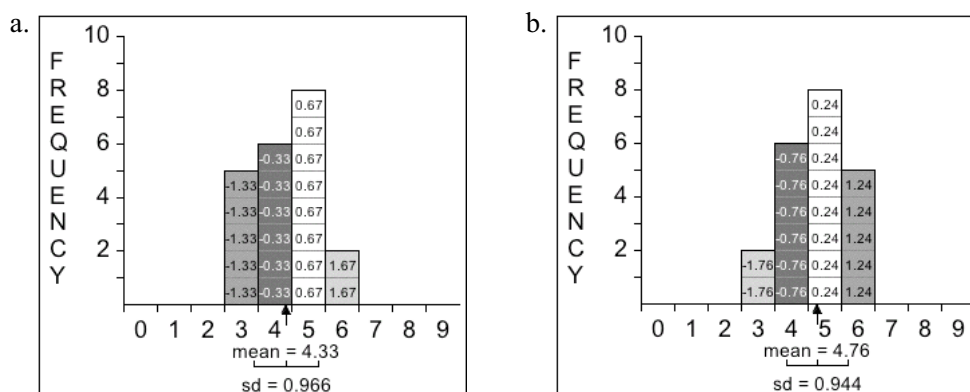


Figure 9. Nancy's arrangements for the smallest standard deviation in Game 4.

Nancy: Yeah. Oh. I've got all of the biggest values together so, there...it was like the one with three where I had them in order. Because I had them kind of in order, and then I had the smallest one. But now I'm just getting all the biggest values closest to the mean. Okay. Before I didn't, so.

The combination of Contiguous with More Bars in the Middle appears to represent a more complex understanding of factors that affect the standard deviation where the students moved from

considering only the closeness of the values in relation to each other to also considering the relative density about a perceived center of the distribution. While not as complete as the completely coordinated conception of the standard deviation outlined earlier, this thinking is a closer approximation than most students demonstrated during the first two games.

Bell-Shaped

Students who made a Bell-Shaped justification stated that they were trying to produce a symmetrical or bell-shaped arrangement. Except for Game 3, none of the bar sets allowed the creation of a perfectly symmetrical distribution. This type of statement was the primary justification offered in Game 5, being made by only one student during Game 2 and two students during Game 4. This is probably the result of the interviewer drawing students' attention to the bell shape of distributions that produced the smallest standard deviation in Game 4.

Students who offered Bell-Shaped justifications typically positioned the tallest bar first, perhaps to function as an anchor or central point. The other bars were placed to the left and right of this central location. A few students were methodical in their placement, alternating the other bars to the left and right of center in order of height. Others seemed to place a few of the tallest bars to the left and right of center, and then checked the size of the standard deviation as they tested out various arrangements.

Mona appeared to use a Bell-Shaped approach in Game 5 when trying to come up with the smallest standard deviation, but she did not have a method that initially produced the most symmetric or bell-shaped arrangement of the five bars (see Figure 10). She started with the arrangement in Figure 10a, saying that "it has to have sort of like a normal, normal shape, and this kind of looks normal to me." The interviewer intervened by asking, "Is that as normal as you can make it look? Is that as bell shaped as you can make it look?" Mona then switched the locations of the two shortest bars to produce the graph in Figure 10b, finding that it produced a smaller standard deviation, and judging that the arrangement was more bell shaped.

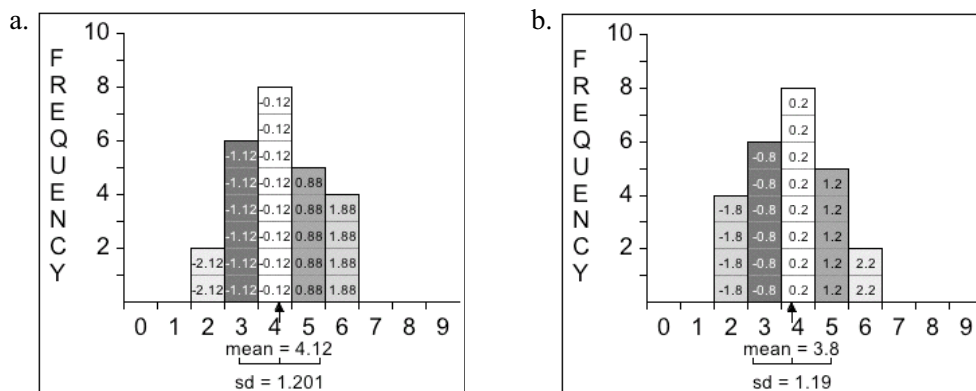


Figure 10. Mona's arrangements for the smallest standard deviation in Game 5.

Balance

The Balance justification was introduced earlier when describing statements students made about arrangements for the largest standard deviation. A balance approach was also used to create an arrangement that produces the smallest standard deviation, as illustrated by Carl in Game 4. Carl made the same type of arrangement as represented in Figure 9b. The interviewer asked why this would produce the smallest standard deviation.

Carl: Um, because, um, the way I have it right now, the red [tallest] bar's in the middle, the blue's [second tallest] on the left, the orange [third tallest] is on the right. And when I had the green [shortest] on the right, there was less numbers in the orange to throw off, um, the two extra numbers that were going to be there at the six value. So, by placing it on the other side, there's more – there's one more number with the blue bar to throw off

one of the numbers in the green bar, so it will bring it a little bit closer.

Intv: Okay. When you say throw off, um, if I were to say balance?

Carl: Yeah.

Intv: Is that a similar thing to what you're saying.

Carl: Yeah. The balance of the scale. It's like having, um, I don't, um, It's kind of like trying to weigh on a scale, or like a teeter-totter. Um, there was – you have a lot in the middle, and some on the sides, but by placing it on one of the sides, it would actually teeter more that way because there's less values in the middle for that one. But if you place it on the other side, it would teeter less because there's more values in the middle to help, yeah, keep it in balance.

The Balance justification may be an extension of the Bell-Shaped justification in that the approach produces a somewhat symmetric and bell-shaped distribution. However, the Balance justification expresses a deeper understanding of the interrelationship between the frequencies of values, their relative placement with respect to the mean, and the overall effect on the standard deviation. The Bell Shape justification seems to be more rule-oriented, using a visual template to produce a distribution that is expected to minimize the standard deviation.

The Same Standard Deviation

Students used two general approaches to make arrangements that produced the same standard deviation.

Mirror Image

The student would often create a mirror image of a distribution to produce a distribution with the same standard deviation. Some students expressed this approach as “swapping” or “switching” the positions of the bars, while others stated that the arrangement was “flip-flopped”. The student would state that since the bars were still as close together or as far apart as before (either relative to each other or relative to the mean), the standard deviation should be the same. Some students demonstrated that they understood that a mirror image would produce the same value for the standard deviation following a prompt from the interviewer. After Jane created the graph in Figure 11, the interviewer posed a situation for her to consider.

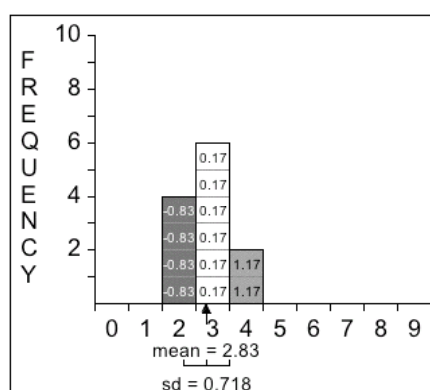


Figure 11. Jane's third arrangement for the smallest standard deviation in Game 2.

Intv: Um, I'm going to move the blue bar off to the side, put the orange bar on the left side of the red.

Jane: OK

Intv: And just ask you to leave those two there.

- Jane: OK
- Intv: Can you still come up with the smallest standard deviation?
- Intv: OK. And what are you doing there?
- Jane: I just moved the blue bar back next to the red one. It is just a mirror, it's one of those mirror imagey things again where you just switch the two. It doesn't matter which side, just as long as you have the biggest bar closer to the mean.

Location

In this approach, the student would move all the bars an equal distance, maintaining the same relative arrangement and standard deviation. The student would note that the standard deviation should be the same because the relative distances between the bars or from the mean were maintained. Troy made the graphs in Figure 12a and Figure 12b for the fourth and fifth tasks in Game 5. After checking and finding that the last graph did produce the smallest standard deviation, the interviewer followed with a question.

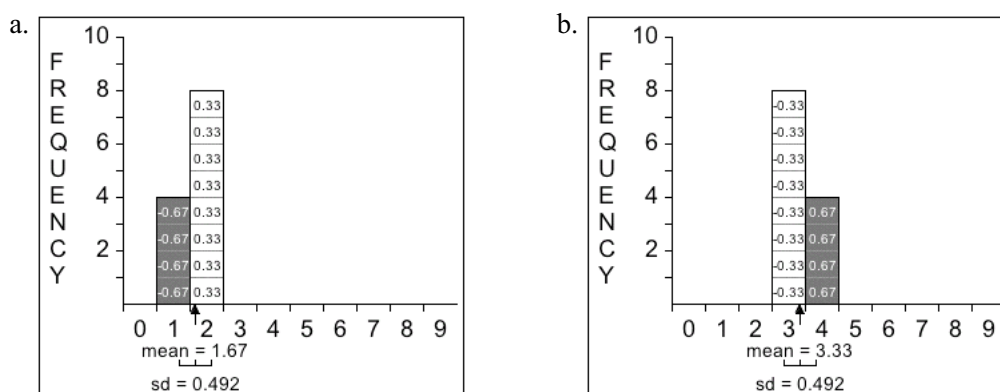


Figure 12. Two of Troy's arrangements for the smallest standard deviation in Game 1.

- Intv: Can you tell me anything about what you think is going on now with, with the relationship of the bars to each other and the size of the standard deviation?
- Troy: Oh, just that the closer the bars are together the lower the standard deviation is.
- Intv: Mm hmm.
- Troy: Vice versa.
- Intv: OK
- Troy: Um. And it doesn't matter where they are, and what the mean is or where they are on the x-axis.

Coordination of Concepts

Review of Carl's More in the Middle justifications during Game 2 indicates that he was able to coordinate the effects of the values and the frequencies represented by the bars on the value of the mean. Carl also appeared to display some coordination between the effects of frequency and deviations from the mean and the value of the standard deviation, although his focus was primarily on having most of the values (the tallest bar) close to the mean. None of the other students demonstrated this type of coordination during Game 2. Carl continued to demonstrate an increasing coordination of concepts as he progressed through the games.

Several more students mentioned deviations from the mean and demonstrated conceptual coordination in their explanations during Game 4. Mary started with the arrangement in Figure 13 for

the task of producing the largest standard deviation. When asked why she thought the arrangement would produce the largest standard deviation, Mary offered a Far Away-Mean explanation:

Mary: I think we did this for another one where we made sure the highest frequencies were as far away from the mean, or as possible, and then try to just do it in order of the biggest to the smallest and keep them all grouped away. So I don't know, 4.175. I bet you it will maybe work this way too.

Mary was generalizing from Game 2 to Game 4, but did not have a full understanding of how to maximize the size of the deviations from the mean. However, she did express a coordination of both value and frequency with respect to distance from the mean and the relative size of the standard deviation.

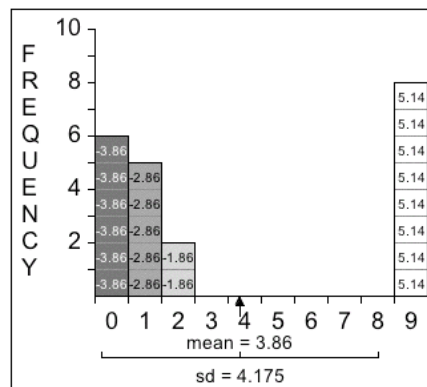


Figure 13. Mary's first arrangement for the largest standard deviation in Game 4.

Linda also demonstrated an understanding that increasing the number of large deviations increases the standard deviation.

Linda: Um, I put like the largest bars on the two outside scores and the smaller ones on the inside just so, like, the bigger chunks of them would be farther away from the mean and the smaller ones might, I didn't think that would affect the standard deviation as much as the larger values.

Adam, who first gave a Balance type of explanation, also observed that there was a pattern to how the bars were placed and connected this to deviations from the mean.

Adam: I'm guessing this smaller, the smaller, the smallest sets of data values have to be towards the mean. Has to be closer towards the mean than the larger sets, because if you want the standard deviation, which is the larger amount of numbers away from the mean, then you are going to have to take out the largest sets of numbers and keep them farther away from the mean and keep the smallest sets closer to the mean.

Here, Adam appears to coordinate the effects of frequency and deviation after the fact.

Alice demonstrated a more complex understanding of how frequency and deviation from the mean combined to affect the size of the standard deviation. As discussed earlier, Alice produced the graph in Figure 5 and justified why it produced the largest standard deviation with a combination of Balance and Far Away – Mean statements. The interviewer followed her statement with a hypothetical question:

- Intv: If we were to switch positions of the gold bar [frequency of 5] and the green bar [frequency of 2], what do you think would happen to the standard deviation?
- Alice: It would get pulled towards the gold and the red bar. Because there is more frequency over there than on the other two. Because they aren't pulling the mean that much more over.
- Intv: OK. What do you think would happen to the standard deviation, then?
- Alice: I think it would get smaller.

When Alice stated “it would get pulled towards the gold and the red bar,” she appeared to be referring to the mean. As such, her prediction would be correct. She demonstrated an understanding that the mean would shift towards the larger cluster of values, creating a larger number of smaller deviations from the mean, and subsequently result in a smaller standard deviation. Carl and Alice were the only students to demonstrate this level of complexity in their justifications during Game 4.

Summary of Exploration Phase Justifications

Table 5 presents a list of the justifications that students gave for bar arrangements produced during the exploration phase. The entries in Table 5 indicate the number of instances of each justification across students. The tallies for the five games provide a sense of when a justification first appeared and if a justification continued to be used. The tallies illustrate distinctions among the justifications with respect to the type of tasks for which they were most likely to be used.

The first five justifications listed in Table 5 represent reasons given primarily for arrangements in the first task, ordered from most to least frequent. The next three justifications are those used primarily for the third task where the student was asked to produce an initial arrangement with the smallest standard deviation, again ordered from highest to lowest frequency. The Mirror Image and Location justifications represent reasons given for arrangements that produced the same standard deviation as a previous arrangement (tasks 2, 4, and 5). The Guess and Check approach is listed last because it was used with almost equal frequency in tasks 1 and 3, and represents a general strategy more so than a justification.

Table 5. Number of Instances of Each Justification Used by Students in the Exploration Phase.

JUSTIFICATION	TASKS										TOTAL
	GAMES					TASKS					
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>Largest SD</u>	<u>Smallest SD</u>	<u>3</u>	<u>4</u>	<u>5</u>	
Far Away	12	8	12	8	5	43	2				45
Balance		7	8	10	10	30		4		1	35
Mean in the Middle	2	2	5	4		8	1	4			13
Equally Spread Out		7	1			8					8
Big Mean	1					1					1
Contiguous	17	11	12	7		1		42	2	2	47
More Bars in the Middle		6	1	9	3			18		1	19
Bell-Shaped	1	1		2	7	1		10			11
Mirror Image	28	27	23	24	22		59		35	30	124
Location	12	13	15	13	14			4	27	36	67
Guess and Check	6	15	3	14	9	25	1	20	1		47

The Equally Spread Out justification was used primarily during the second game with three bars, and then dropped out of use after feedback indicated that the approach did not produce a distribution with the largest standard deviation. The idiosyncratic Big Mean justification was only used once in

the first game, dropping out of use after the student received feedback that it was not effective. Far Away and Balance became the predominant justifications for arrangements intended to produce the largest standard deviation. A statement that an arrangement tended to place the mean in the middle of the distribution accompanied Far Away and Balance statements for a few students, but was not used frequently.

Contiguity was the primary justification used for the first arrangement intended to produce the smallest standard deviation, although it was not used at all for the fifth game. Students also tended to accompany Contiguity justifications with a recognition that an arrangement placed more of the taller bars in the middle of the distribution. This coupling of Contiguity and More Bars in the Middle was predominant during the second game with three bars and the fourth game with four bars of unequal frequencies. The tallies also illustrate that the Mean in the Middle and the More Bars in the Middle justifications appear to be distinct in that the former occurred primarily for the largest standard deviation criterion, whereas the latter was used exclusively for the smallest standard deviation criterion. Statements that an arrangement produced a bell-shaped distribution were made primarily during the fifth game, presumably as a result of prompting from the interviewer.

The predominant strategy for producing a new arrangement with the same standard deviation as a previous arrangement was to create a mirror image of the prior distribution. As would be expected by the constraints of the second task, this was the only justification given for producing a second distribution with the largest standard deviation. The Mirror Image and the Location justifications were used with near equal frequencies for tasks 4 and 5 which required additional arrangements with the smallest possible standard deviation.

A Guess and Check approach was used primarily when producing the first arrangement for either the largest or smallest standard deviation criteria. Once an arrangement was found that met the criterion, the Mirror Image or Location justifications were used to produce subsequent arrangements that met the criterion.

3.2. TESTING PHASE

Table 6 presents the justifications students used in their responses to the 10 test items. Item 8 was the most difficult item with only one student making the correct choice. Nine of the 12 students answered 9 of the 10 test items correctly. The other three students answered 7 of the test items correctly. Given the high correct response rate, correct responses were not associated with any particular justification or combination of justifications.

Table 6. Number of Students Who Used Each Justification to Explain Choices During the Testing Phase.

<u>TEST ITEM 1</u>	<u>TEST ITEM 2</u>	<u>TEST ITEM 3</u>	<u>TEST ITEM 4</u>	<u>TEST ITEM 5</u>					
Location	12	More in Middle	9	More in Middle	6	Bell	3	Contiguous	7
		Bell Shape	3	Far Away	4	More in Middle	3	Mean in Middle	1
		Range	1	Mean	2	Balance	1	More in Middle	1
		Contiguous	1	Value	1			Equal Spread Out	1
				Bell Shape	2			Range	1
				Contiguous	1				
<u>TEST ITEM 6</u>	<u>TEST ITEM 7</u>	<u>TEST ITEM 8</u>	<u>TEST ITEM 9</u>	<u>TEST ITEM 10</u>					
Mirror Image	11	Contiguous	7	Calculation	7	More in Middle	6	Calculation	9
Mean in Middle	1	More in Middle	5	More in Middle	4	Mean in Middle	3	Range	4
		Far Away	5	Mean in Middle	4	Far Away	3	More in Middle	2
		Mean	5	Bell Shape	4	Mean	3	More Values	2
		Mean in Middle	1	More Values	4	Calculation	2	Mean in Middle	1
		Range	1	Range	2	Contiguous	2	Bell Shape	4
				Contiguous	2			Frequency	1

Students' justifications for their responses tended to be similar to those identified during the exploration phase, and tended to reflect relevant characteristics of the graphs. For example, all students indicated that the two graphs had the same standard deviation for the first test item, noting that the arrangements were the same but in different locations. Most of the students' justifications for test items 2 and 3 were More in the Middle statements, although test item 3 also prompted Far Away justifications, probably because the graph on the right was U-shaped. For test item 6, students recognized that one graph was the mirror image of the other and that the standard deviations were the same. Almost all of the students gave a Mirror Image justification for their choice in item 6.

Students found test item 4 a little more difficult than the first three tests. Most noted that the only difference between the two graphs was the placement of the black (shortest bar). Only about half of the students provided a justification for their responses, which tended to be either that the graph on the left was more bell-shaped, or that it had more values in the middle or around the mean compared to the graph on the right. A few students stated that the left-hand graph was more skewed.

Two students (Adam and Lora) judged incorrectly that the graph on the left would have a smaller standard deviation in test item 4. Adam noted the different location of the black bar in both graphs, but did not offer any other justification for his response. Adam did not demonstrate a consistent understanding of how frequency and deviation from the mean combined to affect the standard deviation during the exploration phase, which was evidenced by his extensive use of a Guess and Check approach. Item 4 may have proven difficult due to this lack of understanding. Lora also displayed Guess and Check behavior during the exploration phase, but not to the same extent as Adam. Lora's justification was actually a description of each graph's characteristics and did not compare the two graphs on features that would determine how the standard deviations differed. When asked how the graph on the right compared to the graph on the left, Lora responded, "That, well that's kind of the same thing as this one, but this one is, um, the smallest one is on the end instead, and along with the other one." No other justifications were offered before checking.

The difficulty with test item 4 may have resulted from the nature of the tasks during the exploration phase. The tasks required solutions that had some degree of symmetry and did not require students to produce skewed distributions, although skewed distributions were often generated and checked before an optimal distribution was found for the third task. The arrangements produced for the largest standard deviation were U-shaped, with a large gap through the middle, while the optimal arrangements for the smallest distribution were bell- or mound-shaped. This experience during the exploration phase probably facilitated the comparison in test items 2 and 3, but did not necessarily provide experience directly related to the comparison needed to solve test item 4. Nonetheless, the majority of the students correctly identified the graph on the right as having a larger standard deviation.

The first four test items involved graphs with contiguous bar placement, so that test item 5 provided the first test of sensitivity to gaps between the bars. Almost all of the students' justifications for test item 5 involved statements of contiguity. A few students noted that the range was wider in the left-hand graph so that the graph on the right should have a smaller standard deviation given that the bars were the same and in the same order in both graphs. One of these students noted that the bars were equally spread out for the graph on the left, although the emphasis was on the wider range. Two students, Adam and Jeff, appeared to ignore the gaps between bars in the left-hand graph and responded that the two graphs would have the same standard deviation, suggesting that shape was the main feature they considered in their decisions and an insensitivity to spread or deviation from the mean.

Test item 7 also tested students' understanding of how gaps affected the standard deviation, although in a more subtle way. All students answered test item 7 correctly, using arguments of contiguity and more values in the middle for the left-hand distribution as the reason why the right-hand distribution would have a larger standard deviation. Some students also noted that the right-hand distribution had relatively more values away from the mean. Test item 9 also presented gaps in the two distributions and challenged the misconception that having the bars evenly or equally spaced produces a large standard deviation. All students answered test item 9 correctly and offered

justifications similar to those used in test item 7, indicating that the initial misconception represented by the Equally Spread Out justification had been overcome by those who first displayed it.

Test items 8 and 10 were designed to challenge the idea that perfectly symmetric, bell-shaped distributions always have the smaller standard deviation. For test item 8, all but one of the students incorrectly stated that the graph on the right would have a smaller standard deviation. Linda answered correctly that the symmetric, bell-shaped graph would have a larger standard deviation. Her reasoning was that “there are more bars. There, um, even though like I think they are clumped pretty well, there’s still, um, a higher frequency and so, there’s three more, so because of like the extra three there’s going to be more like room for deviation just because there’s more added values.” Other students noted that there were more bars or values in the graph on the right, or that it had a larger range, but still selected the smaller option as an answer. Students tended to note the bell-shape or the perception that more values were in the middle of the right-hand distribution compared to the one on the left as reasons for their responses.

After students checked their answers to item 8 and found out they were incorrect, the interviewer attempted to guide their attention to the ambiguity that stemmed from a comparison of characteristics between the two graphs. Attention was drawn to the differences in the number of values and the range, pointing out that while the bell shape suggested the right-hand graph would have a smaller standard deviation, the larger range made it possible for it to have a larger standard deviation, and the different number of values (or bars) presented a different situation than was presented either during the exploration phase or in the previous tests. The interviewer suggested that ambiguous situations might require calculation of the standard deviations, and demonstrated how the information provided by the program could be used to do so.

When students came to test item 10, they were more likely to note the difference in the number of bars or values between the two graphs and resort to calculating the standard deviation for the graph on the right. Nine of the students came to a correct decision predominantly through calculation. Two of the students (Troy and Jane) initially predicted the graph on the right to have a larger standard deviation, but then noted the discrepancy in range or number of values, calculated the standard deviation, and changed their decisions. Three students (Jeff, Lora, and Linda) did not perform calculations and incorrectly responded that the right-hand graph would have a larger standard deviation. Because Linda correctly answered test item 8, she did not receive the same guidance from the interviewer as the other students, which may account for her incorrect response on test item 10.

4. DISCUSSION

While some research on students’ understanding of statistical variability has been conducted, little is known about students’ understanding of measures of variability, such as the standard deviation, how that understanding develops, and effective ways to support that development. The study reported here was an attempt to address this lack of information. An exploratory research design was used that employed a multi-phase interview protocol where students interacted with a computer environment designed to help them coordinate underlying concepts that are foundational to an understanding of the standard deviation. The primary instructional goal was to help students develop an understanding of how deviation from the mean and frequency combined to determine the value of the standard deviation. The primary research goal was to develop a better understanding of students thinking about the standard deviation, both after classroom instruction on the concept, and as the concept was developed and elaborated. What has the study contributed with respect to both of these goals? To address this question, a brief critique of some aspects of the research design is offered, followed by a discussion of implications for instruction and future research.

4.1. LIMITATIONS

Definitive statements about the necessity of the instructional design features cannot be made due to the exploratory nature of the investigation. Other approaches may be more optimal and efficient at promoting students’ conceptual understanding of the standard deviation. For example, the series of

games was designed so that the complexity of the bar arrangements increased incrementally. This was expected to facilitate attending to the various factors that affect the standard deviation and coordination of their simultaneous contributions. However, the incremental increase may not be necessary. The use of several four bar and five bar games may be just as effective in eliciting students justifications and, at the same time, may provide sufficient feedback over a series of games to promote an integrated understanding. It may also be just as effective to describe these relationships to students, perhaps introducing one relationship among factors at a time, building up the complexity of the description in steps, and presenting graphical examples as illustrations for each step.

Students' justifications for their responses to the test items were based on comparisons of relevant characteristics of the histograms and were similar to their justifications expressed during the exploration phase. However, it cannot be firmly established that these justifications were learned or developed during the interview. An independent assessment of the characteristics of histograms students attend to and the type of justifications they give prior to interacting with the computer environment would have been helpful in revealing students' prior knowledge and conceptions.

The situation presented in the computer environment was artificial, although none of the students indicated that this was a problem or produced any difficulty. Nonetheless, a more realistic situation might provide the same type of feedback and support conceptual development while promoting reasoning with real data. The process of building up a distribution, one data value at a time, may be more effective at promoting distributional reasoning and thinking (see Bakker, 2004). For example, values from a real data set could be presented one at a time, and the student could be asked to anticipate the effect on the mean, deviations, and standard deviation of adding the next value to the distribution.

4.2. IMPLICATIONS FOR RESEARCH AND INSTRUCTION

The ensemble of justifications, strategies, and concepts found in this study indicate that students in an introductory statistics course form a variety of ideas as they are first learning about the standard deviation. Some of these ideas, such as the Contiguous, Range, Mean in the Middle, and Far Away-Values rules, capture some relevant aspects of variation and the standard deviation, but may represent a cursory and fragmented level of understanding. Others such as the Far Away-Mean, Balance, More Values in the Middle, and Bell-Shaped rules, represent much closer approximations to an integrated understanding. There are still other ideas, notably the prevalent Equally Spread Out rule and the idiosyncratic Big Mean rule, that are inconsistent with a coherent conception of the standard deviation. Some students also demonstrated an ability to coordinate the effects of several operations on the value of the standard deviation, an indication of a more integrated conception.

To what extent did students achieve a fully coordinated conceptualization of the standard deviation through interaction with the computer program and the interviewer? Only one student, Troy, appeared to have a prior belief that the location of the mean was directly related to the size of the standard deviation, but he readily used mirror images and moving distributions to new locations to produce arrangements with the same standard deviation by the third game. Interaction with the computer environment also appeared effective in changing the conception of students who presented the Equally Spread Out justification early in the interview. By the end of the exploration phase, all of the students appeared to have understood that the mirror image of a distribution conserved the value of the standard deviation. They also demonstrated the understanding that the relative, and not absolute, location of the bars determined the standard deviation.

Only a few students provided justifications by the end of the exploration phase that approached a fully coordinated conception of how frequency and deviation from the mean combine to influence the value of the standard deviation. Students' arrangements and justifications initially indicated an understanding that a large number of values needed to cluster around the mean to produce a relatively small standard deviation, while larger numbers of values needed to be placed far from the mean, in both directions, to produce a relatively large standard deviation. Students' justifications became more complex as the number of bars increased. More students made references to the distribution of values relative to the mean, indicating an awareness of deviation. They also tended to combine earlier

justifications, such as presenting both Balance and Far Away arguments to justify a distribution for the largest standard deviation, or combining Contiguity and More in the Middle arguments for why a distribution should have the smallest standard deviation. Therefore, while many students did not have a fully coordinated understanding by the end of the exploration phase, most had developed parts of this conceptualization and began to coordinate some of the concepts.

Most of the students used a rule-based approach to compare variability across distributions instead of reasoning from a conceptual representation of the standard deviation. Even among the students with apparently richer representations, their explanations during the testing phase were usually based on finding a *single* distinguishing characteristic between the two distributions rather than reasoning about the size of the standard deviation through a conception that reflected how density was distributed around the mean in each distribution. This suggests that students tend to take a rule-based, pattern recognition approach when comparing distributions. If this is the case, two questions need to be addressed. What type of experiences are required to move students from a rule-based approach to a more integrated understanding that can be generalized to a variety of contexts? In addition to the interactive experience presented in the current study, do students need additional support to reflect on the relationships between the different factors and to attend to and coordinate the related changes?

A pattern-recognition, rule-based approach is consistent with a goal of finding the right answer by noting characteristics that differentiate one arrangement from another and noting the correspondence with the size of the standard deviation. This orientation was supported by the design of the software in that the value of the standard deviation was always available and there were no consequences for an arrangement that did not meet a criterion. This may have overemphasized the need to be “correct” instead of promoting the exploration and reflection needed to develop an understanding of factors that affect the standard deviation. Is there a way to modify the task so there is less emphasis on a correct solution and more emphasis on exploring the relationships among the factors that affect the standard deviation? One possibility is to modify the software so that the mean and standard deviation are not automatically revealed. This might promote reflection and result in less Guess and Check behavior.

The interviewer attempted to extend students’ conceptual understanding by trying to draw their attention to relevant aspects of the distributions, and by modeling the desired conceptual understanding. The software was designed to help students identify factors that affect the standard deviation. The software and interview were not designed with the promotion of model building in mind. A model eliciting approach may be more likely to produce the “system-as-a-whole” thinking (Lesh & Carmona, 2003) that is needed for a fully coordinated conception of the standard deviation, and to allow students to develop a more integrated representational system (Lehrer & Schauble, 2003), rather than a collection of separate and potentially conflicting rules.

Several changes to the program could be introduced to support model building and study how it affects understanding of the standard deviation. The software currently draws attention to a single bar rather than visually emphasizing how characteristics change simultaneously. A second display above the graphing area of the histogram that presents horizontal deviation bars colored to match the corresponding vertical frequency bar colors may facilitate coordination of simultaneous changes in values, the mean, deviations, and the standard deviation. The interview protocol would also need modification to include model eliciting prompts and probes. This can be done through eliciting conjectures from students about how changes to arrangements will affect the value of the mean and deviations, and how these changes subsequently affect the standard deviation, promoting a coordination of the concepts and a relational structure that models their mutual effects. This contrasts with the current protocol where the interviewer modeled the thinking and reasoning for the student rather than supporting students to produce their own conjectures and test their implications.

ACKNOWLEDGMENTS

The authors would like to thank three anonymous reviewers and the editor for their careful and thoughtful reviews of earlier versions of this manuscript.

REFERENCES

- Bakker, A. (2004). Design research in statistics education: On symbolizing and computer tools. (Doctoral dissertation, Utrecht University). Utrecht: CD-β Press.
- Bubules, N. C., & Linn, M. C. (1988). Response to contradiction: Scientific reasoning during adolescence. *Journal of Educational Psychology*, 80, 67-75.
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Chin, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1-49.
- delMas, R. (1997). A framework for the development of software for teaching statistical concepts. In J. B. Garfield & G. Burril (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 International Association of Statistics Education (IASE) Round Table Conference* (pp. 85-99). Voorburg, The Netherlands: International Statistical Institute.
- delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3). [Online: www.amstat.org/publications/jse]
- delMas, R., Garfield, J., & Chance, B. (2004). Using assessment to study the development of students' reasoning about sampling distributions. Paper presented at the Annual Meeting of the American Educational Research Association, April 13, San Diego, CA. [Online: www.gen.umn.edu/faculty_staff/delmas/AERA_2004_samp_dist.pdf]
- Garfield, J., delMas, R., & Chance, B. (1999). The role of assessment in research on teaching and learning statistics. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.
- Hardiman, P., Pollatsek, A., & Well, A. D. (1986). Learning to understand the balance beam. *Cognition and Instruction*, 3, 63-86.
- Hoerl, R., & Snee, R. D. (2001). *Statistical thinking: Improving business performance*. Duxbury Press: Pacific Grove, CA.
- Lehrer, R., & Schauble, L. (2003). Origins and evolution of model-based reasoning in mathematics and science. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 59-70). Mahwah, NJ: Lawrence Erlbaum.
- Lesh, R., & Carmona, G. (2003). Piagetian conceptual systems and models for mathematizing everyday experiences. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 71-96). Mahwah, NJ: Lawrence Erlbaum.
- Meletiou-Mavrotheris, M., & Lee, C. (2002). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistics Education Research Journal*, 1(2), 22-37. [Online: www.stat.auckland.ac.nz/serj/]
- Moore, D.S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants*. Washington, DC: National Academy Press.
- Nickerson, R. S. (1995). Can technology help teach for understanding? In D. N. Perkins, J. L. Schwartz, M. M. West, & M. S. Wiske (Eds.), *Software goes to school: Teaching for understanding with new technologies* (pp. 7-22). New York: Oxford University Press.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211-227.
- Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer Academic Publishers.

- Roth, K., & Anderson, C. (1988). Promoting conceptual change learning from science textbooks. In P. Ramsden (Ed.), *Improving learning: New perspectives* (pp. 109-141). London: Kogan Page.
- Saldahna, L. A., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51(3), 257-270.
- Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Bidulph & K. Carr (Eds.), *Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 6-22). Rotorua, N.Z.: University of Waikata.
- Shaughnessy, J. M., Watson, J., Moritz, J., & Reading, C. (1999, April). School mathematics students' acknowledgment of statistical variation. In C. Maher (Chair), *There's More to Life than Centers*. Pre-session Research Symposium, 77th Annual National Council of Teachers of Mathematics Conference, San Francisco, CA.
- Snee, R. (1990). Statistical thinking and its contribution to quality. *The American Statistician*, 44, 116-121.
- Snir, J., Smith, C., & Grosslight, L. (1995). Conceptually enhanced simulations: A computer tool for science teaching. In D. N. Perkins, J. L. Schwartz, M. M. West, & M. S. Wiske (Eds.), *Software goes to school: Teaching for understanding with new technologies* (pp. 106-129). New York: Oxford University Press.
- Thompson, P. W., Saldahna, L. A., & Liu, Y. (2004). Why statistical inference is hard to understand. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

ROBERT DELMAS
University of Minnesota
354 Appleby Hall
128 Pleasant Street SE
Minneapolis, MN 55455
USA