

An Investigation of Statistical Thinking in Two Different Contexts: Detecting a Signal in a Noisy  
Process and Determining a Typical Value

Randall E. Groth

Salisbury University

Corresponding Author:

Randall E. Groth  
Salisbury University  
Department of Education  
1101 Camden Avenue  
Salisbury, MD 21801 USA

Email: [regroth@salisbury.edu](mailto:regroth@salisbury.edu)

Key words: statistics, data analysis, measures of center, measures of central tendency

## ABSTRACT

The study describes students' patterns thinking for statistical problems set in two different contexts. Fifteen students representing a wide range of experiences with high school mathematics participated in problem-solving clinical interview sessions. At one point during the interviews, each solved a problem that involved determining the typical value within a set of incomes. At another point, they solved a problem set in a signal-versus-noise context (Konold & Pollatsek, 2002). Several patterns of thinking emerged in the responses to each task. In responding to the two tasks, some students attempted to incorporate formal measures, while others used informal estimating strategies. The different types of thinking employed in using formal measures and informal estimates are described. The types of thinking exhibited in the signal-versus-noise context are then compared against those in the typical value context. Students displayed varying amounts of attention to both data and context in formulating responses to both problems. Suggestions for teachers in regard to helping students attend to both data and context when analyzing statistical data are given.

In discussing the nature of the subject of statistics, Garfield and Gal (1999) noted, “In statistics, data are viewed as numbers with a context. The context motivates procedures and is the source of meaning and basis for interpretation of results” (p. 208). This viewpoint stems from the position that statistics and mathematics are two distinct disciplines, and that statistics is not simply a branch of mathematics (Moore, 1992). The implication is that while mathematical thinking may in some cases take place in highly abstract contexts, statistical thinking can never be divorced from the circumstances in which empirical data arise. While mathematical thinking, in a sense, can transcend physical contexts, statistical thinking cannot and should not attempt to do so.

#### Interplay Between Statistical and Contextual Knowledge

Wild and Pfannkuch (1999) described the interplay between contextual knowledge and statistical knowledge. They stated that when one is engaged in doing statistics, there is “continual shuttling backwards and forwards between thinking in the context sphere and the statistical sphere” (p. 228). This occurs in various stages of a statistical investigation, including formulating the problem, devising a plan, collecting data, analyzing data, and drawing conclusions. They elaborated upon this shuttling process at the stage of analyzing data:

At the analysis stage questions are suggested by context knowledge that require consulting the data – which temporarily pushes us into the statistical sphere – whereupon features seen in the data propel us back to the context sphere to answer the questions, “Why is this happening?” and “What does this mean?” (p. 228).

One must continuously integrate and synthesize statistical and contextual knowledge while engaged in statistical thinking.

## Two Contexts for Statistical Thinking

Since contextual knowledge is as an essential component of statistical thinking, the contexts in which teachers and textbooks set problems have impacts on the breadth of statistical thinking students are given the opportunity to develop. Konold and Pollatsek (2002) argued that students should be given more opportunities to work with statistical problems set in contexts that involve searching for “signals in noisy processes.” This section will elaborate upon the meaning of the “signals in noisy processes” metaphor and also discuss the “typical value” context, which is more commonly encountered in school settings.

### *Signal-Versus-Noise Context*

Konold and Pollatsek (2002) presented the following item as an example of a data analysis problem that involves detecting a signal in a noisy process:

A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student were 6.2, 6.0, 6.0, 15.3, 6.1, 6.3, 6.2, 6.15, 6.2. What would you give as the best estimate of the actual weight of this object? (p. 268).

In the case of the repeated measures problem above, the arithmetic mean of the weights that are bunched closely together could be viewed as a signal that estimates the true weight of the object. The measurement of the object can be viewed as a noisy process that contains variation stemming from various possible sources. Konold and Pollatsek (2002) acknowledged the possible cognitive complexity in using repeated measurement problems with students, pointing out that the mean as a reliable indicator of signal was not universally accepted by scientists during the early development of the discipline of statistics (Stigler, 1986). Hence, they called for more research on students’ thinking in such contexts in order to help advise instruction.

*Typical Value Context*

The typical value interpretation of the arithmetic mean has received a great deal more attention in curriculum materials and in research literature than the signal-versus-noise interpretation (Konold & Pollatsek, 2002). The following is an example of a problem set in a typical value context:

The numbers of comments made by 8 students during a class period were 0, 5, 2, 22, 3, 2, 1, and 2. What was the typical number of comments made that day? (Konold & Pollatsek, 2002, p. 268).

Several studies have provided insights about students' thinking in regard to typical value problems (Mokros & Russell, 1995; Watson & Moritz, 1999; Jones et al., 2000; Mooney, 2002).

Mokros and Russell (1995) reported that students may respond to typical value problems by: (i) locating the most frequently occurring value; (ii) executing an algorithm; (iii) examining the data and giving a reasonable estimate; (iv) locating the midpoint of the data; or (v) looking for a point of balance within the data set. A study by Watson and Moritz (1999) that investigated the development students' thinking in regard to typical value problems placed a developmental structure on the categories of thinking documented by Mokros and Russell (1995). The Watson and Moritz (1999) study found, "A primitive median concept as 'middle' or mode concept as 'same as others' is usually acquired well before students are introduced to the arithmetic mean" (p. 35). Jones et al. (2000) and Mooney (2002) found that the appropriate application of formal measures to typical value contexts constitutes a level of thinking beyond the exclusive use of informal, idiosyncratic strategies. The ability to be thoughtful and critical about applying formal measures to typical value problems marks a relatively sophisticated level of statistical thinking.

### *Summary*

Signal-versus-noise and typical value contexts both provide settings for statistical data analysis. School curricula and research have given a great deal of attention to typical value contexts but not to signal-versus-noise contexts. Subsequently, little is known about how students think in the signal-versus-noise context, which is “perhaps the most fundamental conceptual model for reasoning statistically” (Konold & Pollatsek, 2002, p. 286).

### *Purpose of the Study*

The purpose of the present study was to investigate students’ thinking within both typical value and signal-versus-noise contexts. The specific research questions guiding the investigation were:

- (1) What patterns of thinking are exhibited by students representing a range of experiences with high school mathematics in response to a task set in a typical value context?
- (2) What patterns of thinking are exhibited by students representing a range of experiences with high school mathematics in response to a task set in a signal-versus-noise context?
- (3) How do the patterns of thinking exhibited in the typical value context compare to those exhibited in the signal-versus-noise context?

This research focus was chosen for the purpose of advising future research and teacher education efforts. The answers to the research questions are pertinent to advising further research, as they were designed to make inroads into the little-explored area of students’ thinking within signal-versus-noise contexts. The descriptions of students’ thinking also hold value for teacher education, since they can be used to build teachers’ knowledge of students’ cognition. Such knowledge is an important component of the knowledge base for teaching (Shulman, 1987; Fennema & Franke, 1992; Mewborn, 2003).

### Methodology

A qualitative design was chosen for the study because such designs facilitate the investigation of intricate thinking processes rather than just end products (Merriam, 1988; Strauss & Corbin, 1990; Bogdan & Biklen, 1992). Within the qualitative design, task-based clinical interviews were used as the primary means of data collection. Goldin (2000) noted that such interviews allow researchers to “focus research attention more directly on the subjects’ processes of addressing mathematical tasks, rather than just on the patterns of correct and incorrect answers in the results they produce” (p. 520).

### *Participants*

Purposeful sampling (Patton, 1990) was used in the selection of study participants. The goal of purposeful sampling is to “select information-rich cases whose study will illuminate the questions under study” (Patton, 1990, p. 169). The central questions of this study concerned the identification and description of different patterns of statistical thinking. In order to facilitate the observation of a number of different patterns of thinking, students were chosen on the basis of the mathematics courses they had taken while in high school. The goal was to select students representing a fairly wide range of exposure to school mathematics.

There were three different categories of participants interviewed for the study. The participants in the first category were college freshmen who had completed a semester-long high school statistics course. Participants in the second category were college freshmen who had completed a year-long high school statistics course. The participants in the third category were still in high school at the time of the clinical interviews. Each of the participants in the third category was enrolled in Algebra, Geometry, Advanced Placement (AP) Calculus, or AP Statistics at the time of the interview sessions. The one student taking AP Statistics was

interviewed during the latter part of the first semester of the course. In total, fifteen students participated in clinical interviews. Three of the participants came from the first category, four from the second, and eight from the third. Tables 1, 2, and 3 summarize some of the important information about the students in categories 1, 2, and 3, respectively.

<INSERT TABLE 1 HERE>

<INSERT TABLE 2 HERE>

<INSERT TABLE 3 HERE>

Students were recruited from each of the three categories by contacting several university instructors and high school teachers. Several instructors at a Midwestern university helped in the recruitment of students by mentioning the study in their classes and asking for volunteers. The first two categories of students interviewed included students from three of those instructors' classes. The very first category also included a student from a different Midwestern university who was recruited with the help of her former high school statistics instructor. Teachers at two different Midwestern high schools helped in the recruitment of students in the last category by discussing the study with their classes and asking for volunteers. Four students in the last category came from one of the high schools, and four came from the other high school. Because of the manner in which the sample was drawn, the study does not attempt to make statistical generalizations to larger groups of students. Instead, it describes the interview task responses of the experientially-diverse sample of students chosen and draws inferences for instruction that are sustainable from an examination of the patterns of thinking exhibited.



### *Interview Tasks*

#### *Typical Value Task*

The typical value task posed to students during clinical interviews was designed to be similar in form to the Konold and Pollatsek (2002) example stated earlier. Students were asked to determine the typical income in the following set of individuals' incomes: \$16,000; 56,000; 80,000; 5,000; 25,000; 54,000; 24,000; 53,000; 37,000; 45,000; 110,000; 38,000; 46,000; 2,000,000; 26,000; 64,000; 28,000; 40,000; 14,000; 52,000. The value of \$2,000,000 was an outlier (more than 1.5 times the interquartile range above from the third quartile of the data set), making uncritical application of the mean to determine the typical value problematic. Students were also asked to explain how they came to their conclusion about the typical income. The income information appeared on cards that also included data about each individual's age, favorite color, political party affiliation, level of computer literacy, and hours of sleep per week.

#### *Signal-Versus-Noise Task*

The signal-versus-noise task posed during clinical interviews was patterned after the Konold and Pollatsek (2002) example stated earlier. Students were told that a fish had been weighed on the same scale 7 different times, and the resulting scale readings (in lbs.) were 19.2, 21.5, 10.1, 23.1, 22.0, 20.3, 21.8. Given this information, students were asked to estimate the true weight of the fish and explain their thinking. The value of 10.1 was an outlier (more than 1.5 times the interquartile range below the first quartile of the data set). This made it problematic to estimate the actual weight by simply calculating the mean of all the measurements.

### *Procedure*

The two tasks described in the present study were administered as parts of a larger interview protocol (Groth, 2003a). Each of the interview participants was informed that it would take

approximately 2-3 hours to answer all of the questions on the protocol. Six participants in the first two categories decided to split the total interview time between two interview sessions. Julie was the only student in the first two categories who decided to do the entire interview in one session. Most of the participants in the last category were interviewed during study hall periods, so the total interview time for these students was generally split among three different 50-minute blocks. In each case, the typical value problem was administered during the first interview session, and the signal-versus noise task during the last. Only one of the students in the last category, Daniel, decided to do the entire interview protocol in one sitting by scheduling an interview session after school hours. Students were given access to a calculator and were allowed to use or not use it as they saw fit. Students' responses to tasks on other sections of the interview protocol are reported elsewhere (Groth, 2003a, Groth, 2003b, Groth, in press).

As students were interviewed, the interviewer collected data by taking field notes, audio recording responses, and keeping students' written work. The audio recordings were later transcribed for analysis. The interviewer's role was to read the problems aloud, to encourage students to think aloud while solving the problems, and to ask them to explain their thinking when they did not automatically do so.

### *Data Analysis*

Analysis of student responses to interview tasks was advised by a strategy described by Maykut and Morehouse (1994). Strategies for solving the typical value task were compared against each other. In cases where the strategies shared common characteristics, they were grouped together into a cluster (for instance, strategies that incorporated formal measures of center were initially grouped together). The initial clusters were then examined for differences, and subcategories were formed based upon the differences (for instance, two of the subcategories

for the cluster “incorporated measures of center” were “critical use of the mean” and “uncritical use of the mean”). The same data analysis strategy was used to analyze responses to the signal-versus-noise task. This continuous, progressive coding process continued throughout the representation of the study results in writing (Glesne, 1999). The end result of this stage of analysis was the formation of answers to research questions one and two, which concerned describing patterns of students’ thinking within each different context.

During the second stage of data analysis, the patterns of response for the typical value task were compared against those for the signal-versus-noise task. To facilitate this stage of analysis, a matrix comparing the thinking exhibited for the two tasks was constructed (Miles & Huberman, 1994). The matrix was then examined for relationships between individuals’ strategies in both contexts. The matrix is shown in Figure 1, and a description of the patterns observed within it is contained in the last subsection of the results section.

<INSERT FIGURE 1 HERE>

## Results

### *Patterns of Thinking for the Typical Value Task*

Two broad patterns were evident in responses to the typical value task. Nine of the students interviewed attempted to incorporate formal measures of center in responding to the task, while six did not. The different patterns of thinking exhibited in response to the task are listed in the row headings in Figure 1 and described in detail in this section.

#### *Strategies Incorporating Formal Measures*

*Uncritical use of the arithmetic mean.* Two of the students did not perceive the use of the arithmetic mean of all the incomes to determine the typical value as problematic. Jeff, one of the two, simply executed the add-and-divide algorithm for the mean on a calculator and announced

that \$146,650 was the typical income. Rick, the second of the two, started by examining the data and then stated that the typical income was in the area of \$40,000 to \$50,000. However, he was not content with this estimate, as shown by his use of the calculator to find the arithmetic mean of all the values. After calculating the mean, he re-thought his initial estimate and said that the typical income was in the area of \$140,000. The word “typical” was strongly associated with the use of the arithmetic mean of all the data set values in both cases, much like students in the Mokros and Russell (1995) study who associated the idea of typical value with an algorithm.

*Critical use of the arithmetic mean.* Four students who used the arithmetic mean of all the individual values began by executing the add-and-divide algorithm on all the incomes, but decided that it did not give an accurate estimate of typical value after comparing the result against the data. Hillary, one of the four students, discarded the lowest and the highest value in the data set and calculated the mean of the rest. She then stated that the typical income was \$44,000. Crystal, another of the students, did not discard any values immediately. She thought that taking the mean of the mean incomes from a number of age brackets would produce a more satisfactory answer. When this strategy produced a number higher than the mean of the individual scores, she decided to discard the \$2,000,000 value before executing the average of averages strategy. The other two students, Paul and Julie, abandoned the use of the arithmetic mean after comparing the result of the add-and-divide algorithm against the data. The two then decided that the median was a more accurate indicator of typical value. While students in this category associated the idea of typical value with an algorithm, they did not apply the algorithm in the data-blind manner characterizing strategies in the previous category.

*Exclusive use of average of averages strategy.* Lisa, the last student who made use of the arithmetic mean in responding to the typical value task, used a strategy of calculating the

arithmetic mean of the mean values from different age brackets, much like Crystal had done. She identified the \$2,000,000 value as an “outlier,” but did not consider discarding it. She thought that her “average of averages” strategy would compensate for the pull the \$2,000,000 had on the rest of the values, but did not actually calculate the value to see if it would. As she talked about her solution strategy, she tried to attach a formal name to it, commenting at one point, “I forget what it’s called. It’s like how far they are apart. Is it like the derivative?” As with students in the previous categories, the idea of typical value was strongly associated with a formal algorithm. This association appeared to be even stronger in Lisa’s case, since she felt the need to apply not just one algorithm, but two in order to compensate for the pull of the outlying value.

*Exclusive use of the median.* Two students, Nancy and Bill, focused on trying to determine the median without first using the arithmetic mean to estimate typical income. Nancy started by arranging the data cards in order from lowest to highest income. She then stated, “I’d say about maybe 40,000 average. Um, I just went to the middle of the chart I arranged and took the median.” She had not produced the actual median of 42,500, but she didn’t acknowledge a problem with the procedure she used to determine her answer. Bill used the median in a slightly different manner. After arranging the data cards according to age, and then splitting the ages into groups with ranges of 10 years each, he stated,

About 45, maybe? No, wait a minute. Maybe about 40. First I see that I have more incomes from the younger generation. And even though none of theirs reach 40...they’re all about 26, 28, I used those to outbalance the fact that you’ve got one over here at 50, and 86, an 110,000, stuff like that. I tried to balance them out. And then your median category, the 30’s, they’re all making around 40 or 45.

Instead of just using the median income, Bill used the typical income of the median age category. Nancy and Bill were the only two students attempting to use formal measures who employed the median without trying the arithmetic mean first. Both, however, appeared to associate the idea of typical value with a formal algorithm for median (arranging data values in order and finding the middle value), just as students in the previous categories initially associated typical value with the mean.

### *Strategies Not Incorporating Formal Measures*

*Clumping strategies.* Modal clumps of varying sizes were used by four students in estimating typical income. Luke estimated the typical value to within a range of 5,000. In justifying this decision, he stated, “The majority of the data is [sic] between \$50,000 and 55,000.” Laura also looked for a clump containing a large amount of the data. After dividing the incomes into six groups with ranges of varying sizes, she concluded that the typical income would be about \$54,000, since there were “the most in the 50’s range.” Kristen justified an estimate of \$24,000 for the typical income by stating, “There are lots of 20 thousands.” Jessica gave a broad estimate of \$30,000 to \$50,000, stating, “There are a couple of people making a lot, but mostly in between, I don’t know, between 30,000 and 50,000. That might be a little broad there.” Each of the students using a clumping strategy attempted to identify a range of incomes of some size that contained the greatest amount of data values, much like students in previous studies (Konold & Higgins, 2003).

*Midpoint strategies.* Two students attempted to identify the midpoint of the income data set without making reference to trying to determine the formal measure of median. Daniel, for instance, estimated, “Somewhere around \$25,000. Some of them were lower, and some were higher. It seems like it was about in the middle.” Brooke took a similar approach but produced a

value closer to the actual median of the data set. After arranging the incomes from highest to lowest and inspecting the result, she said, “I would probably say around 48,000 per year, but there are some, like, extremes, but it typically ranges around there.” In each instance, the student attempted to informally estimate the midpoint of the data set. In this way, their strategies reflected a conception Mokros and Russell (1995) described as “average as midpoint.”

### *Patterns of Thinking for the Signal-Versus-Noise Task*

As with the typical value task, two broad patterns were evident in responses to the signal-versus-noise task. Eight of the students interviewed incorporated formal measures of center in responding to the task, while seven did not. The different patterns of thinking exhibited in response to the task are listed in the column headings of Figure 1 and described in detail in this section.

#### *Strategies Incorporating Formal Measures*

*Uncritical use of the arithmetic mean.* Just one student, Daniel, was content to calculate the mean of all seven weights in order to estimate the true weight of the fish. In response to the task, he stated, “It’s probably around 19.7 pounds. I just added up all the weights he came up with and divided by seven.” There was no sign of concern that the outlying value of 10.1 might skew the estimate of the actual weight. This strategy paralleled those in the typical value task that were blind to patterns appearing in the data.

*Critical use of the arithmetic mean.* The most common approach to the signal-versus-noise problem was to discard the outlier measurement of 10.1 pounds and then compute the mean of the rest. Hillary, Paul, Lisa, Crystal, and Bill all took this approach to the problem. The thinking of Lisa, Crystal, and Bill did differ slightly from that of the other two students when it came to justifying why they were not including the outlying value in the calculation of the mean. While

Hillary said she discarded the outlier because “it was so extreme” and Paul did so “because it would throw off the average weight,” the other three justified discarding it by hypothesizing that the measurement was due to human error. Lisa called the outlier “an obvious error,” Crystal said she thought the fisherman “screwed up,” and Bill also said that he didn’t think the measurement was correct. The latter three reflect attention to contextual factors as well as the numerical data, while the former two do not exhibit any explicit attention to the context.

*Average of averages strategy.* Kristen was the only student to calculate both the mean and the median of the given data set. She related her thinking during the interview as follows:

[The typical value is] 20.9. Yeah, I just put those weights in numerical order, and then, uh, oh, wait, no, I counted wrong. 21.5. I thought that the middle one was between 20.3 and 21.5, but the middle number is 21.5. Or you could add them all up and divide them by...(picks up calculator)...that’s quite a bit of a difference. OK, I’m changing my answer again to 20.6, because if you just added all those weights together and then divided by 7, you got 19.7 as the average. If you just looked in the middle, you got 21.5. So, then I just add 21.5 to 19.7 and divided the two, and got 20.6. So, I just did the average of both averages.

Her first impulse was to use the median, but then she also seemed compelled to calculate the mean. When the two turned out to be fairly far apart, she applied the arithmetic mean algorithm once again to the two measures calculated. Her thinking resembled the thinking of the students applying an average of averages strategy to the typical value context.

*Use of the arithmetic mean of maximum and minimum.* The last student to use a formal measure in responding to the signal-versus-noise context task was Jessica. She described her strategy in the following manner:



Jessica: I would probably just find the average of the highest one and the lowest one, but I think he should just get a new scale... (uses calculator) ...If you find the average of the highest and the lowest, it's 16.05.

Interviewer: So you would say the true weight was 16.05?

Jessica: Probably around there, but he should check it on another scale. That's probably just the closest he could get with those measurements.

Jessica did not seem to mind that the value of 16.05 was fairly far from the majority of the measurements, not to mention the fact that it was unreasonably precise for an estimate given the rest of the measurements. She indicated a belief that there may be something wrong with the data, but did not recognize the weight of 10.1 as a likely measurement error that should be disregarded. Instead, she made the outlying value an important part of her estimate.

*Strategies Not Incorporating Formal Measures.*

*Estimation strategies based on numerical data.* Five students used informal estimation strategies that approximated the true weight based on the available data. Two of these students, Luke and Jeff, explicitly justified discarding the outlying 10.1 pound measurement because they believed it to be due to measurement error. A third student, Nancy, said she excluded the outlier but did not express a belief that it was due to measurement error. The last two students in this category, Julie and Brooke, produced estimates near to those of the others using informal estimation, but did not say that the outlier had any effect upon their decision. Brooke, for instance, stated, "I would say the fish is about 21 pounds. I just kind of spot checked it. It seemed like there was a couple of 21s and a low 22, so right around there." All of the other students in this category used similar "spot checking" strategies that placed their estimates within a pound of Brooke's.

*Estimation strategies based largely on context.* Laura produced an answer in the same neighborhood as the students using informal estimates based on the numerical data. However, her strategy was different because her speculation about the context in which the measurements were produced largely drove her thinking. When asked to estimate the true weight of the fish, she responded, “Probably between, like, 21.6 and 21.7, because there are two that are – the 21.8 and 21.5 are so close together, and those two are so close together that it was probably around there.” Instead of basing her estimate on examining where the majority of the data fell, she focused upon the two measurements that turned out to be closest together. The two measurements closest together were presumably the most trusted because the probability they arose from measurement error was perceived as small.

Rick, the second student who based his estimate of the true weight largely upon context, explained his solution to the task in the following manner:

I actually think that the 10.1 was the exact amount of the fish. I would say this just because if you think about it, if the fish was going around on the scale, he probably put down his hands to control the fish, and that probably added a good 10 pounds. It just matters how hard he pushed down on the fish. I think when the fish was still, he got the 10.1. I think that’s how he got the right answer. A 23 pound fish is a pretty big fish.

There’s a lot of 10.1 pound fish caught.

Rick identified contextual factors that would make the majority of the data recorded inaccurate. He drew upon his knowledge of difficulties involved in weighing a fish and also upon his knowledge of the weights of most fish caught in order to justify his answer.

*Comparison of Patterns of Thinking Across Contexts*

Figure 1 illustrates the relationships between the patterns of thinking exhibited by study participant for both tasks. The figure is divided into four quadrants by the double lines. In the upper left hand quadrant (quadrant 1) are students who incorporated formal measures of center in both the signal-versus-noise and typical value task. The upper right hand quadrant (quadrant 2) contains those who used a formal measure for the typical value task and an informal estimate for the other. The lower left hand quadrant (quadrant 3) contains those who used an informal estimate for the typical value task and a formal measure for the other. Finally, the lower right hand quadrant (quadrant 4) contains those who used only informal estimating strategies in both contexts.

*Quadrant 1 Characteristics*

Four of the five students appearing in quadrant 1, where formal measures were incorporated in both contexts, were former or current high school statistics students. The fifth student, Bill, was enrolled in AP Calculus but had not taken a formal high school statistics course. Hence, all of the students in the quadrant had experienced a great deal of formal schooling in the subject area of mathematics.

The most common strategy in quadrant 1 was the critical use of the mean in both contexts. Three students' sets of responses fit this description. All of the students in the quadrant critically applied the mean in the signal-versus-noise task. The only student in the quadrant not producing fairly reasonable answers to both tasks was Lisa. The "average of averages" strategy she applied to the typical value task was an unfruitful attempt to force a formal measure upon the data. Lisa performed differently on the signal-versus noise task, where she critically applied the arithmetic mean to produce a reasonable estimate of the actual weight of the fish.

*Quadrant 2 Characteristics*

Quadrant 2 was more diverse, in terms of students' mathematical backgrounds, than quadrant 1. Two of the students in quadrant 2, Jeff and Julie, had experienced a formal course in high school statistics. The other two students, Rick and Nancy, were both high school sophomores who had not experienced such a course. Therefore, the mathematical backgrounds of students in quadrant 2 varied much more than that of the students in quadrant 1.

Each student in quadrant 2 paired formal strategies on the typical value problem with informal strategies on the signal-versus-noise task. Each pair of strategies in the quadrant is unique, as shown by the fact that no two students fall within any given cell in Figure 1. Jeff and Rick both made uncritical use of the arithmetic mean in responding to the typical value task, strongly associating "typical" with the procedure for calculating mean of all the values in the data set. Jeff and Rick did not associate the arithmetic mean with the signal-versus-noise context, however. Instead, they attended to the data and the context in order to produce an estimate of the true weight. Julie and Nancy, the other two students in quadrant 2, made more reasonable use of formal measures in the typical value context. Each student used median income as an indicator of typical value. Julie did so after calculating the mean and noting the outliers, while Nancy immediately made use of the median. In the signal-versus noise context, both Nancy and Julie made an estimate based on the available numerical data

*Quadrant 3 Characteristics*

Quadrant 3 contained perhaps the most variation in terms of mathematical experience of any of the four quadrants. Kristen had the most extensive mathematical background, as she was a college freshman who had just completed a semester-long high school statistics course. Daniel, another student in the quadrant, was just beginning high school and was enrolled in a geometry

course. The third student, Jessica, was in her second year of high school and was taking an algebra course.

Students in quadrant 3 made use of informal measures in the typical value context and formal measures in the signal-versus-noise context. Two of the students, Kristen and Jessica, associated typical value with the “clump” containing the most data. Both students seemed to force a formal measure of center upon the signal-versus-noise context. Kristen thought it necessary to take the arithmetic mean of the median and mean of the data set. Jessica calculated the mean of the highest and lowest values in the set, even though the lowest value was an outlier. The third student, Daniel, seemed to have difficulty reasoning in each context. He attempted to estimate the center of the data set in response to the typical value task, yet produced an answer far from the actual median. He then uncritically applied the arithmetic mean to estimate the true weight in the signal-versus-noise task.

#### *Quadrant 4 Characteristics*

As with quadrant 3, students in quadrant 4 were fairly diverse in terms of mathematical background. Laura had recently completed a year-long high school statistics course. Luke was a high school senior enrolled in AP Calculus, but had not taken a formal statistics course in high school. Brooke was in her second year of high school and was taking an algebra course.

Each student in quadrant 4 made no use of formal measures in either task. Luke and Laura each associated the word “typical” with the data clump containing the most values. They each took different approaches to the signal-versus-noise task, however, since the numerical data drove Luke’s strategy, while contextual factors largely drove Laura’s. Brooke’s strategy on the signal-versus noise task was similar to Luke’s in that it was an estimate driven by the numerical data, but her approach to the typical value task was unique among those in this quadrant in that

she associated the word “typical” with the midpoint of a data set rather than with the clump containing the greatest number of values.

### Discussion

Near the outset of this paper, it was mentioned that one expected benefit from the present study would be deeper insights about students’ thinking in the two statistical contexts under investigation. The preceding section described several patterns of thinking that were observed and the relationships between the types of thinking employed in the two different types of problems. Those descriptions of students’ thinking hold several implications for instruction pertaining to the use of formal measures of center in different contexts.

Formal measures of center were used at some point by 12 out of the 15 students participating in the study. Nine students felt it was appropriate to do so in the typical value task, while eight did so in the signal-versus-noise task. While measures of center were used to produce reasonable responses in many cases, in other cases the measures seemed to be forced upon the data. The uncritical use of the arithmetic mean was observed in both tasks. Both tasks also prompted some students to apply an inefficient and unnecessary “average of averages” strategy to the data. The signal-versus-noise task was unique in that it prompted one student to calculate the mean of the highest and lowest values in the data set to produce an estimate of actual weight, even though the lowest value was an outlier.

The unreasonable uses of formal measures in the above cases show that the incorporation of formal measures does not always indicate a more sophisticated level of thinking, if one defines sophisticated statistical thinking in terms of taking both data and context into account. Many times the informal estimates observed produced more reasonable solutions for the given tasks. For instance, students who used spot-checking strategies in the signal-versus-noise task often

produced more reasonable responses than those using formal measures. This illustrates a pitfall of high school statistics curricula that focus almost exclusively on helping students become more proficient in applying formal measures. Students also need to become proficient in knowing when it is *not* necessary or desirable to apply such measures.

The fact that some of the students interviewed for this study reached for formal measures quickly and uncritically lends support to the notion that informal experiences in reducing data should lay the foundation for experiences with formal measures of center in later grades (Mokros & Russell, 1995; National Council of Teachers of Mathematics, 2000). In most cases where students misapplied measures of center, it seemed they had little or no inclination to examine the data set first to estimate the value a formal measure of center might produce. Instead, they applied the formal measures rather blindly, hoping to produce the “right” answers to the task. This phenomenon suggests that a key part of early, informal experiences with data reduction should be to emphasize that there are often several “right” answers to problems in data analysis, and that the reasonableness of any given estimate of a typical value or a signal among noise should be judged against how well it represents the data set as a whole.

While teachers of the youngest children in school need to be conscious of providing experiences that will lay the foundation for later work with formal measures of center, teachers of adolescent children need to be conscious of treating the formal measures as additional tools for data analysis rather than the only acceptable ones. This point is illustrated by examination of Rick’s response to the typical value task. His initial reasonable estimate of typical value was rather quickly replaced by the uncritical use of a formal measure. Instead of staying with his sound intuition, he placed his trust in a straightforward application of the add-and-divide algorithm he had learned for dealing with such problems in school. In his case, the acquisition of

a formal tool had done more harm than good, since it trumped sound intuitions that were already in place. A fruitful approach to resolving problems like this may be to have students keep making informal estimates frequently during their first encounters with formal measures. In cases where there is a conflict between intuitive estimates and formal measures, they should be encouraged to find the causes of the conflict rather than simply replacing their intuitive estimates with formal tools that usually produce the “correct” school-sanctioned answers. Too often, children are more concerned with school-taught statistical conventions than with critical analysis of data and context (McGatha, Cobb, & McClain, 2002).

Those responsible for teaching formal measures of center to adolescent students can learn another important lesson by examining the responses that Rick and Laura gave to the signal-versus-noise task. Their responses to the task were driven, in large part, by thinking about the impact contextual factors might have upon data production. In each case, their non-use of the arithmetic mean of the main “clump” of values was justified with plausible contextual considerations. It could be argued that their responses represent a higher level of statistical thinking than that of students who responded to the task by calculating the mean of the non-outlying values with no contextual justification for discarding the outlier, since Rick and Laura exhibited some of the “shuttling back and forth between data and context” (Wild & Pfannkuch, 1999) that characterizes statisticians’ thinking. When challenging a class to solve a task in a signal-versus-noise setting, teachers would be well-advised to elicit thinking from students such as these who think deeply about the context in which the data were produced. McClain and Cobb (2001) found that discussions about the data-production contexts of tasks were essential elements of a teaching experiment they conducted with middle school children. Hearing the reasoning of students like Rick and Laura may help prompt other students to consider the context along with



the data in producing a solution to a problem, even if they ultimately do not agree with the original contextually-based conclusions.

Finally, it should be noted that the present study both resonates with and builds upon some of the categories of thinking documented by previous research (Mokros & Russell, 1995; Watson & Moritz, 1999). It resonates with previous study findings about the typical value context because some of the categories, such as an algorithmic conception of typical value, a midpoint conception, and the use of “clumping” strategies have all been previously observed. It builds upon existing categories of thinking because of the explicit attention paid to investigating the signal-versus-noise context. In particular, this study shows that just as some students immediately associate the typical value of a data set with an algorithm, some also immediately associate the “signal in a noisy process” with an algorithm. Some students, such as Daniel and Jessica, produced rather unreasonable estimates through their applications of the arithmetic mean to the data. Even Hillary and Paul, who made reasonable estimates by discarding the outlier, did not justify their decisions to discard the outlying value by discussing the context in which the data were produced. It could be argued that they still held an algorithmic conception of “signal in a noisy process,” just that their algorithm was more complex than Daniel and Jessica’s because it included identifying outliers as one of its steps. An important area for further research is to determine what types of interventions would help students break free of an algorithmic conception of “signal among noise.” Perhaps instilling in such students some of the contextually-based doubts scientists historically have had about this particular application of the arithmetic mean (Stigler, 1986) would be one fruitful approach.

### Conclusion

The present study describes patterns of statistical problem-solving behavior in different contexts. The patterns of thinking illustrate how some students' pursuit of a school-sanctioned solution to a problem can short-circuit the reflexive interaction between data and context that should characterize statistical thinking. Strategies such as uncritical use of the arithmetic mean, use of unnecessarily cumbersome formal strategies, and the elimination or retention of outliers without consideration of the context that produced them all seem to reflect indoctrination into school statistics rather than a deep understanding of the actual discipline of statistical data analysis. Teachers need to be conscious of building students' statistical intuitions about data and context and connecting them to formal measures without implying that the formal measures should replace intuition. School statistics can bear a closer resemblance to the discipline of statistical data analysis as it encourages students to think carefully about data-production contexts instead of fostering a perception that formal tools and algorithms are necessary in order to produce "correct" answers.

### Acknowledgements

This study is a product of a doctoral dissertation completed at Illinois State University. I wish to thank my dissertation committee co-chairs, Cynthia Langrall and Edward Mooney, for all of their support. I also wish to thank Sharon McCrone and Beverly Hartter, who served on the committee and provided insightful comments about the dissertation research.

## References

- Bogdan, R.C., & Biklen, S.K. (1992). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.
- Fennema, E., & Franke, M.L. (1992). Teachers' knowledge and its impact. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147-164). New York, NY: Macmillan.
- Garfield, J., & Gal, I. (1999). Teaching and assessing statistical reasoning. In L.V. Stiff & F.R. Curcio (Eds.), *Developing Mathematical Reasoning in Grades K-12* (pp. 207-219). Reston, Va.: National Council of Teachers of Mathematics.
- Glesne, C. (1999). *Becoming qualitative researchers: An introduction* (2<sup>nd</sup> ed.). New York: Longman.
- Goldin, G.A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A.E. Kelly & R.A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 517-546). Mahwah, NJ: Erlbaum.
- Groth, R.E. (2003a). *Development of a high school statistical thinking framework*. (Doctoral dissertation, Illinois State University, 2003) *Dissertation Abstracts International*, 64 (04), 1202.
- Groth, R.E. (2003b). High school students' levels of thinking in regard to statistical study design. *Mathematics Education Research Journal*, 3, 252-269.
- Groth, R.E. (in press). An exploration of students' statistical thinking. *Teaching Statistics*
- Jones, G.A., Thornton, C.A., Langrall, C.W., Mooney, E.S., Perry, B., & Putt, I.J.

- (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning*, 2, 269-307.
- Konold, C., & Higgins, T.L. (2003). Reasoning about data. In J. Kilpatrick, W.G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33, 259-289.
- Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. London: The Falmer Press.
- McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, 45, 103-129.
- McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical understandings: developing a conjectured learning trajectory. *Journal of Mathematical Behavior*, 21, 339-355.
- Merriam, S.B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Mewborn, D.S. (2003). Teaching, teachers' knowledge, and their professional development. In Kilpatrick, J., W.G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 45-52). Reston, VA: National Council of Teachers of Mathematics.
- Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

- Mokros, J., & Russell, S.J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26, 20-39.
- Mooney, E.S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning*, 4, 23-64.
- Moore, D. (1992). Teaching statistics as a respectable subject. In F. Gordon & S. Gordon (Eds.), *Statistics for the Twenty-first Century* (pp. 14-25). Washington, D.C.: Mathematical Association of America.
- National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.
- Patton, M. (1990). *Qualitative evaluation and research methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage.
- Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, (1), 1-22.
- Stigler, S.M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Strauss, A.L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Watson, J.M., & Moritz, J.B. (1999). The development of concepts of average. *Focus on Learning Problems in Mathematics*, 21, 15-39.
- Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical inquiry. *International Statistical Review*, 67, 223-265.

Table Captions

Table 1. *Students who had Recently Completed a Semester-long High School Statistics Course*

Table 2. *Students who had Recently Completed a Year-long High School Statistics Course*

Table 3. *Students Enrolled in High School at the Time of the Study*

Table 1

Student	Year in school at time of study
Lisa	College freshman
Kristen	College freshman
Laura	College freshman



Table 2

Student	Year in school at time of study
Jeff	College freshman
Hillary	College freshman
Paul	College freshman
Julie	College freshman

Table 3

Student	Year in school at time of study	Course enrolled in at time of study
Crystal	High school senior	AP Statistics
Bill	High school senior	AP Calculus
Luke	High school senior	AP Calculus
Daniel	High school freshman	Honors Geometry
Jessica	High school sophomore	Geometry
Nancy	High school sophomore	Geometry
Brooke	High school sophomore	Algebra
Rick	High school sophomore	Algebra

Figure Caption

Figure 1. *Summary of Patterns of Response Across Tasks*

Figure 1

Signal vs. Noise	Typical Value	Uncritical use of the mean	Critical use of the mean	Average of averages	Mean of maximum and minimum	Estimate based largely on data	Estimate based largely on context
Uncritical use of the mean						Jeff*	Rick
Critical use of the mean			Hillary* Crystal* Paul*			Julie*	
Exclusive use of average of averages			Lisa*				
Exclusive use of the median			Bill			Nancy	
Clumping strategy				Kristen*	Jessica	Luke	Laura*
Midpoint strategy		Daniel				Brooke	

\*Student had completed or was currently enrolled in a high school statistics course.