DEVELOPING A MODEL TO DESCRIBE THE USE OF CONTEXT KNOWLEDGE IN DATA EXPLORATIONS

Edward Mooney and Cynthia Langrall
Illinois State University, United States
Steven Nisbet
Griffith University, Australia
langrall@ilstu.edu

This study aimed to examine the role knowledge of context plays in supporting or interfering with middle school students' statistical thinking. A model of context knowledge use was developed based on a model of context support developed by Beck, McKeown and McCaslin (1983) to describe students' use of context knowledge. The results of the study showed that students' use of context knowledge fell into three categories.

INTRODUCTION

It is widely acknowledged that proficiency in statistical skills enables people to become productive, participating citizens in an information society and to develop scientific and social inquiry skills (e.g., National Council of Teachers of Mathematics [NCTM], 2000; Secretary's Commission on Achieving Necessary Skills [SCANS], 1991). In response to this need, calls for reform in mathematics education have advocated a more pervasive approach to statistics instruction at all levels (e.g., NCTM, 2000). Researchers examining the development of students' statistical understanding are beginning to produce a coherent body of knowledge that can inform instruction (Ben-Zvi and Garfield, 2004).

Crucial to the development of statistical understanding is the realization that data are numbers in context (Moore, 1990). By *context* we mean the real-world phenomena, settings, or conditions from which data are drawn or about which data pertain. This definition is consistent with the way the term is used by others (e.g., Gal, 2004; Moore, 1990; Pfannkuch and Wild, 2004) with regard to statistics. According to Moore, "data engage our knowledge of the context so that we can understand and interpret rather than simply carry out arithmetical operations" (p. 96). However, researchers have found that context can create obstacles as well as supports in developing students' statistical understanding (Berg and Phillips, 1994; Mevarech and Kramarsky, 1997). Because little is known about the interaction between ones' knowledge of context and ability to analyze and interpret data, this study aimed to examine the role knowledge of context plays in supporting or interfering with middle school students' statistical thinking. More specifically, can we categorize the role of context in students' statistical thinking when solving problems requiring the comparison of two data sets?

THEORETICAL CONSIDERATIONS

The design and analysis of this study was informed by Dapueto and Parenti's (1999) theoretical model for describing the relationship between context and the formation of mathematical knowledge. In adapting aspects of their theory to the domain of statistics, we identified three factors that should serve as considerations when investigating students' statistical reasoning and thinking. The first factor involves the students' field of experience, or familiarity with the context of a problem or task. The second factor involves the mathematics/statistics inherent in the problem context, that is, whether the context necessitates the use of certain mathematical/statistical knowledge. The third factor pertains to the meaningful role the mathematics/statistics plays in understanding or interpreting the problem context. From the perspective of the Dapueto and Parenti model, as students build conceptual models or *mathematize* problem situations, they shift (possibly back and forth) between the use of context knowledge and mathematical/statistical knowledge.

Other researchers (Pfannkuch and Wild, 2004; Shaughnessy, Garfield, and Greer, 1996) have also described the notion of shifting or the interplay between data and context. According to Pfannkuch and Wild, the ability to integrate statistical and contextual information, knowledge, and conceptions is a fundamental element of statistical thinking. They contend that "because information about the real situation is contained in the statistical summaries, a synthesis of statistical and contextual knowledge must operate to draw out what can be learned from the data about the context sphere" (p. 20). Similarly, Watson and Callingham (2003) describe two essential components for statistical literacy as the "mathematical/statistical understanding of the content and engagement with context in exploiting this understanding" (p. 20).

Our work was also informed by Gal's (2004) model of statistical literacy, which describes the types of knowledge (literacy skills, statistical knowledge, mathematical knowledge, context knowledge, and critical questions) and dispositions (beliefs, attitudes, critical stance) that enable a person to "comprehend, interpret, critically evaluate, and react to statistical messages" (p. 50). According to Gal, these knowledge bases and dispositions overlap and interactions among them occur as one engages in statistical situations. For example, he stated that "if a listener . . . is not familiar with a context in which data were gathered, it becomes more difficult to imagine why a difference between groups can occur, what alternative interpretations may exist for reported findings about an association detected between certain variables, or how a study could go wrong" (p. 64). Although Gal's model of statistical literacy was aimed at consumers of statistics, it is also applicable to students in school settings. His discussion about interactions among knowledge bases and the notion of taking a critical stance supports our investigation into the interplay between data and context and informed the analysis of data in this study.

Beck, McKeown and McCaslin's (1983) model of contextual support was used to develop an initial model of context knowledge usage. The model is designed to examine the role story context plays in understanding the meaning of target words. The model consists of four levels of contexts. The first level is misdirective context. At this level the context can lead the reader to an incorrect meaning of the target word, as in the phrase, 'He has a *frog* in his throat.' This setting would not allow the reader to have an accurate interpretation of the target word, frog. The next level is nondirective context. At this level, the context does not lead the reader to any understanding of the target word (e.g., Tommy released the *frog*.) General context, the third level, will provide the reader with an overall setting to understand the target word's meaning. The sentence, 'The *frog* swam away,' provides the reader with the idea that a frog is something that swims. Directive context provides the reader with an implicit understanding of the target word (e.g., The *frog* is an excellent swimmer since it is a semiaquatic amphibian with smooth skin). We hypothesized that a modified version of Beck, McKeown and McCaslin's model could be used to describe the way students use context knowledge when examining data. The category descriptors focus on whether the knowledge interferes or assists students in understanding or completing a task involving data. The categories are described below:

1. *Misdirective context knowledge* – Knowledge of the context that interferes with completing or understanding the mathematical task at hand.

2. *Nondirective context knowledge* – Knowledge of the context that neither assists in nor interferes with completing or understanding the task at hand.

3. *General context knowledge* – Knowledge of the context that assists with understanding the mathematical task at hand.

4. *Directive context knowledge* – Knowledge of the context that helps to complete the mathematical task at hand.

METHOD

Participants

Two sixth-grade students volunteered to participate in the study – Brad and Jordan, pseudonyms. Both students were near the top of their class in regards to mathematics. They were able to share their ideas and worked well together. Prior to the start of the study, the researcher interviewed each student individually to determine areas in which each student may have expertise – i.e., to determine possible contexts based on familiarity.

Brad’s favorite subjects are science and mathematics. He particularly likes the experiments done in science. He plays four different instruments and enjoys playing video and computer games. Brad is involved in several school activities including basketball, baseball, jazz band and scholastic bowl. He likes basketball and baseball though he does not watch either very much. His favorite baseball team is the St. Louis Cardinals. Brad reads adventure and mystery books; listens to rock, classic rock and some rap music and is fond of watching cartoons and movies on television. He often visits family members during the summer.

Jordan’s favorite subject is mathematics. He has played chess since the first grade and has run track since the 3rd grade. He watched some of the track events during the Olympics. Jordan enjoys videogames – his favorite is Halo 2. He enjoys reading Harry Potter, Chronicles of Narnia and Charlie Bloom series books. He enjoys going to museums; doesn’t watch much television; and enjoys listening to Maroon 5, the Beatles, the Eagles, U2 and Green Day.

Data Collection

Across the 3 sessions, Brad and Jordan were presented with 4 tasks. The students were encouraged to work together and to talk about what they were thinking while working on the tasks. The intent was to present the students with a variety of data sets and examine how students use their knowledge of the context to determine which information will be necessary and which information could be ignored. All sessions were audio-taped and video-taped for transcription.

Tasks

Each task consisted of making a comparison of 2 datasets and taking a critical stance based on the data. The datasets contained a variety of information in which the students needed to filter out what was needed and what was not. The first task, Venus vs. Serena, involved examining data about the tennis careers of Venus and Serena Williams. Brad and Josh were asked to use the information to determine the better tennis player. The data included such information as their birth date, birthplace, career single and doubles wins, WTA ranking, brand of tennis clothes, and career highlights. In Elvis vs. the Beatles task, students needed to examine such information as album sales, weeks on the Billboard charts, number of gold and platinum albums, Grammy awards and Rock and Roll Hall of Fame induction to determine if Elvis Presley or the Beatles had the better music career. The American Airlines vs. Delta Airlines task involved deciding which airline was more reliable. The data included corporate information, the number of flights in 2004, the number of cancelled flights, number of late flights, the number of customer complaints and the average time airborne. Finally, the Cardinals vs. the White Sox task involved determining which baseball team was better. The data included the name and capacity of the stadium, team leaders in batting and pitching, the team salary, the championship history and the date the franchise started. (Tasks in their entirety can be seen at www.math.ilstu.edu/langrall).

Analysis

Qualitative methods of analysis were used to analyze data collected from video and audio taped. Transcripts were examined to see when students used context knowledge in completing the task. Context knowledge was considered any information the student stated, related to the task at hand that was not presented or recognized originally in the data sets. Each episode in which context knowledge used was examined in comparison to the students’ solution in order to determine the role the knowledge played in completing the task. Episodes that could not be categorized were grouped to determine if new categories of context knowledge usage arose.

FINDINGS

Across the 4 tasks there were 21 episodes of context knowledge usage that were analyzed. All episodes were categorized using the existing categories. Table 1 shows the distribution of episodes across tasks. Results are discussed by categories.

Table 1: Distribution of Context Knowledge Use Across Tasks

	Category of Context Knowledge Use			
Task	Directive	General	Nondirective	Misdirective
Venus vs. Serena	3	4	2	0
Elvis vs. the Beatles	0	2	1	0
American vs. Delta	0	4	0	0
Cardinals vs. the White Sox	1	1	2	0

None of the episodes were categorized as misdirective. Five of the 21episodes were categorized as non-directive. These statements consisted of related facts to the data and were typically side notes made while examining the data. The information did not play any part in completing the task. In the Venus vs. Serena task, Jordan talked of his surprise that both women are right handed: "… they're both right-handed. Usually players that are left-handed are better players because the spin [of the tennis ball] is different." In the Cardinals vs. the White Sox task, Brad and Jordan are examining the number of World Series titles each team has when they had the following dialogue:

> B: The Cardinals have nine World Series [titles]. The White Sox have two. That's when …his name was…one of the best players for the Cardinals…
>
> J: [Mark] McGwire?
>
> B: No, …Stan, Stan the…Stan 'the Man' Marslow…Marslow I think.

In this case, Brad's attempt to recall Stan Musial's name played no role in the students' discussion of which team was better. In each case, the students provided information that neither interfered nor assisted in their decision.

The most of the episodes, 11, were categorized as demonstrating general context knowledge use. There were two types of general context knowledge usage. First, knowledge of the context was used to clarify or explain the data. Two examples are seen below when Brad and Jordan are discussing the American Airlines vs. Delta Airlines data. In the first example, Brad is trying to understand a category of data presented, in this case number of diverted flights.

> B: Diverted flights? What's 'diverted flights'?
>
> J: I think it's when you get off one plane and move to another one, something like that.

In the second example, Jordan is trying to explain why the average airborne time of flights of the two airlines cannot be compared.

> B: Airborne time.
>
> J: Well, it depends on where you are flying?
>
> B: American Airlines goes across the world.
>
> J: So does Delta probably. [Airborne time of a flight] depends on where you're flying.

It should be noted that in the final comparison of the two airlines, Brad and Jordan did use the average airborne time of flights to help determine the airline that was better.

The second use of general context knowledge was for the purpose of trying to rationalize or explain why the data presented are as they are. In the first example below, Jordan tried to explain why Serena Williams has no WTA ranking for the years 1995 and 1996.

> J: [Serena's] average [ranking] is eighteen.
>
> B: Eighteen?
>
> J: Yeah, 'cause she was 99th in 1997.
>
> B: But, [Venus' 204th ranking] is in 95. So…
>
> J: Oh, [Serena has no ranking] ' cause she did not make it in 1995. You have to be in the top 130.
>
> B: [Venus] was 204th in 1996 and 95.
>
> J: What? 204?

In the next example, Brad and Jordan noticed that Delta Airlines had more mishandled baggage reports than American Airlines as well as more late flights. Here they rationalized what they noticed.

> B: Happier customers are flying American.
>
> J: Yeah, American passengers are happier. [Delta flights] are probably taking longer because they handle baggage so poorly.

There were 5 cases in which directive context knowledge was used. These situations typically described the reason that data were not used in analyzing the data. In the Venus vs. Serena tasks, Jordan explained that when comparing the two tennis players, "we couldn't [use] doubles [wins] because they usually play doubles with each other." In the Cardinals vs. the White Sox tasks, Jordan used his context knowledge to explain why they cannot compare the two teams "since they're in different leagues [and] they don't play the same people." In these cases, the data were not used in the final decision making process. Jordan did give an example of using context knowledge to explain why they needed to use data in the Venus vs. Serena tasks. Jordan explained to Brad that Grand Slam Singles events are "four of the best games [in tennis]."

DISCUSSION

Context knowledge is vital for determining "the relevance of data to the problem" (Pfannkuch and Wild, 2004, p. 38). In this study, two students were presented similar tasks in various contexts to examine how their context knowledge was used to determine the relevant data for comparisons. The results of the study indicate that students' use of context knowledge varies and can be described and categorized. Also, the model of context knowledge usage can describe students' use of context knowledge when working with data to take a critical stance. Usage fell into 3 of the 4 categories.

Teachers should be aware of the role that context knowledge can play in students' statistical thinking. Rarely do questions involving the use of data have the necessary data neatly organized and presented to them in advance. The study showed that students can handle messy data and filter out needed data to make a stance. Often in classroom experiences, students are asked to find measures of center, construct graphs, and read tables. Open-ended tasks such as these provide students with the opportunity to apply these statistical skills in more authentic settings. Teachers may want to investigate student 'expertise' in order to provide statistical tasks that will engage students and allow them to draw upon their context knowledge.

This study was conducted in conjunction with the work of Langrall, Nisbet and Mooney (see report in these proceedings). Only two students were used in the initial examination of context knowledge use. Further research is needed to determine if other categories of context knowledge usage exist, and if descriptors should be modified. Also, the researchers plan to conduct a teaching experiment to determine how context knowledge is used and can be developed in the classroom setting.

ACKNOWLEDGEMENT

The study was conducted in conjunction with Langrall, Nisbet and Mooney (2006).

REFERENCES

Beck, I. L., McKeown, M. G. and McCaslin, E. S. (1983). Vocabulary development: All contexts are not created equal. *The Elementary School Journal*, 83, 177 – 181.

Ben-Zvi, D. and Garfield, J. (Eds.). (2004). *The Challenge Of Developing Statistical Literacy, Reasoning, And Thinking*. Dordrecht: Kluwer.

Berg, C. A. and Phillips, D. G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. *Journal of Research in Science Teaching*, 31, 323-344.

Dapueto, C. and Parenti, L. (1999). Contributions and obstacles of contexts in the development of mathematical knowledge. *Educational Studies in Mathematics*, 39, 1-21.

Gal, I. (2004). Statistical literacy: Meaning, components, responsibilities. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge Of Developing Statistical Literacy, Reasoning, And Thinking*, (pp. 47 – 78). Dordrecht: Kluwer.

Mevarech, Z. R. and Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, 32, 229-263.

Langrall, C., Nisbet, S., and Mooney, E. (2006). The interplay between students' statistical knowledge and context knowledge in analyzing data. In A. Rossman and B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics,* Salvador, Brazil. Voorburg: The Netherlands: International Statistical Institute.

Moore, D. (1990). Uncertainty. In L. Steen (Ed.), *On the Shoulders of Giants: New Approaches to Numeracy*, (pp. 95-137). Washington, D.C.: National Academy Press.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Pfannkuch, M. and Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 17-47). Dordrecht: Kluwer.

Secretary's Commission on Achieving Necessary Skills. (1991). *What Work Requires of Schools: A SCANS Report for America 2000*. Washington, DC: Department of Labor. (ERIC Document Reproduction Service No. ED 332 054).

Shaughnessy, M. J., Garfield, J., and Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, and C. Laborde (Eds.), *International Handbook Of Mathematics Education*, Part I, (pp. 205-238). Dordrecht: Kluwer.

Watson, J. and Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2, 3-46.