

THE INTERPLAY BETWEEN STUDENTS' STATISTICAL KNOWLEDGE AND CONTEXT KNOWLEDGE IN ANALYZING DATA

Cynthia Langrall

Illinois State University, United States

Steven Nisbet

Griffith University, Australia

Edward Mooney

Illinois State University, United States

langrall@ilstu.edu

In this study, we examined the role of context knowledge in data analysis tasks and investigated the interplay between students' use of context knowledge and statistical knowledge. Findings showed that knowledge of context was an important factor contributing to students' engagement in statistical tasks. Students used context knowledge in three broad ways: to rationalize the data or their interpretations, in taking a critical stance toward the data, and in ways that were not necessarily productive or pertinent in addressing the task at hand.

INTRODUCTION

Calls for reform in mathematics education have advocated a more pervasive approach to statistics instruction at all levels (e.g., Australian Education Council, 1994; National Council of Teachers of Mathematics, 2000). Although 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), the contextualized nature of data and the interaction between ones' knowledge of the context and ability to analyze and interpret data has received less attention. Yet the realization that data are numbers in context (Moore, 1990) is crucial to the development of statistical understanding. 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).

In terms of data analysis, we define *context* as 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. We are aware, however, that from a pedagogical perspective the contextual nature of data is only one aspect of context that must be considered. For example, the setting (e.g., school, home, work) in which one engages data is another contextual factor that might influence the way data are analyzed or interpreted. Another related factor is the presentation of data or the way it is encountered, for example, through the media, as a classroom task, or on a formal assessment.

There is evidence in the research literature to suggest that any of these context factors (data, setting, task) may influence the way a student engages in data analysis. Research regarding problem solving in general has shown that students perceive of tasks differently when encountered in different settings (e.g., Bernstein, 1996; Nunes, Schliemann, and Carraher, 1993) and that the presentation or context of a task itself (see Cooper and Dunne, 2000) has bearing on a students' approach to solving a problem and its solution. With regard to data analysis, some researchers have reported that the way data are presented (Watson, 1997) as well as the context of the data themselves (Berg and Phillips, 1994; Mevarech and Kramarsky, 1997) can create obstacles as well as supports in developing students' statistical understanding. Although motivated by all of these research findings, we opted to examine only one aspect of context in the study reported in this paper. That is, we investigated how students draw on their knowledge of the data context (*context knowledge*) when analyzing data.

THEORETICAL CONSIDERATIONS

Dapueto and Parenti's (1999) theoretical model for describing the relationship between context and the formation of mathematical knowledge informed the design and analysis of this study. 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 data being analyzed or interpreted. The second factor involves the statistics or mathematics inherent in the task at hand, that is, whether the data analysis or interpretation necessitates the use of certain *statistical knowledge*. The third factor pertains to the meaningful role statistics plays in understanding or interpreting the data. As students build conceptual models or mathematize problem situations, they shift (possibly back and forth) between the use of context knowledge and statistical knowledge.

Other researchers (e.g., Pfannkuch and Wild, 2004; Shaughnessy, Garfield, and Greer, 1996) have 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 contended 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) described two essential components for statistical literacy as the "mathematical/statistical understanding of the content and engagement with context in exploiting this understanding" (p. 20). In this study, we examined the context knowledge and statistical knowledge students accessed when analyzing data and investigated whether, and how, they shifted between the two.

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 supported our investigation into the interplay between data and context and informed the analysis of data in this study.

METHOD

Participants

The participants in this study were six students (3 males-3 females, ages 11-12) from a large primary school (i.e., catering for students in Grades 1 to 7) in a middle-class suburb of Brisbane, Australia. The school has a good reputation in the community and includes a greater than normal percentage of high-achieving students in the school population. The classroom teacher selected student participants in response to our request for two groups of three students such that one student in each group had great interest in and expert knowledge of some particular topic or pursuit, and hence could be described as having extensive contextual knowledge. The other four students were to have no great interest in the first two students' topics. The groups were comprised as follows: Group A—Larry (tennis fanatic), Tracey, and Don; and Group B—Beth (a female pop-singer fanatic), Mandy, and Joel. All names are pseudonyms.

Data Collection

We constructed tasks that required students to make comparisons between data sets. Two tasks corresponded with the two "expert" students' areas of interest (tennis and pop singers) to help us examine the role that context knowledge played in analyzing the data. A third task was developed on a topic that was thought to be of no special interest to any of the students, namely champion discus throwers. Figure 1 presents a summary of each task.

Data were collected during three sessions, one for each task. Each session lasted about 45 minutes and started with a brief whole-group discussion during which the students were introduced to the task. The students then worked in their assigned groups, and one of us

(Langrall, Group A; Nisbet, Group B) monitored their discussion and encouraged students to contribute their thoughts and suggestions for analyzing the data. We refrained from leading the discussion in any particular direction although we did ask probing questions to better understand students' thinking as they engaged in the task and when necessary to refocus their attention on the task at hand. Both groups were video- and audio-taped, and the students were encouraged to record their answers and opinions on the sheets of paper provided. Each session closed with another whole-group discussion during which students from each group shared their interpretations of the data. Data sources included our field notes, student work, and transcripts of each session constructed from the audio recordings and verified, when necessary, by viewing the videotapes.

| |
|---|
| <p><i>Task 1: Tennis Players</i></p> <ul style="list-style-type: none"> • <i>Does the USA produce better men's tennis players than Australia?</i> [Two tables of data were presented; (a) a listing of the winners and runners-up of the men's Wimbledon championship from 1955 to 2004, including the players' nationality and seedings, the final score, and the match time; and (b) a listing of the top 20 male tennis players with their current ranking, nationality, prize money for 2005 to date, and career prize money.] |
| <p><i>Task 2: Pop Singers</i></p> <ul style="list-style-type: none"> • <i>In a recent pop culture survey, teenagers identified Britney Spears and Delta Goodrem as two of the top female performers. Based on the following data collected from the Top 40 Charts, which of these two singers is more popular?</i> [A table of data was presented for each singer listing seven "Top 40" hits between 2002 and 2005; song title, total number of weeks on the charts, dates and positions for five highest rankings for each song] • <i>Does this extra information help you interpret the data?</i> [Tables presented with additional data on total number of different charts and specific charts for top five rankings.] |
| <p><i>Task 3: Discus Distance</i></p> <ul style="list-style-type: none"> • <i>The tables below show the winning discus throws for men and women in the Summer Olympics from 1928 to 2000. Based on these data, who has shown better performance over the years – men or women?</i> [Two tables were presented, one for men and one for women. Each table included 3 columns of information: year, winner's name, and distance in meters.] • <i>Examine this graph. What do these data tell you?</i> [Double bar graph representing the same data as the tables was presented.] |

Figure 1: Summary of tasks

Copies of the three tasks can be found on the following website: www.math.ilstu.edu/langrall.

Analysis

We used qualitative methods of analysis to identify and characterize students' use of statistical knowledge and context knowledge and the interplay between them. For each task, we documented instances of when students used statistical or context knowledge and when they moved flexibly between the two to analyze the data and interpret their findings. We also identified which students initiated the use of context knowledge and described the nature of both the context and statistical knowledge that was brought to bear on the task at hand. Students' comments or ideas were identified as reflecting context knowledge when they pertained to information that was not explicitly stated in the data sets or when students took a critical stance (Gal, 2004) with regard to the data. Students' comments and written work were taken to represent statistical knowledge when they pertained to such things as: reading and describing the data, computing averages, finding totals, partitioning a data set, or comparing data values.

Using the scheme described above as a basis for coding the data, the first two authors separately reviewed the transcripts and field notes for each session. We deconstructed each session into episodes that were defined by shifts in the focus of the students' work or nature of their comments. Each episode was coded as reflecting the use of statistical knowledge or context knowledge and when appropriate, descriptors were assigned to further characterize the episode. The few episodes that did not reflect the use of either context or statistical knowledge were essentially off-task activities and were not included in the data for analysis. Results of our

independent coding were fairly consistent, any discrepancies were discussed, and agreement was reached for all episodes. By way of example, consider the following excerpt from Task 1 that illustrates how students shifted between context knowledge and statistical knowledge:

Students in Group A work together to review the list of Wimbledon Champions and keep a tally of the winners for Australia and USA. As they count the totals Larry interjects: “But the only problem is most of the [Australian] winners and runners up, all of them were back before 1980 so they don’t play tennis anymore. Connors and McEnroe still play a little bit, they play in the oldie’s tour. . . . So let’s just look at the 1990s. For the champions, you’ve got one Australian...”

In the first part of this excerpt, Larry made a comment based on his reading of the data noting that most of the Australian’s who won Wimbledon championships did so prior to 1980. He transitioned seamlessly into a discussion about the players who are no longer active on the tennis circuit and those who continue to play “in the oldie’s tour.” We consider this shift to reflect the use of context knowledge because the data set does not indicate whether the players are still active. Based on this knowledge, Larry suggested they partition the data set and examine information only for the 1990s and later, shifting back to a statistical focus.

FINDINGS

Prior to collecting data, we hypothesized there would be noticeable differences in the nature of discussion and data interpretation between Group A and Group B, for Tasks 1 and 2, depending on the presence of an expert. However, for Task 3 where the context was neutral (i.e., no identified expert in either group), we conjectured that the work of the two groups would be similar. Our findings generally supported these hypotheses.

The Use of Context Knowledge

Findings for Tasks 1 and 2 (Table 1) indicate that in both cases, the group with the expert for the task at hand engaged in a greater percentage of episodes related to context than the group without an expert. Notwithstanding the fact that Group A was more talkative than Group B, both groups exhibited essentially the same pattern of responses. That is, when the group included an expert for the task at hand, about half of the episodes reflected the use of context knowledge. On Task 3 (Table 1), for which neither group included an expert with respect to the data, the percentages of episodes reflecting the use of context knowledge were about the same, a little more than and a little less than one-third.

Table 1: Percentage of Episodes Reflecting Use of Context Knowledge

| Group | Task | | |
|-------|--------------|-------------|------------|
| | 1 | 2 | 3 |
| A | 13/25 = .52* | 5/19 = .26 | 9/29 = .31 |
| B | 2/9 = .22 | 8/15 = .53* | 7/20 = .35 |

* Presence of “expert” in group

Our findings also show that for Tasks 1 and 2, the respective expert initiated the majority of context-related episodes. For example, on Task 1, 11 of the 13 context-related episodes (85%) were attributed to Larry, the tennis expert. On Task 2, 6 of the 8 context-related episodes (75%) were attributed to Beth, the pop singer expert. For the most part, when the group did not include an expert for a given task, the distribution of students attributed to context-related episodes was generally balanced. The only exception was for Group B on Task 1; one student, Joel, initiated the only two episodes related to context.

We identified three broad characterizations to describe the ways in which the students used context. The most common use was to rationalize the data or support one’s interpretation of the data. This occurred in more than half of the context-related episodes across both groups. For example on Task 3, the students in Group B decided to focus on discuss throwers with winning distances greater than 50 meters. Having identified that only two men had distances less than 50 meters, Mandy and Beth offered the following explanation:

M: Maybe see how in the earlier years, maybe the disc, the disc thingy was heavier because they didn't have the technology that made it lighter. It's big and in the earlier years it was pretty heavy, but now it's a big one and it's lighter than the big one in the earlier years.

B: And maybe a different type of ground and they didn't have the right shoes or something.

The second and third uses of context each occurred in about 23% of the episodes. One of these uses of context was characterized as being nonproductive to the analysis of data or not pertinent to the task at hand. For example, on Task 2 Beth initiated episodes that focused on the immaturity of Britney Spears and the fact that Delta Goodrem now has a boyfriend. The other use of context, however, reflected what we consider to be a desirable approach to data analysis and interpretation. It involves taking a critical stance that includes (a) questioning how the data were collected or what they represent; (b) recognizing the need for more or different data; and (c) bringing new information or insight to the task at hand. For example, on Task 1, Larry initiated two episodes in which he commented on the need for more information about the results of all of the grand slam tennis tournaments, not just Wimbledon.

We conclude that knowledge of context can be an important factor contributing to students' engagement in statistical tasks. At times knowledge of the context can assist in data analysis and positively influence the discussion within a group, but, at other times, context knowledge can be distracting and of little assistance in completing the task at hand. As might be expected, we found that the context expert tended to dominate context-related discussions. We also found that the context expert was able to establish authority within the group in that his or her expertise was recognized and upheld by the group members.

Use of Statistical Knowledge

The presence of an expert with respect to context appeared to have no effect on the statistical knowledge brought to bear on the three tasks in this study. Across tasks, episodes reflecting the use of statistical knowledge were attributed to a mix of students, with no one student dominating. Statistical concepts and skills used by the students included the following: reading the data in tables; visually inspecting the data in graphs; making pointwise comparisons of data values; partitioning data into subsets; describing trends in the data; computing averages; making predictions based on the data; and identifying greatest, least, and extreme values in the data.

DISCUSSION

Although some of the findings of this study may not come as any surprise to experienced teachers, it is important that these issues be investigated and the results documented. The presence of a context expert with respect to the data influenced students' work, specifically increasing the percentage of episodes reflecting the use of context knowledge. Even without particular expertise, all three characterizations of the use of context knowledge were evident in students' deliberations. However, some uses of context were more helpful and relevant than others in terms of completing the task at hand. Interestingly, although context-related discussions were sometimes protracted, we never found more than two consecutive episodes reflecting the use of context knowledge. This was not the case with the use of statistical knowledge. We found that students employed a variety of strategies in analyzing the data, resulting in multiple episodes reflecting the use of statistical knowledge. Nevertheless, there was a periodic shifting between the use of context and statistics, supporting our theoretical view of a potential interplay between context knowledge and statistical knowledge. The findings of this study emphasize the importance of providing opportunities for students to engage in investigations that encourage them to integrate contextual and statistical information (Pfannkuch and Wild, 2004; Watson and Callingham, 2003) in order to develop statistical literacy (Gal, 2004).

In terms of pedagogical implications, teachers should be aware of students' context knowledge, and be skillful in forming groups for data analysis and interpretation. They should recognize that context experts may contribute positively to the discussion, but may also bring up issues that can distract the group or dominate and stifle contributions from other group members.

Teachers should note that students respond differently to different tasks and different contexts, according to the extent of their interest and expertise.

Teachers should be conscious of the fact that tasks such as those described in this study provide good opportunity for students to (a) engage with and interpret real-world data, and (b) mathematize problem situations. Teachers therefore should know their students' interest areas and actively seek out related data sets. The tasks described in this study are open-ended and give students with a wide range of abilities the opportunity to respond at their respective levels. The use of such tasks also may reinforce, for students, the view of mathematics as a relevant, interesting and motivating activity.

ACKNOWLEDGEMENT

This study was conducted in two settings, Australia and the United States, under somewhat different conditions. In this paper we present the study conducted in Australia and refer the reader to a companion paper by Mooney, Langrall, and Nisbet (2006) for the US phase.

REFERENCES

- Australian Education Council. (1994). *Mathematics: A curriculum profile for Australian schools*. Carlton, VIC: Curriculum Corporation.
- Ben-Zvi, D. and Garfield, J. (Eds.) (2004). *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Dordrecht: Kluwer.
- Bernstein, B. (1996). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique*. London: Taylor and Francis.
- 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.
- Cooper, B. and Dunne, M. (2000). *Assessing children's Mathematical Knowledge: Social Class, Sex And Problem Solving*. Buckingham, UK: Open University Press.
- 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. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, 70, 1-52.
- 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.
- Mooney, E., Langrall, C., and Nisbet, S. (2006). Developing a model to describe the use of context knowledge in data explorations. 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.
- Nunes, T., Schliemann, A. D., and Carraher, D. W. (1993). *Street Mathematicians And School Mathematics*. Cambridge: Cambridge University Press.
- 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.
- 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, The Netherlands: Kluwer.
- Watson, J. and Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2, 3-46.
- Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, (pp. 107-121). Amsterdam: IOS Press and the International Statistical Institute.