DOES CONTEXT EXPERTISE MAKE A DIFFERENCE WHEN DEALING WITH DATA?

Cynthia W. Langrall
Department of Mathematics, Illinois State University, United States
langrall@ilstu.edu

The statistical literacy exhibited by students in two Year 6 classes was examined during an 8-day unit of instruction carried out under two conditions. For one class, data sets were designed to be specific in nature and related to the interests of many of the students. For the other class, data sets were generic in nature. Although students rarely engaged the data context, when they did, it influenced their informal inferential arguments and supported them in taking a more critical stance toward the data.

BACKGROUND

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. Thus, calls for reform in mathematics education have advocated a more pervasive approach to statistics instruction at all levels (Franklin et al., 2007; National Council of Teachers of Mathematics, 2000). Critical to the development of statistical understanding is the realization that data are *numbers in context* (Moore, 1990). 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 the research reported here, I define *context* as the real-world phenomena, settings, or conditions from which data are drawn or about which data pertain. Although this definition is consistent with the way the term is used by others (e.g., Gal, 2004; Moore, 1990; Pfannkuch & Wild, 2004) with regard to statistics, from a pedagogical perspective the contextual nature of data is only one aspect of context that can be considered. For example, the setting (e.g., school, home, work) in which one engages data is another contextual factor that can influence the way data are analyzed or interpreted. Another related factor is the presentation of data or the way it is encountered – 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.

In previous studies (Langrall, Nisbet & Mooney, 2006; Langrall, Nisbet, Mooney & Jansem, submitted), colleagues and I investigated the influence of students' expertise of the data context on their statistical reasoning and on group discussions involving the analysis of data. We found that students use knowledge of the data context in six ways: (a) to interpret or make sense of data, (b) to justify the use of particular data for the task at hand, (c) to bring additional information or new insight to a task, (d) to provide a rationale or explanation of data – why the data are the way they are, (e) to provide justification or additional support for conclusions drawn, and (f) in a non-directive manner with regard to a given task. In concert with Moore's (1990) quote presented above, we have concluded that students must possess some level of context expertise to *interact with* data during analysis as opposed to merely *acting on* data in a procedural way. The studies from which these findings were drawn were all conducted with small groups of select students in interview settings. More recently, we expanded our investigations to a whole-class setting to examine the role of context in the regular school curriculum, specifically with regard to the development of students' statistical literacy.

METHODS

Our research team[1] designed a teaching experiment to examine whether students' develop or exhibit greater levels of statistical literacy when instruction incorporates data sets selected to match students' areas of expertise than when instruction utilizes data that are relatively generic in nature. Two classes of students engaged in an 8-day unit of instruction conducted by a classroom teacher as part of the regular Grade 6 mathematics curriculum. In keeping with the school district's curriculum expectations for Grade 6, the focus of the unit of instruction was on understanding

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

measures of center and interpreting data. The participants of the study included the classroom teacher and all students in her first and second period classes. In each class, the teacher selected two groups of four students for targeted case study. According to the teacher, these were high achieving students who could be expected to participate actively in class.

Our intent was that for one class, data contexts would be relevant to the particular interests and areas of expertise of many of the students. For the other class, typical textbook data sets would be used. However, we found that the nature of the lessons introducing measures of center, specifically those presenting the balance and fair share models for the mean, did not lend themselves to such alterations. Thus, data contexts were different for the two classes only during the lessons aimed at data interpretation–the final four lessons in the unit. In this paper, I refer to only two of the tasks that students were assigned. They are presented in Figure 1 to illustrate the difference between what I refer to as specific and generic data sets.

Specific Context Task

Which videogame genre sells better?

Sports Title	Units Sold
FIFA Soccer	5,700,000
Madden NFL	5,700,000
Mario & Sonic	3,400,000
NBA 2K8	670,000
NCAA Football	1,500,000
ProEvo Soccer	3,800,000
Skate	780,000
The Bigs	730,000
TW PGA Tour 07	610,000
TW PGA Tour 80	2,300,000

Action Title	Units Sold
Assassin's Creed	4,900,000
Devil May Cry 4	1,000,000
Drawn to Life	820,000
Crash of the Titans	1,800,000
Legend of Zelda	1,900,000
Pokemon Battle …	850,000
Resident Evil, Wii	900,000
Super Mario Galaxy	4,100,000
Tomb Raider	700,000
Drakes Fortune	730,000

Generic Context Task

Do sodas have more caffeine than other drinks?

Name	Mg of Caffeine in 8 ounces
Soda A	38
Soda B	37
Soda C	27
Soda D	27
Soda E	26
Soda F	24
Soda G	21
Soda H	15
Soda I	23

Name	Mg of Caffeine in 8 ounces
Energy Drink A	77
Energy Drink B	70
Energy Drink C	25
Energy Drink D	21
Iced Tea A	19
Iced Tea B	10
Coffee	83
Hot Cocoa	2
Juice Drink	33

Figure 1. Sample data analysis tasks

Prior to the instructional unit, students were given a statistical literacy assessment, drawn from items developed by Watson and her colleagues (Watson & Callingham, 2003; Watson, Kelly, Callingham & Shaughnessy, 2003). A parallel post-assessment was administered at the conclusion of the unit. Both assessments were conducted as individual interviews that were audio recorded. Other sources of data were collected during the instructional sessions. Each day's lesson was video recorded with one camera panning the entire classroom, audio recorders were placed at the tables of the two target groups (and later in the study at the tables of other groups as well), short video clips of individual students or small groups as they worked independently were periodically recorded using a hand-held recorder, student papers were collected when applicable, and researcher field notes were taken daily. All of this yielded a large data corpus that has been analyzed and interpreted from several different perspectives.

My focus here is directed narrowly on one aspect of statistical literacy–critical disposition–and its relation to context expertise. During the instructional sessions, we noted that students in both classes rarely evaluated the data they were presented in a critical way, even though the data sets and tasks were intentionally ambiguous. Thus, I analyzed the video and audio recordings of whole-class and small-group interactions for instances in which students took a critical stance in their analyses of data. Following Gal (2004), I identified when students exhibited a critical disposition toward the data by (a) questioning the source of the data they were provided, (b) identifying discrepancies or inaccuracies in the data, (c) expressing the need for more or different data to address the task at hand, or (d) acknowledging or looking for multiple interpretations of the data. I then examined the role that knowledge of the data context played in supporting students' critical evaluation of the data.

FINDINGS

The findings of my analysis indicate that regardless of the contextual nature of the data (specific or generic) or their knowledge of the data context, students appeared predisposed to carrying out arithmetical operations as a means of analyzing data. That is, their standard approach when comparing data sets was to find total values or visually inspect the data to determine which set was greater. Sometimes they computed and compared an average value for each data set (typically the mean, but occasionally some students determined the median, mode, and range). In an effort to challenge the efficacy of this approach, we posed a task in which the totals of the data were relatively close and essentially meaningless with regard to the question posed. This task required students to determine which restaurant served relatively healthier sandwiches given the fat content of each of the sandwiches on the menus at two restaurants. The data were presented in graphical form that highlighted the fact that at Restaurant A, three sandwiches were excessively high with regard to fat content. In contrast, at Restaurant B the data were more evenly distributed but with values greater than most of the data for Restaurant A. Even in this situation, students' initial approach to comparing the data sets involved finding the total or average number of fat grams for all the sandwiches at each restaurant. Although the students' approach to this particular task points to interesting findings not necessarily related to context knowledge, I mention it here to emphasize the tenacity of the "comparing totals" strategy exhibited by these students. They generally approached data comparison tasks in a procedural way and showed surprisingly little engagement in terms of actually interpreting the data.

Given this finding, it is not too surprising that the students in this study had not developed a critical disposition toward data and rarely engaged context knowledge spontaneously when working with the data sets we presented. However, I did find evidence of the importance of context knowledge in supporting students' informal inferential arguments. This was most salient in the whole-class discussions for the two tasks presented in Figure 1—Video Game Sales and Caffeine in Drinks. To illustrate this finding, I first present a summary of the discussion pertaining to the video game task to illustrate how the students' knowledge of the data context eventually prompted them to question the data they had been presented. I then present part of the class discussion about whether sodas have more caffeine than other drinks to illustrate how a student's argument can be curtailed when context knowledge is not accessible.

Video Game Sales Task

As the data sheet for this task (see Figure 1) was distributed, the students were told that the lesson for the day would be focused on a data analysis task that they would work on together at their table groups and then there would be a whole-class discussion. Their task was to determine which genre of video game (sports or action) sells better and they were given some data presented in tables for their consideration. Students worked in small groups for about 15 minutes and then shared the results of their analyses in a whole-class discussion. There was general agreement among the students that the sports genre "sold better" than the action genre. Their strategies were all based solely on the data presented in the tables and included (a) comparing the largest data values in each genre, (b) comparing the total number of units sold for each genre, (c) comparing the average number of units sold for each genre, and (d) finding the difference between the total number of units sold for each genre. After about 10 minutes the discussion was winding down and

I asked whether the fact that there were more units sold of the sports games was convincing evidence that the sports genre sells better than the action genre. In response to my question many students nodded an affirmation but one student made the following statement: “I have something to say but it might not be true. There could be more games in action that are more popular than more games in sports.” He was essentially proposing that not all video games were represented in the tables provided. The response to his statement was immediate and many students raised their hands to comment on this idea. The discussion shifted dramatically and became more critical in nature. Students identified specific games that were not included in the data set. When asked if there were other questions they had about the data, students commented that, in addition to the number of units sold, it might be useful to know the prices of each game, the type of system each game is made for, and whether the data represent used games or only new games. Throughout this discussion, students drew on first-hand experience and knowledge of the topic. At the very end of the discussion, a student raised the issue of how the data were selected: “The person who made this [the tables] might like sports games better so they put more popular sports games than action games.”

Caffeine in Drinks Task

As the data sheet for this task (see Figure 1) was distributed, the students were told that the lesson for the day would be focused on a data analysis task that they would work on together at their table groups and then there would be a whole-class discussion. Their task was to determine whether sodas have more caffeine than other drinks and they were given some data presented in tables for their consideration. Students worked in small groups for almost 10 minutes and then shared the results of their analyses in a whole-class discussion. Most of the students had concluded that other drinks (as a category) had more caffeine than sodas. One group reasoned that the caffeine in sodas was more than 20 mg, but some of the other drinks had more than three times that amount, so “there was clearly more caffeine in the other drinks than in soda.” Several other groups compared the total or average amount of caffeine in each beverage category. At least one group computed the mean, median, mode, and range, which prompted a brief discussion about whether the range was a useful measure. Up to this point in the discussion, the general consensus was that other drinks had more caffeine than sodas. However, one pair of students shared that there was no difference in the amount of caffeine found in sodas and other drinks and the discussion shifted to an exchange between two students.

> Speakers are identified by their initials (pseudonyms are used for all students), the letter “T” refers to the teacher.
>
> A: We figured…it depends on what type of drink it would be. Like we don’t know what kind of energy drink it is or what kind of soda it is. For all we know, it could be [inaudible name of drink] or something else. But like for coffee and root beer, what would have more caffeine–coffee or root beer? Most likely coffee, right? (Someone calls out, “Root beer doesn’t have caffeine.”) Yea. So say, we could convince you that they’re close to the same. One soda doesn’t have that much difference cause 5 sodas have more caffeine than the other drinks but the other drinks that do have more caffeine have between 2 and 10 grams more, some of them like way more.
>
> T: Comment on their strategy. . . . Gina?
>
> G: Well, I don’t (inaudible) because we’re not comparing coffee to like root beer cause it’s all a general comparison. And soda’s not gonna, you’re not going to find one soda that’s going to have like a hundred milligrams of it. You’re not gonna find something that’s like way above that. Most sodas and coffees and energy drinks, they’re are all around the same thing. So, it’s kind of like more general.
>
> T: OK. Mallory.
>
> M: I think it is a good way to go because it kind of is comparing. You’re trying to compare these two things.
>
> T: (to Alicia) Do you want to respond to Gina’s concern?
>
> A: Why do you think it’s more general?

G: Because you're not comparing Soda A to Energy Drink A. You're comparing like Soda A to every other drink and not just like, cause you could find probably a soda that's way up there next to like coffee. But then you could find something as low as hot chocolate.
A: Right. But it's like, we were just thinking, we don't know what the drinks are. Different drinks are going to have different things.
G: Yeah, but they're all going to range around 20 or 30.
A: You do make kind of a good point, we didn't think of that. But did you say earlier that you didn't think the range would make that big of a difference, like it didn't matter about that?
G: Yeah, because they're all gonna be around there, so.
(break in transcript)
T: Any other questions, any other information you think would make a difference? Does anyone have a different strategy that we haven't shared yet . . . . Actually, Bobbie I thought you did something that we didn't talk about. Didn't you make some comment about just energy drinks?
B: Yeah, energy drinks have a lot more caffeine in them because they're supposed to keep you awake. They give you a lot of energy. So normally that would be what would make that a lot more higher than the other group.
T: So did that figure in to how you made your comparison about whether sodas have more caffeine than other drinks or not?
B: No, not really but I know that so I figured it [other drinks category] would be a little bit higher.

Without first-hand knowledge of the context, the amount of caffeine in specific beverages, the students were inadequately equipped (and perhaps less inclined) to take a critical stance toward the data. In contrast, the students who analyzed data for the video game task were able to name specific games that they all agreed were popular and would make a difference if included in the data set. Their conversation became quite lively when they began to consider how the data set could be altered. This was not the case with the caffeine task. Although most students acknowledged that different drinks have different amounts of caffeine, no one was able to identify the amount of caffeine in specific beverages. Alicia's argument was essentially hypothetical and she was not able to convince others (certainly not Gina) to change their thinking. Gina adhered to the data that were presented and argued that *all* sodas would have between 20 and 30 mg of caffeine, a conclusion that was based on her interpretation of the data presented. From a somewhat different perspective, although Bobbie recognized that energy drinks were intentionally high in caffeine, she did not take that into account when interpreting the data. Instead she referred to that fact as validation that the other drinks category would have more caffeine than sodas rather than challenging the dichotomy of "sodas" and "other drinks" as appropriate categories for comparison.

DISCUSSION

In many respects, the students in this study are like thousands of others in classrooms across the United States and in other countries as well; that is, discussion and argumentation is not a regular part of their mathematics instruction. None of the students in this study regarded their schoolwork critically. They accepted tasks we presented as mere exercises and saw their role as finding an answer to whatever question was posed, even when the solution process was as unreasonably simple (at least from our perspective) as adding up all the data values. In this way, the students perceived data analysis as nothing more than "school mathematics" (Nunes, Schliemann & Carraher, 1993).

Against this backdrop, I address the question that is the title of this paper: Does context expertise make a difference when dealing with data? I maintain that it does, and return to the quote by Moore (1990) in which he posits that context knowledge provides the capacity to understand and interpret data rather than simply carry out arithmetical operations. Although the two lesson excerpts presented in this paper cannot provide conclusive evidence regarding the role of context knowledge, they do illustrate how knowledge of the data context can support students in making informal inferences and taking a critical stance toward the data or task at hand. The findings of this study point to issues that must be addressed in the type of instruction we provide students. Firstly, data analysis must be approached as a process (Makar & Rubin, 2009) so that students learn to *interact with* data rather than merely apply arithmetical procedures to *act on* data. Knowledge of the data context can encourage and support such interactions. Secondly, teachers need to consider the generic nature of textbook data sets and determine the extent to which the context, or lack thereof, could inhibit students from engaging fully in the analysis of data. This is even more important for students who have not developed a critical disposition toward data and who would not automatically apply what Gal (2004) refers to as *worry questions* to evaluate the various aspects of the data and how they were collected.

AUTHOR NOTE

[1]A number of people were involved in the larger study from which the data for this paper were drawn. Members of the research team included: Edward Mooney, Josh Hertel, Kathy Abry, Jennifer Grandfield, and prospective teachers from Dr. Mooney's middle school methods class.

REFERENCES

Franklin, C., Kader, G, Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaffer, R. (2007). Guidelines for instruction and assessment in statistics education (GAISE) report. Alexandria, VA: American Statistical Association.

Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, *70*, 1-52.

Langrall, C. W., Nisbet, S., & Mooney, E. S. (2006). The interplay between students' statistical knowledge and context knowledge when analyzing data. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Association for Statistical Education. (CD-Rom)

Langrall, C. W., Nisbet, S., Mooney, E. S., & Jansem, S. (submitted). The role of context expertise when comparing data. *Mathematical Thinking and Learning*.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82-105.

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., & Carraher, D. W. (1993) *Street mathematics and school mathematics*. New York: Cambridge University Press.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17-47). Dordrecht: Kluwer.

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

Watson, J., Kelly, B. A., Callingham, R., & Shaughnessy, M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*(1), 1-29.