DESIGNING OPPORTUNITIES FOR STUDENTS TO REASON ABOUT THE RELATIONSHIP BETWEEN SOURCES AND STRUCTURES OF DATA

Jennifer Nickell and Hollylynne Lee
North Carolina State University, USA
jnickel@ncsu.edu

This paper presents a rationale for designing opportunities for students to reason about the relationship between sources of data and their resulting structures. Students completed a survey about their personal information). Responses were entered into TinkerPlots and attribute names were disguised (i.e., A, B, C, etc.). The students' task was to examine data from several attributes to analyze their structures and then propose which source (question) from the survey was the best match for a particular attribute. We describe how students reasoned and supported their arguments with a focus on how they used different perspectives in describing data as a pointer, case-value, classifier, and aggregate (Konold et al., 2004). Additionally, we propose a new perspective, data as measurement, used by students to make connections between the data and its source.

INTRODUCTION

"Statistics requires a different kind of thinking, because data are not just numbers, they are numbers with a context" (Moore & Cobb, 1997, p.801). Our aim was to design a task to allow students to engage with real data and reason within a context. When real data is incorporated into classrooms, students engage more with the context and draw conclusions based on the context (e.g., Garfield & Ben-Zvi, 2008; Lee, 2013). Furthermore the task was designed for students to reason about data that is personal to them since it can increase their desire to learn and their ability for thoughtful conclusions (Connor, Davies, & Holmes 2006). This paper represents an exposition and reflection from a teacher's perspective on designing and implementing the task, and reasoning about students' work.

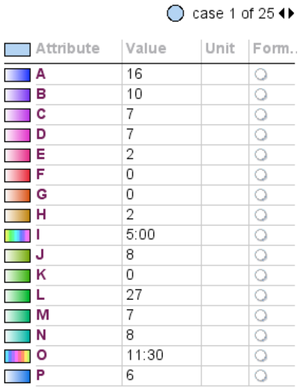
TASK DEVELOPMENT AND IMPLEMENTATION

The original task was published by Garfield and Ben-Zvi (2008) and was designed to assist introductory statistics students in developing an understanding that different statistical questions produce different types of variables. The original task had students complete a personal information survey, then each student was assigned an item and the item number was taped to their backs. Students circulated around the classroom collecting data for the item number on their back. Students had to draw a conclusion and guess their assigned survey item from their peers' responses to the item. We adapted this task for use with age 11-12 students as part of their statistics unit. This task was used with four classes, but examples in this paper are from a single class of 25 students that had field notes from two observers. Students' names used are pseudonyms.

Students were already familiar with writing statistical questions and constructing dot plots, bar graphs, and histograms to represent data. To begin the task, students completed a personal information survey, containing sixteen questions (e.g., "what time did you go to bed last night?"; "what is your shoe size?"). Questions were chosen intentionally to produce different structures of data (e.g., whole numbers, decimals, and time values). Survey responses were used to create a data set in *TinkerPlots* in which 16 attribute names were labeled as A, B, C, etc… rather than names such as "shoesize". The order of attributes was also randomized so as not to match the order of questions from the survey.

Several days after completing the survey, student pairs were seated with access to the original survey the file with their survey data (25 cases of 16 attributes) in *TinkerPlots* (see Figure 1). While examining the data cards, the teacher (first author) asked students what is unusual about the information in the cards. Students responded that the question numbers were not identified. This helped establish that students should not expect to see question numbers or variable names that matched certain questions from the survey. Each pair of students was assigned two attributes and asked to make a conjecture of which questions the data from these attributes most likely came from.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org



case 1 of 25

Attribute	Value	Unit	Form..
A	16		
B	10		
C	7		
D	7		
E	2		
F	0		
G	0		
H	2		
I	5:00		
J	8		
K	0		
L	27		
M	7		
N	8		
O	11:30		
P	6		

Figure 1. Stack of 25 data cards with first case displayed.

DESCRIPTION OF STUDENTS' REASONING

As we reflected on how students reasoned through the task of matching the structure of data to the source, we noticed they demonstrated reasoning that was consistent with Konold, Higgins, Russell, and Khalil's (2004) four different perspectives on data. However, students were using a different type of reasoning. Students also were using the type of number (e.g., rational or whole number) to reason about how such a number would make sense as a measurement of the context of a source of data. We are calling this perspective *data as measurement*. We conjecture that this *data as measurement* perspective may not have been explicitly described by Konold and others for several reasons. One being that in prior studies students may have acknowledged that the data values represent a measurement of some attribute when they explored a statistical question posed and collected data. However, this could have gone unnoticed when researchers have described students' reasoning. Another reason is that this perspective may be an artifact of how we adapted the task from Garfield and Ben-Zvi (2008) since the task is meant to invoke reasoning about the structure of a data given different contextual sources. Other tasks used by Konold and others did not have such a purpose and typically involved students examining data, either given to them or collected by them, representing 1-2 measurements from a context (e.g., see examples of tasks in Russell, Schifter, & Bastable, 2002)

Students' work on our task demonstrates how different perspectives on data can be used to reason about the structure and source of data and illustrates the interpretations they can draw when reasoning using each of these perspectives. For the focus of this paper, we chose to limit the discussion to students' work on six different attributes/questions. Figure 2 provides typical graphical displays we saw students create for these six attributes.

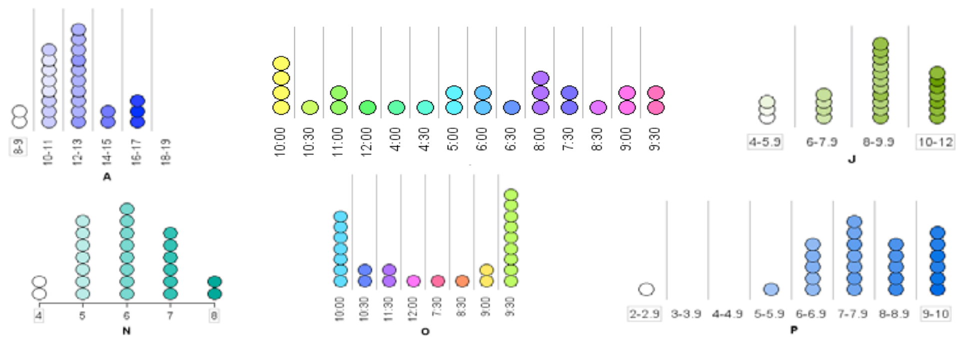


Figure 2. Students' plots for attributes A, I, J, N, O, and P.

Data as a Pointer

All students reasoned as the data as a pointer. This refers to students acknowledging the larger event that produced the data. In the case of this task, students demonstrated this reasoning when they recalled completing the survey earlier in the week during class and made connections that the 16 different attributes had to come from the 16 questions on the survey they took. They also reasoned about the number of cases in the data set (n=25) as matching the number of students in class. Reasoning as a pointer did not help students decide on the source of data but it was necessary to recall completing the survey to reason about data as a measurement or as a case-value. It helped students understand the basic structure of the data set with 16 attributes and 25 cases as a connector to the source of data.

Data as a Measurement

As discussed earlier, we noticed that almost all students engaged in reasoning about how the type of number within a particular attribute told them something about what was being measured in the data source (question posed on survey). Almost all students reasoned by considering data as a measurement to help them make a claim about the source of the attributes they were assigned. One approach students took was in examining attributes I and O. The data values for these two attributes were time measurements (Figure 1). All groups reasoned that these two attributes consisted of time measurements, so they narrowed down the source to two questions that would produce answers in the form of time: “what time did you go to bed last night?” and “what time did you wake up on last Saturday morning?” Another example of students demonstrating a data as measurement perspective was when students examined the type of number (e.g. rational or whole number). There were only two attributes, J and P, for which data had decimal values (Figure 2). All groups recognized that question “what is your shoe size?” could produce responses with decimal values. For example, one student noted that “it [the data values for J] has decimals and decimals are involved in picking your shoe size.”

Data as a measurement allowed students to narrow a focus to possible sources of data to a few questions from the survey. However, generally that perspective did not provide students enough information to reason about a specific source of the data. The survey questions were purposefully written so that several questions produced data with similar data types to ensure students would have to reason about data beyond an obvious measurement match to make strong arguments for claims about a match between the structure and source of an attribute’s data.

Data as a Case-value

It turned out that an effective perspective to reason about this task was data as a case-value. This refers to students locating their own case within the data set to reason about the source of an attribute. Instead of taking a graphical approach like many students in the class, one pair of students, Jamal and Teresa, decided to explore the data cards in *TinkerPlots*. They used the arrows at the upper right of the data cards to examine several cases in the “stack” of data cards for data values for attributes A and J (figure 1). In doing so, Teresa located a card that displayed values she believed matched the responses she had entered on the survey. She examined the values for each attribute and was convinced she had found her card. The teacher challenged Teresa asking her how she knew it was her card. To prove that it, Teresa showed the teacher several data values from the card that had convinced her, specifically the values of attributes she believed matched her responses to questions about the number of letters in her first name, day and month she was born, and her shoe size. The teacher then challenged them to try to locate Jamal’s case card. They believed they were successful in finding Jamal’s card. Teresa and Jamal were able to make claims about the source for 10 out of the 16 attributes before class ended. Looking at the attributes and questions in the collective through the lens of case-values by Teresa and Jamal was an unanticipated approach. This type of reasoning seemed to be an effective way to reason to solve this task. However, reasoning from a data as a case-value perspective would not have been possible if the data had not come from students.

Data as a Classifier

Another effective way to reason through this task was as a classifier, i.e., reasoning about the range of and frequencies of data values. Many groups took the approach as data as a classifier to make conclusions about their attribute. One example was from Stephen and Calvin examination of attribute N. They claimed that the question “how many letters are in your first name?” matched attribute N because “in class the other day we made a histogram about the letters in your first name and the graph was pretty similar." The teacher asked the students to explain what similar meant and they responded, "when we made the graph [in class] it went from 4-8 and most people had six letters just like this one [in *TinkerPlots*]." Although the connection to a prior experience with a similar question was useful, their attention to the range of values from 4 to 8 and noticing the frequency of 6 assisted them in sense making about a reasonable distribution for the letters in students' first names.

Data as an Aggregate

Students who reasoned about the aggregate properties of data described the spread and typical value to determine the data source. Kim and Kelly, when investigating attribute J, had already demonstrated reasoning as a measurement by identifying two possible questions for attribute J: "what is your shoe size", and "how many hours of sleep did you get last night? (round to the nearest half an hour)." Kim and Kelly decided that data from attribute J (figure 2) matched what they considered to be reasonable for hours of sleep because "some people don't get a lot of sleep and some people do. It seems that most kids get 8 to 9 hours of sleep because it says 8-9.9 has the most dots." Kim and Kelly's description that "some people don't get a lot of sleep and some people do" represents that the students were examining the range of the data. By identifying that "8-9.9 has the most dots" they were identifying the typical amount of time their classmates sleep.

Even though it was not necessary for the students to reason about data as an aggregate to complete this task, such reasoning allowed students to identify the source of data. This type of reasoning would not have helped reasoning about all of the attributes. One example is determining attribute A; the range is from 8 to 17 with a typical value of 12 to 13. There are several questions on the survey that could produce that structure of data such as: "how many letters in your first and last name?" or "how many books have you read this year?" Also, students can develop the ability to reason about the aggregate without making connection to the context.

CONCLUSION

Our experience with this task emphasizes the importance of providing students opportunities to explicitly reason about data as a measurement of a context, including anticipation of reasonable values. Since teachers often use pre-collected data, a focus on the *data as measurement* perspective may be helpful before asking students to draw conclusions from data in a context. The observation of students ways of reasoning have led to research on how pre-service mathematics teachers reason about sources and structures of data in a similar task.

REFERENCES

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801-823.

Connor, D., Davies, N., & Holmes, P. (2006). Using real data and technology to develop statistical thinking. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: 2006 NCTM Yearbook* (pp. 185-194). Reston, VA: NCTM.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2004). Data seen through different lenses. Unpublished manuscript. http://www.srri.umass.edu/sites/srri/files/KonoldEtal2004.pdf

Lee, H. S. (2013). Quantitative reasoning in a digital world: Laying the pebbles for future research frontiers. In R. L. Mayes & L. L. Hatfield (Eds.), *Quantitative reasoning in mathematics and science education: Papers from an international STEM research symposium, WISDOM e Monograph #3* (pp.65-82). Laramie, Wyoming: University of Wyoming College of Education.

Russell, S. J., Schifter, D., & Bastable, V. (2002). *Working with data: Casebook.* Parsippany, NJ: Dale Seymour Publishing.