Learning to Use Appropriate Tools Strategically for Bivariate Data Analysis

By: Matthew C. Jones and Mary C. Knaub **Mentor:** Dr. Randall Groth

Introduction:

Middle school students face several challenges in learning to analyze bivariate data. Such analysis is a complex cognitive process that involves:

-Viewing the data as an aggregate rather than just employing case-based reasoning (Bakker, 2004)

-Using context knowledge in ways that support, rather than interfere, with statistical reasoning (Batanero et al., 1996)

-Discerning the direction of trend for data displayed in scatterplots (Cobb, McClain, & Gravemeijer, 2003)

-Coordinating mathematical and statistical ideas in ways that support rather than interfere with one another (Casey, 2015)

Then Common Core State Standards for Mathematics (CCSSM, Common Core State Standards Initiative, 2010) require Grade 8 students to master these cognitive processes. This poses a significant teaching challenge.

Purpose:

The purpose of this study was to investigate how students analyze bivariate data and then design an instructional sequence effective for helping them attain related CCSSM outcomes.

Research Question:

What are the characteristics of an instructional sequence to help students use appropriate tools strategically when analyzing bivariate data?

Theoretical framework:

The Five Strands of Mathematical Proficiency

are defined as five inter-dependent elements "… necessary for anyone to learn mathematics successfully" (Kilpatrick, Swafford, & Findell, 2001, p. 116):

Conceptual Understanding
- Comprehension of mathematical concepts, operations, and relations.

Procedural Fluency
- Skills in carrying out procedures flexibly, accurately, efficiently, and appropriately.

Strategic Competence
- Ability to formulate, represent, and solve mathematical problems.

Adaptive Reasoning
- Capacity for logical thought, reflection, explanation, and justification.

Productive Disposition
- Habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy.

- The five strands help us set goals for student learning.
- A learning progression formulated by Maloney et al. (2004) helped us make conjectures about the optimal sequence for learning goals.
- Practice-based articles (e.g., Kroon, 2016) and previous research (e.g., Casey, 2016) helped us select tasks, tools, and contexts to use to meet our instructional goals.
- For example, a spaghetti bridge building experiment described by Kroon (2016) provided an engaging bivariate data context for our initial lessons.

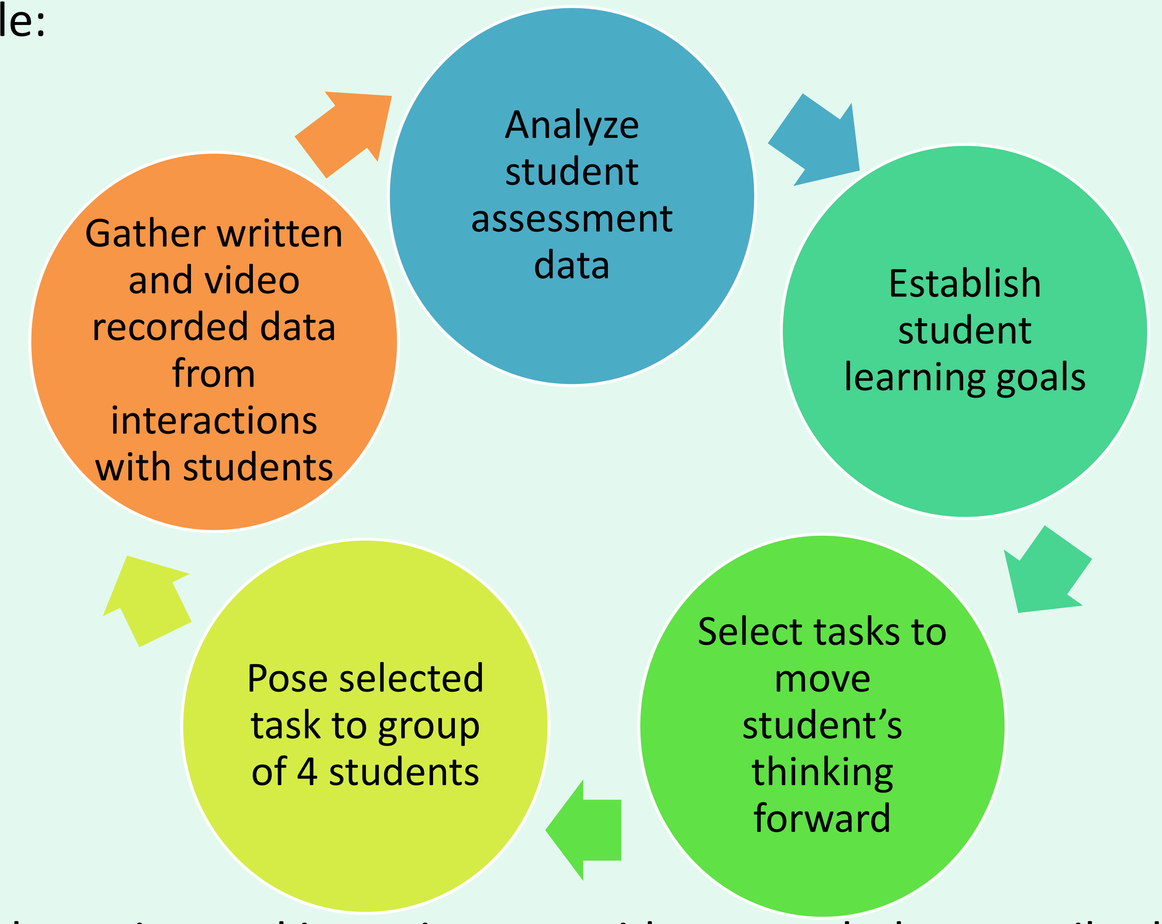
Methodology

Participants:

Four students participated, two male and two female (pseudonyms: Tom, Nancy, Nick, and Kate). All four had completed seventh grade and were preparing to enter eighth. Students participated in seven one-hour sessions as well as 30-minute pre- and post-interviews (one student missed one teaching session).

Lessons:

A sequence of seven one-hour lessons was designed to develop students' abilities to use appropriate tools strategically to analyze bivariate data. We followed an iterative research and development cycle:



Each session and interview was video recorded, transcribed, and analyzed qualitatively to reveal students' strengths and weaknesses along the Five Strands of Mathematical Proficiency.

Methodology

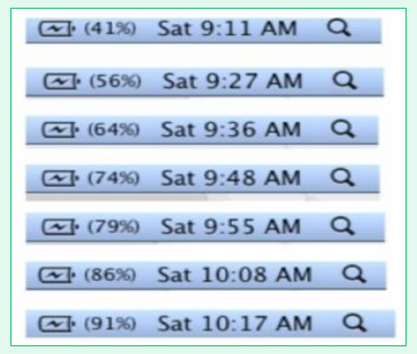
Data Gathering and Analysis:

Pre- and post-assessment interviews were conducted with each student. Interviews lasted 30 minutes and were video recorded. Key items were:

Jerry forgot to plug in his laptop before he went to bed. He wants to take the laptop to his friend's house with a full battery. The pictures below show screenshots of the battery charge indicator after he plugs in the computer at 9:11 a.m.

When can Jerry expect to have a fully charged battery? Justify your answer.

Key Interview Task 1



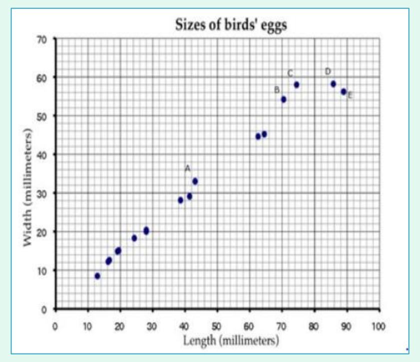
Source: Illustrative Mathematics (modified): https://www.illustrativemathematics.org/content-standards/8/SP/A/2/tasks/1558

This scatter diagram shows the lengths and widths of the eggs of some American birds.

a. A biologist measured a sample of one hundred Mallard duck eggs and found they had an average length of 57.8 millimeters and average width of 41.6 millimeters. Use an X to mark a point that represents this on the scatter diagram.

b. What does the graph show about the relationship between the length of birds' eggs and their widths?

Key Interview Task 2



Source: Illustrative Mathematics (modified): https://www.illustrativemathematics.org/content-standards/8/SP/A/1/tasks/41

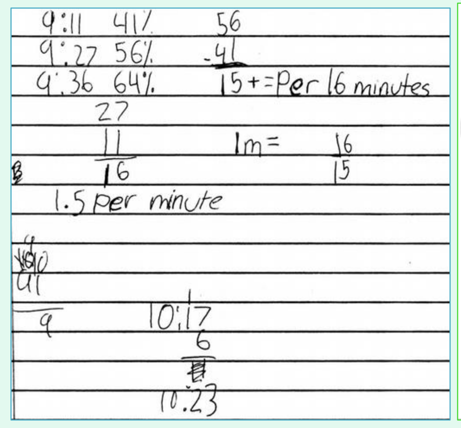
Empirical Teaching and Learning Trajectory:

Initial Assessment Results:

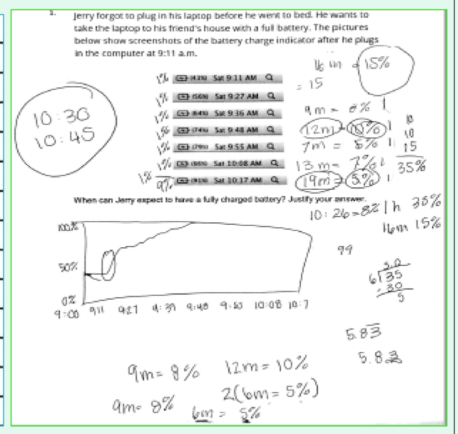
For key interview task 1, none of the students used a scatterplot or best fit line; therefore, it became our initial instructional goal to help the students see the usefulness of these tools.

-Data from the first key task:

Tom: I see kind of a pattern with the big number and a smaller number, big number, smaller number.
Interviewer: Okay, so you see a pattern there in between the times in the screenshots?
Tom: Not really a pattern but different...
Interviewer: You're seeing a big number, a smaller number, a bigger number, smaller number?
Tom: Yeah.
Interviewer: Okay, so does that help you at all to figure anything out for what time it will be charged?
Tom: Not really.



Kate looked for a pattern and tried to determine how much the battery charged per minute on average.



Nancy looked for a pattern and tried to determine how much the battery charged per minute on average.

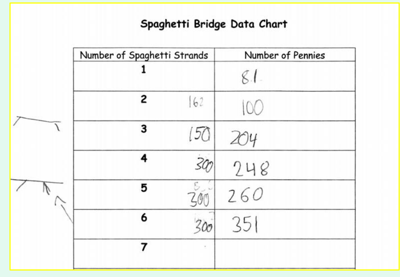
When responding to key task 2, three of the four students struggled to define the bivariate relationship shown in the scatterplot. This was an important concept we decided to build during the first lessons.

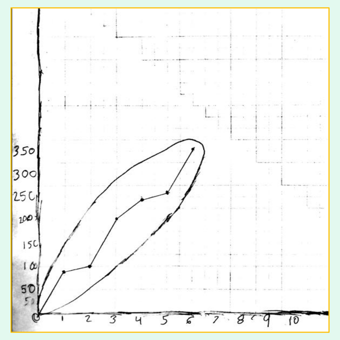
-Summary of data from key task 2:

Both Nick and Nancy misread the scatterplot and plotted points by reversing the coordinates. Nick said the relationship shown was that the length was greater than the width; whereas, Nancy defined it as an average and that it showed how some [eggs] were bigger than others. Tom claimed that he did not see a relationship. Kate correctly plotted the point for part a. She stated that the relationship shown was as the [egg] gets longer it gets wider.

Gathering and Representing Data (Weeks 2-3)

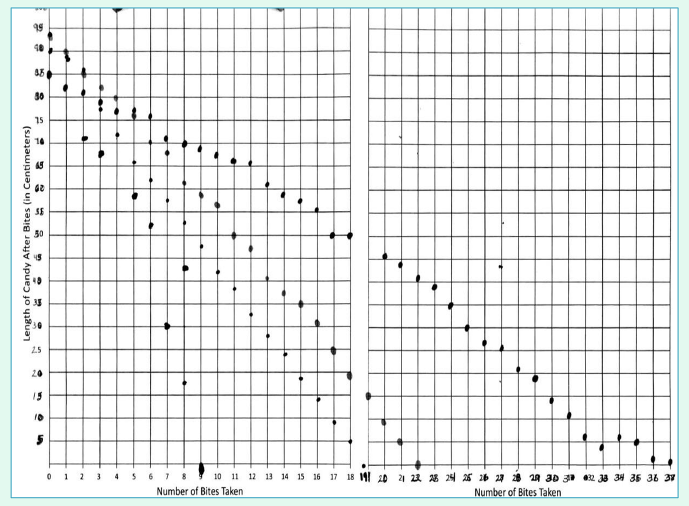
These lessons focused on the students gathering their own data. Students used spaghetti strands, pennies, and a cup to test how many pennies it took to break different numbers of strands bunched together. Students were then asked to represent the data in a manner they believed to be best. Students struggled representing the data, so at the beginning of the second lesson, they were shown a scatterplot displaying the same context with different data points. By seeing this, students then were able to see the table as coordinates to plot within the scatterplot. They each were given then task to create their own scatterplot using the data they had gathered. Although they successfully created scatterplots at this stage, they had difficulty discerning linear trends in the data. In the student work shown below, for example, they connected the dots rather than focusing on linear trend.





Positive and Negative Associations (Weeks 4-5)

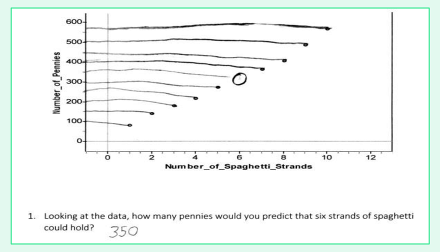
These lessons focused on helping students see linear patterns in relationships between variables. Students were given Fruit-by-the-Foot and took normal bites. They graphed the relationship between number of bites and amount left (all four student data sets are shown on the same graph below).



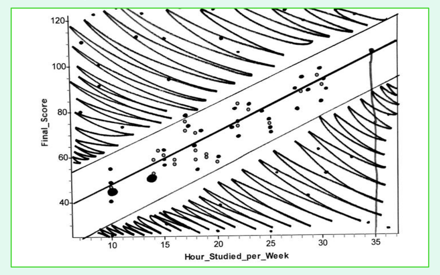
When analyzing their data, students began to discuss linear trends, noticing how some of their lines were steeper than others. This also gave them their first experience with a negative relationship between variables, allowing us to begin to discuss contexts that yield both positive and negative linear relationships.

Strategically Using Tools to Make Predictions (Weeks 6-8)

These lessons focused on predictions from bivariate data sets. One prevalent strategy students initially used to make predictions was to extend horizontal or vertical lines to individual data points, as shown below:

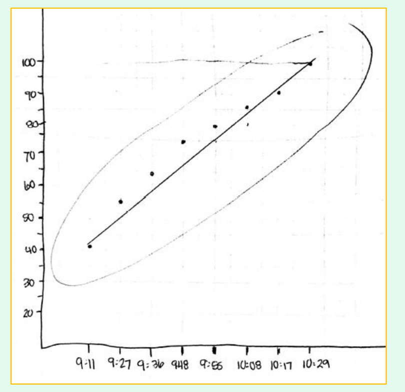


To help students see trend lines as useful tools, we used a data set on time spent studying vs. test score received. We asked them to plot unreasonable estimates for scores you would receive for different numbers of hours of studying. After plotting several points, they shaded the "unreasonable region" and sketched a trend line through a series of reasonable estimates we asked them to make, as shown below:

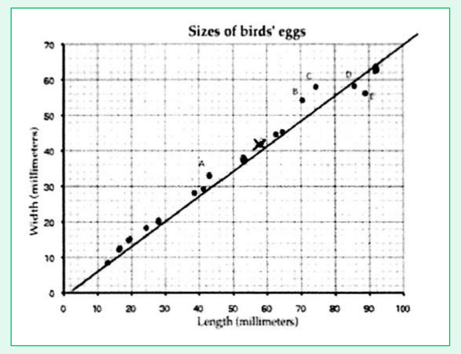


Post-Assessment Results:

For key interview task 1, students used various strategies to accurately predict when the battery would be fully charged. One student constructed a scatterplot and used an informal line of best fit in order to help make the prediction the most accurate. Other students used various strategies such as trying to find a pattern within the data set, but seemed to only focus on one variable.



When responding to key task 2, three out of the four students were able to define the positive relationship shown, and informally placed a line of best fit through the data to assist them in predicting. However, only two of the students used a line of best fit across all interview contexts requiring predictions. Although students generally used the tool as intended, work remained to help them recognize its usefulness across multiple situations.



Reflection and discussion: It was evident that some of the CCSSM Standards were challenging to attain. Students could construct scatterplots and interpret relationships in bivariate data. At the end of the study, however, they struggled to an extent deciding when to use the scatterplots. Our findings indicate that students need experiences across multiple contexts to begin to use trend lines as data analysis tools. Our task sequence suggests that optimal engagement occurs when students deal with data they have collected themselves and when problem contexts are carefully selected by attending to students' thinking.

References

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64 – 83.

Batanero, C., Estepa, A., Godino, J., & Green, D. (1996). Intuitive strategies and preconceptions about association in contingency tables, *Journal for Research in Mathematics Education*, *27*(2), 151-169.

Casey, S.A. (2015). Examining student conceptions of covariation: A focus on the line of best fit. *Journal of Statistics Education*, *23*(1). Retrieved from www.amstat.org/publications/jse/v23n1/casey.

Casey, S.A. (2016). Finding what fits: Students explore six tasks to develop criteria for finding an informal line of best fit. *Mathematics Teaching in the Middle School, 21* (8), 484-491.

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers.

Cobb, P., McClain, K., & Gravejeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1-78.

Kilpatrick, J. Swafford, & Findell, B. (Eds.) (2001). Adding it up: Helping students learn mathematics. Washington, DC: National Academy Press.

Kroon, C. D. (2016). Spaghetti bridges: Modeling linear relationships. *Mathematics Teaching in the Middle School, 21* (9), 563-568.

Maloney, A.P., Confrey, J., Ng, Dicky, & Nickell, J. (2004). Learning trajectories for interpreting the K-8 Common Core State Standards with a middle-grades statistics emphasis. In K. Karp (Ed.), *Annual perspectives in mathematics education: Using research to improve instruction* (pp. 23-33). Reston, VA: National Council of Teachers of Mathematics.