Learning About Statistical Distributions and Typical Values in Sixth Grade

Authors: *Nicholas Bennett, Rachel Dean*
Faculty Mentor: *Dr. Randall Groth*

Introduction

Organizing data and determining typical values are not trivial tasks. In order to do so, students must develop graph sense (Friel, Curcio, & Bright, 2001), view data represented in a graph as an aggregate (Bakker & Gravemeijer, 2004), and avoid strictly procedural conceptions of measures of center (Mokros & Russell, 1995).

In this study, we aimed to learn how students organize quantitative data, perceive aggregates, and discern typical values, and to design an instructional sequence to help these statistical thinking processes develop.

The research question guiding our study was, "How does a group of children entering sixth grade organize data, perceive aggregates, and determine typical values before, during, and after instruction?"

References

Bakker, A., & Gravemeijer, K.P.E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht: Springer.
Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education, 29*(1), 41-62.
Boaler, J., & Humphreys, C. (2005). *Connecting mathematical ideas: Middle school video cases to support teaching and learning*. Portsmouth, NH: Heinemann.
Curcio, F.R. (2010). *Developing data graph comprehension* (3rd ed.). Reston, VA: National Council of Teachers of Mathematics.
English, L.D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics*, *81*(1), 15-30.
Friel, S.N., Curcio, F.R., & Bright, G.W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124-158.
Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2014). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305-325.
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.
Mokros, J., & Russell, S.J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20-39.
Noll, J., & Kirin, D. (2017). TinkerPlots model construction approaches for comparing two groups: Student perspectives. *Statistics Education Research Journal, 16*(2), 213-243.

Literature Review

Problem-based instruction can help students develop conceptual understanding and deal with unfamiliar and complex problems (Boaler, 1998; Boaler & Humphreys, 2005). As students begin studying statistics, appropriate tasks to pose include those that involve organizing data (Curcio, 2010), describing aggregate features of distributions (Konold et al., 2014), comparing groups (Noll & Kirin, 2017), and making predictions from data (English, 2012). These types of tasks lay a foundation for students to make data-based inferences and decisions.

We used the Common Core Learning Progression Model (Maloney, Confrey, Ng, & Nickell, 2014) to structure our lessons and set learning goals for students. Our initial goals were to address portions of standards 6.SP1-6.SP.5 in the learning progression.

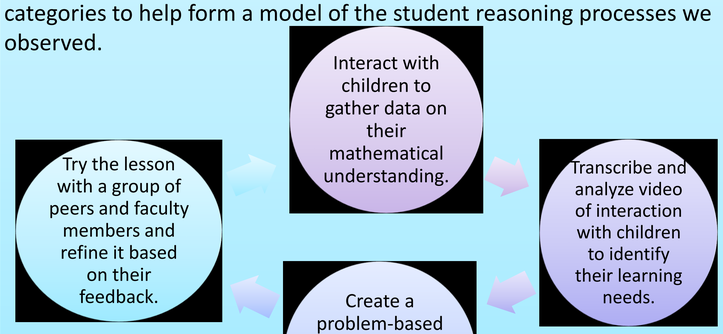
- 6.SP.1 – Recognize statistical questions as associated with variability in the data.
- 6.SP.2 – Understand distributions of data as described by center, spread, and overall shape.
- 6.SP.3 – Recognize measures of center and measures of variation as single values representing data sets.
- 6.SP.4 – Display numerical data in plots on a number line (dot plots, histograms, and box plots)
- 6.SP.5.a – Report the number of observations.
- 6.SP.5.b – Describe attribute, how it's measured, and units of measurement.
- 6.SP.5.c – Describe measures of center (median, mean) and variability (interquartile range, mean absolute deviation) in relation to context.
- 6.SP.5.d – Relate choice of measures of center and variability to shape of the data distribution.

Methodology: Participants

Four students, two male and two female, who had completed fifth grade and were advancing to sixth grade participated in the study. The pseudonyms assigned to the students were Allison, Andrew, Brian, and Claire. Three of the students attended all sessions; one of the students missed two lessons.

Instructional Cycle

Over nine weeks, we utilized the cycle below to analyze each student's reasoning to identify the learning needs of the group and design seven weekly problem-based lessons. We began the cycle by completing individual pre-interviews with students and ended with individual post-interviews. After each interview and lesson, we collaboratively analyzed the data, assigning codes to describe students' reasoning about ideas relevant to the intended learning progression. We clustered codes into categories to help form a model of the student reasoning processes we observed.



Methodology: Key Interview Items

Key Item 1:
Below are the 25 birth weights, in ounces, of all the Labrador Retriever puppies born at Kingston Kennels in the last six months.
13,14,15,15,16,16,16,16,17,17,17,17,17,17,17,18,18,18,18,18,18,18,18,19,20

a. Use an appropriate graph to summarize these birth weights.
b. Describe the distribution of birth weights for puppies born at Kingston Kennels in the last six months. Be sure to describe shape, center, and variability.
c. What is a typical birth weight for puppies born at Kingston Kennels in the last six months? Explain why you chose this value.

URL: https://www.illustrativemathematics.org/content-standards/tasks/1026

Key Item 2:
The table below shows the number of customers at Malcolm's Bike Shop for 5 days, as well as the mean (average) and the median number of customers for these 5 days.

Number of Customers at Malcolm's Bike Shop	
Day 1	100
Day 2	87
Day 3	90
Day 4	10
Day 5	91
Mean (average)	75.6
Median	90

Which statistic, the mean or the median, best represents the typical number of customers at Malcolm's Bike Shop for these 5 days?
Explain your reasoning.
NAEP Question ID: 2007-8M9 #8 M073501
URL https://nces.ed.gov/nationsreportcard/nqt/

Empirical Teaching and Learning Trajectory:

Pre-Interview Results

Key Item 1:

For Key Item 1 Brian decided to create a line plot with the data given. Claire and Allison also made a line plot; Andrew made a graph that resembled a histogram. When asked to determine typical value, Claire, Andrew, and Allison focused on the tallest stack of data; Brian did not choose a specific value in the graph.

Key Item 2:

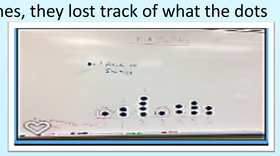
For Key Item 2, Andrew and Brian relied solely on context knowledge, and not the data distribution (see interview excerpt to the right). Claire chose the highest number in the data set. Allison chose the median because it was in close proximity to many data points.

Andrew: (thinking) I think there would be 87 average for the customers coming.
Interviewer: Okay 87, so why do you say 87?
Andrew: Because bike stores are kind of popular
Interviewer: They're kind of popular? Especially around here, yeah?
Andrew: (nods)

Lessons 1-3: Organizing and Interpreting Data

Lesson 1; How Many Red Skittles?:

- The group was able to create a dotplot together.
- Students were inconsistent in reading values from the dotplot; at times, they lost track of what the dots represented.



Lesson Two; Typical Number of Siblings:

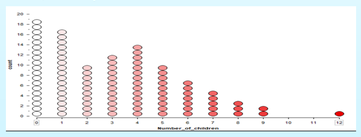
- After collaboratively creating a dotplot from Post-It notes, students were more consistent in reading dotplot data values.
- When asked to analyze the data and determine typical values, students tended to rely strictly on context knowledge rather than data, or to focus only on the tallest stack.

Lesson Three; Comparing Typical Numbers of Different Colored Candies in Skittles Packs:

- When asked to use dotplots showing the numbers of different colored Skittles in several packs, students associated typical values either with the tallest stacks or some other stack in the plot.
- When asked if there were typically more yellow or purple Skittles in a pack, Claire focused on how many packs had 4 purple or yellow Skittles, even though 4 was not beneath the tallest stack in the dotplot.

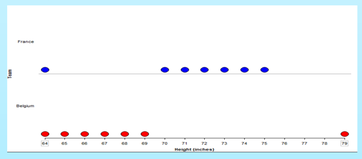
Lessons 4-5: Discerning Aggregate Characteristics

Lesson Four; Family Data in TinkerPlots:



- Students were asked to use informal words like 'mountain', 'cluster', and 'gap' to describe graph features.
- Students began to move past focusing only on stacks to consider aggregate features comprised of multiple stacks and gaps in the data; for example, given the word 'hillside', Brian circled multiple stacks to describe it.

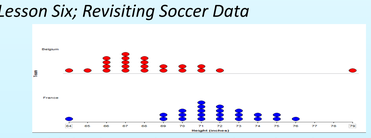
Lesson Five; Soccer Data:



- Students were asked to compare the heights of soccer players on two different teams.
- The students started to identify outliers as points separated from most of the data by a gap.
- The students shifted from thinking only about vertical stacks to grouping horizontally across data points to identify clusters.

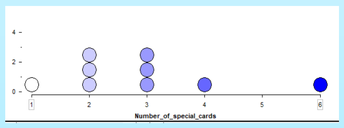
Lessons 6-7: Using Aggregate Characteristics to Make Comparisons and Predictions

Lesson Six; Revisiting Soccer Data



- Multiple-size stacks were reintroduced to students.
- Students successfully relabeled the soccer graph using the words "outlier," "cluster," and "gap".
- Brian, Allison, Andrew were able to identify typical values as those that were within the main cluster of data.

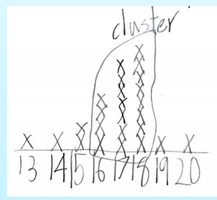
Lesson Seven; Pokémon Data in TinkerPlots



- Students were introduced to new statistical terms, "mean" and "median", to determine typical value.
- Students compared what they thought was typical with the median.
- Students identified that the typical value is usually located inside of the main cluster of data.
- All students recognized that multiple values in the main cluster could potentially describe typical value.

Post-Interview Results

Key Item 1:



Andrew decided to create a graph that resembled a histogram (top left work sample), while the other students chose to create line plots. Students identified some aggregate features such as central clusters (top right work sample). All four students reasoned that the typical value was shown by the tallest stack of data.

Key Item 2:

- In contrast to the pre-interview, Andrew and Brian were able to look at the data to choose the median as the typical value because of its proximity to many data points instead of relying solely on context knowledge.
- Allison and Claire used the same strategies for this item as they did during their pre-interviews.

Reflection and Discussion

During our study, we found that one of the most difficult reasoning skills for students to develop was to identify the main cluster of data in a distribution. Our students' idea of the main cluster was initially limited to the tallest modal stack, but they did begin to focus on multiple stacks after two lessons. We spent most of our study helping students identify central clusters that contained typical values. Near the end, we were able to address the portion of our learning progression that dealt with recognizing measures of center as single values representing data sets. Our experience suggests that teachers should not rush to introduce formal measures of center such as mean and median. It is important for students to identify aggregate features of distributions, such as central clusters, and then begin to associate them with formal measures of center. Doing so puts students in a better position to learn concepts such as mean and median with understanding.