AN ANALYSIS OF K-8 PRE-SERVICE TEACHERS AS DATA STORYTELLERS

Heather Barker and Emily Elrod
Elon University, North Carolina, USA
hbarker2@elon.edu

A required course at our university, Geometry and Data for K-8 teachers, focused primarily on geometry. A course redesign was completed to place more emphasis on the Data and Statistics Unit. This study seeks to understand how this change impacted the way pre-service teachers perceive real data and how they might use data in their own classrooms. Participants created a data story using an online statistical software tool, CODAP. The data stories were analyzed to determine how the tool and focal statistics concepts were utilized to present data to a hypothetical audience of students. Participants also took a pre- and post-test to determine changes in levels of statistical conceptual understanding. These results show a statistically significant increase in conceptual understanding of statistical topics. Findings from the analysis of the data stories created highlight the preservice teachers' abilities to develop data stories to guide students through real and messy data explorations.

INTRODUCTION

In a world that is constantly collecting data on its people, whether it be through tracking of online spending habits, health data, or outcomes of elections, it has become increasingly important that students are statistically literate consumers of data. This increasingly data driven world has led many in education to realize the importance of including more statistics standards in K-12 curriculum so that students are covering statistical standards as early as first grade. This inclusion of new standards began in 2010 in the Common Core Standards (CCSSM, 2010) and was also adopted by many states, including the standards in our state, which are followed in preparing pre-service teachers (PSTs) at our university. Unfortunately, this shift to teaching more statistics standards has led to PSTs who do not feel confident enough to educate students in statistics topics (Lovett & Lee, 2018).

The Statistical Education of Teachers (SET) Report (Franklin, et al., 2015), outlines recommendations for preparing prospective teachers at all levels. For prospective elementary school teachers, the SET report recommends that coursework include an introductory section of statistics geared toward elementary education, an entire course on statistical content, and more attention given to statistics in existing mathematics content. In addition to these recommendations, prospective middle school teachers should take a second course strengthening their conceptual understanding of ideas in the middle-school curriculum. The SET report outlines the goals of statistical preparation for elementary- and middle-school teachers to include: 1) learn statistics and reasoning skills for teaching through meaningful problem-solving experiences, 2) understand the connection between statistical concepts in elementary and middle grades and their relation to other subjects, and 3) develop pedagogical content knowledge for effective statistics teaching.

Using current technology tools to teach statistics is essential in promoting statistical thinking in students (Biehler et al., 2013; Pratt et al., 2011). Research has shown that students who engage in using real data and dynamic statistical software packages, instead of calculators, show not only better understanding of concepts, but also more enthusiasm during statistical investigations (Lee & Hollebrands, 2008, 2011; Pratt et al., 2011). Dynamic statistical software utilizes a drag and drop environment where students can take data that is often found in a table and drag and drop it onto graphs or other visual displays (Finzer, 2000). The participants of this study use CODAP (Common Online Data Analysis Platform), a free, online dynamic statistical software tool (CODAP, n.d.).

THEORETICAL PERSPECTIVES

Many perspectives have been used to analyze statistics and data analysis tasks. Here we used two different theoretical perspectives to describe data investigations shared by the PSTs, the data investigation process framework (Lee et al., 2022), and data moves (Erickson et al., 2019).

Data Investigation Process Framework

Many frameworks for engaging in the statistical investigation cycle have been shared over the past decades. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) framework (Bargagliotti et al., 2020) outlines four phases of a statistics investigation cycle: pose questions, collect data, analyze data, and interpret results. These four phases are also emphasized in the SET report (Franklin, et al., 2015). Building on the work of frameworks such as these, as well as using findings from research in data investigations from those in data science, Lee et al. (2022) proposed a *Data Investigation Process Framework* that includes six phases. These phases include framing the problem, considering and gathering data, processing data, exploring and visualizing data, considering models, and communicating and proposing action. These six phases can be visualized as pieces of a puzzle that come together to build the data investigation process, instead of a cyclic or linear process. The phases can occur in any order and phases can be revisited throughout the process to work together, like puzzle pieces, to build the data investigation.

Data Moves

Erickson et al. (2019) use CODAP as a tool to define data moves students can use to engage in statistical analysis. Erickson et al. (2019) define data moves as "an action that alters a dataset's contents, structure, or values" (p. 4). The data moves they define are filtering, grouping, summarizing, calculating, merging/joining, and making hierarchy. *Filtering* produces a subset of data, including removing extraneous data. *Grouping* is used to set up a comparison among different subgroups, which can include *binning* that occurs with continuous variables. *Summarizing* is the process of producing an aggregate value or statistic. *Calculating* is creating a new attribute in the table or data from existing data. *Merging* is when analysts join different tables or datasets into one and *joining* adds attributes to existing cases from another dataset. *Making a hierarchy* occurs when a grouping is created over the entire dataset, giving a large view of the organization of the data within the table. Erickson et al. (2019) propose that data moves can be used as a supplement for other frameworks within statistics education.

Building off of these frameworks and prior work, we attempt to answer the following research questions: 1) What impact does an increased inclusion of statistics topics in a mathematics content course for K-8 PSTs have on the PSTs' understanding of statistical concepts?, 2) Which phases of the data investigation process framework are present in data stories created by K-8 PSTs?, and 3) What data moves do K-8 PSTs use when creating a data story?

CONTEXT

Course Description

The participants of this study were enrolled in a course titled *Geometry and Data for K-8 Teachers*, one of two mathematics content courses required for prospective elementary and middle school teachers. The prerequisite for this course is either an introductory statistics course or Calculus 1. The topics for these courses include those that are outlined in the state mathematics standards. Those standards are grouped by five main topics, two of which are covered in this course, *Measurement and Data* and *Geometry*. There are more Geometry and Measurement standards in kindergarten through eighth grade, but with the increasing push to include statistics into the curriculum, statistical understanding (titled *Data)* is introduced starting in the first grade and remains in each grade level through eighth grade within our state's mathematics content standards.

The first author taught this course in Fall 2022 and Winter 2023, redesigning the course content to address more statistics standards. In prior offerings of this course, of the 41 course outcomes listed, only three addressed statistics. This resulted in statistics being covered for approximately one week in a 14-week semester. The Winter semester is a three-week term where students take only one course for three hours a day. Each day translates to roughly one week of a 14-week semester. The course redesign included three weeks of statistics content (with an additional day for basic probability), while deleting three course outcomes related to geometry that are no longer part of the current K-8 state mathematics standards. The statistics instruction emphasized the statistics investigation cycle included in the GAISE framework and the use of technology to learn statistics.

CODAP Story Builder Plug-In

Within CODAP, there is a plugin called Story Builder. The Story Builder plug-in was created as part of the *Writing Data Stories* project (Wilkerson, et al., 2021). The goal of the tool was to provide an "infrastructure for students to a) integrate their data investigations with relevant motivational and contextual information; b) record the step-by-step process and rationale of their data analysis; and c) reflect on and share the contextual and process-oriented elements of data analysis work with others" (Wilkerson et al., 2021, p. 504). The plug-in creates what is called a "moments bar", where each moment acts as a page for the user to show what they have done with the data such as creating graphs, inserting text boxes with explanations, and making changes to the dataset(s). Example of PST created moments will be shared in the results section. The *Writing Data Stories* project uses the tool to analyze middle grades students' data stories (Wilkerson et al., 2021). Our project will instead target PSTs preparing to teach elementary and middle grades students.

Participants

The participants for this study consisted of 44 PSTs enrolled in two sections of a mathematics content course for elementary and middle grades majors. In Fall 2022, there were 12 students enrolled in the course. In Winter 2023, there were 32 students enrolled in the course. The students ranged from first year to fourth year students. Though this course is meant for both students who are completing licensure requirements for elementary and middle grades, one was a middle grades licensure student and 43 were elementary education majors. (The elementary grades licensure for this state includes grades K-6, and middle grades licensure is grades 6-9.)

METHODOLOGY

After receiving IRB approval at our institution, the following data were collected with the expressed permission of participants. At the beginning and end of the statistics unit, students took the Levels of Conceptual Understanding of Statistics (LOCUS) assessment (Jacobbe et al., 2023) as a pre- and post-test to gauge the impact of the unit. The LOCUS assessment questions were Beginning/Intermediate level meant to assess conceptual understanding of statistics consistent with the GAISE framework.

The final week of the unit on data, students used CODAP to create a "data story". Students were tasked with choosing a dataset of interest from the website CORGIS: The Collection of Really Great, Interesting, Situated Datasets (Bart et al., 2022). Participants were tasked to create a data story, using the Story Builder plug-in, that would be appropriate to engage middle grades students in a data exploration. Students were required to include at least 5 moments in their data story. Students were assessed based on an introduction to their dataset, introducing 2-3 statistics that could be used to describe the data, 2-3 graphics to share their viewpoints, questions that could help their audience think deeper about the data, and the clarity and appropriateness of this data for a middle-grades audience. Data stories were created in groups of two to three students, for a total of 17 data stories (7 from Fall 2022 and 10 from Winter 2023). Students shared their data stories via a URL link created in CODAP.

The LOCUS pre and post results as well as the data stories were blinded using pseudonyms. Dedoose (a qualitative software tool) was used to qualitatively code the data stories using the frameworks presented above. To code the data stories, screenshots of each moment were captured, making sure to include key parts of each moment (such as when a new variable was created in a table). The two authors coded each data story then met to check for inter-rater reliability, discussing any discrepancies in coding until a consensus was reached.

RESULTS

LOCUS Results

There was a total of 42 participants who completed both the pre- and post-test. The mean pre-test score for the 42 participants out of 100 points was 55.45 (SD 16.63) and the mean post-test score was 60.83 (SD 17.10), with a mean difference of 5.38 (SD 16.01). These results indicate that there was a significant increase in conceptual understanding of statistical topics before and after the statistics unit, $t(41) = 2.17$, $p = 0.0179$.

Data Stories Analysis Example

As an example, we present a data story with explanations of the codes that were applied. Then a summary of the codes for all data stories according to the data investigation process framework and data moves will be shared. Figure 1 shows a complete data story exploring data about astronauts' missions in space. The data story includes 5 moments; the highlights of each are shown in Figure 1.

Figure 1. A complete data story about a dataset containing missions of astronauts.

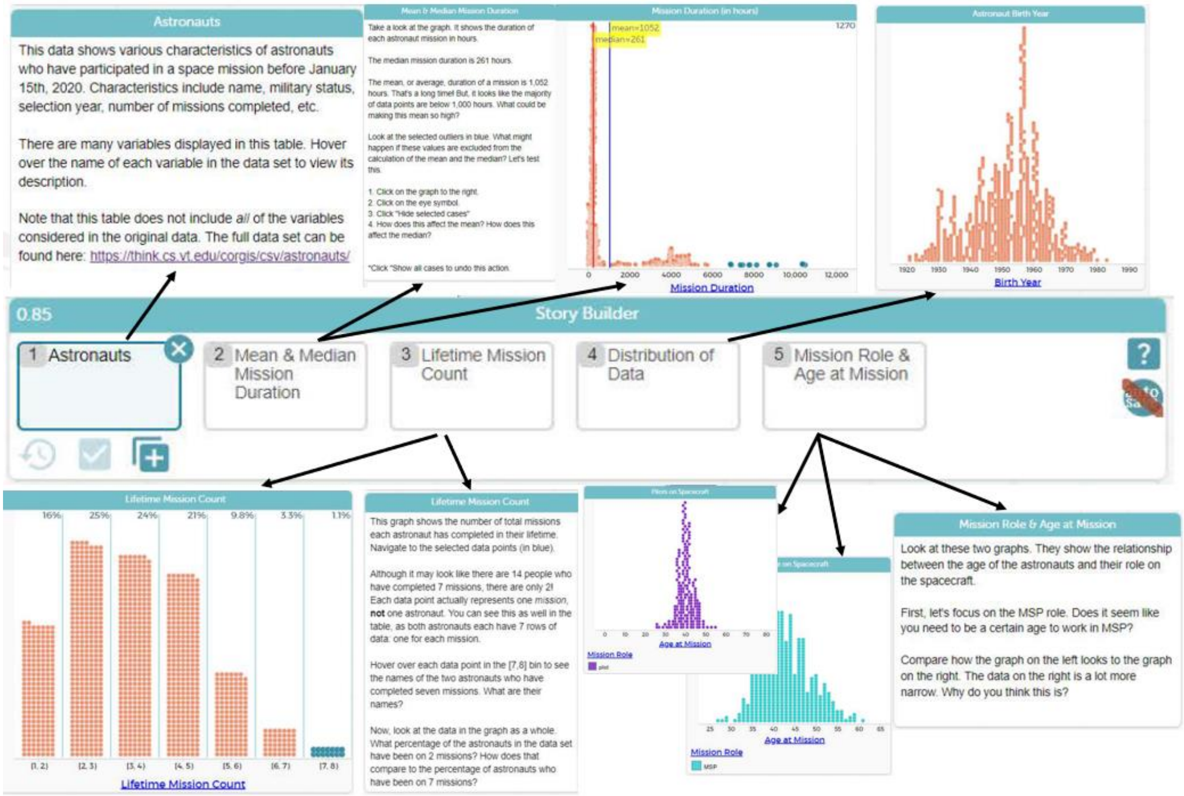


Figure 1 includes pieces of the data investigation process framework. In Moment 1, the PSTs are asking the user to *Consider and Gather Data* by clicking on variables to read their description. Throughout the moments the PSTs are *Exploring and Visualizing Data* by creating graphs and asking the user to identify patterns, statistics, or shapes in the data. In Moment 5, the PSTs are *Framing a Problem* by asking the user to consider why there may be a difference in age based on mission role. Several data moves were coded here as well. In Moment 2, users are asked to *filter* the data by hiding outliers as well as *summarizing* the data by finding a statistic. Moment 3 shows an example of *grouping* done by binning data. (Note: not all examples of codes were shared.)

Data Stories Code Occurrence

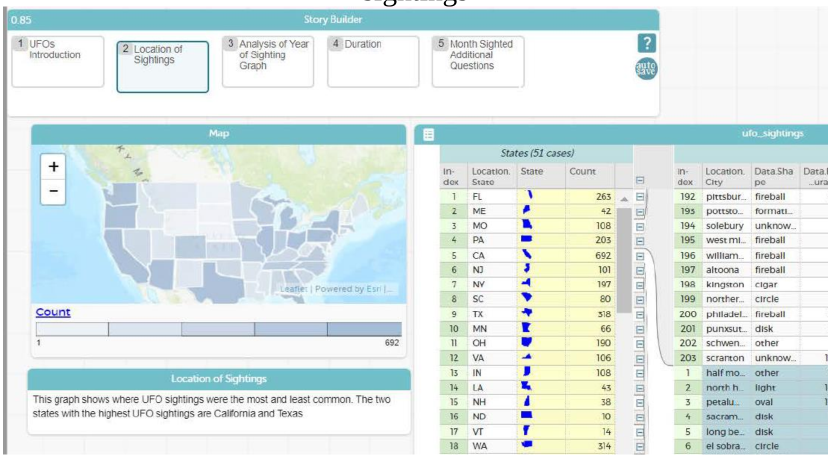
The results in this section highlight the number of data stories that were coded for components from the theoretical perspectives previously shared. Often a data story would have aspects of several moments coded separately, with some codes reoccurring throughout the story. In this section, we will share how often the included data stories were coded as representing a given theoretical framework attribute.

The data stories were initially coded to determine what components of the data investigation process framework (Lee et al., 2022) were included by the PSTs. All 17 student pairs showed evidence of engaging in *explore and visualize data* within their data stories. The next most frequently occurring code was *consider and gather data*, which was coded in nine data stories. The code *frame the problem* occurred in five data stories, *communicate and propose* was evidenced in two data stories, and *process data* occurred in a single data story. None of the data stories included aspects that were coded for *consider models*.

All but one data story engaged in at least two data moves. *Calculating* was coded in six of the data stories. The most common *calculating* data move was to add a "boundary" attribute to the table to create a map. CODAP has a lookup function that can be used to create boundaries using geographical identifiers, such as countries or states, to create opportunities to visualize data by countries or states. *Filtering* was coded in seven data stories while *grouping* was performed in 13 of the data stories. *Summarizing* was coded the most often, in 15 out of the 17 data stories. Only three groups attempted to make a *hierarchy* and no groups were found to use *merging/joining*. This is not a surprise since the PSTs were taught how to import one table into CODAP and not shown how to import more than one.

Figure 2 shows part of a data story about UFO (unidentified flying object) sightings in the United States. Figure 2 highlights how PSTs created a boundary value (using a Location.State variable) and a count variable for the number of sightings in each state. Once the boundary value is created, then the Map feature can be used. Here they overlaid count onto the map to show a scale of UFO sighting occurrences. In CODAP, calculated variables are highlighted in yellow, making it straightforward to code for the *calculating* data move. The PSTs also used the *hierarchy* feature in CODAP to sort the data in the table by state.

Figure 2. Part of a data story showing how PSTs created a boundary and count value for UFO sightings



DISCUSSION

We feel that with the increase in statistics standards addressed in this course, we are supporting the curricular recommendations outlined in the SET report by providing prospective teachers opportunities to engage in authentic statistics investigations within a mathematics content course (Franklin et al., 2015). The LOCUS results show an increase in conceptual understanding of statistics topics providing evidence for the effectiveness of the changes made to this content course. Future sections of this course will continue investigating the impact of the inclusion of more robust statistics content.

Engaging PSTs in activities like the one noted in this study can help them gain confidence in not only learning key skills in data science but may prepare them to support their own students' learning of statistics (Dolor & Noll, 2015). Being an active participant in creating a data story allows PSTs to interact with large and messy datasets to help increase the likelihood that, as K-8 teachers, they may lead the way to foster more data literate citizens. Though none of the students created a data story that engaged in all six pieces of the data investigation process, Lee et al. (2022) state that creating activities that engage in all parts of the process is not necessary to help guide young students through various statistics ideas.

Students were given very little instruction on data moves that they could use in CODAP since the unit covering statistics in the course itself was limited. If more instruction regarding this framework were provided, the data moves may have been more robust. Despite this fact, many students were able to engage in meaningful data moves which helped further their data investigations. Next steps will include investigating the co-occurrence of data moves and pieces of the data investigation process.

What may be more interesting for future research, which was not investigated here, was the data preparation that these preservice teachers engaged in prior to creating their data stories. The students were given free reign to choose a dataset of their own interest from the CORGIS site. Though these datasets are provided in a "cleaned" .csv file, most were still large and messy. Many of the groups chose to filter out some of the data, for example one group investigating data about billionaires decided to only look at data collected in one year, though the dataset spanned many years. Other groups had several variables that they found were not helpful for their investigation and decided to delete the variables. This part of cleaning the data is a very important step that often takes up a large part of data scientists' time when working on projects (Lee et al., 2022; Wilkerson et al., 2022). In future course offerings, more time will be given to preservice teachers to engage in this crucial step. Future researchers should consider ways to record this process and draw connections between steps classroom teachers take to prepare data for their students to use in data investigations and how that coincides with the work of data scientists.

We want to ensure these opportunities are not focused just on secondary prospective teachers, but that building a foundation in statistics in early grades is just as important to creating data literate students. Studies have shown that there is a lack of growth in statistical understanding from elementary to middle grades then from middle grades to secondary (Callingham & Watson, 2017). If elementary grade students are not properly introduced to statistical thinking early in their development, they may find it hard to fully develop statistical thinking in later grades. We hope changes to preservice curriculum efforts to develop statistical thinking in teachers can lead to robust statistical thinking in students starting at a young age.

REFERENCES

Bargagliotti, A., Johnson, S., Franklin, C., Perez, L., Arnold, P.,, Spangler, D.A. & Gould, R. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II).* American Statistical Association, Alexandria, VA.

Bart, A.C., Kafura, D., Shaffer, C.A., Tibau, J., Gusukuma, L., & Tilevich, E. (2022). Corgis: The Collection of Really Great, Interesting, Situated Datasets. Retrieved from https://think.cs.vt.edu/corgis/csv/.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In *Third International Handbook of Mathematics Education* (pp. 643–689).

Callingham, R. & Watson, J. (2017). The development of statistical literacy at school. *Statistics Education Research Journal*, 16(1), 181 – 201.

Common Core State Standards for Mathematics (2010). Retrieved from http://www.corestandards.org/.

CODAP (n.d.). *Common Online Data Analysis Platform (CODAP),* https://codap.concord.org/

Dolor, J. & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal*, 14(1), 60 – 89.

Finzer, W. (2000). Design of Fathom, a dynamic statistics environment, for the teaching of mathematics. Paper presented at the *International Conference on Mathematics Education*. Copenhagen, Denmark.

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., & Spangler, D. (2015). *The statistical education of teachers*. Alexandria, VA: American Statistical Association.

Jacobbe, Tim, Bob delMas, Brad Hartlaub, Jeff Haberstroh, Catherine Case, Steven Foti, and Douglas Whitaker. Establishing the validity and reliability of the LOCUS Assessments. *Numeracy* 16, Iss. 1 (2023): Article 5. DOI:

Lee, H.S. & Hollebrands, K. (2008). Preparing to teach mathematics with technology: An integrated approach to developing technological pedagogical content knowledge. *Contemporary Issues in Technology and Teacher Education*, 8(4).

Lee, H.S. & Hollebrands, K. F. (2011). Characterizing and developing teachers' knowledge for teaching statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education: A joint ICME/IASE study* (pp. 359-369). New York: Springer.

Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. In *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 97-107). Springer, Dordrecht.

Wilkerson, M.H., Lanouette, K., & Shareff, R.L. (2022). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public data sets. *Mathematical Thinking and Learning*, 24(4), 312-330.