Designing and implementing data lessons in secondary education

Emily Thrasher, Michelle Pace and Bruce Graham
North Carolina State University, U.S.A.
epthrash@ncsu.edu

This study examines how secondary teachers incorporated the six-phase Data Investigation Process (Lee et al., 2022) into classroom lessons following a professional learning experience. Analysis of 13 lesson plans, interviews, and survey responses revealed that while most lessons addressed multiple phases, few supported full engagement across all six. Framing the Problem was the most consistently attended-to phase, often grounded in authentic contexts and clear investigative questions. In contrast, Explore and Visualize was less developed, frequently limited by specific learning goals constraints. Lessons that integrated technology, especially CODAP, were more likely to support deeper student exploration and reasoning. Findings underscore the need for thoughtful instructional design.

INTRODUCTION

The integration of statistical concepts and data analysis has become increasingly prevalent in secondary education curricula. In response to evolving educational needs, numerous nations including Canada, Germany, New Zealand, the United Kingdom, and the United States have initiated curriculum reforms that emphasize the incorporation of data science skills within K-12 educational frameworks (Sukol, 2024). Educational researchers recommend that schools foster students' data literacy through hands-on problem-solving activities centered on data investigations, which engage learners in meaningful exploration of statistical concepts (e.g., Bargagliotti et al., 2020, Ben-Zvi et al., 2018).

However, despite the growing emphasis on statistical education in K-12 settings, Lovett and H.S. Lee (2017) found that educators frequently struggle with confidence and competency when teaching statistical content. Research indicates that educators recognize the importance of incorporating real-world data into their instruction (Delaney & V. R. Lee, 2024) and that teachers require enhanced preparation in understanding the methodologies and analytical processes involved in data investigation practices (Burgess, 2018). In response to these needs, we designed a 10-month professional learning experience to support secondary teachers in designing and implementing data investigations as part of the InSTEP project to develop teachers' expertise to teach statistics and data science (NSF #1908760). We explored both the teachers' experiences and the lessons developed to answer the following research question: *How do teachers plan to engage STEAM learners in key phases of a data investigation process while planning data investigation lessons for their classrooms?*

FRAMEWORK

The Data Investigation Process (H. S. Lee et al., 2022) was used to frame the professional learning experience for our teachers and the data investigation lesson plans that our teachers created within this project. The Data Investigation Process involves six interconnected and non-linear phases (see Figure 1). The first phase is Framing the Problem to understand the context and formulate the investigative question(s). In Consider and Gather Data, investigators select/collect and understand their data to answer investigative questions. After gathering data, the investigators Process the data by structuring the data into a usable format. Exploring and Visualizing Data is when multiple visualizations and statistical measures are created and explored through exploratory data analysis using appropriate technology. Closely related, investigators Consider Models when they to choose models that address the problem or answer the questions. Finally, Communicating and Proposing Action shares results and recommendations to the appropriate audience. It is through engaging in these phases, often multiple times, that investigators can make data-based decisions and claims related to their problem.

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*. © ISI/IASE.

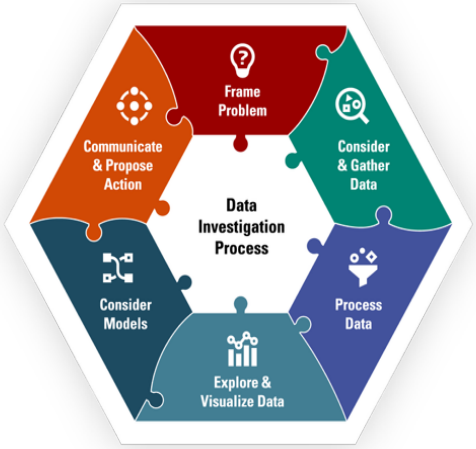

Figure 1. Data Investigation Process

METHODS

This study employed a case study approach (Yin, 2018), a methodology well-suited for examining how a group of individuals engages with a real-world context event using various sources of evidence. Grounded in the Data Investigation Process, our case study triangulates multiple sources of data, enabling analytic generalizations (Yin, 2018). Multiple data sources were analyzed and used to triangulate themes: created lesson plans (n=13), final interviews (n=15), and final surveys (n=13).

Our case comprises 15 teachers that created 13 lesson plans during the professional learning experience throughout the 2023-2024 school year. Our hybrid professional learning experience (in-person, synchronous, asynchronous) involved bringing together various experts: teachers, data science education experts, and data scientists from industry. Teams of teachers, with support from data scientists and data science education experts from the InSTEP team, created data investigation lesson plans, implemented lessons in their classrooms and refined their lesson plans. The professional experience also included engagement with resources from the InSTEP professional learning platform related to data investigations. Our case consists of four teachers that identify as male and 11 identifying as female. Two teachers taught middle school (ages 11-14) while the rest taught high school (ages 14-18).

Data investigation lesson plans were analyzed in iterative cycles, with all authors participating. Through cycles of individual coding and whole group discussion, criteria for what constituted a level of engagement for each phase within a lesson plan were created. A *Not Present* rating represents a lack of engagement in the phase within a lesson plan. A *Partially Established* rating represents limited cognitive demand for students or indicates missing parts of the specific data investigation phase. An *Established* rating represents a lesson plan that proficiently provides opportunities for learners to engage in that phase. An *Exemplar* rating indicates that a lesson plan was an ideal example of engagement within that phase. For example, for the Explore and Visualize phase, a *Partially Established* rating indicates a very prescriptive exploration of data for students and requires them to make one data visual or calculate one statistic. An *Established* rating indicates a less prescriptive exploration for students and gives them the freedom to look at multiple visuals of their data; however, may take away freedom through limiting variable exploration or choice of graphs created or statistics calculated. Finally, an *Exemplar* rating provides students the freedom to explore and create visuals they deem appropriate to answer their investigative questions. Similar descriptions were made for each phase of the Data Investigation Process. The authors then engaged in iterative coding cycles in which they coded individually and met to address discrepancies until consensus was reached.

Interviews were transcribed and coded in three cycles, combining a priori codes based on the research question and the Data Investigation Process framework with open coding. Survey responses were analyzed using descriptive statistics for eight relevant Likert scale questions, and open-ended

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*. © ISI/IASE.

responses were coded using the interview codebook. Codes and statistics were grouped into themes to address the research question.

FINDINGS AND DISCUSSION

The following section brings together analysis from coding the lesson plans, transcripts, and using the survey results when possible. First, findings of general patterns across lesson plans will be reviewed followed by a deep dive into two phases: Framing the Problem and Explore and Visualize.

General Lesson Plan Findings

Analysis of the lesson plans revealed patterns in how students were supported to engage in the Data Investigation Process as a whole. Levels of engagement varied widely across the six phases within individual lessons. Table 1 summarizes each lesson's context, disciplinary focus, and engagement level for each phase. Fewer than half of the lesson plans gave students opportunities to engage at the *Established* or *Exemplar* level in most phases (4 phases or more). The remaining seven plans offered *Established* or *Exemplar* engagement in three or fewer phases; one of these included only a single *Exemplar* rated phase, with no others rated *Established*. These findings suggest that fully addressing all six phases of the Data Investigation Process in a single lesson is challenging.

Table 1. *Context and Levels of Engagement in Phases of Data Investigation Process by Lesson Plan*

Context	Discipline	FP	CGD	PD	EVD	CM	CPA
Consumer Price Index	History	EST	PE	NP	EST	PE	EST
Solar Energy	Science	EST	PE	NP	EST	PE	PE
Solar Energy	Math	EST	EST	NP	PE	PE	PE
Screen Time	Math	EX	NP	NP	NP	NP	PE
NCAA Basketball Tournament	Math	EX	EX	PE	EX	EX	EX
# Licks to Tootsie Pop Center	Math	EX	EX	NP	PE	EST	EST
College Basketball Rivalry	Math	PE	PE	NP	EST	EST	EX
Restaurant Nutrition	Math	EST	EX	EST	EST	EST	EST
Specific Heat Capacity	Science	EST	EX	PE	EST	EST	PE
Air Pollution	Science	EST	EST	NP	PE	EST	PE
Student Performance	Math	EST	PE	NP	EST	EX	EST
Census at School	Math	EST	EX	NP	EX	EX	EX
AP Enrollment	Math	PE	PE	NP	PE	PE	EST

Note. Phases of the Data Investigation Process are abbreviated as follows: Frame the Problem (FP), Collect and Gather Data (CGD), Process Data (PD), Explore and Visualize Data (EVD), Consider Models (CM), Communicate and Propose Action (CPA). Levels of Engagement are abbreviated as follows: Not Present (NP), Partially Established (PE), Established (EST), Exemplar (EX).

Teacher interviews offer additional insight into their experiences with implementing all phases of the Data Investigation Process. Many described the challenge of balancing meaningful data investigations with the structural demands of their curriculum. As one of the teachers who contributed to the development of the screen time lesson plan explained,

> *And it's like one standard, but there's never a question on it on the [standardized state test] that they take. And so a lot of people don't even do that standard. But it was hard to do an entire project about one standard that literally is taught in one day.*

This highlights the tension between teachers' desire to cultivate meaningful student inquiry and the pressure to align with pacing guides driven by standardized tests within the project. Other teachers simply stated the issue as not having enough time. As one participant succinctly noted: "*Time, time, time, time, time is the biggest challenge, of course.*" These challenges align with other researchers'

findings around implementation of professional learning in their classrooms (e.g., Morgan & Bates, 2018).

Looking across all lesson plans by phases revealed patterns in the execution of different phases of the Data Investigation Process (see Table 1) by these teachers. The Framing the Problem phase was the strongest component, with 11 out of 13 lesson plans scoring *Established* or *Exemplar*, while Process Data was the weakest, with only one lesson plan scoring *Established* and none scoring *Exemplar*. As expected from the framework, the analysis suggests a connection between the Explore and Visualize Data and Consider Models phases with eight out of 13 lesson plans obtaining the same levels of engagement for those phases. Below we highlight findings through the lens of two phases of the Data Investigation Process (Frame the Problem & Explore and Visualize Data).

Frame the Problem Phase

The most attended-to phase across all lesson plans was Framing the Problem, with 11 of the 13 lesson plans rated as either *Established* or *Exemplar*. It was one of the only two phases – along with Communicate and Propose Action – in which no lesson was rated *Not Present.* Over half of the lesson plans earned their highest engagement rating in this phase, highlighting the success of the professional learning experience in helping teachers design lessons that center student inquiry in authentic and relevant contexts, aligning with other studies on real data and contexts (e.g., Neumann et al., 2013).

To assign these ratings, we examined whether each lesson: (1) provided a meaningful context; (2) introduced a clear investigative question (either implicitly or explicitly); and (3) supported student readiness to engage with the data. *Exemplar* lessons did all three, while *Established* lessons tended to have minor gaps, often in the strength of connection between context and question or in the clarity of what students were expected to explore. *Partially Established* lessons included some elements but lacked cohesion or depth in Framing the Problem under investigation.

Lesson plans drew from a wide range of STEAM contexts (see Table 1), including historical trends in fluctuations of the consumer price index, restaurant nutrition, and air pollution, as well as more unique scenarios such as *How many licks does it take to get to the center of a Tootsie Pop?* Even among lessons that used similar datasets, teachers approached problem framing in notably different ways. For instance, two lessons utilized college basketball data, but with divergent aims: one explored factors leading to team success and recruitment strategies, while the other analyzed a well-known local rivalry selected to boost student buy-in. A similar contrast appeared in the two solar energy lessons, which worked from overlapping datasets but grounded the context in different disciplines. In the science lesson, students examined how weather patterns impacted solar panel output, whereas in the math lesson, students engaged in a broader challenge to design a solar-powered park using solar data to inform the feasibility of their choices.

These variations highlight how teachers tailored the Framing the Problem phase to both content areas and students' interests aligning with key considerations given by H. S. Lee and colleagues (2022). In interviews, teachers expressed excitement about framing lessons in ways that allowed students to see themselves in the data. One teacher reflected on his AP Enrollment lesson:

> *So, the general question was does the demographics of those courses reflect demographics of our school? And then from there, there was a question, why, or why not? … I came to the realization that they found themselves in one of those cells in a table. I figured that that would be an activity that is catered to everybody in a classroom that I have. So, that was quote unquote the hook, or selling point, of what made students interested in doing this activity.*

Like in similar studies (e.g., Casey et al., 2021; Delaney & V. R. Lee, 2024), many of the teachers selected datasets comprised of real data, where students solved a problem and/or asked questions about a real phenomenon within their disciplines. In the final survey, most teachers (84.62%) agreed or strongly agreed that their students were interested in the datasets from their lessons. Ultimately, while most lessons were teacher-directed in how investigative questions were introduced, one lesson (Census at School) invited students to pose their own.

In addition to analyzing how engagement for Frame the Problem was rated in these lesson plans, the amount of time dedicated in the lesson plan to this phase stood out. The lessons on Screen Time and Solar Energy (science) provide a helpful contrast in how time was allocated to this phase, since both

lessons dedicated nearly an entire class period to Framing the Problem. Though the screen time lesson was rated as an *Exemplar*, this strength was isolated to the Frame the Problem phase. The lesson plan received a *Partially Established* rating in Communicate and Propose Action and was rated *Not Present* in the four remaining phases of the Data Investigation Process. This suggests that even a deeply contextualized launch can fall short of supporting students through a meaningful investigation. While Jackson and colleagues' (2012) assert the importance of a deeply contextualized launch, this provides evidence that it does not necessarily lead to continued engagement in a data investigation. By contrast, the Tootsie Pop lesson reached *Exemplar* status in Framing the Problem with a very brief launch – just enough to spark curiosity and establish purpose for exploring data. This lesson moved quickly into data collection and analysis, balancing student engagement with instructional momentum. Taken together, these examples highlight an important takeaway: a successful Framing of the Problem sets the stage for rich data exploration by dedicating just enough time to create an entry point while also maintaining cognitive demand and a clear path forward into data exploration.

Explore and Visualize Phase

Explore and Visualize was one of the least attended-to phases in the lesson plans created by our teachers. Almost half (6 out of 13) of the lesson plans were rated *Established* for this phase, and two were rated *Exemplar*. This means that in over half of the lessons, students were encouraged to create multiple visualizations and/or compute statistics. In the two *Exemplar* lesson plans, students were given complete freedom to explore their data in order to answer their investigative questions. The following directions provide an example of this open invitation to explore from the College Basketball Rivalry lesson plan:

> *Complete your analysis below by creating graphs and copying them along with an explanation for what you have identified using those graphs. You should create at least 3 graphs that provide evidence for why a certain team may have performed better than other teams.*

Lesson plans rated *Established* often gave students some freedom but still prescribed specific variables or types of graphs to explore. A notable trend among these Explore and Visualize phases rated as *Established* was the alignment between the prescribed elements and the learning goals of the lessons, suggesting that the structured nature of the exploration was, in part, a response to those goals. This reveals the tension between focusing on statistical ideas and focusing on rules and procedures, which was discussed by Ben-Zvi and colleagues (2018). For example, in one lesson plan about Restaurant Nutrition, the learning goal was: *Students will be able to construct visual representations of data using dot plots and boxplots.* This led to instructions such as:

> *Individually, each of you need to choose one of the other dietary needs (total fat, sodium, fiber, ...) and analyze it across the four restaurants using boxplots that show outliers. You can do this by dragging the dietary need you are analyzing to the x-axis of your current graph of calories or by clicking calories on the x-axis and changing the variable. You and your partner must choose different needs.*

These instructions required multiple visualizations and offered a choice in which variables to attend to in order to answer the investigative question, but limited the graph type to boxplots, in line with the lesson's learning goal.

In the four lesson plans where Explore and Visualize was rated *Partially Established*, students were only instructed to create a single, specific visualization or calculate one specific statistic – often by hand (3 out of 4 lessons). Although these tasks aligned with the learning goals, students' engagement with the Explore and Visualize phase was limited. In one lesson, students were not asked to explore data at all.

Data from the interviews highlighted that teachers often described a connection between the use of technology and their ability to engage students in the Explore and Visualize phase (Table 1). Nine out of 13 lesson plans used technology to support this phase, and all lessons rated *Exemplar* or *Established* incorporated technology. One lesson rated *Partially Established* used technology as well. Overall, there was a clear connection between more established implementations of Explore and Visualize and the use of technology, as noted by other researchers (e.g., Erickson et al., 2019).

In eight of the lesson plans, CODAP supported students in exploring their data and maintaining engagement, while R was used in one lesson plan. In interviews, teachers described how CODAP's

ability to provide visualizations and statistical analysis supported students in reasoning with data. The following quote from a teacher describing their lesson on the Consumer Price Index typifies this idea:

> *I had the students then take the data sets, put it into CODAP, and then map it over comparing to historical events...The students all of a sudden realized that rubber used to be really cheap in the early 1900s. And then by World War I, all of a sudden they were like, "Holy cow, Mr. T, in 1916, the rubber's through the roof."...And the kids found it interesting.*

Many teachers noted they plan to use CODAP more regularly in their classrooms showing the importance of appropriate technology in completing statistics and data science tasks. This increased use of technology for teaching statistics and data science aligns with calls from the field for more technology integration (Gil & Gibbs, 2016; Lovett & H. S. Lee, 2017). Teachers often commented on how CODAP provided a "low entry" point for students to engage in the Data Investigation Process and allowed for more opportunities for discussion and conceptual understanding. One teacher described:

> *I think the efficiency of using the tool like CODAP and just constructing and making these visualizations...15 years ago I was just like, "All right, everyone, let's draw a box plot by hand. Here we go."...So really leveraging this technology to allow the conversations to be the priority, and the concept to be the priority.*

These findings suggest that while some lessons provided opportunities for rich data exploration, the learning goals can be a limiting factor to full engagement with the Explore and Visualize phase. The integration of technology, particularly CODAP, emerged as a key factor in supporting more open-ended and conceptually focused explorations.

CONCLUSION

These findings highlight the possibilities and complexities of incorporating the full Data Investigation Process into STEAM lesson plans. Framing the Problem emerged as the most consistently well-established phase, with attention of the teachers to authentic and engaging contexts. However, other phases were less consistently addressed, such as Explore and Visualize. For well-established Explore and Visualize lesson plans, the integration of technology, often CODAP, played a role in supporting engagement and conceptual understanding aligning with findings from prior implementation research. Teachers cited limited time, curricular constraints, and assessment pressures as barriers to full implementation of the Data Investigation Process in classrooms. While few lessons fully integrated all six phases of the framework, several exemplars point to what is possible with strategic support and planning.

As is the nature of case studies, these findings are intended to shed light on possible experiences of teachers in developing lesson plans involving data investigations for their classrooms. Findings suggest a need for curricular flexibility and professional development to support the planning and implementation of data investigations within secondary classrooms.

ACKNOWLEDGEMENT

The work described in this paper is supported by the National Science Foundation under Grant DRL 1908760 awarded to NC State University. Any opinions, findings, conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation. Thank you to Gemma Mojica for support on this project. All three co-authors provided to the development of this paper equally.

REFERENCES

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II)*. American Statistical Association and National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Ben-Zvi, D., Gravemeijer, K. & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K., Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473–502). Springer. https://doi.org/10.1007/978-3-319-66195-7_2

Burgess, T. A. (2018). Teacher knowledge of statistical investigations. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher*

education. A joint ICMI/IASE study (pp. 259–270). Springer. https://doi.org/10.1007/978-94-007-1131-0_26

Casey, S. A., Harrison, T., & Hudson, R. (2021). Characteristics of statistical investigations tasks created by preservice teachers. *Investigations in Mathematics Learning, 13*(4), 303–322. https://doi.org/10.1080/19477503.2021.1990659

Delaney, V., & Lee, V. R. (2024). High school teachers' data set aesthetics. *Information and Learning Sciences, 125*(7-8), 475–490. https://doi.org/10.1108/ILS-06-2023-0063

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education, 12*(1). https://doi.org/10.5070/T5121038001

Gil, E., & Gibbs, A. L. (2016). Introducing secondary school students to big data and its social impact: A study within an innovative learning environment. In J. Engel (Eds.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association for Statistics Education,* ISI/IASE.

Jackson, K. J., Shahan, E. C., Gibbons, L. K., & Cobb, P. A. (2012). Launching complex tasks. *Mathematics Teaching in the Middle School, 18*(1), 24–29. https://doi.org/10.5951/mathteacmiddscho.18.1.0024

Lee, H. S., Mojica, G. F., Thrasher, E. P., & Baumgartner, P. (2022). Investigating data like a data scientist: Key processes and practices. *Statistics Education Research Journal 21*(2). Article 3. https://doi.org/10.52041/serj.v21i2.41

Lovett, J. N., & Lee, H. L. (2017). New standards require teaching more statistics: Are preservice secondary mathematics teachers ready? *Journal of Teacher Education, 68*(3), 299–311. https://doi.org/10.1177/0022487117697918

Morgan, D. N., & Bates, C. C. (2018). Addressing the barriers of time. *The reading teacher, 72*(1), 131–134. https://doi.org/10.1002/trtr.1716

Neumann, D. L., Hood, M., & Neumann, M. M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal, 12*(2), 59–70. https://doi.org/10.52041/serj.v12i2.304

Sukol, S. (2024, March 18). *Beyond borders 2024: Primary and secondary data science education around the world.* Data Science 4 Everyone. https://www.datascience4everyone.org/post/beyond-borders-2024

Yin, R. K. (2018). *Case study research and applications: Design and methods (6th ed.)*. Sage.