LEARNING TO MAKE SENSE OF DATA IN A CODAP-ENABLED LEARNING ENVIRONMENT: INTERACTIONS MATTER

Gemma F. Mojica
NC State University
gmmojica@ncsu.edu

Heather Barker
Elon University
haallmon@ncsu.edu

Christina N. Azmy
Catawba College
cnazmy19@catawba.edu

Hollylynne S. Lee
NC State University
hollylynne@ncsu.edu

In this study, we investigated how sixth and seventh grade students used CODAP to make sense of roller coaster data while engaged in Exploratory Data Analysis (EDA). Using instrumentation theory, we examined students' instrumentation approaches, as well as the types of instrumental orchestration utilized by teachers as they interacted with student pairs during EDA.

Keywords: Data Analysis and Statistics, Technology, Instructional Activities and Practices

Introduction

Statistics has gained prominence in school curricula in the US (Franklin et al., 2007; National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practice & Council of Chief State School Officers, 2010), which includes a focus on reasoning about data. One way to encourage students to reason about data is providing opportunities to engage in Exploratory Data Analysis (EDA). EDA first developed by Tukey (1977), involves exploring data to summarize main characteristics. EDA is the "art of making sense of data by organizing, describing, representing, and analyzing data, with a heavy reliance on informal analysis methods, visual displays" (Ben-Zvi & Ben-Arush, 2014, p. 197). While approaches often use visual methods, such as graphs and other representations, statistical measures are sometimes calculated to make sense of data. Ben-Zvi (2004) points out that exploring data involves examining features such as shape, center, and spread; it involves considering graphs and looking for other characteristics of data like clusters, gaps, and outliers. Cobb and McClain (2004) recommend that EDA should be the focus of early experiences with instruction because of the emphasis on finding trends and patterns.

EDA often involves the use of technology, and there is evidence that innovative technology tools aide students in developing statistical thinking (e.g., Biehler et al., 2013). We are interested in understanding students' engagement with the Common Online Data Analysis Platform (CODAP), (https://codap.concord.org/), which has many powerful dynamic visualization and calculating capabilities that make it an ideal tool for engaging in EDA. Specifically, we investigated the following research questions:

- *RQ1: How do students use CODAP to make sense of data while engaged in EDA?*
- *RQ2: What types of orchestration emerge as teachers interact with students as they engage in EDA using CODAP?*

Theoretical Perspectives

Our study draws on two theoretical perspectives from instrumental theory: instrumental genesis and instrumental orchestration. To understand students' learning processes as they made sense of data during EDA using CODAP, we used Ben-Zvi and Ben-Arush's (2014) types of instrumentation. Instrumental orchestration was used to understand teachers' interactions with students as they explored 157 US roller coasters using CODAP (Drijvers et al., 2010).

Instrumental Genesis

Five components comprise instrumental genesis (IG) (Ben-Zvi & Ben-Arush, 2014). The *subject* is a learner who accomplishes a task using an instrument. An *object* is a specific task. An *artifact* (a component of a tool) is a physical or virtual device that is used by the subject, which has no meaning for the learner in isolation. A *utilization scheme* is a cognitive scheme that the subject uses to accomplish a task using one or more artifacts. When the subject has successfully used the utilization scheme to accomplish a task, the artifact becomes an *instrument* for the learner to use. The authors indicate that IG occurs when a subject uses utilization schemes to transform an artifact into an instrument that can be used as a meaningful tool to achieve a particular goal.

There are two components of IG, *instrumentalization*, the ways in which the subject's prior knowledge acts on the tool, and *instrumentation*, the way the instrument influences the subject's learning process. In this work, we are interested in instrumentation. Ben-Zvi and Ben-Arush (2014) identify three processes of instrumentation that learners use to investigate data: unsystematic, systematic, and expanding. An *unsystematic* approach to investigating data involves actions that are not intentional or systematic, where learners make sense of a few basic artifacts and associated actions. *Systematic* instrumentation involves intentional and somewhat organized exploration, occurring after the learner has become familiar with artifacts, and may be more focused on the tool rather than the task. The third process involves *expanding* emerging instrumentalization (i.e., ways in which students' prior knowledge acts on the tool) of an artifact and associated actions that transform into a more usable and powerful instrument that can be used in a meaningful way in new contexts and situations.

Instrumental Orchestration

Instrumental orchestration is the teacher's intentional and systematic organization and use of various artifacts in a learning environment to guide the learners' instrumental genesis in relation to a mathematical task (Drijvers et al., 2010; Trouche, 2004), or in our case a statistical task. The three elements within instrumental orchestration include the following: a) *didactical configuration*, referring to the design of the teaching setting and artifacts, b) *exploitation mode*, referring to the ways the teacher makes decisions to exploit the didactical configuration to achieve the learning goals, and c) *didactical performance*, referring to the in the moment decisions made by the teacher on how to act on the didactical configuration and enact the exploitation mode. While Drijvers et al. and Mojica et al. (2019) identified orchestration types of teachers' purposeful use of technological tools during whole class instruction, we are interested in teachers' orchestration as they interact with pairs of students.

Participants and Context

Participants in this study were 19 sixth grade and 25 seventh grade students between the ages of 11 and 12-years old from a small urban school in the southeastern US. The school is racially diverse, and 48.6% of the students receive free/reduced lunch. Less than half of the students are proficient in mathematics (40.2%), as compared to 63.2% in their district.

We report on the same 72-minute mathematics lesson, implemented in both a sixth and seventh grade classroom, where students engaged in EDA using CODAP to make sense of roller coaster data. This lesson took place during the second week of the school year, prior to any formal instruction on statistics. This was students' first experience with CODAP. Both classes were taught by an experienced researcher, from a large research university in the southeastern US, with expertise in the teaching and learning of statistics, as well as using technology tools. The regular mathematics classroom teachers were also present during the lesson and interacted with students while they engaged in EDA. Since this paper focuses on how the teachers interacted with students only as they worked in pairs (not as a whole class) during EDA, we refer to all as teachers.

Each lesson consisted of four parts: 1) teacher launching the investigation (whole class); 2) teacher introducing CODAP as a tool using a small data set (whole class); 3) student pairs investigating larger data set using CODAP (pairs); and, 4) teacher facilitating discussion as student pairs present interesting noticings (whole class), the results of their EDA. The teacher launched the lesson by asking students to consider aspects of roller coasters that might make the ride thrilling or scary and then showed a video of a wooden roller coaster from the data set, from the point of view (POV) of a rider, to introduce the context of the data. Students discussed attributes of coasters they thought might be thrilling or scary, and then the teacher introduced students to CODAP by facilitating the exploration of a small data set of 31 US roller coasters using a CODAP document. Our analysis focuses on part 3 of the lesson where students worked in pairs to explore a larger data set of 157 US roller coasters, with 15 numerical and categorical attributes (e.g., name, location, design, top speed, maximum height, etc.). Students were encouraged to ask their own questions and find interesting things they could share about the coasters using features in CODAP, such as graphs. While student pairs engaged in EDA, all teachers monitored student work and interacted with students.

Methods

Data collected for this study is part of a larger project. Classes were video recorded using three cameras from multiple perspectives. While student pairs used CODAP to investigate the roller coaster data, all cameras recorded the teachers' interactions with student pairs or focused on student pairs as they worked. Six student pairs' laptop screens were recorded as screencasts throughout the entire class. The regular mathematics classroom teachers selected pairs to represent divergent student thinking. We used a deductive approach to selecting video for analysis (Derry et al., 2010). To examine how students used CODAP to make sense of data, we selected video recordings from the screencasts of students' laptops while they were engaged in EDA with the 157 roller coaster data set using CODAP, as well as video recordings from cameras that showed students' and teachers' interactions. All selected video was initially viewed to identify episodes, our unit of analysis. Episodes were defined as an action or group of closely related actions that resulted in a process of instrumentation. After multiple researchers had viewed the video, episodes were established after arbitration and agreement was reached.

Once episodes were identified, we created content logs to provide a time-indexed description of the events on the video (Derry et al., 2010). Each episode was coded by two different researchers. Episodes of student pairs' screencasts were coded to identify the processes of instrumentation that learners used to investigate data (Ben-Zvi & Ben-Arush, 2014): unsystematic, systematic, and expanding. To identify the types of instrumental orchestration that emerged, we first identified all questions and interactions between the teachers and students as they worked in pairs. We used open coding until themes emerged to identify orchestration types. When disagreements between coders occurred, the authors arbitrated until consensus was reached, and in some instances a third researcher made the final decision.

Olanoff, D., Johnson, K., & Spitzer, S. (2021). *Proceedings of the forty-third annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Philadelphia, PA.

Results

Students' Use of CODAP to Make Sense of Data

To investigate how students used CODAP to make sense of data, we identified the instrumentation processes that six student pairs used to reason about 157 US roller coasters. The number of episodes, where an action or group of closely related actions resulted in a process of instrumentation, varied across pairs. Table 1 shows the instrumentation process identified for each pair as they engaged in EDA. Unsurprisingly, all pairs initially engaged in unsystematic instrumentation. While pairs 1, 2, 5, and 6 moved between unsystematic and systematic instrumentation, only pairs 3 and 4 engaged in unsystematic, systematic, and expanded instrumentation. Pair 2 is the only pair that worked unsystematically for most of their EDA.

Table 1: Instrumentation Processes of Student Pairs

Pair	Ep 1	Ep 2	Ep 3	Ep 4	Ep 5	Ep 6	Ep 7	Ep 8	Ep 9
1	Unsys	Unsys	Sys	Sys	Sys				
2	Unsys	Sys	Unsys	Unsys					
3	Unsys	Unsys	Sys	Exp					
4	Unsys	Unsys	Unsys	Sys	Sys	Exp	Unsys	Unsys	Unsys
5	Unsys	Sys	Unsys	Sys	Sys	Sys	Sys		
6	Unsys	Sys	Sys	Sys	Sys	Unsys	Sys	Unsys	

Example of pair that used unsystematic and systematic instrumentation. Pair 6 is an example of a pair that engaged only in unsystematic and systematic instrumentation. They began their exploration by clicking on different features of CODAP, including the map feature, slider, and opening graphs. It is important to note that opening a new graph window results in cases being displayed as a configuration of randomly scattered data points. Additionally, the map and slider features of this CODAP document were not linked to the data. This unsystematic approach enabled them to identify CODAP features that were available to them that could be potentially used to make sense of data. The pair quickly took a systematic approach by adding different attributes to a graph. Figure 1a shows the graph that was created after one student asks if the *maximum drop* is affected by the *number of inversions*. After answering a few questions about the data, the pair is then curious about how many attributes they can include on the graph, which leads them to unsystematic instrumentation as they create a new graph, see Figure 1b. Using a trial and error approach, they add three attributes, *state, year opened,* and *top speed* to the graph to conclude that at most three attributes can be added.

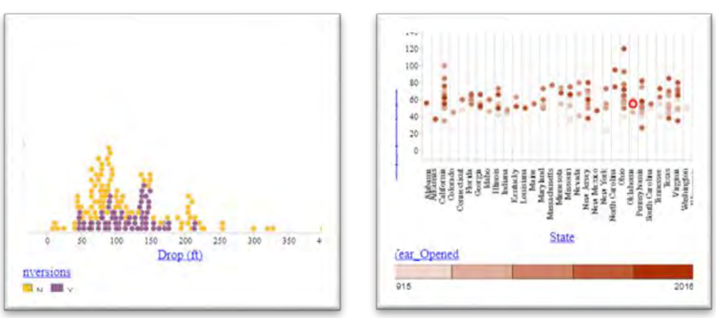

Figure 1a and b: Examples of Systematic (1a) and Unsystematic Instrumentation (1b)

Olanoff, D., Johnson, K., & Spitzer, S. (2021). *Proceedings of the forty-third annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Philadelphia, PA.

Example of pair that used unsystematic, systematic and expanded instrumentation. Pair 3 not only used features in CODAP in an intentional way to make sense of data, they engaged in expanded instrumentation. For example, they created a scatter plot comparing the *maximum drop* to the *top speed*, and then overlaid *type* on the graph to investigate if the material a roller coaster is made of affects the relationship between *top speed* and the *maximum drop* (see Figure 2). This made the use of the graph more powerful for them by allowing them to pose and answer a new question while using more features of CODAP. One of the students concluded that a lot of wooden coasters are slower and have a "shorter" drop, and the fastest ones are steel.

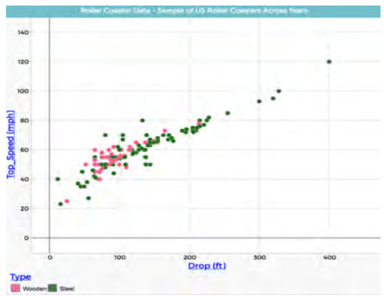

Figure 2: Student Created Scatterplot

Types of Teachers' Orchestration that Emerged

We identified 42 instances of orchestration by the teachers that were categorized into eleven different types. Table 2 illustrates the types, provides a definition and example, as well as indicates the percent of time each type occurred. Several of these orchestration types seem applicable to contexts beyond statistics and data analysis and using technologies other than CODAP, such as inserting terminology (2.38%) and providing technical assistance (4.76%). However, most of the orchestration types related specifically to teaching statistics, such as noticing trends or relationships in data and suggesting data moves. Suggesting a data move (28.57%) and assessing students' progress in their EDA (21.43%) accounted for a majority of the orchestration types. Four of the orchestration types occurred only one time (2.38%): insert terminology, clarify, focus on a case, and link multiple representations. Noticing trends and/or relationships (11.90%), making a claim or inference (9.52%), and explaining statistical reasoning or supporting a claim (7.14%) made up 28.56% of the orchestration, which are all significant in designing learning environments to support students in developing productive statistical thinking. It is beyond the scope of this paper to provide examples of every orchestration type. Therefore, we will focus on suggesting a data move and inserting terminology.

Table 2: Orchestration Types for Interactions with Student Pairs

Type	Percent	Definition	Example
Assess Progress	21.43	Assess where students are in their exploration or statistical investigation cycle (pose, collect, analyze, interpret).	What are you exploring? What are you looking at in your graph?
Relate to Context	7.15	Discuss own experiences or students' experiences related to the context.	I've never been to Carowinds, but I go to Busch Gardens a lot.
Insert Terminology	2.38	Introduce statistical terminology.	Officially that is called a scatterplot. You'll learn a little bit more about those later. It is where you are looking at two variables at the same time.

Provide Technical Assistance	4.76	Provide technical assistance – help student perform specific data move or use features of CODAP.	Click on that and remove it.
Clarify	2.38	Clarify directions or information	In case you don't know, you're supposed to be answering question number 8.
Focus on Case	2.38	Focus students' attention on specific case.	Who is that?
Suggest Data Move	28.57	Suggest a data move.	I'm gonna throw a twist into your graph, and see if you guys can make sense of this. Ok? I want you to grab wood vs. steel. I think it is type. Grab type. Put it in the middle of your graph.
Link Multiple Dynamic Representations	2.38	Draws students' attention to the dynamic nature of multi-linked representations in CODAP.	Did you notice that when you click on the case here in the other graph it shows up and in the table it will show up, too.
Notice Trends and/or Relationships	11.90	Encourage students to notice trends or patterns, which may include relationships between multiple attributes.	Ok, and do you think that it matters whether you are inverted or not and how fast you go?
Make Claim or Inference	9.52	Encourage students to make a claim or inference.	What states tend to have coasters that go really fast?
Explain Statistical Reasoning/Support Claim	7.14	Provide opportunity for students to explain their reasoning and/or support a claim/inference with evidence	What does the graph tell you?

Suggest a Data Move

Almost 30% of the orchestration types were identified as a suggest a data move. Within this type, we noticed two distinct themes, which resulted in different learning opportunities for students. An example of each kind will be illustrated below. The first shows the way a teacher interacted with Pair 3, a seventh-grade pair, whose scatterplot was previously shown in Figure 2.

Teacher 1: So, I'm gonna throw a twist into your graph, and see if you guys can make sense of this. Ok.
Student 1: Yeah.
Teacher 1: So, I want you to grab wood versus steel. I think it's type. Here we go. Grab type. Put it in the middle of your graph.
Student 1: Yeah.
Teacher 1: Yep. What did it do?
Student 1: That's pretty cool. It's telling us right now which parts are wooden and which coasters are steel.
Teacher 1: There we go. Take a look at that, and see in a little bit if you could tell the class anything that you might notice that's interesting.
Student 2: How about …
Student 1: A lot of wooden ones are slower and have a shorter drop, and the fastest ones are steel.

Pair 3 had already constructed a scatterplot while exploring the relationship between *drop* and *top speed*, and the teacher suggested that they drag and drop the attribute *type* in the center of the graph, coloring wooden coasters pink and steel coasters green. After the teacher suggested adding the categorical attribute to the graph, she followed up with a question that encouraged students to notice a relationship. Almost immediately, Student 1 was able to reason about the relationship between three attributes. When this teacher suggested a data move, she almost always followed it by a question encouraging students to notice a trend in the data or a relationship when exploring multi-variate data.

The following shows the way another teacher interacted with a different seventh-grade pair.

Teacher 2: Here, drag this a little bit so you can see. Where's your graph?
Student 3: Our graph is down here.
Teacher 2: [Takes control of the mouse.] Oh, ok, so what you can actually do is drag a category here and one right here so you can compare two things. So, compare, like, the max height to the max speed, see if they correlate.
Student 3: You have to create another graph though.
Teacher 2: No, you don't. Let me show you. Now you can see there's a trend, that has the height increases the speed increases.

In this instance, the teacher also suggested a data move. In contrast to the previous example, the teacher did not merely make a suggestion but took control of the mouse and created a scatterplot to show the relationship between *maximum height* and *top speed*. Rather than encouraging the students to notice a relationship between these two attributes, the teacher describes a positive relationship. We argue that this type of orchestration limited students' opportunity to reason statistically. It is plausible to infer that the students thought the teacher was suggesting that *maximum height* and *top speed* be graphed as dotplots on two separate graphs, since Student 3 indicated she thought they needed to create another graph. Perhaps these students may not have been ready to reason about the relationship in the way that the teacher suggested and ultimately constructed for them.

Insert Terminology

After the seventh-grade Pair 3, described earlier, had created a scatterplot (Figure 2), comparing *drop* and *top speed*, the teacher asked students what kind of graph they created. One student responded that it is a "spaceship", and the other student responded that it is an "aurora". The teacher then explained, "Officially that is called a scatterplot. You'll learn a little bit more about those later. It is where you are looking at two variables at the same time. So, what does that graph tell you?" As indicated in the section above, the students were able to reason about the relationship. While this only occurred one time, it provides an example of an appropriate way to introduce statistical terminology. Students were able to reason without knowing the name of the graph and learned new vocabulary. We conjecture that the second example in the previous section shows a way that using new terminology may have limited students' thinking. While we do not have evidence as to whether or not the students knew what correlate meant, we argue that using this terminology likely did not provide an opportunity to support students' reasoning.

Discussion

Our analysis of student pairs conducting an EDA using CODAP has provided evidence of how students make use of artifacts in CODAP to create instruments to answer meaningful questions of their own interest. We found that students who were able to transform the artifacts in CODAP to meaningful tools (i.e., going from unsystematic to systematic to expanding instrumentation) were able to pose and answer more robust questions that surfaced during EDA. All of the student pairs, except one, were able to move from using an unsystematic to a systematic approach to making sense of data. Two of the six pairs were even able to use an expanded approach that transformed features in CODAP into a more usable and powerful instrument that were used in a meaningful way in new situations. While this did not occur many times and for all student pairs, we hypothesize that this was likely due to the fact that this was students' first exposure to CODAP, as well as many students' first experience engaged in EDA. Nonetheless, this provides evidence that even students' initial experiences with using CODAP during EDA can support them in developing statistical reasoning as they make sense of data. An important implication for designing learning environments is that given an appropriate tool and well-designed task that uses real data, students can learn to use a tool while engaging in EDA. While teachers often acknowledge the affordances of using technology to support student learning, they sometimes argue they have insufficient time to do so. We suggest that these findings indicate that teachers do not need to teach students to use a tool first and then provide opportunities to engage in statistical thinking later.

Additionally, we found that students' interactions with teachers often impacted how they moved between *different* types of instrumentation. In some cases, students move from unsystematic to systematic approaches was preceded by an orchestration by the teacher. In fact, in all cases of students using expanded instrumentation, the approach was always preceded by an interaction with the teacher. We were not surprised that most orchestration types categorized as suggest a data move since this was students' first experience with CODAP. Nor were we surprised that merely suggesting a move and then the teacher making explicit their own conclusions about relationships between attributes limited students' opportunities to reason statistically. However, this work provides direct evidence of what we know anecdotally. Our findings indicate that at least one way a teacher can support students moving from unsystematic to systematic or systematic to expanding instrumentation is to explicitly encourage them to notice a trend or relationship. Further, this work shows that different orchestration types provided different learning opportunities for students to develop statistical thinking. Future work should examine this relationship between students' instrumentation and teachers' orchestration more closely.

In conclusion, we believe that providing opportunities for students to engage with well-designed tasks that use real, motivating data are fundamental aspects of designing learning environments that support students' statistical thinking. We also argue that providing opportunities for students to reason about data using dynamic statistical tools, like CODAP, is a fundamental component of learning environments that develop students statistical reasoning. Interactions with such technologies and teachers' orchestration impact learning opportunities for students.

Acknowledgments

This study was supported by the National Science Foundation under Grant No. 1625713 awarded to NC State University. Any opinions, findings, and conclusions or recommendations expressed herein are those of the principal investigators and do not necessarily reflect the views of the National Science Foundation.

References

Aguirre, J. M., Mayfield-Ingram, K., & Martin, D. B. (2013). *The Impact of Identity in K-8 Mathematics: Rethinking Equity-based Practices*. National Council of Teachers of Mathematics.

Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning, and thinking* (pp. 121–145). Kluwer.

Ben-Zvi, D., & Ben-Arush, T. (2014). EDA instrumented learning with TinkerPlots. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 193–208). Springer Spektrum.

Ben-Zvi, D. Gravemeijer, J., & Ainley, K. (2018). Design of statistic learning environments. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 473–502). Springer.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clemments, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). Springer.

Cobb, P., & McCain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning, and thinking* (pp. 375–395). Kluwer.

Derry, S. J., Pea, R. D., Barron, B, Engle, R. A., Erickson, F., Goldman, R., Hall, R., Kischmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences, 19*(1), 3–53.

Dick, T. P., & Hollebrands, K. F. (2011). *Focus in high school mathematics: Technology to support reasoning and sense making*. National Council of Teachers of Mathematics.

Drijvers, P., Doorman, M., Boon, P., Reed, H., & Gravemeijer, K. (2010). The teacher and the tool: instrumental orchestrations in the technology-rich mathematics course. *Educational Studies in Mathematics, 75*(2), 213–234.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) Report: A Pre-K-12 curriculum framework*. American Statistical Association.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students use of modal clumps to summarize data. In B. Philliips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics,* Cape Town, South Africa. Voorburg: International Statistical Institute. Available from, https://www.stat.auckland.ac.nz/~iase/publications/1/8b2_kono.pdf

Makar, K. & Confrey, J. (2014). Wondering, wandering, or unwavering? Learners' statistical investigations with Fathom. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 351–362). Springer Spektrum.

Mojica, G. F., Baker, H., & Azmy, C. N. (2019). Instrumented learning in a CODAP-enabled learning environment. In J. M. Contreras, M. M. Gea, M. M. López-Martín & E. Molina (Eds.), *Proceedings of the Third International Virtual Congress on Statistical Education*.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Author.

National Governors Association Center for Best Practice & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Author.

Trouche, L. (2004). Managing complexity of human/machine interactions in computerized learning environments: Guiding students' command process through instrumental orchestrations. *International Journal of Computers for Mathematical Learning, 9*, 281–307.

Tukey, J. (1977). *Exploratory data analysis*. Addison-Wesley.