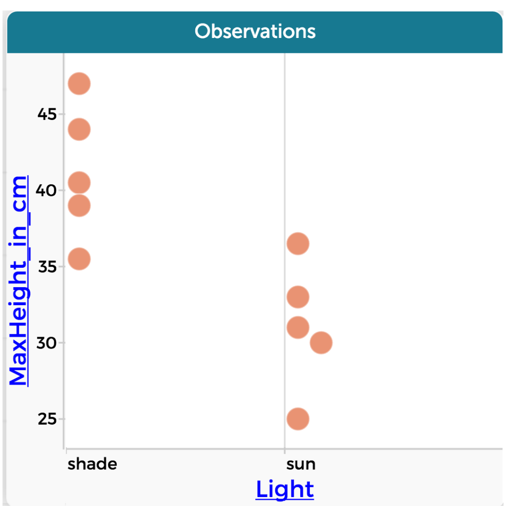
Students Making Sense of Multilevel Data

1. INTRODUCTION

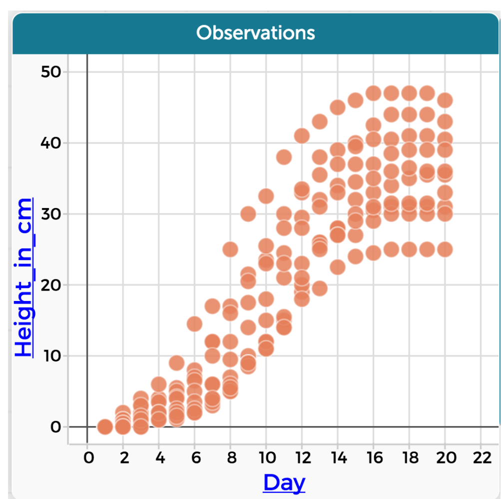
A distinction has traditionally been made between the skills of graph construction and graph interpretation (e.g., Glazer, 2011). Most of this research has looked at students' abilities to create graphs by hand or to interpret what we will refer to as "static" graphs. Increasingly, however, graphs are created with software, and the displays of data we encounter are often dynamic representations with which we can interact.

There are important differences between graphs students create with software and the ones they create with paper and pencil. With paper and pencil, there are relatively few constraints on the form graphs can take. Indeed, young students often produce representations that are quite unique (Lehrer, 2007). In contrast, students using a software tool may make some limited decisions about the attributes to use, the graph type, and title and labels. But the graph that eventually appears is subject to conventions built into the tool (Hancock, Kaput, & Goldsmith, 1992). Because of this, the situation frequently arises in computer environments where students create a graph that they, at least initially, are not able to interpret (Konold, Pollatsek, Well, & Gagnon, 1997). When this occurs, the work of graph creation and graph interpretation become closely tied.

As an example, imagine creating the two graphs in Figure 1. These data are height measurements of ten plants, half grown in the shade and half in the sun. Consider what conclusions you might draw from these graphs about how these plants grew.



a. Maximum Height vs. Light graph (Graph 1)



b. Height vs. Day graph (Graph 2)

Figure 1. Graphs of MaxHeight vs. Light (1a) and Height vs. Day (1b) in *CODAP*

Certainly, you would want to know more about how these data were collected before offering definitive statements, but tentatively, it would seem reasonable to conclude from Graph 1 that the plants in the shade grew taller on average than those grown in the sun. From Graph 2, you might conclude that all the plants showed the same logistic growth pattern, starting off relatively slowly, growing more rapidly from about day 5 to day 12, and then growing more slowly and leveling off over the final few days.

Interpreting graphs draws on various knowledge and skills (Friel, Curcio, & Bright, 2001; Carpenter & Shah, 1998). One critical domain involves understanding conventions used in graphing, such as the use of two axes to simultaneously plot the values from two attributes. Required also is knowledge about the context involved, in this case about plants and how they grow. But you could not have arrived at the conclusions above without solving a puzzle of sorts. Why are there many more points in Graph 2 than in Graph 1? And though this can be puzzling in this case, it highlights a required step in interpreting all graphs, and that is mapping the graphic elements back to their real world referents. To do so not only requires domain knowledge of plants' growth, but also knowing (or inferring) the structure of the experiment that produced the data.

This experiment involved measuring the heights of 10 plants over 20 days. The points in the two graphs actually refer to different things. In Graph 1, the points (or cases) represent measures (maximum heights) of individual plants, and there are 10 of them. But the points in Graph 2 do not refer to plant-level measurements. Each of those points is an observation of a plant's height on a particular day. Because each of the 10 plants was observed over 20 days, there are a total of 200 points, though some of them are not visible due to overlapping. Furthermore, to make the generalization about the logistic growth pattern of the plants over days, as shown in Graph 2, you need to perceive a series of 20 points as belonging to the same plant. If we add lines in Graph 2 that connect the 20 points from daily observations of each of the 10 plants, these lines would make salient the two levels of cases in Graph 2, the cases connected by lines referring to plants and the individual points referring to daily observations of plants.

We argue that figuring out what the points in these graphs refer to in the real word is a critical step in making sensible interpretations of these graphs. We suggest that this skill entails more than simply a "literal reading of the data" (Curcio, 1987, p. 384) by, for example, citing two axes. That is, a student looking at a particular point in Graph 2 (Figure 1b) might be able to say that that point is on day 20 with a height of 25. While this statement might be correct, it fails to capture that the point corresponds to an observation made for a particular plant on day 20. In our view, correctly interpreting a graph at the basic level involves more than just reading off values. Rather, one must be able to associate elements in the graph with the corresponding observations that were made in the real world.

In this article, we focus on this rudimentary ability because if students are unclear about what a point in a graph represents, they cannot make any sensible higher-level observations about what that graph reveals about the data. In Graph 2, for example, students who interpret each point as an observation of a plant on one day will draw different conclusions than students who interpret each point as one plant. In this instance, the students are interpreting what the case is differently. Indeed, several researchers have emphasized the importance of students grounding their exploration of graphic displays in the meaning of cases (Cobb, McClain, & Gravemeijer, 2003; Konold, Higgins, Russell, & Khalil, 2015; Lehrer & Schauble, 2004).

What we call a case could also be referred to as a "unit of analysis" or "unit of observation." A number of factors can make interpreting a case more or less difficult (Friel, Curcio, & Bright, 2001; Konold & Higgins, 2003). When we are involved in collecting data ourselves, we are in a better position to determine what graphic elements refer to in the real world. But often, as in this plant example, we need to interpret graphs of data that we have not ourselves collected. In these situations, we need to reconstruct from the graph and other information what those observations are and how they were collected.

Notice that we did not provide any information about how the plant experiment was conducted when we asked you to interpret the graphs in Figure 1. To proceed, you needed to reconstruct from the graphs the basic story of how those data were collected. As we mentioned, this is a critical skill in graph interpretation, and while we often have additional information in accompanying text or a title of a graph that helps us do this, we still need to seek out that information as part of the process of understanding what we are looking at in a graph.

Another possible confusion in interpreting the cases in a graph arises when, as in Graph 2, a study involves more than one level of case. In these situations, which are common, it can be easy in looking at a graph to get confused. In our example, a particular graph can either be showing information about the 10 plants or about the 200 daily observations, and we must figure out which it is to correctly interpret the graph.

In this study, we explore how students build their understanding of what the cases are when 1) they did not collect the data themselves and 2) those data represent two levels of cases or units of analysis. But unlike the situation we placed you in earlier, we presented these same plant graphs in the context of the data-analysis software called Common Online Data Analysis Platform (*CODAP*). This software is being designed at The Concord Consortium to help students organize, visualize, and analyze data. It is also designed to host multilevel data, and even most professional tools do not make this structure explicit. As we describe below, we expected that various features of *CODAP* would be helpful to students in figuring out what the cases are when they occur at more than one level.

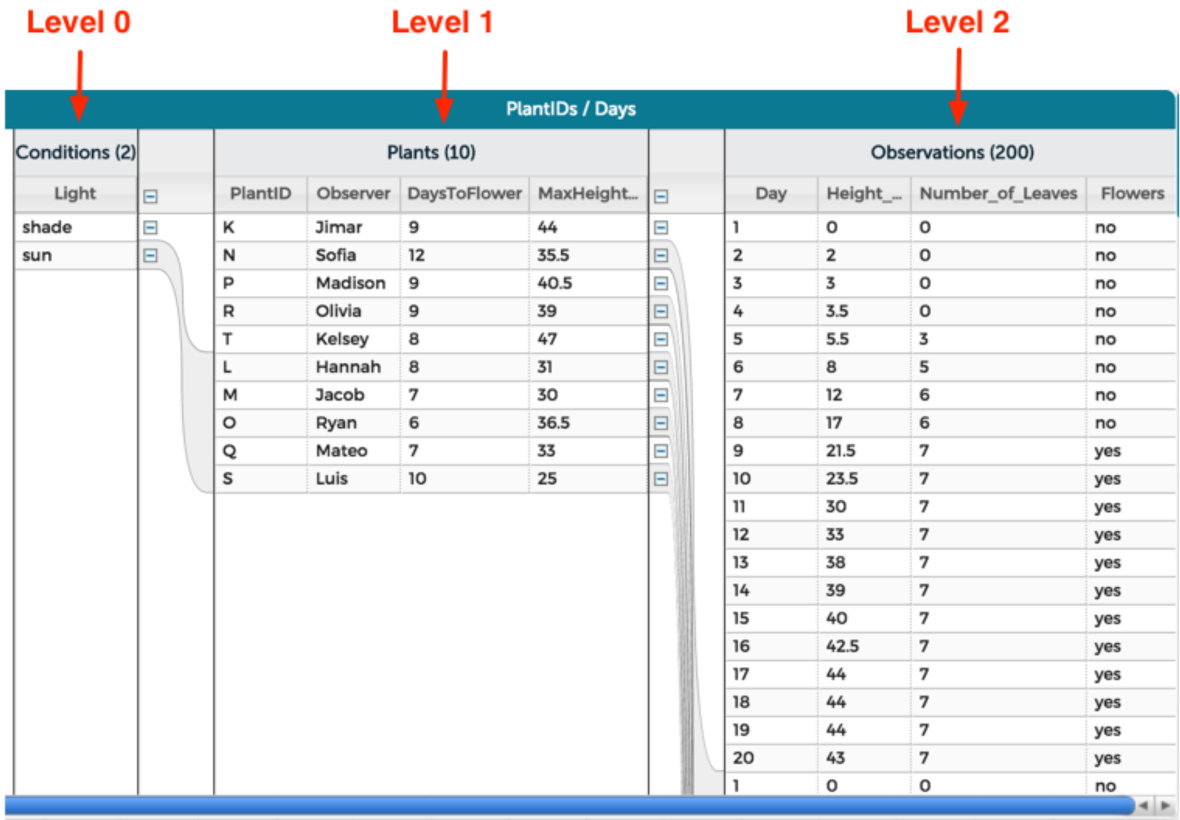
Working with multilevel or hierarchically structured data is fundamental and ubiquitous in science and engineering, yet the development of students' comprehension of this data structure remains understudied. In fact, the literature focuses primarily on four components of data analysis: forming a question, collecting data, analyzing data, and drawing conclusions (Wild & Pfannkuch, 1999). Virtually no attention has been given to how students record and organize data, and these are becoming increasingly important skills in the age of Big Data, where data scientists work with massive sets of multilevel data.

Prior research by Konold, Finzer, and Kreetong (2014) showed that students from middle school through university invented nested structures to hold complex data. They gave students "snapshots" depicting vehicles traveling along sections of highway and asked them to record and organize the information on blank sheets of paper. Fewer than 25 percent of the students organized the data in the flat, row-by-column format that most software employs. Instead, students created nested structures, which organized the data into tables at different levels, typically indicated by headings or different spatial locations on the page. Thus, students seem to have an intuitive understanding of hierarchical or multilevel relationships, but not necessarily the means to formally represent and analyze hierarchically structured data sets. *CODAP* provides students with those very features. We designed the current study as a follow-up on Konold, Finzer, and Kreetong (2014) to see whether students could understand hierarchical structures that they themselves had not created, and furthermore, to learn about the types of supports students might require to come to this understanding.

Prior research has demonstrated the promise of interactive visualization software to help students work with flat, row-by-column data structures, such as that shown in Figure 2 (Chance, Ben-Zvi, Garfield, & Medina, 2007; Hancock, Kaput, & Goldsmith, 1992; Konold & Miller, 2011). Our research extends that work to multilevel data that are structured hierarchically (see Figure 3), representing attributes and outcomes at a higher level and raw measurements at a lower level.

	PlantID	Observer	Light	DaysToFlower	MaxHeight	Day	Height	Number_of_Leaves	Flowers
17	K	Jimar	shade	9	44	17	44	7	yes
18	K	Jimar	shade	9	44	18	44	7	yes
19	K	Jimar	shade	9	44	19	44	7	yes
20	K	Jimar	shade	9	44	20	43	7	yes
21	L	Hannah	sun	8	31	1	0	0	no
22	L	Hannah	sun	8	31	2	0.5	0	no
23	L	Hannah	sun	8	31	3	3	0	no
24	L	Hannah	sun	8	31	4	3.5	0	no
25	L	Hannah	sun	8	31	5	4	0	no
26	L	Hannah	sun	8	31	6	5	2	no
27	L	Hannah	sun	8	31	7	6	4	no
28	L	Hannah	sun	8	31	8	6	5	yes
29	L	Hannah	sun	8	31	9	9	5	yes
30	L	Hannah	sun	8	31	10	12	5	yes
31	L	Hannah	sun	8	31	11	15	5	yes
32	L	Hannah	sun	8	31	12	19	5	yes
33	L	Hannah	sun	8	31	13	26	5	yes
34	L	Hannah	sun	8	31	14	28	5	yes

Figure 2. Flat structure of observations of ten plants in *TinkerPlots*



Light	PlantID	Observer	DaysToFlower	MaxHeight...	Day	Height_...	Number_of_Leaves	Flowers
shade	K	Jimar	9	44	1	0	0	no
sun	N	Sofia	12	35.5	2	2	0	no
	P	Madison	9	40.5	3	3	0	no
	R	Olivia	9	39	4	3.5	0	no
	T	Kelsey	8	47	5	5.5	3	no
	L	Hannah	8	31	6	8	5	no
	M	Jacob	7	30	7	12	6	no
	O	Ryan	6	36.5	8	17	6	no
	Q	Mateo	7	33	9	21.5	7	yes
	S	Luis	10	25	10	23.5	7	yes
					11	30	7	yes
					12	33	7	yes
					13	38	7	yes
					14	39	7	yes
					15	40	7	yes
					16	42.5	7	yes
					17	44	7	yes
					18	44	7	yes
					19	44	7	yes
					20	43	7	yes
					1	0	0	no

Figure 3. Hierarchical structure of observations of ten plants in *CODAP*

Figures 2 and 3 show different representations of the same multilevel plant dataset. Figure 2 represents this multilevel data in a flat structure, while Figure 3 represents the same data in a hierarchically structured table. The flat structure includes one row for every observation and repeats values for attributes such as PlantID and Light. The table in Figure 3 organizes the plant data into three different levels: Conditions (Level 0, whether grown in sun or shade), Plants (Level 1, for plant-level attributes), and daily Observations (Level 2, measures of various attributes for each plant on each day). There is a nested, one-to-many relationship between the rows in Level 0 and Level 1 and between the rows in Level 1 and Level 2. In between the levels, swooping brackets with areas of alternating gray and white shading link particular rows to one another. For example, the white area connecting the first row under Conditions to the first five rows under Plants shows that Plants K, N, P, R, and T were grown in the shade. Similarly, the white area connecting the first row under Plants to the first 20 rows under Observations shows the 20 daily measurements for Plant K.

Typically, students work in software like *TinkerPlots*, *Fathom*, *Microsoft Excel*, or other programs that use flat data tables. However, software using flat data structures limits the types of graphs that one can create. For example, it is difficult to use the flat data table (Figure 2) to create a sensible graph to answer the question, “Did the plants in the sun grow taller than those grown in the shade?” A graph with MaxHeight on the y-axis and Light on the x-axis results in a confusing display with 20 points for each plant (see Figure 4a). In contrast, the hierarchically structured table (Figure 3) limits repetition and includes plant-level attributes, which results in a graph that is easier to interpret with one point for each plant (see Figure 4b). *CODAP* allows students to graph multilevel data in ways not possible with traditional data analysis software.

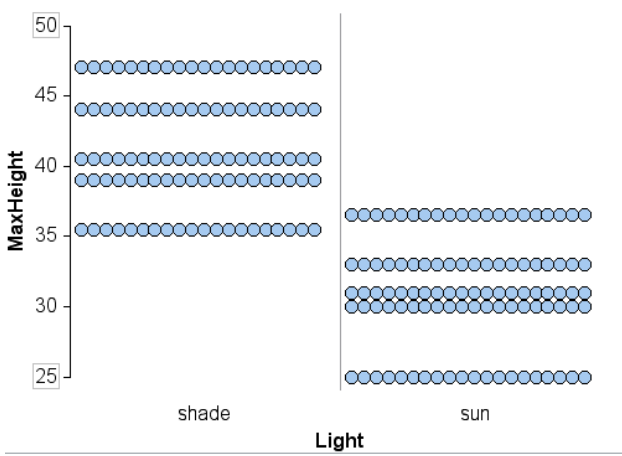
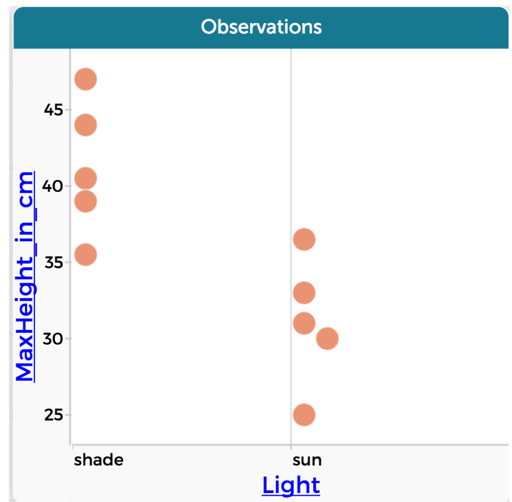
	
a. *TinkerPlots* graph generated from flat table	b. *CODAP* graph generated from hierarchical table

Figure 4. Comparison of graphs of MaxHeight vs. Light generated using a flat data table in *TinkerPlots* (4a) and a hierarchical data table in *CODAP* (4b)

In this article, we examine one student’s interpretations in *CODAP* of the two graphs in Figure 1 along with the hierarchical table in Figure 3. We focus on what appeared to confuse her and what helped her move her understanding forward. By “understanding” we mean her capacity to articulate what the individual points in the graphs represented — that is, to identify what they referred to in the real world.

Prior research has shown that students tend to have difficulty making connections between corresponding elements in multiple representations (Ainsworth, 1999; Seufert, 2003; van Someren, Reimann, Boshuizen & deJong, 1998). One of the affordances of *CODAP* is dynamic linking, or linked selection. When points or rows are selected in one representation, they are also selected in all other visible representations. This is shown in Figure 5. When a student clicks on a row or multiple rows in the table, the corresponding points in all the graphs become highlighted in blue. Similarly, when a student clicks on a point or selects multiple points in a graph, the corresponding rows in the data table become highlighted in pink. This linked representational model has origins in Kaput's (1989) work. Kaput applied a linked representational model to functions and graphs. In this form, a student could have three different windows in a computer environment: a graph window, table window, and an equation window. A change in any one window would result in changes in the other windows. KCP Technologies, as the developer of *The Geometer's Sketchpad* and *Fathom*, built on Kaput's work. They were pioneers in the exploration of the potential of dynamic manipulation and linked selection in improving students' conceptual understanding (Finzer & Jackiw, 1998; van Someren, et al., 1998). *CODAP*'s use of these technologies is an evolution of this prior work.

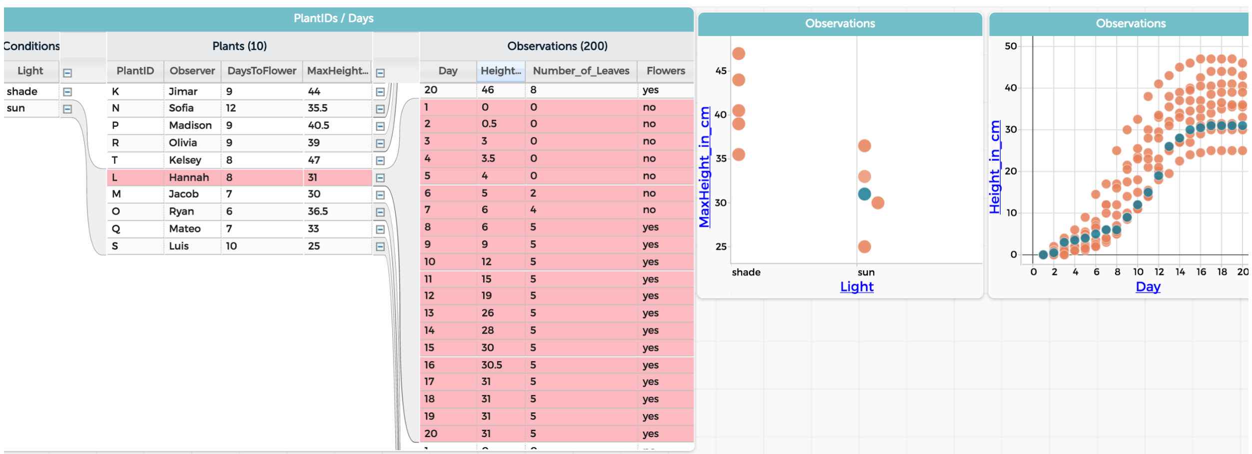


Figure 5. Three representations in *CODAP*: a table with three levels of data, and two graphs of data in the table. Note the dynamic linking among representations occurs when the user clicks on a table row or on a point in a graph. The row of plant-level data highlighted in Level 1 corresponds to the highlighted point in Graph 1, and the 20 highlighted rows of daily observations in Level 2 correspond to the curve of 20 highlighted points in Graph 2. Furthermore, the 20 rows highlighted in Level 2 are observations of the plant highlighted in Level 1, and the 20 highlighted points in Graph 2 represent observations of the plant highlighted in Graph 1.

In this paper, we examine the affordances of dynamic linking and other representational features for making hierarchical data structures accessible and understandable for students. Hierarchical or nested structures may be intuitive for students, but it is also important to understand the extent to which students can make sense of data that they have not collected themselves, in representations that use conventions such as those used in CODAP that they may not be familiar with. By tracing one student's shifting understanding as she interacts with the data and the software, we illuminate the difficulties she encountered, key features of the data analysis environment, and how the interaction among the student, the representations, and the interviewer's questions reveal and support her shifting understanding. This close analysis of one student's reasoning is part of a broader research program that focuses on the following research questions:

- What conceptual difficulties do students encounter in identifying what the cases are when working with multilevel, hierarchically structured data that they have not structured themselves?
- What strategies, representations, and reasoning processes do students draw on to make sense of the cases in multilevel, hierarchically structured data?

● What are the affordances of dynamic linking and other representational features in data analysis software for making hierarchical data structures accessible and understandable for students?

Answers to these questions can contribute to the field of data science education and inform the design of data-analysis learning environments.

2. METHODS

In this paper, we present a descriptive case study (Yin, 2002) that follows one 7th grade student's sense-making during her first encounter with the *CODAP* software. The detailed analysis of our interview with "Catherine" exemplifies some of the struggles we frequently observed as students worked with hierarchically structured data and highlights some of the strategies and representations that can support students in addressing those struggles. We selected Catherine's case because, while her sense-making processes were similar to those of other students, she was uniquely vocal and able to clearly articulate what she was thinking as she made sense of the data. By the end of each interview, nearly all of the students in the study came to understand the correspondences among the representations and their referents in the real world. We use Catherine's case to show how the representations and software features, like dynamic linking, helped students come to these understandings.

We selected Catherine's case from interviews with 31 7th and 8th grade students from the San Francisco Bay Area. We conducted a 45–60 minutes-long, semi-structured cognitive interview with each student, including planned questions and spontaneous probes to elicit student thinking (Appendix A). We recorded the interviews using software that captured each student's face, the student's and interviewer's voices, and the computer screen. We interviewed students as they used *CODAP* to analyze the data from observations of 10 Wisconsin Fast Plants®**,** which a fictitious class grew over 20 days. We used simulated data that both mimicked real plant data and met particular criteria for our interview study (we wanted the data to be relatively clean and to contain discernible patterns).

Each interview began with an introduction that provided students with a sense of the classroom context and the way the data were collected (see interview protocol in Appendix A). The interviewer showed students a time-lapse video of the growth of Wisconsin Fast Plants® from the time just after planting until the plants died (https://www.youtube.com/watch?v=JumEfAbjBjk). Then, the interviewer showed the students a sample data collection sheet (see Appendix B) and demonstrated how students in the fictitious class used the data collection sheet to record data for one plant over many days.

The body of the interview consisted of two main parts. First, each student interpreted a data table and two graphs (see Figure 6) that were specified in the interview protocol. The table was hierarchically structured with three distinct levels of data. The two graphs are the same as those in Figure 1 in the introduction. Graph 1 plots attributes from Level 0 and Level 1 of the table, and each point in Graph 1 corresponds to a measurement for one plant. In contrast, Graph 2 plots attributes only from Level 2 of the table, and each point in Graph 2 corresponds to an observation of one plant on one day.

In the second part of the interview, the students created and interpreted their own graphs. In this paper, we focus on the first part of the interview to understand how Catherine interpreted the researcher-generated data representations and how she drew on the representational features of *CODAP* to make sense of the data.

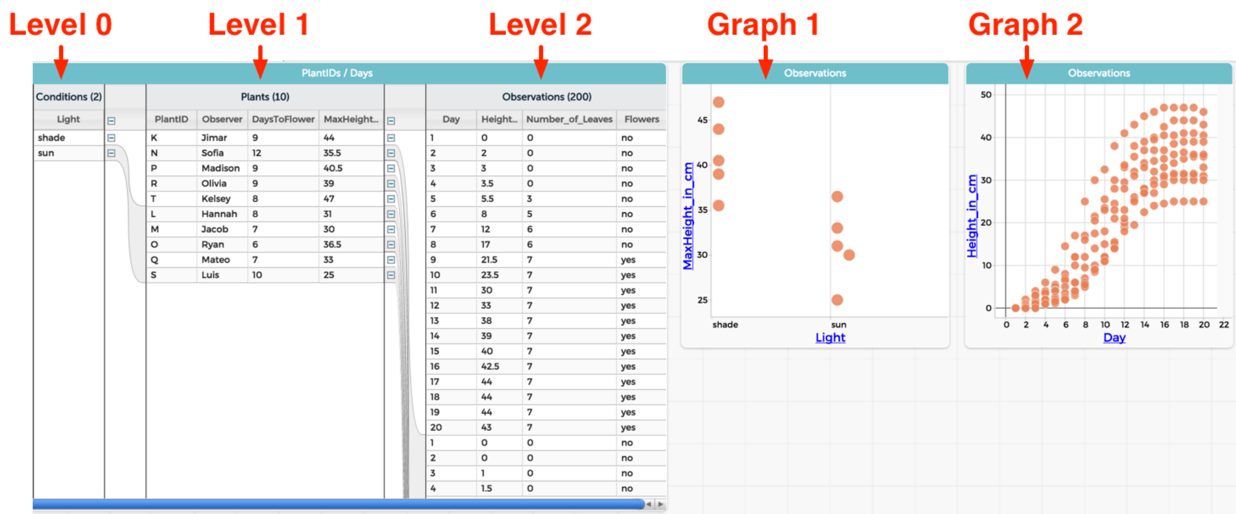


Figure 6. Three representations used in Part 1 of the interview protocol: a table with three levels of data, "Graph 1," and "Graph 2"

Catherine's interview began with an introduction to the three-level data table seen in Figure 2. The interviewer first demonstrated how to make graphs by creating Graph 1. (References to Levels 0, 1, and 2, and Graph 1 and Graph 2 are used here to support the reader's understanding of this analysis, but were not used during the interview.) Then, Catherine created Graph 2 under the guidance of the interviewer. The interviewer then asked Catherine to discuss some discoveries she could make about how the plants grew. She was prompted to "click on different things, scroll through the table, you can click on points, do whatever you need to do, and just explore all of this data here in front of you, and. . . tell me some of the discoveries you can make about what is happening here, and how these plants are growing."

In the sections that follow, we trace Catherine's shifting understanding of the meaning of the two sets of points: the plant-level data points in Graph 1 and the daily observation-level data points in Graph 2. We did not expect that she would immediately understand these representations. In our analysis, we treat her engagement with the representations as a sense-making process. We identify her strategies, the visual cues in the data analysis environment, and other factors that contribute to her evolving understanding.

Our analysis suggests that Catherine initially generated a series of unclear, tentative claims about what the points in the graphs represent. She later concluded her exploration of the graphs by confidently and clearly stating her correct interpretation of the points and their relationships to the table. These correct interpretations include:

- Each point in Graph 1 corresponds to a row in Level 1 of the table and to data about one plant in the experiment.

- Each point in Graph 2 corresponds to a row in Level 2 of the table and to an observation of one plant on one day during the experiment.

3. RESULTS AND DISCUSSION

The verbal responses, gestures, and actions within the *CODAP* environment during the interview were the primary data sources. The videotaped interview was transcribed and both the videotape and transcript were collaboratively analyzed in small research group meetings to draw inferences about Catherine's developing understanding of the data. We focus on how Catherine utilized features of the *CODAP* software to construct an understanding of the plant data and what the points in the graphs referred to in the real world.

The three episodes below demonstrate Catherine's shifting understanding of the meaning of the plant-level data points in Graph 1 and the daily observation-level data points in Graph 2. Episode 1 provides evidence of Catherine's initial, confused interpretation of the meaning of the points in Graph 2 prior to discovering interactive features of *CODAP*. In Episode 2, Catherine used the interactive features of *CODAP*, particularly dynamic linking, to identify correspondences across the graphs and the hierarchical table to correctly interpret the meaning of the points in the graphs. Episode 3 illustrates Catherine's progression from a static to dynamic interpretation of the representations in *CODAP*.

Catherine's sense-making process was similar to others in the sample of 31 students. By the end of the interview, 28 other students had worked towards and articulated a clear understanding of the points in Graph 1 and Graph 2, after exhibiting varying degrees of initial confusion. Like Catherine, these students made sense of the graphs by drawing correspondences between the levels of the table and each graph and with aspects of the experiment. The remaining two students in the study were inconsistent in their interpretation of the points, and even by the end of the interview, there was not clear evidence of a correct understanding of the meaning of the points. Students utilized dynamic linking in a variety of ways. Sixteen students made some accurate interpretations of the points in Graph 1 or Graph 2 without utilizing dynamic linking. Catherine was among the 15 students who relied more heavily on dynamic linking to develop a correct, stable and coherent interpretation of the points.

3.1 Episode 1: Challenges in Interpreting the Meaning of the Points

Before Catherine discovered the dynamic features of the *CODAP* software, her interpretation of the meaning of the points vacillated. She began the interview by exploring Graph 2, the Height vs. Day graph, where each point represented an observation, or measurement, for a plant on one day.

When the points first populated Graph 2, Catherine immediately reacted to the curve of points. She swept her cursor over the curve from left to right (from Day 0 to Day 20) and said, "You can see like how it grows over time." [6:55]

Then, with her mouse hovering over the points near the bottom of Graph 2 (see Figure 7), Catherine interpreted the graph to make her first claims about whether or not the plants grew on particular days.

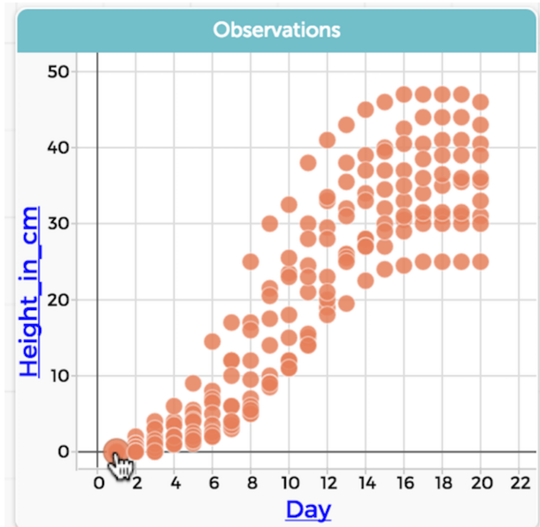


Figure 7. Catherine hovering her cursor over the single visible point on Day 1 in Graph 2

Catherine's first interpretation of Graph 2 suggests that only one plant's growth is represented when she uses the singular "it" to describe what is growing. However, she quickly followed this incorrect initial reading with a series of statements that reflect the accurate understanding that heights for multiple plants were collected over time. She noted that the points under consideration represented the heights of plants on the first and second day, and her reference to "they all" indicates she believed that multiple plants are represented in the graph. She correctly identified areas on the graph that represent Day 1 and Day 2, and she compared the heights of three different plants on Day 2. Note, however, that she switched from using the pronoun "it" when she described the single visible point at Day 1, to using "they" when interpreting the multiple visible points on Day 2. Because of Catherine's shift from a singular to plural pronoun, it is unclear whether she understood that the data were recorded for and represented *all* ten plants on Day 1.

Catherine's response to the interviewer's question of what each point represents also highlighted her confusion over the meaning of the points in Graph 2. Her response of "plants" suggests that she did not yet understand that the different points correspond to different measurements, not necessarily different plants. However, Catherine did begin to articulate an understanding of how the representations relate to the experiment that took place. She drew correspondences across the features of Graph 2 to the plant context to discuss how the points show the growth of the plants, but she did not yet explain where she would find one particular measurement or the set of 20 measurements for any given plant. Instead, she spoke more vaguely with inconsistent pronouns about how the points show heights.

We see additional evidence of Catherine's struggle in the next excerpt as she attempted to reconcile her expectation that the graph should show all of the daily observations with the fact that she saw only a few points on Day 1 and Day 2.

[12:48–13:15]

> C: And if you keep on counting it, this is it, this is it, and this is it [moves the cursor over a curve of points on Graph 2]
> I: So, all those points are for one plant?
> C: I don't . . . I'm guessing it is. It's kind of, it's really confusing for me because I don't know if this [gestures across many points on Graph 2] is ALL the plants.
> C: 'Cause this could NOT be all the plants.
> C: 'Cause I don't think they grew. . . That wasn't how many [hovering cursor over the few points shown at Day 1 and Day 2] –
> C: . . . Or maybe it's only the ones that grew that day. Hmm. Could be that.
> I: So it sounds like you're still trying to figure out what each point represents.
> C: Yeah, sort of. Because. . . . one of these [mouse over a point] could represent a couple—like three that grew the same height, but we wouldn't know.
> I: Oh, you mean one point could. . .
> C: . . . represent. . . Like this [points to point on Day 1] probably means that none of the plants grew that day, 'cause there's only one.

In this excerpt, Catherine again encountered some difficulty in interpreting the meaning of points in Graph 2 and how each point maps onto a real world referent from the plant experiment. Catherine considered how each point could correspond to one or to many plants for each day. First, she traced her cursor over a curve of points in Graph 2 as she said, "this is it, this is it, this is it," identifying individual points that seemed to her to represent consecutive measurements of the same plant. The interviewer prompted her to clarify her statement, "So, all those points are for one plant?" Catherine expressed her confusion and pointed to the handful of points near Days 1 and 2, stating, "that wasn't how many," indicating that there were fewer points than she had expected to see in the graph. She concluded that there were not enough points on the graph to represent all of the daily observations.

The single visible point representing a height of 0 centimeters on Day 1 was the source of confusion for several students in the study. The overlapping points became an obstacle for Catherine as she tried to make sense of what each point referred to in the real world. In *CODAP*, data points with the same values, like the ones on Day 1 of Graph 2, pile directly on top of one another, which gives the appearance of a single point on a static representation. In *CODAP,* the user can temporarily drag a point away from its location. After releasing the point, it animates slowly back into place. Thus, when a user drags away the point on the top of the pile at Day 1 of Graph 2, she sees another point under the original one, indicating that multiple data points can occupy the same location within the graph. Students who did not experiment with dragging points in this manner often misinterpreted what appears to be an individual point.

Throughout this episode, Catherine maintained a static orientation towards the representations, not utilizing *CODAP*'s dynamic capabilities. She did not drag points away to discover the overlapping points in Graph 2, nor did she click to highlight any of the representational elements, which could have prompted her to draw correspondences across the graphs and table and gain insight into the meaning of the points. From a static perspective, she made a reasonable guess that the point for Day 1 may represent more than one observation. In Episode 2, we trace

Catherine's progress after the interviewer encouraged her to click on the points, which revealed some of *CODAP*'s dynamic features.

Episode 1 highlights the fact that there are many relationships to understand within this representation-heavy data analysis environment, particularly when working with multilevel data: Students must figure out how the tables and graphs relate to each other, how the various representations relate to referents in the real world, what levels of data are involved, and how the different levels of data are represented in the tables and graphs. At various moments, students may have the correct understanding of some relationships but not others, and their understanding may change quickly as they work with the representations. Catherine moved quickly through several conflicting interpretations of what was represented in Graph 2 during Episode 1 alone. Due to the complexity of the representational system, students might not realize that they do not have a stable-enough or complete-enough understanding of the representations to use them productively. For example, Catherine and several other students in our study stated confidently that the points in Graph 2 represented "the plants," which is correct, but vague. When we probed further for an explanation of what each point on Graph 2 represented, students often struggled to state that each point represented the measurement of height for a single plant on a single day. This pattern suggests that students may benefit from spending time asking themselves focused questions about the precise mappings between features of the representations and their real world referents, particularly when the data include multiple levels.

3.2 Episode 2: Dynamic Linking Supported Catherine's Understanding of the Representations

Immediately following Catherine's misunderstanding about the overlapping points at Day 0, the interviewer told her that points can overlap and demonstrated how to view the points underneath by dragging the top point away. Before the interviewer finished the explanation, Catherine began her exploration of *CODAP*'s dynamic highlighting by clicking on, and thereby selecting, the point for Day 1 in Graph 2. This was a turning point in the interview, as it prompted Catherine's experimentation with dynamic features of *CODAP,* which supported her interpretation of the points in each graph.

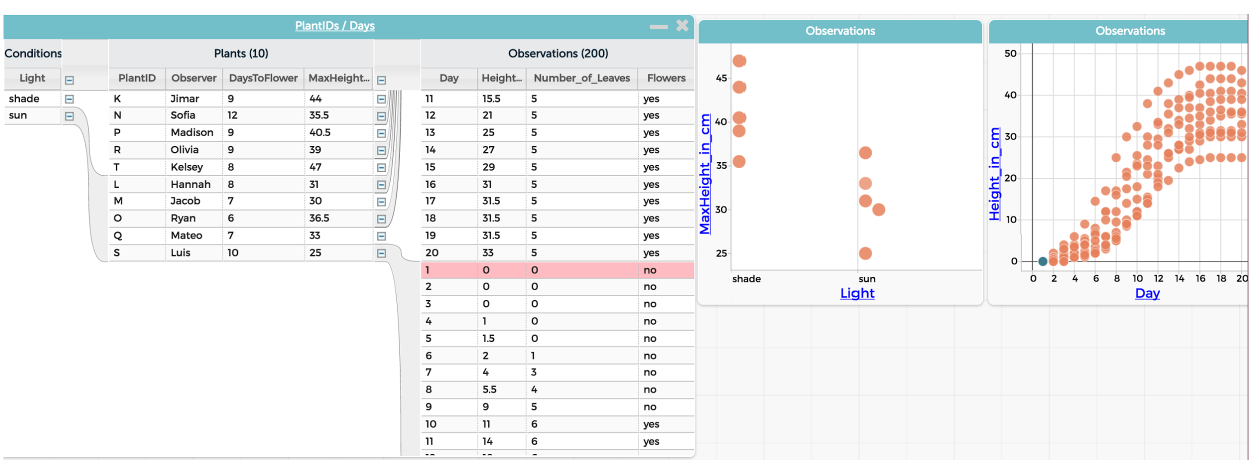


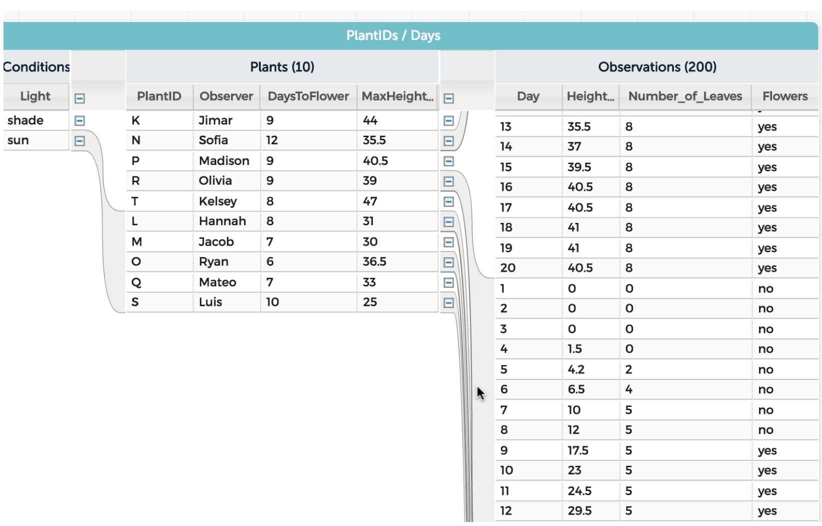
Figure 8. Catherine clicked on the point on Day 1 of Graph 2, which turned the point blue and simultaneously highlighted in pink the corresponding row in the table

[14:57–15:15]

C: This one is, this one on Day 1 [points to pink row in the table, see Figure 8], I don't know which, whose this one is or which plant, but I know it's this point right here [referring to blue point in Graph 2].
C: So I guess if I clicked on here [C clicks on the next row of the table] then I know that that's the one that grew 0 [looks at the corresponding blue point on Day 2 of the graph with a height of zero]. . . .
C: So if you click on these [points to the rows in Level 2] then you just know which one, how much it grew that day.

Catherine's use of clicking on the graph and table represent her first major shift from merely reading information from the graphs to using the *CODAP* software to interact with the data dynamically. In doing so, she discovered the correspondence between Graph 2 and Level 2 of the table. However, she continued to struggle with her understanding of how the representations relate to the experiment that produced the data. Here, she made correspondences across features of the representations without drawing on any information about the plants' growth, and she acknowledged this in some of her vague language, like "I don't know. . . whose this one is or which plant." Thus, Catherine's discovery of the dynamic linking feature supported her developing understanding of the meaning of the points in Graph 2, but she still needed to do additional work to continue to make meaning of the points in the context of the plant experiment.

In the next excerpt, Catherine learned how to scroll through the table. Like many students we interviewed, scrolling helped her recognize the function of the brackets connecting adjacent levels of the table as well as the one-to-many relationship between rows in adjacent levels.


PlantIDs / Days

Conditions

Light
shade
sun

Plants (10)

PlantID	Observer	DaysToFlower	MaxHeight...
K	Jimar	9	44
N	Sofia	12	35.5
P	Madison	9	40.5
R	Olivia	9	39
T	Kelsey	8	47
L	Hannah	8	31
M	Jacob	7	30
O	Ryan	6	36.5
Q	Mateo	7	33
S	Luis	10	25

Observations (200)

Day	Height...	Number_of_Leaves	Flowers
13	35.5	8	yes
14	37	8	yes
15	39.5	8	yes
16	40.5	8	yes
17	40.5	8	yes
18	41	8	yes
19	41	8	yes
20	40.5	8	yes
1	0	0	no
2	0	0	no
3	0	0	no
4	1.5	0	no
5	4.2	2	no
6	6.5	4	no
7	10	5	no
8	12	5	no
9	17.5	5	yes
10	23	5	yes
11	24.5	5	yes
12	29.5	5	yes

Figure 9. Catherine using her cursor to identify "lines dragging down" on the table

[17:46–19:05]

C: [Scrolling through Level 2 of the table after interview first informed her of the feature]. Okay, I think what they did here [Level 2] was like say, I might be getting this right, so. . . Olivia [moving cursor over Olivia's row in Level 1], this could be her data for her 20 days [referring again to Level 2].

I: Okay.

C: And then they marked it here on the points [referring to highlighted points in Graph 2]. And then, 'cause if you look here on this side [Level 1 of table]. . . then you can see these lines that are dragging down [see Figure 9] and I think that's dragging down to their data and what they got.

Catherine scrolled through the table to make sense of the relationship between Level 1 and Level 2, and through this exploration, she also began to understand the relationship between Graph 2 and Level 2 of the table. She first recognized the relationship between Level 1 and Level 2 by identifying the daily observations of Olivia's plant, and she explained that the brackets between the levels, which became interpretable as she scrolled through the table, supported this interpretation. At this time, Catherine also began to articulate an understanding of the hierarchy of the data by identifying the nested relationship between one row in Level 1 and 20 rows in Level 2 of the table. Catherine then elaborated on the relationship between Level 2 and Graph 2 by describing how the daily observations are indicated "on the points" by highlighting in Graph 2. At this point, she seemed to have a clearer understanding of the meaning of each point in Graph 2 and how each point corresponded to a daily observation of one plant.

As with Catherine, the dynamic linking and scrolling-table brackets features of *CODAP* encouraged many of the students in our study to look for correspondences across the tables and graphs, which helped them see the hierarchical or nested structure of the data. In Catherine's case, learning about the relationship between Level 1 and Level 2 helped her articulate her understanding of the data in Graph 2.

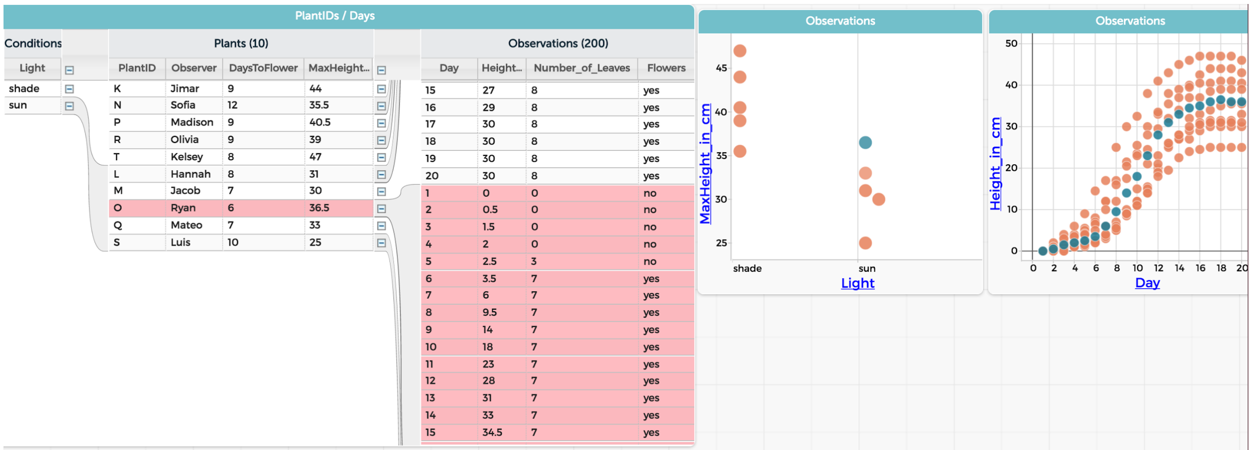


Figure 10. Dynamic linking across Level 1, Level 2, Graph 1, and Graph 2 that resulted after Catherine clicked on a point in Graph 1

In the next excerpt, Catherine gained a clear understanding of what each point represents in Graph 1 when she clicked on a point in that graph.

> [21:52–21:55]
>
> C: And here [Graph 1] I think they just generalized what people got.

Earlier in the interview, she had interpreted each point in Graph 1 as a “generalization” of sorts from Graph 2, but she could not explain what she meant. She was also unclear as to why there were fewer points in Graph 1.

> [21:56–22:23]
>
> I: What do you mean they generalized?
>
> [C clicks on data point in Graph 1. See Figure 10.]
>
> I: Oh, what just happened there?
>
> S: Oh wait, NOOO, they’re doing these people [moving cursor over Level 1 of the table]. OHHH, so this one [Level 2] is for this one [Graph 2]. This makes sense. Like these observations on here [Level 2] is on here [Graph 2]. And then, these observations here [Level 1] are over here [Graph 1]. Like this [referring to point in Graph 1] is for each person and I think, so this is their max height. This is what Ryan got. His max height [referring to blue point in Graph 1 in Figure 10 above].

When Catherine clicked on a point in Graph 1, the corresponding rows and points in Level 1, Level 2, and Graph 2 all became highlighted. At this time, she was able to look across the representations and identify the correspondence between Level 1 and Graph 1, as well as Level 2 and Graph 2. For the first time, she also referred to the points in Graph 1 as showing maximum height rather than a “generalization,” and recognized why there are fewer points in Graph 1.

In this example, a row from Level 1 corresponded with a point in Graph 1. Similarly, the hierarchical relationship between Level 1 and Level 2 corresponded to the hierarchical relationship between Graph 1 and Graph 2. The dynamic linking made these relationships visible to Catherine, which helped her describe what each point in Graph 1 represented and appreciate the relationship between Graph 1 and Graph 2. In doing so, she demonstrated an understanding of how the points in Graph 1 and Graph 2 showed data from different levels of the data.

> [22:27-22:50]
>
> C: [after clicking on the point in Graph 1 corresponding to Mateo's plant] This is Mateo's. And then you see that this is his data [moving cursor to the highlighted rows in Level 3 of the table] and then all the dots here [moving cursor over highlighted points in Graph 2] start lighting up because this is his data for how it grew. So his max height [moving cursor to the highest highlighted point] is the same [moving cursor to highlighted point in Graph 1]. If you look, his height, that's where it would be. It's the same as here [moving cursor back to highest highlighted point in Graph 2] if you look. So this is on his last day, that's how it grew.

Before this point in the interview, Catherine had used dynamic linking to discover the correspondences across the levels of the table and the two graphs. Drawing on those correspondences helped her understand how the data were structured and what the individual points represented. In this last excerpt, she discussed the correspondences across the representations *in relation* to the plant context itself. She clearly explained that the highlighted point in Graph 1 showed the maximum height that Mateo's plant reached, the set of 20 points in Graph 2 showed the growth of Mateo's plant, and the highest point in Graph 2 that corresponded to the highlighted point in Graph 1 showed the tallest height that Mateo's plant reached in the experiment. We claim that only at this point was Catherine actually "reading the data," and that before that she was mostly just associating points in the graphs to corresponding values.

In Episode 2, Catherine showed strong understanding of three key, interrelated aspects of the representational system: 1) she successfully related the symbols to their referents by describing the points on the graphs in the context of the plant growth experiment, 2) she moved across representations by describing the correspondences among the data table and the two graphs, and 3) she described the hierarchical nature of the data. While she may have shown limited evidence of these understandings during Episode 1, it was only with her active use of the dynamic linking tools in Episode 2 that her explanations became clear and more consistent. Her use of these tools not only supported her own understanding, but they also seemed to support her communication about what the representations showed.

3.3 Episode 3: Catherine Shifted from a Static Orientation to a Dynamic Orientation Toward the Representations

After experimenting with clicking on rows in the table and points in the graphs, Catherine described how treating Graph 1 as a static representation can be confusing.

[29:24–29:35]

> C: That's the only confusing part. It doesn't say who it is, so if this was printed out, I wouldn't know whose point was what, unless they were in different colors.
> I: Oh, the graph doesn't show whose it is. I see.
> C: Yeah, you have to 'cause if—this is on a computer, so you have to click it [clicks a point on Graph 1] and then you would know which one it is.

Catherine expressed confusion as she began describing Graph 1. Her description of a hypothetical scenario, in which she had a hard copy of the graph, signaled her static orientation toward the graph at that particular moment: one cannot manipulate a paper copy of a graph and there are no linked resources. From that viewpoint, she felt that the graph was confusing because it did not include information about the plants' observers. She seemed to realize in the moment that she was talking to the interviewer that she could click on a point to find the information she was looking for, and by clicking, she resolved her confusion about Graph 1. This exchange reveals Catherine's momentary orientation toward Graph 1 as a static object and illustrates her spontaneous shift to a more dynamic style of interaction with the data, facilitated by the *CODAP* dynamic environment.

4. CONCLUSION

4.1 Summary of Findings

In our analysis, we focused on Catherine's interpretation of the points in Graph 1 and Graph 2. She began the interview with an unclear and unstable understanding of what each point represented in Graph 2. Part of her difficulty stemmed from the way *CODAP* displays points with the same value. Initially, she did not notice, or at least use, the dynamic linking feature. However, after we encouraged her to click on the points and scroll, she discovered the dynamic linking feature, which made correspondences across the representations visually salient, and supported her understanding of the hierarchical relationships in the data table and two graphs. Catherine then used the details of the linked information from Level 2 of the table to interpret the meaning of the Graph 2 points. Similarly, when she then clicked on a point in Graph 1, she could describe the relationship between Graph 1 and Level 1 of the table, and she recognized that each point represented the maximum height for one individual plant.

This case study highlights the importance of *CODAP's* software features in supporting students in interpreting and analyzing data with interactive visualization software. Catherine began the interview with what we described as a static orientation to the table and graphs, evidenced by her difficulty in interpreting the overlapping points in Graph 2. In Episode 3 above, she even explained how treating Graph 1 as a static representation could be confusing for students. When Catherine discovered the linked selection, scrolled through the data, and interpreted the bracketing that linked levels across the table, her static orientation shifted to one that was more dynamic. This shift helped her identify and understand the correspondences across the table and graphs, which supported her in interpreting the points and what they corresponded to in the real world context of the plant experiment.

As we claimed in the introduction, “reading the data” involves more than being able to correctly read off the values of graphic elements. In our view, the work of coming to a basic understanding of a graph involves reconstructing the events that produced the data. In the plant scenario, that involves the fact that there were 10 different students collecting the data for their individual plants for twenty consecutive days. Furthermore, half the plants were grown in the shade and half in the sun. When first looking at Graph 1 and Graph 2, it is not possible to understand what each point represents without figuring out how they relate to the story of how the data came to be. The highlighting among the graphs and table helped interviewees reconstruct this story, in particular by locating in the hierarchical data table the observations that a particular student made. Indeed, many of the interviewees switched after they understood this structure, from talking about plants to talking about students. Before that, many interviewees had difficulty giving a name to the points. Even when they seemed to understand what the points in Graph 2 represented, they often struggled to find a meaningful way to describe them. But once they understood the data story, they tended to talk about these as the data each student collected.

In Catherine’s case, her interpretation of the graph moved from reading off values from graphic elements to a contextual-based interpretation of the graph. She successfully related the points to their referents, she drew correspondences between the graph and the table, and she described the hierarchical nature of the data in the context of the graph. She drew on CODAP’s dynamic features to improve her understanding of the graph and went beyond reading off the corresponding x and y coordinates for any given point to talking about the plants and how they grew. Thus, in our view, when Catherine was simply “reading the data” in the strict sense that Curcio (1987) defined it, she was not yet linking the information in the graphs to the actual experiment that was conducted. We recommend that a distinction be made between being able to “read off values from a graph” and “reading the data,” where the latter means linking those values to their real world referents.

This study also extends our prior research that showed students invented their own nested structures to organize complex data (Konold, Finzer, & Kreetong 2014). In the current study, we learned that students can come to understand data held in a hierarchical structure that they did not produce themselves and which makes uses of conventions they have likely never seen, such as the use of the table brackets to indicate nesting or linked selection to show relationships.

4.2 Limitations and Suggestions for Future Research

The scope of this paper was limited to one 7th grade student working with a particular data set, so we limit the claims we make and recognize that different students and data sets may result in different interactions in *CODAP*. For example, students with more or less experience with data or technology may interact with *CODAP* in other ways. Similarly, the context and type of data set may also play a role in students’ understanding. The size of the data set, for instance, may be related to how supportive the features of *CODAP* can be. In this case study, we do not consider variables like the size or type of the data set. Instead, we provide a detailed analysis of

Catherine's interpretation and analysis of data in *CODAP* to illustrate the potential power of linked representations, particularly for hierarchically structured multilevel data.

Additionally, the probe questions *may* have led Catherine down a particular path of thinking that she may have not otherwise explored. For example, the interviewer asked Catherine in Episode 1, "So, all the points are for one plant?" The interviewer asked the question to clarify what Catherine meant in her previous statements and gestures. It is difficult to say, though, when questions like this were used for clarifying purposes and when they may have guided student thinking, and this could be a threat to validity.

Further research must be done to address these limitations and to learn more about student thinking and how to best integrate software like *CODAP* into math and science classrooms. Some important next steps for research may include the following: (a) interview studies with different types of data sets and with different grade level students, (b) design-based research studies that inform the development of data science curricula that use data analysis platforms, and (c) an investigation of linked selection in other contexts.

4.3 Implications

Data literacy has been highlighted as an important component of math and science instruction, and this research advances our understandings of how students reason with multilevel data. Catherine's case highlights some of the difficulties students encounter when making sense of hierarchically organized data in a complex data analysis environment, and it identifies some of the strategies and supports students can use to do this work. Catherine's case also exemplifies the affordances of an environment that supports interaction with multiple, linked representations, in contrast to the limitations of static representations. The findings from this work can inform curricula, software development, and teachers' knowledge for teaching.

First, this study highlights the importance of shifting students' static orientation towards graphing to a more dynamic orientation. A static orientation to graphs may be common for students who are accustomed to simply displaying the results of their studies as opposed to using graphical displays to pose questions and explore relationships in data. Catherine initially approached the data with a somewhat static approach and even articulated the difficulty that such an approach presents. Once Catherine was instructed in point selection, she was able to use this feature to resolve her confusion about how overlapping data points were represented in the graph, and her work in *CODAP* shifted toward a more dynamic style. With the increasing role of technology in the classroom, it will be important to design curricula that help students come to see programs like *CODAP* not only as graph-making software, but more importantly, as empowering data analysis environments.

Second, this paper illustrates how *CODAP* features, such as dynamic linking and scrolling brackets, can support students in drawing correspondences across multiple representations in a data analysis environment and thereby build up an understanding of the larger story of how the data were created. As previously discussed, researchers have found that drawing such correspondences can be difficult for students (Ainsworth, 1999; Seufert, 2003; van Someren, Reimann, Boshuizen & deJong, 1998). Like *The Geometer's Sketchpad* and *Fathom*, *CODAP* continues to build on and extend Kaput's (1989) linked representational model in innovative ways. In this case, armed with the dynamic features within *CODAP*, Catherine was empowered to use the data analysis platform as a resource for constructing a basic understanding of the data. Software developers should therefore consider integrating similar features as they design data analysis environments.

Finally, instruction should build on students' understandings, and our findings corroborate prior work that suggests students have intuitive understandings of hierarchically structured data. Introducing students to hierarchically structured data in a variety of ways, with and without *CODAP,* could then support their developing understandings. For instance, not all teachers have access to computers, and those teachers should explore other ways to represent and discuss hierarchically structured data without *CODAP*.

Though there has been an increasing focus on data science education, data science remains understudied and additional contributions to the literature must be made to move forward. It is important that future research continues to include interviews and fine-grained analyses of students' thinking to better inform our understanding of how students reason with data.

REFERENCES

Ainsworth, S. (1999), “The functions of multiple representations,” *Computers & Education, 33(2-3),* 131-152.

Carpenter, P. A., and Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4(2),* 75.

Chance, B., Ben-Zvi, D., Garfield, J., Medina, E. (2007), “The role of technology in improving student learning of statistics,” *Technology Innovations in Statistics Education Journal, 1(1).*

Cobb, P., McClain, K., Gravemeijer, K. (2003), “Learning about statistical covariation,” *Cognition and Instruction*, *21(1)*, 1–78.

Curcio, F. R. (1987), “Comprehension of mathematical relationships expressed in graphs,” *Journal for Research in Mathematics Education*, *18(5)*, 382–393.

Finzer, W. & Jackiw, N. (1998), “Dynamic manipulation of mathematical objects,” From the NCTM Standards 2000 Technology Planning Conference, Washington, DC.

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.

Glazer, N. (2011), Challenges with graph interpretation: A review of the literature, *Studies in Science Education*, *47(*2), 183-210.

Hancock, C., Kaput, J. J., Goldsmith, L. T. (1992), “Authentic inquiry with data: Critical barriers to classroom implementation,” *Educational Psychologist*, *27(3),* 337-364.

Kaput, J.J. (1989), “Linking representations in the symbolic systems of algebra,” In S. Wagner & C. Kieran (Eds.), *Research agenda for mathematics education: Research issues in the learning and teaching of algebra*. Reston, VA: National Council of Teachers of Mathematics.

Konold, K. Finzer, W., Kreetong, K. (2014), “Students' methods of recording and organizing data,” Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Konold, C. and Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Higgins, T., Russell, S. J., Khalil, K. (2015), “Data seen through different lenses.” *Educational Studies in Mathematics*, *88(3),* 305-325.

Konold, C. & Pollatsek, A., Well, A. Gagnon, A. (1997), "Students Analyzing Data: Research of Critical Barriers," In J. B. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics: 1996 Proceedings of the 1996 IASE Round Table Conference* (pp.169-190). Voorburg, The Netherlands: International Statistical Institute.

Konold, C. and Miller, C.D. (2011), *TinkerPlots: Dynamic data exploration,* Emeryville, CA: Key Curriculum.

Lehrer, R. (2007). Introducing students to data representation and statistics. In J Watson & K Beswick (Eds.), *Mathematics: essential research essential practice (Proceedings of the 30th Annual Conference of the Mathematics Education Research Group of Australasia)* (Vol. 1, pp. 22–41). Adelaide: MERGA.

Lehrer, R., and Schauble, L. (2004), “Modeling natural variation through distribution,” *American Educational Research Journal*, *41(3),* 635–679.

Suefert, T. (2003), “Supporting coherence formation in learning from multiple representations,” *Learning and Instruction*, *12(2),* 227-237.

Van Sommeren, M. W., Reimann, P., Boshuizen, H. P. A., de Jong, T. (Eds.) (1998), *Learning with Multiple Representations*. Oxford: Elsevier.

Wild, C.J. and Pfannkuch, M. (1999), “Statistical thinking in empirical enquiry,” *International Statistical Review,* 67(3), 223-265.

Yin, R. (2002), *Case study research: Design and methods*. Thousand Oaks, CA: Sage Publishing.

APPENDIX A: INTERVIEW PROTOCOL

1. Explain the purpose and process of the interview

1	I study the way people use computer tools to make sense of things, especially how the information is visually organized on the screen. I'm going to have you work on the computer while I ask some questions. As you do the work, I want you to do what we call a "think aloud," which means you'll talk about what goes through your mind as you are doing the activities. For example, if you were clicking on something, you might say, "I'm not sure what this part means; I'm clicking here to see if it will tell me more." I'll ask you questions to prompt you as you go along. This isn't a test or a quiz, and there are no grades. We want to learn about out *how people think* and what *strategies* they use to figure out new information. I have software running that records the screen as you work, so if you're describing something, and if it's helpful for you to point, try to use the mouse to point with the cursor instead of just pointing with your finger, so when I watch the video later, I'll know exactly what you were pointing to. Anytime you want to stop doing the interview, you can just let me know.

2. Introduce the classroom context

2a	A middle school biology class grew several Fast Plants for a project. Fast Plants are a type of plant that grow really quickly—it only takes about 28 days for them to go through the complete life cycle. Here is a sped-up video showing the growth of Fast Plants that are the same type of plant that the class grew. You can see that they grow taller, the seed leaves or cotyledons grow (point them out), then it gets true leaves--those are the real leaves--and flowers (skip forward in video), and they eventually start to die. *[Show Time Lapse Video of Fast Plant Growth] (close the window)*

2b	This class wanted to study plant growth, so they put *some* Fast Plant seedlings in the sunlight and *some* of the seedlings in the shade, and they recorded data about each of the plants as they grew. Each student was assigned to one plant. They used data collection sheets like this to collect their data for each plant. *[Show data collection sheet sample.]* Let's fill this data collection sheet out for one day. I'm going to put made-up information here, just to show how the form works. Let's pretend this is the plant I'm observing [Nicole/Lina]; It's 5th period [5th]; And let's say the plant I am observing was grown in the shade condition [circle shade]. Let's say the first day of measuring is today, so we'll put today's date in [10/24/15]. On the first day of planting, the plant wouldn't have grown yet, so we'll put 0 centimeters for height [0]. Now let's move to number of leaves. That's the number of leaves the plant had on that day, which would also be 0 if this is right after we planted the plant [0]. And let's say 'no' to flowering because the plant wouldn't have any flowers on it yet either [N].

3. Introduce graph construction in *CODAP*

[Show CODAP]

The teacher collected students' data sheets and entered the data into this table for analysis. The students made graphs to help them understand *how* these plants grew.

3	*[Interviewer makes first graph while narrating its construction]* Let's make two graphs that the students came up with to see how graphing works in this software (open one new graph). I'll make one and then you'll make one. This one is going to have Light on the x-axis, so I'm going to click and drag "Light" from here (case table) and drop it near the bottom (drag variable to graph). See how that box pops up? . . . and then I'm going to drag MaxHeight onto the y-axis (drag variable). *[Interviewer invites student to create second graph, but interviewer continues to direct the content of the graph & provide support.]* Ok, now it's your turn. Click on this icon to create a new graph. And now let's click and drag "Day" onto the x-axis at the bottom (wait). . . And place "Height" on the y-axis.

4. Invite student to interpret existing graphs and table

4a	*[Prompt students to explore the data and make interpretive statements]* Imagine that you're a student in the class, trying to figure out how the plants grow. Feel free to use the mouse to click around to explore. As you explore all of these data *[gesture across the whole screen]*, tell me some of the discoveries you make about how the plants grew. What can you say about how these plants grew?
4b	*[Probe to understand student's strategies for making sense of the data. Use these probes throughout the interview.]* • How can you tell? How do you know? How did you figure that out? • I noticed that you were (clicking on rows in the table). Can you tell me more about what you were doing? • What are/were you looking at? • What are you trying to figure out? • How does this graph (or table) show you that? • What did you hope it was going to show you? [*if something didn't happen as expected*]
4c	*[Prompt for multiple interpretive statements about the data.]* • What other discoveries can we make about how these plants grew? *[Use more directed prompts for each graph below if the student does not spontaneously talk about each plant. Use the sub-prompts as suggestions only if students appear to have difficulty making interpretive statements about the graph. Otherwise, continue to use more generic prompts.]* What can we learn from this graph (MaxHeight vs. Light)? • How is the data similar or different for the sun and shade groups? What can we learn from this graph (Height vs. Day)? • How does the height change over time? • Do all the plants grow to the same height? • Is there a way to tell if the plants all grew at the same speed?

4d	[*Prompt for specific understanding of the hierarchical organization of the table, including the meaning of the case at each of the three levels.]* • How are the data organized in this table? • What do these three sections show? • Tell me about this part [*point to the swooping column between Plants and Observations]*. What does it do? • Where can I find the data for the plants in the sun condition? • Where can I find the data for plant P? • Where can I find the height of plant P on day 5? [*Ask the following questions for each level.]* • What does this row [*select any row*] tell us? • What does each row represent? [*Probe for specific understanding of the meaning of the variables only if it seems unclear or if there are problems.]* Let's go through, column by column, and tell me what these mean to you. (Point to the top cell in each column, moving from right to left. Let's start with Light: ● What does it mean? ● What are these numbers below that word? (if applicable) ● How can you tell?
4e	[*Probe for student understanding of the case in the graphs]* • Why are there fewer points in this graph (compare graph 1 to graph 2)? • What does each point represent in this graph? • Is this data about one plant or all the plants?

5. Introduce new features in *CODAP*

	I'm going to have you create some of your own graphs, but first, I want to show you a few more features of the graphing tool. [*Guide student through how to do the following:]* • *create a new Height vs. Day graph* • *add Flowers to the middle of this new Height vs. Day* • *marquee select all the plants in the Sun group in the MaxHeight vs Light graph, point out highlighting/linking* • *remove Height from the y-axis* • *add PlantID to the y-axis* • *close graph*

6. Invite student to generate and explore own questions

6	Imagine that you are in this science class, and you're trying to learn something else about the plants. Make a few graphs to help you figure out something else about how the plants grow. *[Let them play around for a while. Provide technical support as needed. Choose the most sensible graph (use your judgment) and ask follow up questions below.*] • What does this graph show you about how the plants grow? • What does each point represent? • Did you use the table to help you think about what graph you wanted to make? If so, how? • What made you decide to do that? • Pick a particular variable (include prompts about the choice of level if they attend to that) • Drag something to a particular place
	[Support students with directives if they can't come up with any other ideas.] • Direct the student toward a variable s/he has not yet explored (e.g., Days to Flower or # of Leaves) • Create and invite student to interpret a hard-to-interpret 1-variable graph, like the distribution of days to flower (sensible) or height (really hard). • "Students made this graph and were trying to figure out what it meant."

7. General prompts about technology (for use throughout interview)

7	*[Use technical support prompts, nudging them to use features that we care about, if users don't spontaneously use these features]* What happens if you…What happens if you. . . ● Click on different points? ● Select a group of points? ● Scroll through the table? ● Select a row in the table? (left/right) What do you notice about the. . . ● Highlighting? (What happens to the other graph? The Table?)

APPENDIX B: DATA COLLECTION SHEET SAMPLE FROM INTERVIEW PROTOCOL

Fast Plant Data Collection Sheet

Name:

Class Period:

Growth Condition: Sun or Shade (circle one)

Plant ID:

Observation Date (mm/dd/yy)	Height (cm)	Number of Leaves	Flowers? (Y/N)	Notes