Engaging Everyday Science Knowledge to Help Make Sense of Data

Susan B. Kelly, University of Illinois Urbana-Champaign, sbkelly2@illinois.edu
LuEttaMae Lawrence, University of Illinois Urbana-Champaign, llawrnc2@illinois.edu
Emma Mercier, University of Illinois Urbana-Champaign, mercier@illinois.edu

Abstract: Making sense of data to inform decisions is an important skill emphasized in current curriculum documents (NRC, 2012). Making sense of data through personal experiences and prior knowledge is one way that students can begin to understand multiple and unfamiliar data sources. This paper examines how middle school students used different data sources when engaged in a collaborative problem solving activity using a multi-touch table during classroom science instruction. In this study, we found that students made personal connections when talking about data. Students engaged in data talk across all conversation quality levels, but the ways students interacted and talked about data varied. Connecting to students' everyday experiences could provide an access point for more complex science content understanding and synthesis and improve student data literacy skills.

Keywords: collaborative learning, data literacy, contextualizing science instruction

Introduction

Researchers report that students struggle to make sense of data. They have difficulty making sense of graphs and patterns (Schauble et al., 1995), they draw conclusions without evidence, and do not use data to support their claims (Sadler, 2004). Students' preferentially use personal knowledge and experiences to explain scientific phenomenon, rather than data (Germann & Aram, 1996). Researchers also identify the need to connect school science to everyday experience, recognizing that learning in school can be irrelevant or abstract (Aikenhead, 2006). This issue can be addressed through place based, problem based, and contextualized curriculum efforts (Rivet & Krajick, 2008; Warren, et al., 200l) that aim for more "connected science" (Bouillion & Gomez, 2001). Engaging students in the analysis of data is one way to help students make connections between school and everyday life, and improve data literacy and the relevancy of science information. This paper examines how students make sense of data by examining the conversations students have around a contextualized science task with everyday implications—how food choices impact the environment.

Involving students in tasks that address real-world problems that they can authentically connect with may engender interest and motivation; there is also evidence that constructing understanding using technology in groups improves learning outcomes (e.g., Mercier & Higgins, 2013). Collaborating provides learners with opportunities to identify patterns and communicate with others to understand a process, create a product, or reach consensus. Group activities, when properly structured, enable students to discover deeper meaning in the content and improve thinking skills. Effective collaborative activities draw on social constructivist frameworks, and often use ill-structured problems—those tasks which engage students with higher-level content that is thought-provoking, difficult to understand, and have multiple possible answers (Barron & Darling-Hammond, 2008).

Computer-supported collaborative learning is one way that students can access and make sense of multiple data sources at the same time. Multi-touch tables allow multiple users to manipulate data directly, and permit more equitable participation, supporting the construction of joint knowledge about a problem (Mercier & Higgins, 2014). Multi-touch tables can provide access to multiple data sources simultaneously, so students have the opportunity to make sense of patterns and relationships between data that is otherwise difficult to synthesize.

We hypothesized that connecting to students' everyday experiences through the use of a contextualized data-driven activity could be one way to use students' prior knowledge as an access point for understanding more complex science content. This study was designed to identify and characterize the conversations around data when working on a collaborative task using a multi-touch table, after six days of classroom instruction. The research questions addressed in this paper are:

1) How do students talk about data when working with multiple sources of data on a multi-touch table?
2) What data topics do students discuss when working on a task focused on the footprint of food?
3) What is the level of data synthesis reached when students engage in a data-driven collaborative task?

Methods

This study was designed as the first phase of a design-research project. Members of the research team led seven days of classroom activities focused on climate change. The activity that is the focus of this paper, took place on the sixth day of the intervention.

Study participants were drawn from 63 students from three 7th and 8th mixed-grade science classes at a local selective public school. All students participated in all activities; data was collected from 11 groups of students, where every student in the group had parental consent to participate.

The *Food for Thought* app was created to be used on large, horizontal multi-touch screens, and designed as an ill-structured problem, with many possible answers to encourage discussion within groups. The task centered around the creation of an environmentally sound meal. Twenty-two different foods were visible on the screen (see Figure 1). For each of the foods, the water footprint, carbon footprint, calories, and cost were calculated. As students placed the food on a dinner plate, the metrics for each category appeared on individual bar graphs for 3 types of data, and as a list for price. All members of the group could interact with multiple data sources at once. The dinner plate remained anchored in the center of the screen while the individual graphs and foods could be moved anywhere on the table. The graphs could be reduced or enlarged in size, and rotated.

Figure 1. Food For Thought App Screenshot

During the preceding class sessions, students covered content related to the carbon and water footprint of food through a variety of activities. While using the application, students were led through three activities. Students were asked to assess and coordinate the various production costs of food using the data provided in the app and to apply that information to each task. Tasks, which increased in complexity, were 1) create your favorite meal; 2) create a dinner that includes a protein; students were asked to swap out proteins and evaluate the data; 3) create a meal that you think is best for the environment. For the purpose of this paper, only task 3 will be analyzed.

Data Sources and Analysis

Data was collected in a lab classroom that was equipped with video recording equipment, due to technical issues with one video, only 10 videos were used; videos were transcribed in playscript form. Emergent coding schemes were developed to account for students' discourse around data. The analysis proceeded in five steps. First, turns were coded to identify data-focused talk. Next, the data talk was coded as either derived from the information contained in the app, or from students' prior knowledge (Table 1). Two researchers coded 20% of the transcripts for data talk with an inter-rater agreement of 98% and Cohen's Kappa of .96.

Table 1: Data definitions as applied to turns

Code	Definition	Examples
App Data	• Values of the food from the app with or without unit designations • Information from within the app • Direct responses after app data statement	• "Beans are 142. What about eggs?" • "Steak is high", "The price went up", "Steak, oh no" (pointing at the graph and looking at the values) • "Steak has a lot of carbon", "No it doesn't"
Prior Knowledge	• Comments based on data not included in the app • Direct responses after prior knowledge statement	• "Beans are a good source of protein" "Beans are healthy", "Spaghetti is bad for you" • "Bananas have to be imported", "Yeah, I know"

In the third stage, data talk was organized by episodes of data talk; episodes were defined as discrete conversation turns about the same topic. Conversations that were happening concurrently were considered to be within the same episode. Data episodes were chosen as a unit of analysis in order to examine when and how personal data was incorporated into conversations and to identify instances of data synthesis. For the fourth stage, data episodes were grouped by the data type contained in each episode; app data only, prior knowledge data only, or as mixed data, when a data episode contained both prior knowledge and app data. Two researchers coded 30% of the transcripts for reliability, with a Cohen's Kappa of 0.87. Because the length of a data episodes varied, a variable was created to account for the proportion of total turns. Each data topic from the table (water, carbon, calorie and cost) that was referred to during the task was counted each time it was used explicitly, identified either by name or its associated value.

A second emergent coding scheme was applied to data episodes to characterize how students referenced data in conversation, and the highest level of synthesis achieved in each episode. This coding scheme identified data synthesis as Low, Medium or Medium-High (Table 2). A synthesis designation, achieved by tabulating episodes, includes both the most frequent and the highest level each group achieved in combination. Thirty percent of the transcripts were coded independently for episode by a second researcher; inter-rater agreement was 82% and Cohen's Kappa was 0.79.

Table 2: Data synthesis coding framework

Category	**Description**	**Examples**
Low	No explicit reference to data in conversation, or the data value is read directly from the table without reflection (an extension of an idea from the data)	"…rice doesn't have much"; "bananas were high" "605"
Medium	Explicit use of data from table or personal experience without specific values; some reflection using data	"So the thing that needed the most water was steak"
Medium High	Data talk is explicit and connected to more than one data type. Some data synthesis.	" if we are going to make it for three meals we need more than 600 calories, and it uses a lot of water"

Results

The total proportion of data talk by turn varied among groups, with a range between 14% (Group 8) and 54% (Group 6). Data talk comprised a little more than one quarter of the turns of group discussion for half of the ten groups. Results from data topic (CO_2, H_2O, cost, calorie) tabulation indicated that all but one group refered to at least one data topic explicitly during task. Four out of ten groups referenced two topics, with half of the groups using three data topics while building an environmentally friendly meal. None of the groups referenced all four data topics. Data topic(s) discussed varied across groups. CO_2, and H_2O data were referenced by six groups, while cost was mentioned by two groups, and referenced least. Only one group (Group 6) referred to a data topic (CO_2) twice during the task. None of the groups used the unit of measurement associated with either water (gallons) or carbon (CO_2 equivalents), in discussion during the task.

Groups engaged in between 3 and 5 episodes of data talk while participating in the task (Mean = 3.90, SD = 0.74). Data conversations that resulted from information from the app alone characterized almost half of the 39 total data talk episodes (49%). Eight of ten groups used prior knowledge when making sense of the data, either in a stand alone statement (15% of data talk by episode) or as part of discussion which integrated prior knowledge with the data from the table (36% of data talk by episode). When taken together, data from prior experience, alone or in combination with table data, constituted 51% of the total data talk when analyzed by episode. One group (Group 8), did not use prior knowledge at all during the activity, instead relying solely on information provided within the app to make decisions. Three groups (4, 6, and 7) did not reference the table data explicitly in conversation, unless it was used in combination with prior knowledge when building a meal. Two of these groups (4 and 6) were the only groups that achieved medium-high synthesis of the information during the task. The remaining group, (Group 7) reached medium synthesis during data conversation.

Nine of the ten groups engaged in low synthesis data talk, which made up 41% of the total data episodes identified. While all groups participated in one or more instances of medium quality data talk (51% of all episodes), only two of the ten groups (Groups 4 and 6) engaged in medium high data synthesis,which was identified in only three episodes, comprising 8% of the total. Group 4 did not engage in any low synthesis data talk, and instead employed medium and medium high talk in discussion. All episodes of data talk for this group involved mixed data talk, where data from both the app and prior knowledge were used in conversation during the task. Group 6 used prior knowledge in one of three data episodes, and mixed data talk in the remaining two episodes.

On average, groups that engaged in higher levels of synthesis also engaged in more data talk; groups that reached lower synthesis designations talked less. Four groups were identified as low-medium synthesis, and mean percent of data talk across these groups was lowest (M = 28.80, SD = 8.79). While all groups had conversations with at least one episode of medium data talk, the four groups characterized by the largest proportion of medium synthesis data episodes also comprised the group with the intermediate amount of data talk (M = 41.00, SD = 10.03). The two groups that were classified as attaining medium-high synthesis also sustained the highest percentage of data talk on average (M = 48.69, SD = 12.29).

Conclusions and Implications

In this study, results indicated that amount of data discussion, the explicit use of data in discussion, and synthesis across data topics was low. We found that prior knowledge was an important component of the data discussions that did take place, and that eight of the ten groups used prior knowledge when talking about data, across all conversation synthesis levels. This aligns with prior research that indicates that connections to everyday experience may be one way that students interact with complex data (Rivet & Krajcik, 2008; Warren, et al., 200l). It is possible that the students who were less experienced in making claims from data used prior information as an access point for understanding the novel data, and that those students that reached higher data synthesis levels also used prior knowledge, or the combination of prior knowledge and data from the app, to grapple with a socioscientific issue, although further research is needed to examine this finding. We also found that while the number of data episodes was similar across groups, the amount of time spent in data conversations was correlated with the level of data synthesis achieved; groups that talked longer also reached higher levels of synthesis.

These results indicate that some groups of students engaged in some complex discussion of data sources related to the impact of food on the environment, using both the data provided to them and their own prior knowledge. Future research will examine how an individual student's prior knowledge supports and sustains, or hinders, a groups' conversation with and about data. This study will inform further development of the task to support the incorporation of prior knowledge, and how the task can more fully support students' engagement with data, while still maintaining an ill-structured format.

References

Aikenhead, G. S. (2006). Science education for everyday life: Evidence-based practice. Teachers College Press.

Barron, B., & Darling-Hammond, L., (2008). How can we teach for meaningful learning? In: Darling-Hammond, L Barron, B., Pearson, P. D., Schoenfeld, A. H., Stage, E. K., Zimmerman, T. D., ... & Tilson, J. L., *Powerful learning: What we know about teaching for understanding (pp. 11-70).* John Wiley & Sons.

Bouillion, L. M., & Gomez, L.M. (2001). Connecting school and community with science learning: Real world problems and school–community partnerships as contextual scaffolds. *Journal of research in science teaching, 38*(8), 878-898.

Germann, P.J., & Aram, R.J. (1996). Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. *Journal of Research in Science Teaching, 33*(7), 573–798.

Mercier, E. M., & Higgins, S.E. (2014). Creating joint representations of collaborative problem solving with multi-touch technology. *Journal of Computer Assisted Learning, 30*(6), 497-510.

Mercier, E. M., & Higgins, S.E. (2013). Collaborative learning with multi-touch technology: Developing adaptive expertise. *Learning and Instruction, 25,* 13–23. doi:10.1016/j.learninstruc.2012.10.004

National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academy Press.

Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. Journal of Research in Science Teaching, 45(1), 79-100.

Sadler, T.D. (2004) Informal reasoning regarding socioscientific issues: A critical review of research. Journal of Research in Science Teaching, 41(5), 513-536.

Schauble, L., Glaser, R., Duschl, R.A., Schulz, S., & John, J. (1995). Students' understanding of objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Science, 4*(2),133-166.

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of research in science teaching, 38*(5), 529-552.