Journal of Pre-College Engineering Education Research (J-PEER)

Volume 7 | Issue 1 Article 5

2017

Students' Successes and Challenges Applying Data Analysis and Measurement Skills in a Fifth-Grade Integrated STEM Unit

Aran W. Glancy
Purdue University, aglancy@purdue.edu

Tamara J. Moore
Purdue University, tamara@purdue.edu

Selcen Guzey
Purdue University, sguzey@purdue.edu

See next page for additional authors

Recommended Citation

Glancy, A. W., Moore, T. J., Guzey, S., & Smith, K. A. (2017). Students' Successes and Challenges Applying Data Analysis and Measurement Skills in a Fifth-Grade Integrated STEM Unit. *Journal of Pre-College Engineering Education Research (J-PEER), 7*(1), Article 5.
https://doi.org/10.7771/2157-9288.1159

Students' Successes and Challenges Applying Data Analysis and Measurement Skills in a Fifth-Grade Integrated STEM Unit

Abstract

An understanding of statistics and skills in data analysis are becoming more and more essential, yet research consistently shows that students struggle with these concepts at all levels. This case study documents some of the struggles four groups of fifth-grade students encounter as they collect, organize, and interpret data and then ultimately attempt to draw conclusions or make decisions based on these data. The activities in which the students engaged were part of an integrated science, technology, engineering, and mathematics (STEM) unit that had students collecting and analyzing data both in the context of learning science concepts and in the context of evaluating prototypes for an engineering design challenge. Students were observed to struggle in a variety of ways, specifically having difficulty (1) properly using certain measurement devices, (2) coordinating quantitative data with the phenomenon being measured, and (3) properly interpreting the significance of variation, uncertainty, and error in the data. Implications for teaching and curriculum design are addressed.

Keywords

Data Analysis, Measurement, STEM Integration, Engineering Design

Document Type

Invited Contributions: Best Papers from ASEE Pre-College Engineering Education

Cover Page Footnote

This paper is part of a set of papers recognized by the American Society for Engineering Education's Pre-College Engineering Education Division as one of the division's best papers from the 2015 and 2016 conferences. This is the first set of ASEE Pre-College Engineering Education Division best papers J-PEER has republished, with permission from the American Society of Engineering Education.

Authors

Aran W. Glancy, Tamara J. Moore, Selcen Guzey, and Karl A. Smith

Available online at http://docs.lib.purdue.edu/jpeer

Journal of Pre-College Engineering Education Research 7:1 (2017) 68–75

Students' Successes and Challenges Applying Data Analysis and Measurement Skills in a Fifth-Grade Integrated STEM Unit

Aran W. Glancy,[1] Tamara J. Moore,[1] Selcen Guzey,[1] and Karl A. Smith[2]

[1]Purdue University
[2]University of Minnesota

Abstract

An understanding of statistics and skills in data analysis are becoming more and more essential, yet research consistently shows that students struggle with these concepts at all levels. This case study documents some of the struggles four groups of fifth-grade students encounter as they collect, organize, and interpret data and then ultimately attempt to draw conclusions or make decisions based on these data. The activities in which the students engaged were part of an integrated science, technology, engineering, and mathematics (STEM) unit that had students collecting and analyzing data both in the context of learning science concepts and in the context of evaluating prototypes for an engineering design challenge. Students were observed to struggle in a variety of ways, specifically having difficulty (1) properly using certain measurement devices, (2) coordinating quantitative data with the phenomenon being measured, and (3) properly interpreting the significance of variation, uncertainty, and error in the data. Implications for teaching and curriculum design are addressed.

Keywords: Data Analysis, Measurement, STEM Integration, Engineering Design

An ability to collect, organize, make sense of, and interpret data has long been an essential skill not just in engineering but in almost every science, technology, engineering, and mathematics (STEM) field. With the increasing availability of large data sets quantifying everything from world health and poverty to baseball statistics or shopping habits, however, data handling skills have earned a place as an essential 21st-century skill (Lohr, 2012; Wilkins, 2000). Statistics educators have long been working on understanding how best to help students develop these skills in mathematics and statistics classes; however, recent initiatives to increase opportunities for K-12 students to engage in engineering as well as integrated STEM activities (Committee on Integrated STEM Education, National Academy of Engineering, & National Research Council, 2014; NGSS Lead States, 2013) may provide another venue through which students can learn and apply data handling techniques. Unfortunately, there is much evidence to suggest that our students are not learning these skills at the level necessary to be able to apply them in their careers or daily lives. TIMSS (Trends in International Mathematics and Science Study) data indicate that data analysis and statistics are areas of weakness for U.S. students (Wilkins, 2000), and Kuklianksy and Eshach (2013), for example, found that undergraduate students in science and engineering courses had difficulty with everything from choosing appropriate representations of data to understanding and accounting for measurement error.

Measurement, data collection, and data analysis are essential elements of many science and engineering activities; thus, when students engage in laboratory investigations or test and evaluate engineering designs, they must apply what they know about data analysis in realistic situations. Because these contexts are more applied and more realistic than what students encounter in typical mathematics and statistics classes, however, data analysis tasks embedded within STEM activities can

http://dx.doi.org/10.7771/2157-9288.1159

create different obstacles for students as well as potentially creating new opportunities to learn. Thus, understanding the ways in which students engage with data in applied engineering and science activities is an important step in helping to maximize the learning opportunities inherent in integrated STEM settings. In order to gain some insight into this process for fifth-grade students, this case study follows four groups of students through several data analysis tasks during an integrated STEM unit centered around an engineering design challenge.

Literature review

In recent decades, statistics educators have made much progress in determining what students need to know and how best to develop their abilities in the domain of data analysis. As noted by Garfield et al. (2008), statistics began to rise in prominence within K-12 classrooms following its addition to the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics (1989). Since then, educators and researchers have learned much about how students think and learn data analysis skills (Garfield, 1995; Garfield & Ben-Zvi, 2007; Shaughnessy, 2006; Watson, 2006). According to these researchers, along with many others, the primary goal of K-12 statistics education should be *statistical literacy*. Gal (2004) provides one articulation of the concept, describing statistical literacy as

> (a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions. (p. 49)

Although other definitions of statistical literacy differ in important ways, the key elements remain largely the same.

In investigating how to support and develop statistical literacy, research within the field of statistics and mathematics education has identified important pedagogical principles and established the importance of some central concepts in statistics. Recommendations include engaging students in authentic activities (Moore, 1998), and allowing them to make sense of data on their own through hands-on activities (Garfield et al., 2008; Watson, 2006). For example, Lehrer, Kim, and Jones (2011) showed that having students design their own statistics for data that they had collected themselves encouraged deep consideration of measurement and variability, and these authors recommended this as a pedagogical approach for teaching statistics.

Science and engineering educators make similar recommendations for developing data analysis skills in applied settings. Bybee (2011) argues that "planning and carrying out investigations should be standard experiences in K-12 classrooms" (p. 36), and Hofstein and Lunneta (2004) found that the literature consistently supports these kinds of tasks saying, "well-designed science laboratory activities focused on inquiry can provide learning opportunities that help students develop concepts" (p. 47). However, Hofstein and Lunneta also noted that the success of this approach is highly dependent on the nature of the task itself and recommended that more research be done into identifying the characteristics of tasks that make them successful. Additionally, Kuklianksy and Eshach (2013) were able to support students' understanding of data analysis during a college physics course by integrating statistics instruction with laboratory investigations throughout the semester. Similarly, with regard to engineering, Hjalmarson, Moore, and delMas (2011) were able to support and develop students' understanding of statistics by requiring them to create their own statistical measures. Data analysis tasks prove most successful when students engage deeply with the data and are involved with the planning and development of the analysis.

Despite or perhaps even because of the wide range of applications of data analysis and statistics, learning these concepts in context can be difficult for students. As Moore (1990) explains, "data are not merely numbers, but numbers with a context" (p. 96). Interdisciplinary applications of concepts in statistics provide students with many opportunities to engage with these ideas, yet the contextual nature of the data means that each application will be unique thus presenting its own challenges for students. Educators should strive to help students to see the connections between context-specific applications of data analysis and the big ideas and concepts that make up the discipline of statistics, but before we can do that we need to identify exactly how students engage with data analysis concepts in applied settings. This case study intends to contribute to that understanding.

Method

The journey that begins with identifying a problem or question that can be answered with data, continues with the collection and organization of data, and concludes with students interpreting and drawing conclusions about the data, requires students to make many connections and logical leaps. This case study documents that path for four groups of fifth-grade students during the data analysis tasks included in an integrated STEM unit focused on engineering and physical science. Specifically, this paper answers the following research questions: *How do students navigate the process of collecting and analyzing data as they work toward drawing conclusions supported by evidence in*

http://dx.doi.org/10.7771/2157-9288.1159

applied contexts? and *What obstacles and successes do they encounter as they engage in this process?*

This research is a descriptive, qualitative study that utilizes a case study approach. According to Yin (2009), "a case study is an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" (p. 18). For the present study, the case is defined as the work and experience of the student groups as they engage in the data analysis tasks embedded in an integrated STEM unit.

Setting

The students, teachers, and curriculum in this study were selected from teachers participating in the *EngrTEAMS: Engineering to Transform the Education of Analysis, Measurement, and Science* project. This project provides professional development and year-long support to teachers as they first learn principles of effective STEM integration and then develop their own integrated curriculum to be used in their classrooms. Forty to fifty teachers per year representing grades 4 through 8 participate in the project, where they work in teams of two to four to develop integrated units in life, earth, or physical science. During the professional development portion of this project, teachers are encouraged to fully integrate data analysis and measurement concepts within their science and engineering lessons, and they are instructed in pedagogical principles that support each of these concepts in integrated settings.

From the group of teachers involved in this project, a team of two fifth-grade teachers who developed a unit in physical science was chosen as the focus of this study. Fifth grade was chosen because, in the state in which these teachers teach, fifth grade is the first grade (according to the state academic standards) in which students are asked to reason and draw conclusions about sets of data within science classrooms. Prior to fifth grade, students learn data analysis concepts in mathematics classrooms. Additionally, in science they are asked to give evidence to support claims starting in third grade, but evidence in third and fourth grade generally consists of a single observation. It is not until fifth grade that students are asked to apply data analysis techniques such as graphing or finding measures of center that they learned in mathematics to their science investigations. Fifth grade is therefore one of the earliest times to find students engaging in authentic, applied data analysis tasks.

From the group of fifth-grade teachers within the project, the two teachers chosen for this study were picked because of the nature of the unit they created as well as the content area the unit covered. Especially at younger grade levels, physical science concepts typically lend themselves to easier and more direct measurements, so physical science was chosen as a starting point for this type of inquiry. Future investigations are planned to examine data analysis tasks in other content areas. Additionally, the unit designed by the teachers chosen for this study included data analysis tasks in the context of scientific inquiry and engineering design that, at least according to their written plan, demonstrated many of the principles for effective data analysis tasks described above in the review of literature.

The two teachers teach at different schools within the same urban, mid-west school district. The district itself enrolls approximately 39,000 students of which about three quarters are eligible for free and reduced lunch. Additionally, approximately one third are English language learners. The largest demographic groups include Asian American (31.4 percent) and African American (29.6 percent).

Participants

Four groups of students were chosen from the students taught by these two teachers in the classes in which they implemented the unit designed for the *EngrTEAMS* project. Two groups were chosen from each teacher, and these groups were selected based on a combination of mathematics ability and classroom dynamics. All students were given a pre-test developed for the *EngrTEAMS* project, which assessed knowledge and skills in engineering, physical science, and mathematics. For this study, only the scores on the mathematics sub-section were considered. Additionally, the classroom teachers placed the students in groups based on their usual classroom procedures. From these groups, two groups from each teacher were selected to make up students in this case study. One group was selected for each teacher such that each student in the group scored above the class average on the mathematics portion, and one group was selected such that the students all scored below average. Additionally, among the groups that met those criteria, the group whose scores were most similar was chosen. In this way, a higher and a lower ability group was selected from the students in each teacher's classes.

Data sources and data analysis

The data for this case study included audio and video recordings of student group work sessions as well as whole class discussion. Additionally, observation notes were recorded during the class meeting times, and digital photographs were taken of all student written or typed work. Because this study is investigating student thinking and learning, it was important to encourage students to articulate their thinking throughout the group work sessions. In many cases, the group dynamic required this as students attempted to communicate with each other about the task, but when this was not the case the researcher asked probing questions such as "why did you do that?" or "can you explain what you did there?" to encourage students to "think out loud." The researcher did not provide guidance or direction beyond answering simple procedural questions during the group work or class discussions.

Once collected, the data sources were coded using qualitative techniques taken from grounded theory (Auerbach & Silverstein, 2003). First repeated ideas were identified, then these repeated ideas were group together into themes. Themes were then examined and interpreted to form theoretical constructs.

Case description

Unit description

The unit that the students participated in was situated within the context of an engineering design challenge. At the beginning of the unit, students were introduced to the problem of land mines in Laos. Un-detonated land mines are a serious threat both to large animals such as elephants and to the people who live in these areas. One technique for dealing with them is to lob objects into areas with land mines to safely detonate them without harming animals or individuals. The students were tasked with the challenge of designing a cheap and portable "launcher" for throwing clay (Play-doh) at land mines. Specifically, the fictitious client in this scenario asked the students for a launcher that (1) could launch a projectile 10 m, (2) could land it within 0.5 m of a target, and (3) incorporated levers in order to do this.

After learning about the challenge, students began a two-part investigation into levers. In part one, all student groups investigated the effect of position of the effort force on the force required to balance a particular load. The groups constructed a lever from a ruler, a dowel rod, and a binder clip, and then recorded the effort required to lift and balance a load for five different positions of the effort force. Each student group constructed an identical lever, and all groups made measurements for the same five positions. Students then used graphing software to create a line-graph of these data, and attempted to use those data to draw conclusions about the lever. Using one data value from each group, the teacher created a class data set that was used for class discussion.

In the second part of the lever investigation, the class generated a list of other variables that they might test (such as position of the load, mass of the load, length of the entire lever arm, etc.) then each group chose one of those variables to explore. Two of the target groups in this study chose to investigate the effect of the position of the load on the effort force, and one group chose to investigate the effect of the size of the load on the effort force. The final group chose to investigate the size of the fulcrum, by which they were referring to the size and/or strength of the binder clip that was used as a fulcrum. In order to test this, they used three different sized binder clips (small, medium, and large), and attempted to keep the load and effort forces in the same position. With guidance from the teachers, groups designed their own investigations to test their chosen variable. Once they had collected data and recorded them in a data table, they again generated a graph and attempted to draw conclusions from the data and the graph. Classes then either created a poster for their results and did a gallery walk to view their classmates' results, or the classes shared their results with the class through discussion. The teacher then provided some summary comments about the results of each of the various experiments.

Once the second part of the lever investigation was complete, the groups began designing and building their launchers. Each of the target groups' initial designs consisted of a long flexible arm with some sort of cup or bowl affixed to one end. The Play-doh was placed in the cup/bowl, and the students held the other end as they threw. In this manner, the launchers were similar to lacrosse sticks. Once each group had completed their prototype, the class went either outside or to the gymnasium to test their designs. Students were given three chances to throw a ball of Play-doh at a target pre-positioned to be 10 m away. Additionally, a 0.5 m radius circle was placed around the target. Groups were given a large tape measure and asked to determine both how far the projectile went and how far it landed from the target.

Based on the results from their initial test students were given a chance to redesign. For the second design and test, however, the testing requirements were modified slightly for each group. None of the target groups attempted to address the new constraints in their new designs. All design modifications were meant to improve reliability, durability, and/or strength. One group briefly switched to a completely different design, trying to construct a bow-and-arrow or sling-shot-like device, but when they were unable to create a functioning prototype, they went back to their original design. The new design constraints only became a factor during the testing phase as they attempted to modify their throwing technique (rather than the design of the launcher itself). In the second test, some groups were asked to launch farther (20 m) or shorter (5 m), some were asked to land the Play-doh closer to the target (0.25 m), and some were asked to launch heavier or lighter projectiles (i.e. more or less Play-doh). Students had to consider their new design constraints as they modified and improved their original designs. After the second test (which reflected the modified requirements) students made a final design recommendation to the client and were asked to justify their designs using the results of the tests.

Themes

Although the target groups were able to successfully complete the experiments and tests and in most cases were able to draw correct conclusions about what they had done, they had great difficulty collecting and making sense of

http://dx.doi.org/10.7771/2157-9288.1159

their data. Several patterns emerged in their difficulties, and these themes are described below.

Measuring devices

In this unit, students used rulers, spring scales, an electronic balance, and measuring tape to collect their data. Only one group, the group who chose to vary the mass of the load in their second investigation of levers, used the electronic balance, and this tool caused no issue for them. The students made sure to zero the scale before using it, they correctly read and recorded the measurement from the digital read-out, and they were able to correctly interpret the meaning of those numbers both verbally and in their writing. On several occasions, they did mislabel the numbers with the wrong units; however, this did not seem to inhibit their ability to correctly interpret the numbers after the fact. They were aware that the measurements concerned the weight (mass) of the object even though they sometimes labeled the data as centimeters.

The other tools, however, caused a variety of difficulties for the students, many of which were related to the scale on the device. The most consistent errors were in using the measuring tapes. Two of the groups used measuring tapes that employed labeling conventions that appeared to confuse the students. Every centimeter on the tape was labeled; however, they were labeled relative to the nearest lower decimeter rather than to zero. Similarly, every 10 cm were labeled relative to the nearest lower meter. This convention is shown in Figure 1. Because of this labeling convention, students frequently misread the measuring tape. In one instance, for example, they recorded 72 m when, in fact what they had actually measured was 9.72 m. In another example, they recorded 8 m as their measurement, when in fact they were looking at the 8th centimeter between two deciles. Although the researcher was unable to record a more accurate measurement before they picked up the measuring tape, what they were trying to measure was clearly between 5 m and 6m. The other two groups used tape measures that only put labels on the meters. Centimeters were only marked with tick marks. These two groups both independently decided to round to the nearest meter making these measurements quite inaccurate. In one instance the students recorded a half meter, but this was only because one student wanted to round up and the other wanted to round down, so they compromised.

Students also had difficulties making measurements with the rulers that they used to create the levers for their lever activities. In order to assist the students in setting up the experiment and to speed up the process, both teachers covered the original scale on the ruler with masking tape. On the masking tape, they marked the positions that the students would need for the investigation, namely the center (marked as 0) and 2.5cm, 5cm, 10cm, 15cm, and 20cm on either side, as shown in Figure 2. Unfortunately, the original scale was still partially visible through the masking tape. Students initially attempted to read the original scale through the tape and had difficulty centering their lever arm as well as locating the positions that they needed to place the load and effort. Additionally, the presence of the same numbers on both sides of the center caused some confusion for the students as well. In most cases, they were able to resolve these issues on their own or with minimal guidance from the teacher.

The spring scale suffered from similar errors in reading the scale, but also exhibited some unique difficulties of its

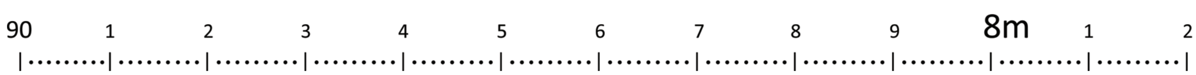

Figure 1. Diagram of the labeling scheme for the tick marks on the tape measure used by the students. The diagram shows measurements from 7.90 m to 8.02 m.

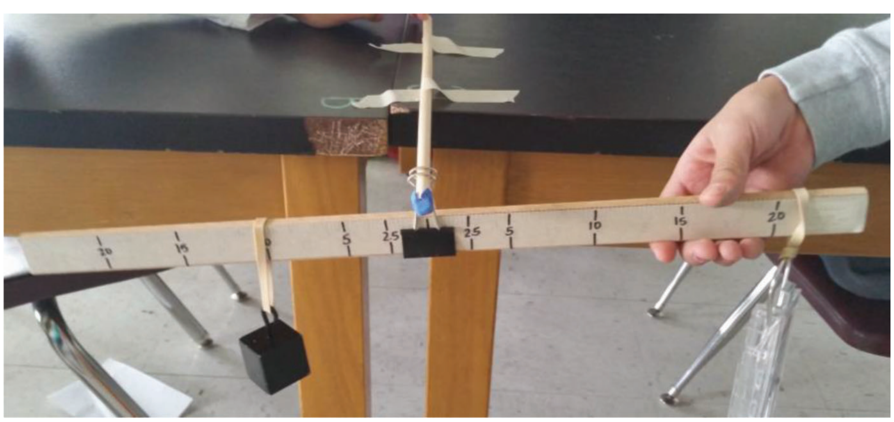

Figure 2. The set-up used in all the lever investigations. The numbers written on the masking tape indicate distance from the fulcrum, measured in centimeters.

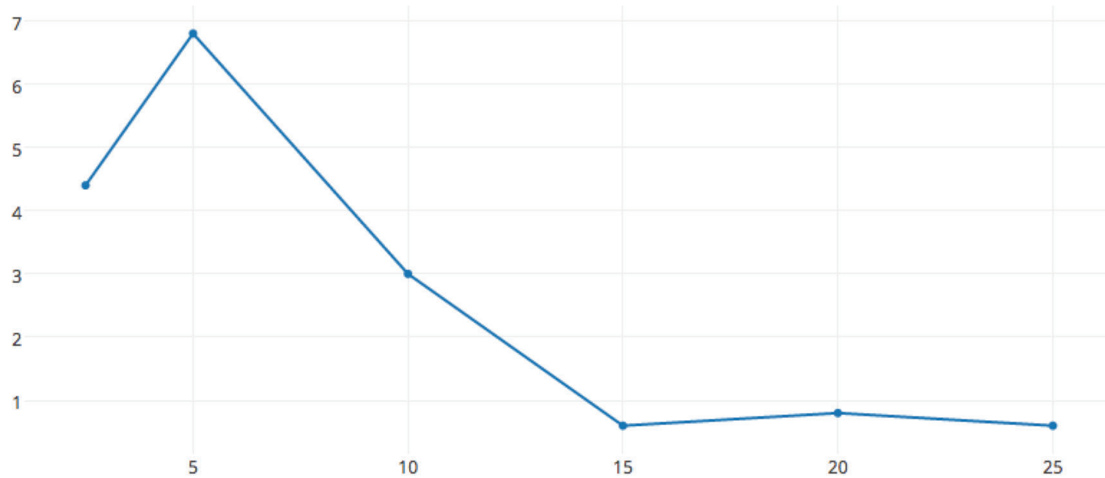

Figure 3. Plot of force required to balance a load vs position of the effort from the fulcrum generated via plot.ly (https://plot.ly/) by a student group.

own. The scale on the spring scale went from 0 to 10 N with major tick marks at each newton and four minor tick marks in between (0.2 N each). The 0.2 N minor tick mark caused some issues as the students read them incorrectly as 0.1 N, at first; however, in all groups at least one student correctly read the scale and was able to convince the rest of the group. More significantly, however, the forces they were measuring ranged over the entire scale. Near the middle of the scale, the students consistently recorded accurate measurements, but for large (i.e. near 10 N) or small forces, they had a variety of difficulties. First, when the effort was place at 15 and 20 cm from the fulcrum, the mechanical advantage was high, so the required force was quite small. In fact, it was so small that the weight of the spring scale itself was enough, or nearly enough, to balance the load. This resulted in a measurement of 0 N for that position, but the majority of the students did not believe this was correct. In some cases, they wrote down 1 N, thinking it couldn't be any smaller than that. Another issue for the small measurements was that not all the scales were properly zeroed. The spring scales they were using were equipped with a sliding scale, which allows for zeroing before making a measurement. Prior to making any measurements, one should, with nothing on the spring scale hook, slide the scale so that it reads zero. Unfortunately, students did not do this prior their first measurements. Of the four groups, most were close to zero, however, one group was 0.4 N off. This did not affect the relative size of their individual measurements; however, when the students compared their measurements to the rest of the class they felt like their data did not agree, especially for small measurements.

The final difficulty with the spring scale was due to the fact that for larger forces many of the spring scales got stuck, thus giving false measurements. Measurements that should have read close to 10 N were reading much lower. Besides being an incorrect measurement, this single data point ended up obscuring the trend from the students. One group's computer-generated graph is shown in Figure 3. Note that they were unable in the time allotted to decide what to label the axes, thus they remain unlabeled. The y-axis is the force (in N) required to balance the given load at the given position, and the x-axis is the distance (in cm) from the fulcrum of the effort force. This graph should show an inverse relationship, but at 2.5 cm from the fulcrum they measured 4.4 N instead of something closer to 10 N, and this single data point made it difficult for them to identify the trend.

These issues with the measuring devices seem to be the result of a combination of several factors. First, the rulers, measuring tape, and spring scale all employ a linear scale, essentially making these scales number lines or at least half number lines. Cramer, Ahrendt, Monson, Wyberg, and Miller (2017) have documented that elementary students exhibit several misconceptions regarding the number line and consistently have trouble accurately locating fractions of a unit on a number line. As this is precisely what they must do to read a measurement on a linear scale, any difficulties that students have understanding number lines themselves will likely be issues when using these measurement tools. This is evident from their confusion over the numbers on either side of the fulcrum, and their difficulties accurately reading the measurements to the centimeter on the tape measure. Additionally, the students showed weaknesses in general number sense, which can be seen in their inability to recognize that 72 m was not a reasonable measurement, or in their inability to recognize that 4.4 N at 2.5 cm from the fulcrum did not make sense even though they acknowledged that it was "harder to pull" at 2.5 cm than at 5cm where they measured 6.8 N. The final issue seems to stem from the measurement devices themselves. The sticking spring scale is one clear example, but labeling schemes on both the rulers and the tape measures also made it difficult for the students to make accurate measurements and interpret the results of those measurements correctly.

Quantitative data for qualitative questions

The second theme that emerged was a mismatch between the type of data the students collected and the data they used to draw conclusions or make decisions. Although the activities in this unit all required students to collect authentic data and to attempt to use those data to justify claims about the investigation or prototype test, in most cases the question driving the investigation or test did not actually require the data. In the first lever investigation for example, the ultimate question was, "how does the position of the effort force effect the force required to balance a load?" In a more advanced class, say in high school, students might actually answer this question with a mathematical model describing the relationship between effort force and distance from the fulcrum; however, in this fifth-grade class, where students are encountering levers perhaps for the first time in science class, this was not the ultimate goal. In this class, the goal was simply to realize that the farther from the fulcrum the effort force is, the less force is required. Although the data for most groups showed this very clearly both via the numbers in their data tables and visually in the inverse relationship apparent in their graphs, when pressed to explain why they knew this was case students invariably went back to their qualitative experiences with the lever itself and not the quantitative data they had collected. In other words, they were able to say that it was "easier" to lift the load when they moved farther from the fulcrum and "harder" when they were close, but they were basing this on how it felt to lift the load and not on what the spring scale had told them. When pressed to connect these feelings to the data, students either responded with "I don't know" or with a description of the shape of the graph without being able to identify how the graph related to the actual measurements.

Similarly, when testing their launchers they collected data to tell them how far they threw the Play-doh and how close it landed to the target, but when asked how well their design had done on the test, they did not refer to the measurements. It was clear from where the Play-doh landed if it was close to the target or not. They did not need to measure the distance to know if they had been successful. This may be part of the reason that they did not realize how inaccurate their measurements often were.

Difficulty interpreting small changes

A third theme that emerged involved students' difficulty in assessing the relative importance of variation in their data. In several cases, the students had trouble interpreting the data because measurement error was of the same order as the actual difference between measurements. For example, the group that decided to measure the effect of the fulcrum size on the lever could not make sense of their data even when the teacher worked individually with this group for an extended period of time. The size of the fulcrum (at least in this set-up) has nothing to do with the mechanical advantage due to the lever, thus we would expect the students to observe no difference between the three different fulcrums. In reality, however, we would not actually expect zero difference between the measurements, but merely a difference that is within the measurement error of the experimental set-up. When this group actually did the experiment, their smallest measurement was 7.4 and the largest was 8.2. Although a 10 percent measurement error is rather large, considering the experimental set-up they were using, this is not unreasonable. But because the measurements were not exactly the same, the students were unable to determine on their own that the fulcrum size had no effect. Even when the teacher tried first to coach them to this idea, and then to directly tell them that this was the case, they did not believe it. Throughout they maintained a belief that the largest fulcrum was the best because it was "the strongest." Similarly, as shown in Figure 3, when varying the position of the effort, the students measure the force required to balance the load at 20 cm to be slightly more than that required at 15 cm despite the fact that it should actually be slightly less. Again, this error is not unreasonable considering the precision and accuracy of the set-up, the instruments, and the students themselves; however, the students in this group had considerable difficulty explaining why 20 cm required slightly more force, and they considered the slight increase to be part of the trend they saw in the graph rather than a result of natural variation in the data.

Conclusion

The data analysis activities in the unit observed for this study possess many of the characteristics identified as contributing to students' abilities to develop skills in making sense of data, yet the students still had considerable difficulty doing just that. Issues of measurement error and difficulties using measuring devices lead to data that made it difficult for the students to interpret. Additionally, students' lack of strong number sense and weaknesses in understanding the number line also contributed to their difficulties in successfully reasoning from their data. As a result, the teachers involved in this study spent much of their interaction time with the students on helping them overcome these challenges rather than on the science or engineering concepts they were meant to be engaging with. It is possible to conclude from this that students in fifth grade, at least in the schools observed here, are not ready for this type of activity, but the authors would advocate for a different conclusion. Ultimately, all groups were able to articulate conclusions or claims consistent with the science or engineering concepts in question. Students noticed and explained, both verbally and in writing, that the farther the effort force is from the fulcrum the less effort is required, or that the larger the load the more force required to lift it; and the groups knew when their designs had been successful at meeting the criteria or not.

The breakdown occurred when students attempted to connect those observations and conclusions to the quantitative data that they had collected. The fact that the students were able to correctly interpret the investigations and tests qualitatively means that the activities themselves are in fact accessible to the students. Thus the challenge for teachers and curriculum writers is to find ways to help students make the connections to their data. Even if students exhibit the weaknesses shown here, if they are able to see the connections between their experiences with the materials and the data they collect, then activities like this might have the potential to be used to help students to develop number sense, make sense of the number lines, and learn to use measurement devices and deal with measurement error more skillfully. In other words, rather than relying on the data to help teach the science, teachers and curriculum designers might be able to structure activities and provide supports for students to help them bridge the gap between what they qualitatively observed and the data that they collected. In that way, science, engineering, measurement skills and data analysis can develop in parallel. Future research is needed to determine what supports and structures are most effective at helping students make sense of their data and connect them to the phenomenon, and also how activities like this can help students learn statistics concepts and develop data analysis and measurement skills. The participants in this case show, however, that simply asking students to collect and interpret data, even when the activity itself is accessible to the students, does not guarantee that students will make meaningful connections between the data they measured and the science and engineering concepts with which they are engaged.

References

Auerbach, C. F., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York: New York University.

Bybee, R. W. (2011). Scientific and engineering practices in K-12 classrooms: Understanding A Framework for K-12 Science Education. *The Science Teacher*, *78*(9), 34–40.

Committee on Integrated STEM Education, National Academy of Engineering, & National Research Council. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. In M. Honey, G. Pearson, & H. Schweingruber (Eds.). Washington, DC: National Academies Press.

Cramer, K., Ahrendt, S., Monson, D., Wyberg, T., & Miller, C. (2017). Making sense of third-grade students' misunderstandings of the number line. *Investigations in Mathematics Learning*, *9*(1), 19–37.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 47–78). New York: Kluwer Academic Publishers.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, *63*(1), 25–34.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372–396. http://doi.org/10.1111/j.l751-5823.2007.00029.x

Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). Research on teaching and learning statistics. In *Developing students' statistical reasoning: Connecting research and teaching practice* (Vol. 75, pp. 21–43). Springer.

Hjalmarson, M. A., Moore, T. J., & delMas, R. (2011). Statistical analysis when the data is an image: Eliciting student thinking about sampling and variability. *Statistics Education Research Journal*, *10*(1), 15–34.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28–54. http://doi.org/10.1002/sce.10106

Kukliansky, I., & Eshach, H. (2013). Evaluating a contextual-based course on data analysis for the physics laboratory. *Journal of Science Education and Technology*, *23*(1), 108–115. http://doi.org/10.1007/s10956-013-9456-6

Lehrer, R., Kim, M.-J., & Jones, R. S. (2011). Developing conceptions of statistics by designing measures of distribution. *ZDM: The International Journal for Mathematics Education*, *43*(5), 723–736. http://doi.org/10.1007/s11858-011-0347-0

Lohr, S. (2012, February 11). The age of big data. *The New York Times*. New York.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.

Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association*, *93*(444), 1253–1259.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: NCTM.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (pp. 77–98). Reston, VA: National Council of Teachers of Mathematics.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilkins, J. L. M. (2000). Preparing for the 21st century: The status of quantitative literacy in the United States. *School Science and Mathematics*, *100*(8), 405–418.

Yin, R. K. (2009). *Case study research: Design and methods* (Fourth). Los Angeles, CA: Sage Publications.