Empowering Students through DATA
Author(s): Paige A. Barnes
Published by: National Council of Teachers of Mathematics
Stable URL: http://www.jstor.org/stable/20876452
Accessed: 22/03/2013 17:08

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of
content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms
of scholarship. For more information about JSTOR, please contact support@jstor.org.

National Council of Teachers of Mathematics is collaborating with JSTOR to digitize, preserve and extend
access to The Mathematics Teacher.

http://www.jstor.org
This activity helps students understand data analysis concepts and acquire skills to take control of their learning.

Paige A. Barnes

As home, school, and community partnerships continue to work toward the vision articulated by the No Child Left Behind legislation and the Equity Principle set forth in NCTM’s Principles and Standards for School Mathematics (2000), teachers continue to find that some students’ education needs are still overlooked. Many high school students experience real-world demands, such as financially supporting themselves by working or helping care for family members, all while maintaining the social life of a teenager. Many have trouble seeing the importance of what they learn in school and how it pertains to their lives. Often, teachers find it difficult to show how abstract mathematical topics, such as polynomials or factoring, relate to students’ nonacademic lives. Data analysis may be a way to help bridge the gap between what students learn in school and their everyday experiences and concerns and thus to enrich their school experiences. This link is important when dealing with data analysis because “data are not just numbers, they are numbers with a context … in data analysis, context provides meaning” (Franklin et al. 2007, p. 7).

The expectations described in NCTM’s Principles and Standards for School Mathematics (2000) state that high school students should be able to understand histograms, box plots, and scatter plots as well as be able to analyze and explain basic statistical concepts related to these representations. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report highlights the need for students to understand enough about statistics to be successful in life: “Every high school graduate should be able to use sound statistical reasoning to intelligently cope with the requirements of citizenship, employment, and family to be prepared for a healthy, happy and productive life” (Franklin et al. 2007, p. 1). To achieve this level of competency, students need some basic understanding of statistical concepts.

The activity that follows is a response to an unfortunate situation in my Basic Algebra class: Because of a teacher shortage the previous school year, some of my students had not yet had a permanent, certified algebra teacher. One of my goals was to see whether these disadvantaged students struggled more than
those classmates who had such certified algebra teachers. Another goal was for my students to learn basic statistical concepts by using real data that might inform their attitude toward school now and their ability to be wise consumers later.

To begin, a survey containing basic demographic questions and other questions based on students' current and past performance in algebra was administered. Students then discussed issues related to data collection, such as appropriate questions and sources of bias. Next, they explored and analyzed the data by using dynamic statistical software. Ultimately, students created posters and presented their findings to their classmates. This sequence was designed to empower students to tell their story through data by using Exploratory Data Analysis (Konold and Higgins 2003, pp. 193–215). Further, the spirit of NCTM's Standards and Principles for School Mathematics (2000) demands that we provide students with the opportunities and tools needed to understand and communicate their circumstances via data analysis and address the issue of access to an equitable mathematical education.

COLLECTING DATA
As a starting point, I considered questions that I wanted answers for:

- Does a lack of teachers hinder students' learning?
- Does it really matter whether teachers are certified?
- Which students struggle most?

With questions like these in mind, I created a survey (see fig. 1). The students who were surveyed included most of the students in Basic Algebra. Students taking Basic Algebra have completed Foundations of Algebra and, on completing Basic Algebra, will have completed the algebra 1 curriculum. Of the 166 surveys given to Foundations of Algebra students, 83 were compiled for use in the data set. Students were not required to complete the surveys and were assured that their responses would remain anonymous. When creating the survey, I kept in mind the guidelines for the survey to be useful: It should be short, lack jargon, be easy to read, have clear instructions, contain interesting and nonthreatening questions, include a rationale, and avoid bias or leading questions (Gall, Gall, and Borg 2003, p. 226).

Once the survey was completed, my classes discussed potential bias, the difficulties associated with missing data, and falsified data. Students' concerns about bias centered around what type of student chose to complete the survey and whether the sample was representative of the student population. It was suggested that only students who cared about school would complete the survey honestly. In some surveys, data were missing or filled in incorrectly. Students noted that we needed to decide how to handle such situations and were eager to discuss how to handle false data. They asked what the media did with respect to false information. We finally decided to include all complete surveys because there was no way to know for sure whether students had been dishonest. Students felt that the media would handle false polling information in the same way. We also standardized the questions about grades. If students entered
a letter grade, we would use the lowest numeric grade for that specific letter grade; for example, a B would equal an 85.

We then had to decide how to enter the data into the dynamic statistics software. Because of its availability, we used Fathom 2.0 (Finzer 2001), although TinkerPlots (Konold and Miller 2005) also would have been an appropriate tool; the author used TinkerPlots for figures 3–5 for reasons that will become clear later in the article. Transforming survey responses to data proved to be a nontrivial task, as the questions were asked in such a way that students needed to determine which attributes were categorical and which were numerical (cf. Konold and Higgins [2003] for more about abstracting) and needed my assistance in making these decisions. This discussion resurfaced later when we studied appropriate representations.

ANALYZING THE DATA

With the information entered in the software, students were able to explore the relationships among the data and discuss important statistical concepts with real data. The GAISE report (Franklin et al. 2007) gives many guidelines for what students need to know about data analysis. The main concepts I discussed with the class included co-variation, center and spread, differentiation between variability and error, graphical representations, and individual data versus aggregate data.

The discussion of co-variation began with questions to help the students start their exploration. How might variation in one quantity relate to variation in another? Does homework have any effect on grades? Do absences have any effect on grades? What other relationships exist among the data? As a class, we discussed these questions and found relationships in these data that co-vary. For example, when looking at absences and grades, students noticed a slight negative correlation. After further discussion, students began to understand that when two attributes co-vary, a positive or negative relationship exists. Regarding absences and grades, for example, this discussion helped them see that as school absences increase, grades decrease.

When describing center and spread, students were familiar with the terms median, mean, and mode, but they were unsure about how to use these terms with this activity. We discussed the median as the center of a set of numerical data; in a graphical representation, the number of data points to the right of the median would be the same as the number of data points to the left of the median. Students wanted to discuss the median as the middle value when data were written in ascending order. The mean was our way of looking at a fair share. To help students better understand the role of mean, I asked them what their grades would be if everyone’s grade had to be the same. By thinking about the mean this way, students were better able to understand what the mean represented in each relationship. Students had no problem locating and understanding the smallest and largest values of the data set, and these measures of center also helped them look at the variation in and the spread of the data. Using the mean to help describe variation and spread was difficult for students. I briefly discussed the relationship between the mean and the median when determining whether data were skewed. With more time, this could be a topic for more investigation and elaboration.

Discussion of mean led to a discussion of the differentiation between variability and error. As noted earlier, students were very interested in bias and missing or false data. At this point, students were able to talk about the differences that exist in the data. Variability helps them interpret and explain causes for outliers, clusters, or gaps in representations. Once we started talking about gaps, clusters, and outliers, students started looking more closely at the graphical representations. They wanted to discuss and find reasons for all outliers and used the clusters as a way to find similarities among their classmates.

While interpreting data, students were given the choice of using dot plots, histograms, and box plots, but they tended to use box plots most of the time.

---

**Foundations of Algebra Survey**

Please answer the following questions honestly. Your name will not be used.

1. Have you ever taken Algebra 1A or Algebra 1B (Basic Algebra) before? __________
   If so, how many times? __________

2. Who was your Algebra 1A (Foundations of Algebra) teacher? __________

3. What was your grade in Algebra 1A? __________

4. How many days were you absent in Algebra 1A? __________

5. What middle school did you attend? __________

6. Race __________

7. Gender: M or F (circle one)

8. According to the following scale, how hard did you work in Algebra 1A (circle one):
   1. I didn’t do anything.
   2. I did only what I wanted to do.
   3. I did most of what was assigned.
   4. I did everything I was assigned.
   5. I did all my class work and worked hard on my homework to help me understand the material.

9. How many hours per week did you spend on Algebra 1A homework? __________

10. What is your current average in Algebra 1B (Basic Algebra)? __________

---

Fig. 1 Survey used in the activity

Vol. 102, No. 8 • April 2009 | MATHEMATICS TEACHER 617
Their choice seemed to be based on their need to explain their preconceived ideas of education: They wanted their graph to tell their story. Unfortunately, the box plot sometimes misled students, a point that will be explained in more detail later.

Because the students had a hard time interpreting the box plots, we discussed what each part of the graph represented. The data were divided into four equal parts, and the five-number summary technique enabled students to create the box plot (Franklin et al. 2007, p. 47). The five numbers include the minimum data value, first quartile, median, third quartile, and the maximum data value. Students picked up on this information easily but tended to focus on these five points exclusively. By mentally omitting the rest of the data set, students were looking at individual cases, not at the data set as a whole. The GAISE report states that interpreting the results through analysis of the box plot requires “comparison based on global characteristics of each distribution” (Franklin et al. 2007, p. 23). We needed to globalize the data, so this was a perfect opportunity to discuss the need to move from looking at individual cases to looking at the data as an aggregate. Konold and Higgins state that “students need to make a conceptual leap to move from seeing data as an amalgam of individuals each with its own characteristics to seeing the data as an aggregate, a group with emergent properties that often are not evident in any individual member” (2003, p. 202). In other words, students need to step back from the data to see all they have to offer.

**THE ACTIVITY: USING EXPLORATORY DATA ANALYSIS (EDA)**

Exploratory Data Analysis (EDA) allows students to become actively involved in the investigation of data. Konold and Higgins (2003) present EDA as an exploration of data that can lead to meaningful discussions and a deeper understanding of data analysis concepts. This exploratory technique was used throughout the activity as we looked at different relationships among data.

My students were particularly interested in the relationship between race and Foundations of Algebra grades. When they looked at the box plot (see fig. 2), they concluded that the African-American population was smarter than either the Hispanic or white populations. However, looking at the first and third quartiles of the box plots, we can interpret the data more accurately. The boxes for each of the races are located in about the same place on the graph. This arrangement should have told students that the majority of students of each race performs at the same level in Foundations of Algebra, but in this instance students misunderstood the box plot representation. They focused on the outliers and wanted to infer that African-American students were performing better than students of other races. This problem could easily be solved by using TinkerPlots, software that allows a student to overlay a box plot onto a dot plot (see fig. 3). Representing the data in this way allows students to see that an outlier in the white population caused the whisker to be pulled further left than the Hispanic and African-American students. Again, this realization alone could help with discussion of error and variability as well as bias and falsified data. The outlier student should not have been promoted to Basic Algebra with an average of 60 in Foundations. This same problem occurred with the relationship between gender and Foundations grades (see fig. 4).

Konold and Higgins emphasize that “understanding this data representation and analysis involves many complex issues, from sorting through what different numbers on a graph mean, to choosing appropriate measures to summarize and compare groups, to identifying relationships between variables” (2003, pp. 212–13). Students used the information from this discussion to analyze more relationships in their data.
As Konold and Higgins (2003) observed, it seemed that students were questioning their data and that their data were generating further questions. Another question students asked was whether or not relationships existed between the amount of homework and a student's grade and between the number of absences and a student's grade. As students offered their findings to the class, we discussed spread, variation, discrepancies, and interpretations of the graph. Students' individual choices of graphical representation led to a discussion about comparing different types of graphical representations, including the benefits and drawbacks of each. In fact, several students commented on the ways in which different graphs could be used to hide different characteristics of the data. This was a good opportunity for discussing how the media uses different representations in data analysis. Students are now better prepared to question graphs that they see in the media and determine whether a certain representation was chosen to achieve a hidden agenda.

Students were grouped according to interest and worked together to create a poster presentation. With their poster as a tool, they had to explain their choice of representation and interpret their findings to the class. The students' presentations addressed several questions:

- Do the students need to miss less school or do more homework?
- Did boys or girls do better in this algebra class?
- Did students of one race do better in this algebra class?

As the students worked on their posters, I circulated throughout the room to ensure that they were on task and to see what they found interesting. Some relationships the students investigated were unexpected and brought up discussions about specific prejudices that exist and why.

Students worked hard to answer their previous questions by looking at the relationship between hours spent on homework and grades. Although students would never admit it, they do feel that homework is important. One group made the observation that students need to complete between one and six hours of homework per week to pass the course. At the same time, the group noticed that students who reported doing more than six hours of homework per week had lower grades. This observation, initially confusing, became an opportunity for discussion, which in turn generated new questions for student investigation: Do these students listen to music or watch television while they study? Do they lack good study skills? Do they suffer from test anxiety? Although we did not have the data to investigate these questions, the students' exploration followed the EDA cycle of conclusions leading to more questions. Teachers using real, student-collected data must be aware of such unintended consequences and the pedagogical challenges associated with them.

Students still seem to be misled by the box-and-whisker plot. Konold and Higgins (2003) note that it is very difficult for students to look at data as an aggregate, an observation that became apparent during this activity. Students did not attend to center and spread, even after having discussed them before the investigation. Further, this sequence of activities provides real data and helps motivate students to overcome this epistemological hurdle.

Still striving to answer their questions about the data, students looked at the relationship between the number of absences and grades (see fig. 5). Unfortunately, there turned out to be no correlation between these two attributes (although all teachers would hope for students to see such a correlation in such an investigation). Although students had not been exposed to the formal measure of correlation, they analyzed the
association between the two attributes and whether that association had a positive or negative trend. They concluded that a student did not have to go to school every day to pass algebra. Also, most students missed five days or fewer, an acceptable number of absences; the grades for these students varied greatly. The data did answer the students’ questions, but the answers are not always what teachers hope for.

REFLECTIONS
The intended outcome of this activity was to give an empirical voice to students who had been disadvantaged. However, students chose to investigate and explore attributes of more interest to them. If I were to use this activity again, TinkerPlots would be a better tool to use, because students can look at box plots and dot plots simultaneously. Also, ample time should be given for students to discuss in depth the concepts of data analysis before the activity to help them assess their findings more accurately.

Although the outcome was not the intended one, two issues became apparent after my students completed this activity. First, students struggle with analyzing data as an aggregate because they want to focus on individual cases. Konold and Higgins (2003) see this as an important issue in data analysis education. I hope that this activity can help students see the need to describe data as a whole group rather than simply as individual cases. Second, students chose their graphical representation on the basis of preconceived ideas. They already had a presupposition of what the data would support and wanted to find a representation that showed that presupposition to be true. Fortunately, we were able to look at this issue as a class and discuss many different representations and the message that each sent to observers. The latter issue is very helpful for students as they interpret daily media presentations. The ability to read, analyze, and understand data is a skill all of us need.

CONCLUSION
While working with this activity, I found that the possibilities for teaching students about data analysis are endless. This activity was used as a way to introduce students to data analysis, but it could also be used to raise students’ levels of awareness and understanding of data analysis, address the epistemological challenges of describing the aggregate, and ensure equity for students who often lack parental advocates. This activity synthesizes Content Standards with the Process Standards of Reasoning and Communication and gives students a wealth of information beyond statistical reasoning. I had also hoped that these students could present their findings to the administration as a way to make this data analysis their own. To communicate their findings in their own voice to decision makers could show students the power of being informed consumers—a profound experience.

Overall, students were very interested in learning data analysis concepts in the context of making statements about their educational experiences. Although data analysis is not a part of the algebra curriculum in many states, this activity provided my students an opportunity to understand how they could affect their own education and gave them an overview of the data analysis concepts they will need to be statistically competent in the real world. As a teacher, I have found that students have difficulty grasping data analysis concepts fully without adequate experience or exposure.

FUTURE DIRECTIONS
This activity can be expanded by providing a richer mathematical base. The activity offered numerous teaching moments to delve into the mathematics involved, but I was unable to take advantage of many of these because of time constraints. In the future, I hope to find more opportunities to build on students’ previous knowledge to make them more competent in data analysis.

REFERENCES

PAIGE A. BARNES, PaigeBarnes@ johnston.k12.nc.us, teaches mathematics at Smithfield-Selma High School in Smithfield, North Carolina. She is interested in using professional development to help teachers reach all students.