DOCUMENT RESUME

ED 391 661 SE 057 608

AUTHOR	Friel, Susan N.; Bright, George W.
TITLE	Graph Knowledge: Understanding How Students Interpret Data Using Graphs.
PUB DATE	Oct 95
NOTE	23p.; Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (17th, Columbus, OH, October, 1995).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Cognitive Processes; *Data Analysis; Elementary School Students; *Graphs; Intermediate Grades; Middle Schools
IDENTIFIERS	*Representations (Mathematics)

ABSTRACT

This paper discusses a research study that focused on middle grades students' abilities to read and to move between different graphical representations before and after instruction. The data analyzed were collected from a group of 76 sixth grade students who were in 3 different mathematics classes in a middle school located in central North Carolina. The ways that students made sense of information presented through graphical representations and made connections between related pairs of graphs were investigated. General conclusions indicate that students need to talk more about graphs, including talking about the structure of graphs and how this informs the statements that can be made about: (1) the information depicted by the graphs and/or (2) the predictions or inferences that can be drawn from the graphs. Contains graphs and questions used on written pre- and posttests. Contains 26 references. (MKR)

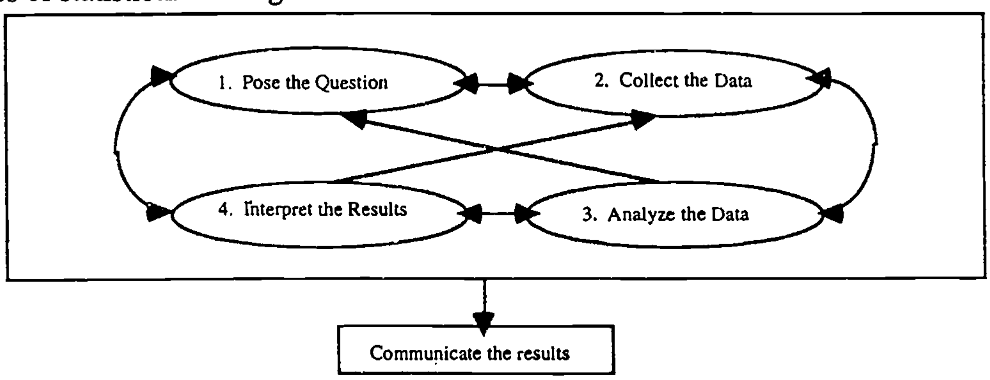
Graph Knowledge: Understanding How Students Interpret Data Using Graphs

Susan N. Friel, UNC-Chapel Hill, sfriel@email.unc.edu
George W. Bright, UNC-Greensboro, brightg@steffi.uncg.edu

Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education
Colombus, OH: October, 1995

Purpose of Study

Statistics is a major content strand across the K-12 curriculum (NCTM, 1989). Central to work in statistics is an understanding of the statistical investigation process (Graham, 1987). A statistical investigation typically involves four components (1) posing the question, (2) collecting data, (3) analyzing data, and (4) interpreting the results, in some order (Graham, 1987). Perry, Kader, and Holmes (1994) suggest a fifth stage of a statistical investigation: communication of results. The model of five interrelated components (Perry, et al., 1994, Lappan, et al.1996). is very helpful in articulating the process of statistical investigation.



Concepts such as measures of center or graphicacy[1] can be linked to the "analyze the data" component of the statistical investigation process. While a central goal is to understand how students make use of the process of statistical investigation within the broader context of problem solving, it also is necessary that we look at students' understanding related to concepts linked to this process. This has led us to consider what it means to understand and use graphical representations as a key part of what it means to know and be able to do statistics. Specifically, we have engaged in a process of developmental research that permits examination of middle grades learning of concepts related to the use and interpretation of representations. We have looked at how such

1 The atility to read graphs (Wainer, 1980)

understanding changes over time and with instructional intervention provided by knowledgeable teachers in order to develop a framework both for looking at students' knowledge of graphing (in the statistical sense) and for developing a research agenda related to this area.

Perspective

The current literature tells us very little about how knowledge develops with respect to the understanding and use of graphs. Our theoretical perspective for understanding students' thinking is built on analysis of the structure of graphs and on cognitive perspectives about graph comprehension. The various results associated with reading and interpreting graphical representations as addressed in the National Assessments of Educational Progress (Carpenter et al., 1971, 1978; Dossey et al., 1988; Lindquist, 1989) highlight students' difficulties with interpretations of relationships presented in graphs. The ability to read and interpret graphs is a "basic skill" and apparently is not being effectively taught (Kirk, et al., 1980). It consists of two components: reading data presented in graphic form and forming appropriate generalizations which describe the depicted relationships.

Students' ability to read graphs is receiving attention by a select number of authors (Curcio, 1987; Gallimore, 1991; Pereira-Mendoza, 1991; Pereira-Mendoza, 1992; Pereira-Mendoza, 1995; Rangecroft, 1991a; Rangecroft, 1991b) who have begun to look more closely at the concept of graphicacy in the curriculum, particularly at the elementary and early childhood levels. In addition, we do have anecdotal and written evidence[2, 3, 4] obtained through developmental research[5] associated with three different curriculum development projects which highlight some of the complexities associated with graphicacy and help frame a set of issues related to understanding graphical representations.

<u>Data Reduction and the Structure of Graphs</u>

The process of data reduction and the structure of graphs are factors that influence graphicacy. Data reduction is the transition from tabular and graphical representations which display raw data to those which present grouped data or other aggregate summary representation. Different graphical representations of numerical data reflect different levels of data reduction. A representation may display the original raw data or a graph may

2 Russell, S.J. and Friel, S. F. (Principal Investigators) *Used Numbers Project*. Funded by the National Science Foundation, 1987-1990.

3 Lappan, G., Fey, J., Fitzgerald W., Friel, S., and Phillips, E. (Principal Investigators) *The Connected Mathematics Project*. Funded by the National Science Foundation, 1991-current.

4 Friel, S. and Joyner, J. (Principal Investigators) *Project TEACH-STAT*. Funded by the National Science Foundation, 1991-1994.

5 See Gravemeijer, K. (1994.) Educational development and developmental research in mathematics education. *Journal for Research in Mathematics Education*, 25, 5, 443-471, for discussion of a research model that integrates curriculum research and design as "developmental research".

display grouped data. Most graphical representations used in the early grades (e.g., picture graphs, bar graphs) involve either just the original data or tallied data from which the original observations may be obtained. Students in upper grades more often use graphical representations of grouped data (histograms, box plots) from which it is usually not possible to return to the data in its original form.

The structure of graphical representations of data may also impact understanding. For example, graphical representations utilize one axis or two axes or, in some cases, may not have an axis. For graphical representations that use both axes, the axes may have different meanings. In some simple graphs, the vertical axis may display the value for each observation while the vertical axis for more typical bar graphs and histograms provides the frequency of occurrence of each observation (or group of observations) displayed on the horizontal axis. Confusion may develop if the different functions of the x- and y-axes across these graphs are not explicitly recognized.

Components of Graph Comprehension

The rudiments of a theory of graphicacy need to address the broader issue of what kinds of questions graphs can be used to answer (Wainer, 1992). Curcio (1987) conducted a study of graph comprehension assessing fourth and seventh grade students understanding of four traditional "school" graphs: pictographs, bar graphs, circle or pie graphs, and line graphs. She identified three components to graph comprehension that are useful here in framing questions:

1. *Reading the data* involves "lifting" the information from the printed page to answer explicit questions for which the obvious answer is right there in the graph
2. *Reading between the data* includes the interpretation and integration of information that is presented in a graph. This includes making comparisons (e.g., greater than, greatest, tallest, smallest, etc.) as well as applying operations (e.g., addition, subtraction, multiplication, division) to data.
3. *Reading beyond the data* involves extending, predicting, or inferring from the representation to answer implicit questions. The reader gives an answer that requires prior knowledge about a question that is at least related to the graph.

The first two components focus on elementary levels of questioning that involve data extraction. The latter component is tied to questioning that involves not only interpreting a graph but utilizing the graph to help make realistic predictions or assess realistic implications from the data (Pereira-Mendoza, 1995).

Sequencing of Graphs

Researchers have also attempted to identify ways of ordering graphs as this relates to comprehension. Studies have established that bar graphs are easier to comprehend than line graphs and that horizontal and vertical bars are equally understood. In addition, bar graphs found to be more difficult to understand than circle graphs but easier to understand than line graphs (Padilla et al., 1986.). Very little, if any, research has been done with respect to the newer statistical graphs, e.g., stem-and-leaf plots, box plots.

Rangecroft (1991a., 1991b, 1994) has indicated the need for attention to be paid to a well-thought-out and detailed progression in graphwork. Over several years, she has developed an articulated sequence for graph introduction that addresses both the sequencing of representations and an apparent developmental/maturity level for readiness. Much of her reasoning actually addresses issues of data reduction and graph structure in addition to learning theory. She has not addressed the apparent relationships among graphs and the possible benefits of focusing on transitions among graphs as a way to promote understanding.

Research

This paper discusses a research study that focused on middle grades students' abilities to read and to move between different graphical representations (i.e., line plots, bar graphs, stem-and-leaf plots, and histograms) before and after instruction. During Fall, 1994, we conducted a study of the ways that students in grades 6 and 8 made sense of information presented through graphical representations and made connections between related pairs of graphs. Students were tested both before and after an instructional unit developed specifically to highlight a particular sequence of graphs that took into consideration increasing degrees of data reduction and building connections between pairs of graphs (and reflect much of thinking voiced by Rangecroft, 1994). Small samples from each grade were also interviewed before and after the unit. Data from the interviews and the tests of these samples of students are used to illustrate the difficulties and successes that students experienced in attempting to understand the material from this unit.

Using Curcio's (1987) three components of graph comprehension as an organizing framework for reporting results, specific attention in this paper is given to the nature of the responses students made to selected written problems presented by pre- and post-instruments. This research provides a framework both for looking at students' knowledge of graphing (in the statistical sense) and for developing a research agenda related to this area.

The attached Figure 1 shows both the graphs and the questions asked on the four parts of the written pre- and post-test instruments. Looking at student responses to these questions provides rich opportunities for raising issues that we want to address with respect to graphicacy. This paper includes data on only two questions from the first section of the written test. These questions focus on a line plot showing numbers of raisins in half-ounce boxes:

1. Are there the same number of raisins in each box? How can you tell?

3. If the students opened one more box of raisins, how many raisins might they expect to find? Why do you think this is so?

These questions have been chosen because they provide insights into the richness of the data and the subtle complexities that appear to exist "just below the surface" when we consider what it is that students *really* understand about representations. Further, we will report only on the analysis as it relates to data collected from a group of Grade 6 students in this paper; subsequent papers will include Grade 8 data and appropriate comparisons across grade levels.

The data analyzed were collected from a group of 76 sixth grade students who were in three different mathematics classes in a middle school located in central North Carolina. The students were a heterogeneous group and were divided into three classes of approximately 25 students each; they all were taught by the same teacher. They had had little experience with statistics prior to this year. Their teacher was part of the *Teach-Stat Project*[6] and was a statistics educator, as were ali the teachers who participated in the study. The study was conducted over a six-week period from mid-October to the end of November, 1994.

This study was not designed to assess a particular instructional model or curriculum. Rather, the authors reasoned that taking a "snapshot" of what students knew about representations at one point in time would not be as productive as trying to assess what students knew both before and after having an opportunity to gain some experience with the process of statistical investigation and with some of the key concepts (including graphs) in statistics. Since statistics is a new curriuclum area at the middle grade levels, many prior studies have not addressed the question of change as it may be related to statistics learning because most students have had few, if any, relevant learning experiences in this area.

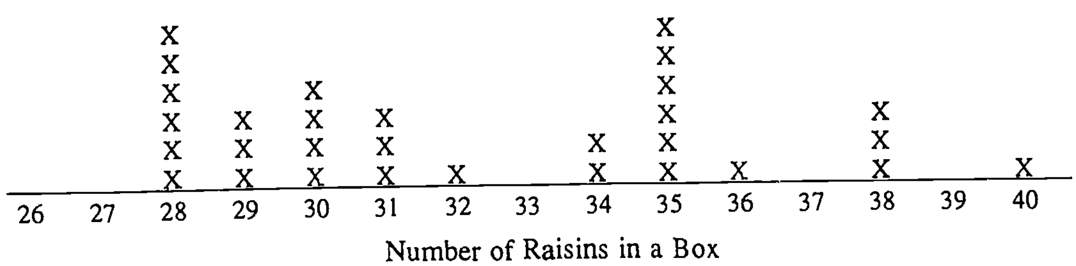
The first two parts of the written instrument were administered both as a pre- and a post-test; the authors reasoned that line plots and bar graphs were graphs with which students have had some exposure. The second two parts (stem plots and histograms) were administered only as post-tests because students have had few, if any, experiences with either of these representations (verified through an earlier pilot test). The authors did not

6 The *Teach-Stat Project* was a three-year teacher enhancement program that prepared over 450 K-6 teachers to teach statistics. Year 1 and Year 2 teachers (300 teachers) participated in 3-week summer institutes. From the Years 1 and 2 teachers, 84 teachers were selected to receive additional professional development to prepare them to be statistics educators who could help support and train other teachers. See Friel, et al, (1996) and Gleason, et al, (1996) for additional information.

want to force students to try to interpret data using representations with which they were unfamiliar, believing that such a situation may unwittingly motivate misundertandings.

In looking at Figure 1, Part I of the written instrument focused on questions related to a line plot that showed data presented in the following context:

> Students brought several different foods to school for snacks. One snack that lots of them like is raisins. They decided they wanted to find out just how many raisins are in $\frac{1}{2}$ ounce boxes of raisins. They wondered if there was the same numbers of raisins in every box. The next day for snacks they each brought a small box of raisins. They opened their boxes and counted the number of raisins in each of their boxes.
>
> Here is a line plot showing the information they found:



From this display, we can see that approximatley half of the boxes have 28-32 raisins and half of the boxes have 34-40 raisins. Although the range is 28-40 raisins, a large clump of data falls in the interval of 28-32 raisins, with a smaller clump in the interval of 34-36 raisins. An unusual value, or outlier, occurs at 40 raisins. The three boxes at 38 raisins may also be outliers. There is a gap between 32 and 34 raisins. The median of these data is 31 raisins and the mean is 31.3 raisins.

As is indicated in Figure 1, Question 1 is a "read between the data" question:

1. Are there the same number of raisins in each box? How can you tell?

This question was the first that students' addressed in this portion of the written test. A "read the data" question could have been, "How many raisins are in the smallest box?" or "How many boxes of raisins had 30 raisins in them?" We chose to move directly to the "read between the data" question because we believed that the "read the data" questions would not show much diversity in response (earlier pilot testing had substantiated this hypothesis). However, given the results below, one wonders if the "read the data" questions might have served as a way of clarifying the structure of the graph for students prior to having them move onto the "read between the data" and "read beyond the data questions".

The data were analyzed by grouping responses to the question into categories used to reason about the question in relation to the data and to the features of the graph. These categories label reasoning strategies that are based on

- properties of the graph (considers both range of data and frequency)
  No, because the x's are not all on one number.
  No because the line plot shows X's on different numbers of raisins
  No, because there are a number of X's spread out showing different numbers
  No, because the X's show how many boxes had that many of raisins. Like 28 had 6 and 29 had 3.
  No, because they are different numbers along the bottom. The X shows how many students that found that number.
  No, because the X's are on different amounts or raisins
  No. If there were the same number in each box there would be X's all above the same number.

- literally "reading" the data from the graph.
  No, because there was 6 boxes of 28, 3 boxes of 29, 4 boxes of 30, 3 boxes of 31, 1 boxes in 32, 2 boxes in 34, 6 boxes in 35, 1 boxes in 36, 3 boxes in 38, and 1 boxes in 40.
  No there aren't the same number of rasins in each box, I found my answer by looking at the data, 6 boxes have 28, 3 have 29, 4 have 30, 3 have 31, 1 has 32, 2 have 34, 6 have 35, 1 has 36, 3 have 38, and 1 has 40.
  No, the data is all scattered out. There were 6 boxes with 28 raisins in them, 3 with 29 raisins in them, 4 with 30 raisins in them, 3 with 31 raisins in the, 1 with 32 raisins in them, 0 boxes with 33 in them and so on.

- properties related to the context or to the data.
  No, because they weigh the boxes until they equal 1/2 ounce. They don't count the raisins.
  No Because some raisins can be smaller and that means you can have more.

- range of the data (considers only range and does not include frequency)
  because it says the number of raisins goes from 26 to 40.
  No, there are not. All you have to do is look at the numbers on the bottom and it tells you how many raisans were in each box.

- frequency of occurrence/height of bars
  No, the X's have different numbers, so there are different numbers of X's in each box.
  No. The 28 boxs is not the same as 26 boxs. It only gose up to 6
  No, Because some do not have as much X's and some have more.
  No, only two because they are not the same height.
  No. If it was it would be evan across all the same number.
  No, because they Do not have the same number of X's.
  No. Because there isn't the same number of X's above each number.
  No. Because some of the numbers have a different number of X's.
  No, because there is more X's in some and less in others.

- other (includes incomplete, unclear, incorrect, or not statistically-reasoned responses)

  No. the chart tells you that the boxes don't have the same number of raisins in them.

  No, the graph shows that there are different numbers of raisins in each box

  No there are not. They all have different amounts.

  No! Because in each column there is different, and because when you add them they come out differently.

  No! Because you can look at them and tell that all of them are not the same because each one has a different number in each one of them

  No, I read the information on the line plot

A matrix (see Figure 2) has been used to provide the percent frequency of occurrence each category for pre- and post-test responses and the paired pre- with post-test responses. There are a number of different observations that may be made about these data:

Figure 2: Percent Category Responses - Question 1, Raisins

Category (Post ⇒) (Pre ⇓)	1	2	3	4	5	6	% TOTAL (Pre)
1. properties of the graph	14.5			7.9	3.9	1.3	27.6
2. reading the graph							
3. properties of the context/data						1.3	1.3
4. range of the data	2.6	1.3		2.6	1.3	1.3	9.2
5. frequency/ height of bars	7.9				3.9	5.2	17.1
6. other (incorrect)	3.9	2.6	2.6	3.9	9.2	22.3	44.7
% TOTAL (Post)	28.9	3.9	2.6	7.9	18.4	38.1	100.0

n=76

- A little over $\frac{1}{4}$ of the students offered explanations for their responses that indicated that they had an understanding of the role of both the data displayed on the axis and the frequencies noted by the X's displayed above the axis.
- A large number of students on both the pre-test (45%) and the post-test (38%) offered explanations that were coded as "other", meaning that they were not judged acceptable in terms providing appropriate reasoning for the answer given.

- On the pre- and post-tests, a similar percentage (17% - 18%) of students indicated in their explanations that their focus was the frequency or the numbers of X's; these students seemed to be saying or actually did say that to have the same number of raisins in each box meant that the columns of X's needed to be the same heights.
- 22% of the students' reasoning appeared to be stable across time, i.e., students who used properties of the graph (15%), range of the data (3%), and frequency/height of the bars (4%)

Our analysis in looking at this problem and student responses involved three parts: (1) to describe students' reasoning, (2) to consider the patterns that emerge in light of statistical reasoning behaviors that may be useful and/or desirable, and (3) to hypothesize other tasks or problems that may give further insights into or provoke changes in students' thinking. In this problem, a limited number of students (roughly 28% pre/post) were able to reason using information about the data values themselves (from the axis) and the frequencies of occurrence of these data values (the X's). The number of students who seemed to focus on the frequency or number of X's as the data values indicates that there may well be confusion even when using line plots about the role of data values and frequencies. We have found such confusions exist with students' reading of bar graphs; we attribute some of these confusions to having to read the frequency using the vertical axis. Here, this is not the case.

What is interesting about this problem is that students (with the exception of 2 out of the 76) answered the question correctly, i.e., no, there are not the same number of raisins in each box. However, once an explanation is given, it is clear that many students are looking at this graph in ways that provide incorrect reasoning for their answer. In addition, a large number of students provide vague or incomplete responses that seem to say "you know...the graph says this!" Part of this may result from our own lack of clarity on how we expect students to be able to talk about graphs. Still another part of this may reflect the usual emphasis in mathematics on "getting an answer" and seldom on the follow-up questioning that insists on clear explanations of why for the answers are given.

Using student responses to this question, it is possible to propose other kinds questions that address needs to clarify reasoning strategies further or to provoke the focus on both the data values and their frequencies as components involved in reading a graph. Consider the three problems shown below:

1. You are given the line plot labeled with the numbers of raisins possible in boxes of raisins. Someone has opened 5 boxes of raisins and each box has the same number of raisins. What is one way that the line plot that shows these data might look ? Why?

2. You are given the line plot labeled with numbers of raisins in the box. Someone has opened 5 boxes of raisins and two of the boxes have the same number of raisins. Of the remaining 3 boxes, each has a different number of raisins. What is one way that the line plot that shows these data might look ? Why?
3. Using the original line plot of the number of raisins in 30 small boxes, what do the four X's above 31 mean? Explain your reasoning.

Questions such as these or other similar questions fit within the scheme of "read the data", "read between the data", and "read beyond the data". In this case, they are based on the students' responses and are designed to begin to highlight the kind of clear thinking about graphs that is the goal of students' work in reading information using representations.

As is indicated on Figure 1, Question 3 is a "read beyond the data" question:

3. If the students opened one more box of raisins, how many raisins might they expect to find? Why do you think this is so?

This question involves inferring from the representation to make a prediction about an unknown case, i.e., opening another box of raisins. It draws on students' abilities to think about such topics as measures of center or clustering of the data.

The data were analyzed by grouping responses to the question into categories used to reason about the question in relation to the data and to the features of the graph. These categories include the use of

- modal reasoning
  35 or 28 because they have had a lot with that number.
  28 or 35 the most typical numbers.
  I think it would be 35 raisins in a box. More people had 35.

- data clusters
  They might find anywhere from 28-32 because that is where most of them are found.
  From 28 to 31 because they have been the most frequent.

- absence of data/"fill in the holes"
  26. It did not have any.
  Somewhere between 33 and 39 because there are not a lot through there and they are all spread out.

- range of the data/range of labels on the horizontal axis
  Probaly between 28 and 40 because 40 is the highest and 28 was the lowest. It is likely it would be there.

- frequency/height of bars
  I would estimate 41 because alot of boxes had 1.
  7. The other boxs has 6 or 3 in them.

- middle of the range
  I think that they would find 34 raisins because that the average I got.
  34, because it is in the middle of all the findings and it has more one.

- sequencing of horizontal scale
  I think 41 will be the number. From the start to finish it is in order from 26-40 next will be 41.
  Maybe 41. The numbers go up on more just like this 41 42 43 44
  They might find 41 because the number goes up in each box.

- identifying the median
  I think that they would find 31 rasins. I think this because 31 is the average [median].
  31 raisins, because I used the median step to help me. First I counted all the X's and then divided 2 into 30 and got 15. Last I counted 15 from the start and got my answer.

- other (includes incomplete, unclear, incorrect, or not statistically-reasoned responses)
  I think 6 because is the high number. I jut think it. I don't know if it just is low.(frequency/ height of bars)
  15, because 15 is the mode and it appears more frequently than the others. (median)
  If the students open one more box they might find not any raisins in that box or they might find a lot of raisins in that box.

A matrix (see Figure 3) has been used to provide the percent frequency of occurrence each category for pre- and post-test responses and the paired pre- with post-test responses. There are a number of different observations that may be made about these data:

- On the pre-test, 55% of the students used the modal response as a way to respond to the question; on the post-test 70% of the students used the modal response.
- Use of measures of center (the mode, median (or mean, which students did not use here)) or clustering of the data are two of the more appropriate strategies to use in responding to this question; 60% of the students on the pre-test and 82% of the students on the post-test used one of these strategies. Median is used by 5% of the students on the post-test and not at all on the pre-test, very possibly reflecting the content of the instructional module. On both pre- and post-tests, no students used the mean as a way to reason about a response to this question although all three measures of center were included in the instructional module.
- Category 9 responses decreased from 18% to 8% of all student responses. Category 7 responses decreased from 13% to 5% of all student responses.

- 54% of all students' reasoning appeared to be stable across time, i.e., students who used modal reasoning (49%), data clusters (1%), and sequencing of horizontal scale (4%) maintained these strategies. 4% of students demonstrating a category 9 response did so both pre- and post-test.

Figure 3: Percent Category Responses - Question 3, Raisins

Category (Post ⇒) (Pre ⇓)	1	2	3	4	5	6	7	8	9	% TOTAL (Pre)
1. modal reasoning	48.6	1.3	2.6					1.3	1.3	55.2
2. data clusters	2.6	1.3		1.3						5.2
3. absence of data/ "fill in the holes"		1.3							1.3	2.6
4. range of the data	1.3									1.3
5. frequency/ height of bars										
6. middle of the range	1.3							2.6		3.9
7. sequencing of horizontal scale	6.5	1.3					3.9		1.3	13.1
8. median										
9. other (incorrect)	9.2	1.3			1.3	1.3		1.3	3.9	18.4
% TOTAL (Post)	69.7	6.5	2.6	1.3	1.3	1.3	3.9	5.2	7.8	100.0

n=76

Again, our analysis in looking at this problem and student responses involved three parts: (1) to describe students' reasoning, (2) to consider the patterns that emerge in light of statistical reasoning behaviors that may be useful and/or desirable, and (3) to hypothesize other tasks or problems that may give further insights into or provoke changes in students' thinking. We have identified a number of different categories as they relate to this problem. We expect that some of these categories will emerge in other problem situations [as in Question 1 with a category that addressed students' focus on the range and, in this question, students' focus on the middle of the range]. We also anticipate that there may be additional categories of responses that did not surface with these students in responding to this question but that may do so in other circumstances.[7]

Part of using statistics involves being able to reason "sensibly" in situations. Students in this study were involved in learning about the three measures of center;

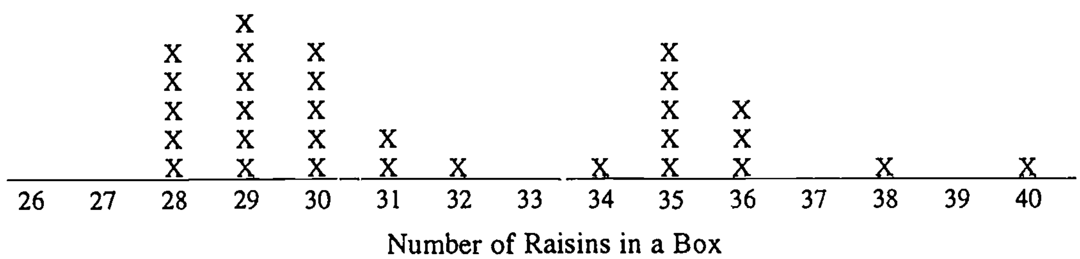
7 See Berenson, et al, (1993) for more detailed discussion of areas in graphical representaions on which elementary teachers fixated.

however, little time was actually devoted to exploring in what ways such measures are helpful in such activities as making predictions or in comparing two or more data sets. For this problem situation and question, one might argue that "clustering the data" is a very sensible strategy for making predictions. The challenge is to sort out what the cluster might be. At least one student selected the entire range as a way to cluster the data and make a prediction about the number of raisins in an unopened box. Another student indicated that from 28-32 was where most of the raisins were found; this is an obvious cluster in which over half the data are located. Both the median (31) and mean (31.3) are found within this cluster.

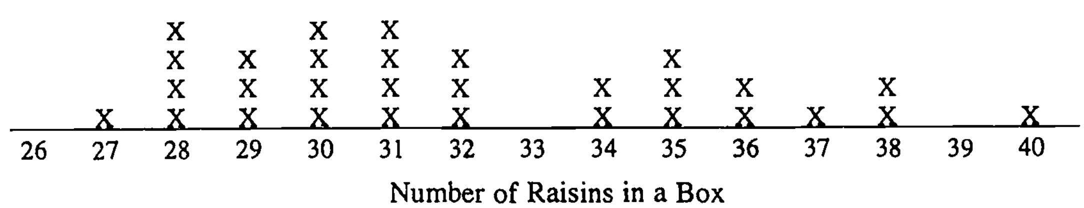
Using student responses to this question, it is possible to propose other kinds of [read beyond the data] questions that address needs to clarify reasoning strategies further or to provoke alternative reasoning strategies such as the use of clustering. Consider the three problems shown below:

1. Students opened another set of 30 boxes of raisins. Make a picture of what you think the line plot for these data might look like. Why do you think this is so?
2. If the students opened one more box of raisins, how many raisins might they expect to find? Why do you think this is so?

a.

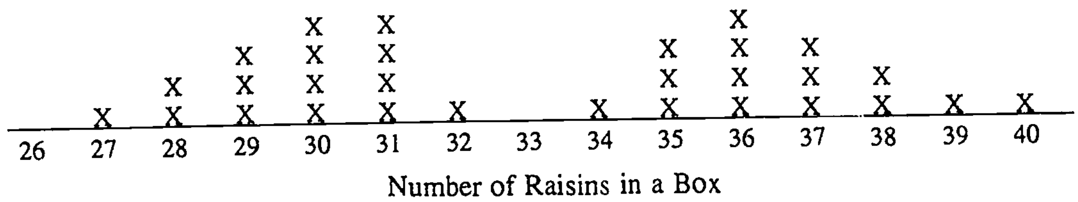


b.



3. Bill answered the question, "If the students opened one more box of raisins based on these data, how many raisins might they expect to find?" by writing, "I would expect to find the median number of raisins which is 33 raisins because the median is the middle number in the data."

   Do you agree with Bill? Explain why or why not. (See next page for graph)



Conclusions

Throughout the discussion of the results, several comments reflecting what these results mean have been made. For purposes here, we close with a more general conclusions:

- Students need to talk more about graphs, including talking about the structure of graphs and how this informs the statements they can make about the information shown by the graphs and/or about making predictions or drawing inferences.
- What do we consider a good "explanation" in the case of question 1? The data suggest that students had many different reasons for responding. An explanation that considers both the data elements and the frequencies as part of the reasoning reflects an understanding of the graph structure. Possibly, students think that the answer is so obvious that they do not know what to say by way of explanation. On the other hand, possibly we are unclear about what we consider an appropriate explanation, and students are left with "fuzzy" ways to reason about this kind of situation because of our own lack of clarity.
- Many of the experiences that students had with data in this study involved describing data and describing what's typical about the data. Measures of center were introduced as tools for description. Beyond the focus of what's typical about the data, there were not many situations that addressed the question of prediction. In this case, student the majority of students focused on mode as a way of predicting the number of raisins that would be found in the next box that was opened. However, thinking about clusters in the data may be a more useful strategy in this situation. Students need opportunities to reason about data from a "prediction" stand point.

With the increasing inclusion of statistics content across the K-12 curriculum, it is possible to begin to explore the development of thinking in this area. What has been discussed in this paper begins to focus this discussion on the role of graphs as part of the process of statistical investigation.

References

Bamberger, C.F. (1942.) *Interpretation of graphs at the elementary school level.* Washington, DC: The Catholic Education Press.

Berenson, S.B., Friel, S., and Bright, G. (1993.) *Elemetnary teachers' fixationson graphical features to interpret statistical data.* Atalanta, GA: Paper presented at the Annual Meeting of the American Educational research Association.

Carpenter, T., Coburn, T.G., Reys, R.E., and Wilson, J.W. (1978.) *Results from the first mathematics assessment of the National Assessment of Educational Progress.* Reston, VA: The National Council of Teachers of Mathematics.

Carpenter, T., Corbitt, M.K., Kepner, H.S., Jr., Lindquist, M.M. and Reys, R.E. (1971.) *Results from the first mathematics assessment of the National Assessment of Educational Progress.* Reston, VA: The National Council of Teachers of Mathematics.

Culbertson, H.M. and R.D. Powers (1959.) A study of graph comprehension difficulties. *AV Communication Review.* 7 (1), 97-110.

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

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M., and Chambers, D.L. (1988.) *The Mathematics report card: Are we measuring up?* Princeton, NJ: Educational Testing Service.

Feliciano, G.D., Powers, R.D., and B.E. Kearl. (1963.) The presentation of statistical information. *AV Communication Review.* 11(3), 32-39.

Friel, S.N. and Bright, G. (1995). *Assessing students' understanding of graphs: Instruments and instructional module.* Chapel Hill, NC: UNC Mathematics and Science Education Network.

Graham, A. (1987.) *Statistical Investigations in the secondary school.* New York: Cambridge University Press.

Kirk, S.K., Eggen, P.D., and Kauchak, D.P. (1980.) *Generalizing from graphs: Developing a basic skill through improved teaching techniques.* St. Louis, Missouri: Paper presented at the International Reading Association Annual Meeting.

Lappan, G., Fey, J., Fitzgerald, W., Friel, S., and Phillips, B. (1996.) *Data about us.* Palo Alto, CA: Dale Seymour Publications.

Lindquist, M.M. (Ed.) (1989.) *Results from the first mathematics assessment of the National Assessment of Educational Progress.* Reston, VA: The National Council of Teachers of Mathematics.

Friel, S.N., and Joyner, J. (Eds.) (1996.) *Teach-Stat for teachers: Professional development manual.* Palo Alto, CA: Dale Seymour Publications.

Gallimore, M. (1991.) Graphicacy in the primary curriculum. in Vere-Jones, D. (Ed.) *Proceedings of the Third International Conference on Teaching Statistics.* New Zealand: Otago University Press.

Gleason, J., Vesilind, E., Friel, S.N., and Joyner, J. (Eds.) (1996.) *Teach-Stat for statistics educators: Staff development manual.* Palo Alto, CA: Dale Seymour Publications.

Padilla, M. J., McKenzie, D.L., and Shaw, E.L. Jr. (1986.) An examination of the line graphing ability of student in grades seven through twelve. *School Science and Mathematics.* 86 (1), 20-26.

Pereira-Mendoza, L. and Mellor, J. (1991.) Students' concepts of bar graphs - some preliminary findings. in Vere-Jones, D. (Ed.) *Proceedings of the Third International Conference on Teaching Statistics.* New Zealand: Otago University Press.

Pereira-Mendoza, L. (1992.) Students' concepts of pictographs and bar graphs. *Proceedings of the National Diagnostics and Prescriptive Mathematics Conference.*

Pereira-Mendoza, L. (1995.) Graphing in the primary school: Algorithm versus comprehension. *Teaching Statistics.* 17, 1, 2-6.

Perry, M., Kader, G. and P. Holmes. (1994). Learning statistics with technology. *Mathematics Teaching in the Middle School*, vol. 1, 2.

Rangecroft, M. (1991a.) Graphwork - Developing a progression Part 1 - The early stages. *Teaching Statistics.* 18, 2, 44-46.

Rangecroft, M. (1991b.) Graphwork - Developing a progression Part 2 - A diversity of graphs. *Teaching Statistics.* 18, 8, 90-92.

Rangecroft, M. (1994.) Graphwork - Developing a progression. in Green, D. (Ed.) *Teaching Statistics at its Best.* Sheffield, England: The Teaching Statistics Trust.

Wainer, H. (1980.) A test of graphicacy in children. Final report of research (NSF SOC 76-17768). *Educational Researcher.* 21(1), 14-23.

Wainer, H. (1992.) Understanding graphs and tables. *Educational Researcher.* 21(1), 14-23.

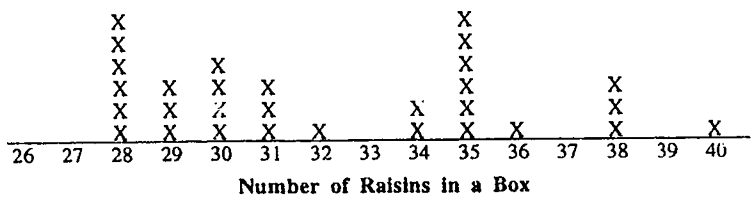
Figure 1 - Graphs and Questions Used on Written Pre-Post Tests

	Part I	Part II	Part III	Part IV
Graph	See attached copy of line plot: Number of Raisins in a Box	See attached copy of bar graph: Lengths of Cats in Inches	See attached copy of stem plot: Minutes to Travel to School	See attached copy of histogram: Allowances for 60 Students
Reading the data	(How many boxes of raisins have 35 raisins in them?)[8]	How many cats are 30 inches long from nose to tail? How can you tell?	Write down are the three shortest travel times that students took to get to school. Write down the three longest travel times students took to get to school.	Jack wanted to know how many students had an allowance of $4.25. Can he use these data to find the answer? Explain why or why not.
Reading between the data	Are there the same number of raisins in each box? How can you tell? How many boxes of raisins had more than 34 raisins in them? How can you tell?	How many cats are there in all? How can you tell? If you added up the lengths of the three shortest cats, what would the total of those lengths be? How can you tell?	How many students are there in the class? How can you tell? How many students took less than 15 minutes to travel to school? How can you tell?	What do you know about the allowances of students in the two tallest bars? Explain your answer.

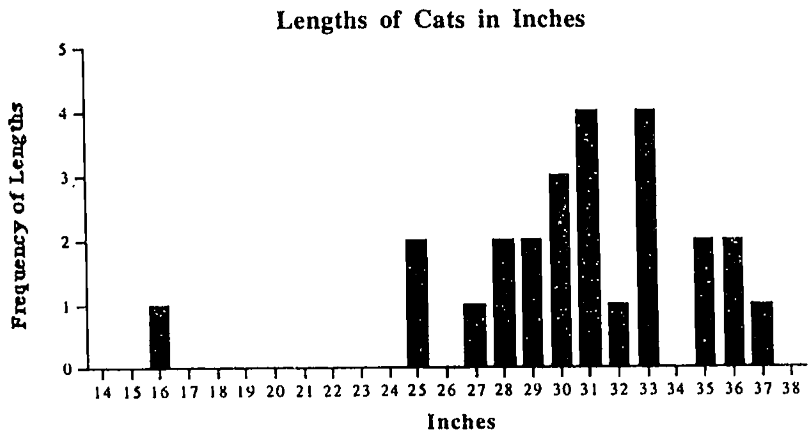
[8] Note: On this written test, a specific question related to "reading the data" was not asked; the question shown is a sample of the kind of question that can be asked.

Reading beyond the data	If the students opened one more box of raisins, how man raisins might they expect to find? Why do you think this?	What is the typical length of a cat from nose to tail? Explain your answer.	What is the typical time it takes for students to travel to school? Explain your answer.	Another students' parents have agreed to look at data about kids allowances. Then they will decide what allowance to give the student. Using the histogram showing this information for 60 students, what allowance do you think the student should make a case for? Why?
Graph Construction	Make bar graph from the line plot.	Make line plot from the bar graph.	Make a histogram to show the information about travel time that is displayed on the stem-and-leaf plot.	

PART I



PART II



PART III

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 1 3 5 6 6 8 9
2	0 1 3 3 3 5 5 8 8
3	0 5
4	5

PART IV