An Authentic Assessment of Students' Statistical Knowledge

Joan B. Garfield

Probability and statistics are increasingly being given an important place in the K–12 mathematics curriculum. According to the NCTM Curriculum and Evaluation Standards (1989), students should learn to apply probability and statistics concepts to solve problems and evaluate information in the world around them. The statistics standards suggest using hands-on activities involving collecting and organizing data, representing and modeling data including the use of technology, and communicating ideas verbally and in written reports. Teachers are encouraged to help students develop important ideas (for example, about distributions, randomness, and bias) and gain experience in choosing appropriate techniques to use in analyzing data.

Many teachers are currently using materials from recent projects or projects in development that have developed curricula and software to implement the NCTM Standards (e.g., the Quantitative Literacy Project, the Reasoning under Uncertainty Project, and the ChancePlus Project). These new materials encourage teachers to have students work on statistical projects: formulate research questions, collect and analyze data, and write up the results. Working on statistical projects individually or in groups engages students in learning about statistics and helps them to integrate the knowledge they have learned.

The Need for New Assessment Methods for Statistics

Historically, most mathematics testing has focused on students' computational skills and few tests have measured higher-order thinking. Statistics items that appear in traditional tests typically test students' ability to calculate correctly the mean and median for a set of numbers or to read a number from a graph. Tests composed of these items not only test skills in
isolation of a problem context but do not test whether or not students understand how these statistical measures are interpreted or know when one is a better summary measure to use than another. They also fail to assess students’ ability to integrate statistical knowledge to solve a problem and their ability to communicate using the statistical language. Alternative forms of assessing statistical learning are needed that will inform teachers about how well students can communicate using the statistical language, understand statistics as an interrelated set of ideas, and how well students are able to interpret a particular set of data.

In reviewing the NCTM Standards for assessment, Webb and Romberg (1988) provide criteria for new assessment instruments for mathematics. These criteria can also be applied to the development of statistical assessment materials. Such instruments should—

1. provide information that will help teachers make decisions for the improvement of instruction;
2. be aligned with instructional goals;
3. provide information on what students know;
4. supplement other assessment results to provide a global description of what mathematics students know.

The “Practical Project” as an Assessment Method

The “practical project” was originally designed as a learning activity to help students integrate what they had learned in preparing for a cumulative exam. In grading these problems, I learned that they are extremely useful indicators of students’ understanding of statistical ideas and their ability to apply these ideas in analyzing data and that they offer valuable insights to the teacher about where additional instruction is needed.

There are two versions of the practical project. In the first version, students collect a set of data of interest to them (consisting of 20 to 40 values) that they will be describing and exploring. Many ideas for data sets are presented and discussed in class, and students are strongly encouraged to gather data in which they are really interested. Examples are given of different types of sports data (data for various teams or individual team members) or data related to popular music (number of minutes of songs, CD prices at different stores). Students are encouraged to look for data sets in magazines (e.g., Car and Driver, Consumer Reports), almanacs, and the newspaper. Some students choose to collect their own data rather than use an existing data set. One student decided to find out the cost of a dozen roses at a variety of florist shops in the city. Other students have collected data at part-time jobs, using sales receipts, tips earned waitressing, or gasoline sales. This is a good opportunity to exploit student diversity.

In the other version of this project, students collect information about themselves every day for three to five weeks. Types of data collected have been the number of minutes spent each day on homework, talking on the telephone, or watching television or the amount of money spent or earned each day. Before these projects are begun, it is important to have each student submit an explanation of what data she or he has selected or decided to collect along with a sample of the data values. The teacher should examine the data values to make sure they are quantitative data and have enough variability to be worth analyzing. Students are instructed to take accurate measurements of time, money, and weight, and not round off their data values to the nearest whole number.

Guidelines for Data Analysis

The analysis of the data can be done individually or in groups (the group selects a particular data set to work on together). Students should use calculators for calculating statistical measures. They should use the following guidelines both in analyzing the data and in reporting the results:

1. Describe the data you collected. What do the data represent? Why did you choose these data? How did you collect them?
2. Summarize the data in a table and use different types of plots to graph the data.
3. Calculate appropriate summary statistics for your data including measures of center and variability. Show your calculations.
4. Write a description of your data set.
   a. Describe the information revealed by the plots, the shapes of the distributions, and how different plots give different information about your data set.
   b. Interpret the summary statistics in terms of the data. Specifically, explain what the measures of center indicate and how the different measures compare.
   c. Does your distribution look like a normal distribution? Why or why not?
5. What did you learn about your data? What questions do you have about your data now that you have analyzed it? What would you do differently next time? What other variables that relate to this one might you choose to analyze?

Evaluating the Practical Project

An analytic scoring method, adapted from the holistic scoring method offered by Charles, Lester, and O’Daffer (1987) for evaluating student solutions
to mathematical problems, can be used to grade the projects. A score of 0 to 3 points is assigned in each of the following categories:

- Communication
- Visual representations
- Statistical calculations
- Decision making
- Interpretation of results
- Drawing conclusions

According to this rating scheme, 3 points indicates correct use, 2 points indicates partially correct use, 1 point indicates incorrect use, and 0 means there was not enough information to evaluate or that this part of the project was missing or incomplete. Because this is a project that requires a large amount of student effort both in analyzing data and in writing up the results, this modified scheme still allows students to earn 1 point for each component if the effort was made but was incorrect. Explanations of the categories and scores along with samples of students’ work are given below.

- **Communication**: The appropriate use of statistical language and symbols

<table>
<thead>
<tr>
<th>Score</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inappropriate use: student misuses words or symbols or uses statistical words in a context that does not make sense.</td>
</tr>
</tbody>
</table>

**Example**: “This data that I collected tells me that the central value measurements were the best way to formulate these data. Most of the information given was very close to the average and was a better way to represent this material.” It is hard to figure out what the student is trying to communicate here. Several different ideas seem to be confused.

2  
Partly appropriate use of language or symbols

**Example**: “The average and the trimmed average are quite comparable. I feel the average works best for me. The median is slightly higher, it is the center of all my times but on most nights I got less than 7 hours sleep; not more.... The center of the data can be different depending on how you find it.” There is some confusion in describing the measures of center and using them in interpretive statements.

3  
Correct use of language and symbols

**Example**: “The histogram is skewed to the right or to the higher values. The histogram and the box plot give me the same type of information. There are no outliers and the majority of the values are at 20 or below. The box plot does give me more information, showing that 50% of the values are below 10 and 25% are between 10 and 20.”

- **Visual representations**: The appropriate construction and display of tables and graphs

<table>
<thead>
<tr>
<th>Score</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Errors in constructing or labeling tables and graphs. For example, leaving out a vertical scale for a histogram or failing to use an equal interval scale for a box plot as shown below:</td>
</tr>
</tbody>
</table>

**Incorrect scale**: 

| 5 | 25 | 32 | 60 | 69 |

**Correct scale**: 

| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 |

**Note**: Although the medians and quartiles might be correctly calculated for the box plot, the score of 1 reflects errors only in graphs. Calculations receive a separate score. Other types of errors might be using bars with unequal widths in histograms or having sequential bars not touching each other.

2  
Tables and graphs have some errors but are mostly correct.

Figure 22.1 shows the introduction, stem-and-leaf plot, histogram, and box plot for Student A. This sample indicates that the student is somewhat weak in constructing graphs. Although the stem-and-leaf plot is correct, it is missing a code and label that tell how to interpret the numbers and what they represent. The traditional histogram is lacking a vertical scale and the horizontal scale is not correctly labeled. The box plot appears to be correct, although it is sloppily drawn. I would give this student a score of 2 for visual representation.

3  
Tables and graphs are correctly constructed.

Figure 22.2 shows a sample of another student’s work. All graphs are correctly constructed and labeled. I would give this student a score of 3 for visual representation.

- **Statistical calculations**: Statistical measures are calculated correctly and appear reasonable.

<table>
<thead>
<tr>
<th>Score</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Errors in calculation lead to answers that are unreasonable, or a formula is not used correctly. For example, students use the median rank instead of the actual median data value, fail to average the middle values in finding the median, give a negative value for a variance or standard deviation, or calculate a mean that is obviously larger or smaller than the rest of the data values.</td>
</tr>
<tr>
<td>2</td>
<td>Some correct, some incorrect calculations</td>
</tr>
<tr>
<td>3</td>
<td>Calculations appear to be done correctly.</td>
</tr>
</tbody>
</table>
I am using data from the men's US Open golf tournament. This was held at Hazeltine Golf Course outside Minneapolis. The data consists of the men's first day scores of the final Championship weekend. I have a tremendous interest in golf and since I attended every day of the golf tournament I thought it would be fun to see the data set used in the particular exercise.

![Histograms](image)

Fig. 22.1. Portions of Student A's practical project

*Decision making:* The tables or graphs selected are appropriate for representing the data; the appropriate summary measures are calculated.

<table>
<thead>
<tr>
<th>Score</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inappropriate tables or graphs are constructed, and decisions do not appear to be based on obvious clues in the data. For example: a histogram is made with 20 short bars indicating a bumpy pattern and no apparent shape instead of condensing it to a fewer number of bars to better reveal the shape of the distribution.</td>
</tr>
<tr>
<td>2</td>
<td>Some decisions seem inappropriate, whereas others are appropriate.</td>
</tr>
<tr>
<td>3</td>
<td>All decisions appear to have been made correctly.</td>
</tr>
</tbody>
</table>

![Box plot](image)

![Leaf and Stem](image)

Fig. 22.2. Portions of Student B's practical project

*Interpretation of results:* The ability to use information from representations and summary measures to describe a data set.

<table>
<thead>
<tr>
<th>Score</th>
<th>Interpretation</th>
</tr>
</thead>
</table>
| 1     | The student seems unable to interpret the plots and measures. For example, the student fails to recognize the shape of the distribution (when it is obviously skewed, bell-shaped, rectangular, etc.), makes unreasonable or incorrect statements (e.g., "The standard deviation and the interquartile range are too close to the average" or "The average is a little smaller than the median, which would mean there are more smaller numbers than $50 than numbers larger than $50."). describes a distribution as normal when the analysis indicates it is
not normal, or merely restates information without interpreting what it means.

```
2 Interpretation is too brief, the student fails to interpret some important information, or the interpretation is partly correct and partly incorrect.
3 Good interpretation of data set using all appropriate information
```

Here are examples of good interpretations from students' papers:

```
"From the shape of the histogram and the fact that the mean, median, and mode are all close together, the data seem to have a normal distribution."

"An outlier with a large value influenced the mean."

"The trimmed average and the average are very close, meaning there are no outliers."

"My data set consisted of how much money I spent each day for 29 days. My graphs show that my data values are skewed to the right, which means I spent a lot of money just a couple of times while I spent a little money many times. The mean says I spent about $14 a day, but I didn't. I spent maybe no money one day while I spent $50 the next day. The median seems more reasonable, saying that I spent less than $4.50 a day for half the days and more than $4.50 a day for half the days."
```

* Drawing conclusions: The ability to draw conclusions about the data, point out missing information, or relate this study to other information

Score Interpretation

1 Student fails to draw conclusions or draws conclusions not substantiated by the data. Or the student fails to notice inconsistencies in conclusions made.

Example:

A student who studied the number of minutes slept each night reported a range of 375 minutes and a variance of .05 minutes. The student concluded, "Although the range was quite large, the variance and standard deviation prove that I get a very similar amount of sleep (.05 variation) over a three-week period." It is clear that the student did not see the contradiction implied by a range of 375 minutes and a variance of .05 minutes. The student also erroneously believes that the measures "prove" something: that he got a similar amount of sleep each night.

2 Weak conclusions are made, but some attempt is made to look beyond the data set.

3 Conclusions are made on the basis of the data analysis. Comments are made about the study as it relates to the real world.

AN AUTHENTIC ASSESSMENT OF STUDENTS' STATISTICAL KNOWLEDGE

Example:

"My data set suggest that I usually get either 3–4 hours of sleep or 7–8 hours of sleep. These are the two peaks in my data. I find this interesting and it also makes sense to me—since I work every weekend morning, I rarely get sleep on the weekends. I try to catch up more during the week."

Scoring Students' Projects

I have designed a rating sheet that I use to score student projects. This sheet is stapled to each student's paper and provides information both on the scoring categories and on why students lost points. Figure 22.3 shows a student's rating sheet.

```
Practical Project Rating Sheet

Name: Sarah H.

1. Appropriate use of statistical language and symbols: (3 pts.) 3
   Comments:

2. Appropriate construction and display of tables and graphs: (3 pts.) 2
   Comments: Missing code and label for stem and leaf plots.

3. Correctness of statistical calculations: (3 pts.) 3
   Comments:

4. Appropriate choice of tables, graphs, and summary statistics: (3 pts.) 2
   Comments: Should have used a regular boxplot to show outliers.

5. Reasonable descriptions and interpretations of data: (3 pts.) 3
   Comments:

6. Appropriateness of conclusions: (3 pts.) 2
   Comments: You did not discuss the outliers and how they affected your analysis and results.

Total score: (18 pts.) 16
```

Fig. 22.3. Sarah's rating sheet

How the "Practical Project" Meets the Criteria for New Assessment Methods

The project provides information on what students know. Scores may be given to students for each category as well as a final score ranging from 0 to 18 points. The individual scores indicate to students their strengths and
weaknesses, so that they can improve weak areas. Teachers wishing to assign a letter grade on this problem would be cautioned not to base this grade exclusively on the percentage of total points earned.

It provides information that will help improve instruction. Teachers can use the category scores to judge where more activities and instruction are needed.

It can be aligned with instructional goals. The “practical project” as an assessment method is aligned with the instructional goals stated in the NCTM Standards. Students are engaged in the basic activities of data exploration by conducting investigations involving higher-order thinking and communication.

The project helps describe what mathematics students know. This method provides useful information about how students use their statistical knowledge and skills in solving an applied problem. The results can supplement other assessment information (such as quizzes and tests) to furnish a better understanding of what statistics students know, and the results can be combined with other assessment information to provide a broader description of the mathematics students know.

An added benefit is that students often learn something interesting about the world or about themselves as a result of working through the project. This is especially true if they have analyzed data about themselves. Students have commented on discoveries made about their habits of spending money, their sleep patterns, and their television watching. Students who have collected data on car performance or flower prices have commented that they learned something interesting they did not know before about the variability of those data. Another benefit is that although students initially may express reactions about the problem’s being too long, too hard, or too much work, they often comment afterward that the process was very helpful in preparing for a test, in figuring out what to study, and in pulling together what they had learned. Students also show a sense of pride in being able actually to “do” statistics.

REFERENCES

