From Phillips, Brian (Ed.), (1997) Papers on Statistical Education presented at ICME-8 (International Congress on Mathematics Education-8) Seville, Spain, July 14-21, 1996. Hawthorn, Australlia: Swinburne Press. ISBN 0 85590 753 3

ASSESSING STUDENTS' INTERPRETATION OF DATA

Iddo Gal
University of Haifa
Israel

Abstract

This paper explores some conceptual and practical challenges that assessment of students' interpretation of statistical data poses to teachers and researchers. Issues involved in eliciting and evaluating students' opinions about data are examined, with illustrations related to students' analysis of data in simple 2-by-2 tables. Implications for instructional practices in statistics education are discussed.

This paper is based in part on work funded by a National Science Foundation grant #90-50006 to the University of Pennsylvania, USA. The opinions expressed do not reflect those of NSF or the University. Thanks to Caroline Brayer Ebby and Patricia Mahoney for their help in qualitative analysis which informed the writing of this paper.

Introduction

The key skills that statistics education aims to develop (NCTM, 1989; Moore, 1990) can be grouped into two overlapping clusters, labelled "generative" and "interpretive". *Generative skills* are involved in acting upon data and "doing" statistics, usually as part of a statistical project, and appear to be the target of the bulk of instruction in statistics. *Interpretive skills* are invoked when students evaluate and communicate about the meaning of data or findings. Interpretation takes place either in *reporting (active) contexts,* where students make sense of patterns in the data they analyzed and discuss implications of their findings, or in *listening (passive) contexts,* where students have to make sense of, or react to, messages with embedded statistical elements or which are based on statistical studies, such as those encountered in the media or the workplace.

While the skills required in generative, reporting, and listening contexts only partially overlap (Gal, in press), in both contexts students are involved in creating or communicating *opinions.* Unfortunately, little has been written about issues involved in assessing students' opinions about data, even though they are a key gateway into students' reasoning about data. Issues involved in eliciting and evaluating the quality of students' opinions about data are illustrated below with regard to students' interpretation of data in tables.

Opinions about data in tables

Tables can be used to organize raw data or to display summary indices, in both generative and interpretive contexts. When examining tables, students should be not only able to find needed details in a table, but also to identify and make sense of differences between groups, associations between variables, or trends over time, and present or discuss their or others' opinions about any such trends. Thus, interpretation of data in tables can be a fertile context for instruction and assessment.

When asking students to interpret information in tables (and other types of displays, such as graphs), at least two different types of questions can be posed, termed here "literal reading questions" and "opinion questions". This distinction corresponds to one first introduced by Curcio (1989) and Wainer (1992) in the context of graph comprehension, though these authors have not extended their frameworks to information in tables, nor have they elaborated on assessment issues.

Literal reading questions require students to "lift" numbers from specific cells in a table (or locations in a graph), or compare certain numbers or data points. Answers to literal reading questions (see example later) are simple and can be unambiguously classified as either "right" or "wrong". In contrast, opinion questions aim to elicit students' ideas about overall patterns of data in a table (or a graph); such questions lead to judgments that cannot necessarily be classified as right or wrong. When forming an opinion about data, students may have to refer to mathematical relationships between data elements in the display, consider other information about the problem context, or consult world knowledge they may have, to help in ascribing meaning to the data.

Issues involved in eliciting and evaluating students' opinions about information in tables are discussed below, using examples from a research project funded by the National Science Foundation. Among other things, students were presented with extended problems involving actors who need to make decisions based on some data. One of the contexts involved a story about a toy company that wants to sell one of two new educational games in schools. Students were told that the company surveyed children to find which of the games may be more popular. After asking students about sampling issues, they were shown results from several surveys that the company presumably conducted, presented in tabular displays as illustrated in Figures 1-2, and presented with literal reading and opinion questions.

	Game A	Game B
School 1	20	5
School 2	15	10

Figure 1. Toy company survey results game by school.

	Game A	Game B
Boys	18	17
Girls	12	13

Figure 2. Toy company survey results game by sex

Posing questions.

We have developed the following guidelines for shaping an initial opinion question:

1. It should not provide hints as to what parts of the table or graph students should examine.
2. It should suggest to the student that a judgment is required, rather than a precise computational response, i.e., a single number.
3. It should represent what an actor in the given problem context may reasonably want to know for functional reasons.

These guidelines aim to lead to development of questions that enable discovery of students' ideas about data analysis, and create a context where students are inclined to look for several possible pattern in the data. If students are not presented with a reasonably realistic context, they may not bring into the task all their reasoning skills and world knowledge. We also note that the need for opinions (as opposed to more impoverished answers), will naturally emerge when students are presented with tasks involving comparisons, e.g., of datasets, subgroups, or measurements at different points in time.

Below are two examples for opinion questions implementing these guidelines:

1. About Figure 1: *In general, do you think that students from the two schools differ in their preferences, or do they have about the same kinds of preferences for the two games?*
2. About Figure 2: *In your opinion, do boys and girls have more or less the same kinds of preferences, or do they have different preferences?*

In comparison, literal reading questions, such as, "*How many children in school 1 preferred game A?*", (see Figure 1) often violate all three principles. They direct students to specific cells. Their phrasing does not invite the formation of a judgment, but rather implies the need to provide a single number as an answer. They do not create a sense that the answer given can serve any functional purpose within the given context.

The sample questions above are open-ended yet each directs students' attention to a specific "view" of a table. In piloting table interpretation questions we initially tried a generic, open phrasing, *What in your opinion can the toy company conclude from the results in this table?*, which seemed to embody our guidelines, yet led to too many uninformative responses. For example, students just reiterated the numbers they saw in each cell, or claimed that "some children liked each game." These responses were not improper in any (mathematical) sense, yet revealed little about students' understanding of the data. We realized that a more direct form of question posing should be used.

Choosing data values.

The nature of data values used in a display, and the number and arrangement of such values, are important as they shape what statistical ideas students have to address when interpreting data. Figures 1 and 2 illustrate two of several data displays we used to explore two topics of interest in our own work:

- how students compare data coming from groups of unequal sizes, which is similar to comparing results from samples of unequal size, a situation common in quantitative research; and
- how students evaluate the meaning of differences between groups, an issue related to the notion of statistical significance. (see Gal, in press, for more details).

Note that these topics are raised by the data in Figure 2 even though only simple counts are used. Percentages or averages could be used instead (if the cover story is changed and information about sample sizes added). The choice of different types of data, in simple or in multi-way tables, offers ways to adapt the level of complexity of the interpretive task to the students' level.

Evaluating reasonableness of opinions

When assessing students' knowledge in many domains in mathematics, many teachers use problems with one or with multiple right answers. Even when there are multiple ways of reaching an answer, since answers can still be either "correct" or "wrong," a teacher does not have to explore the reasoning process or strategy that led to the answer in order to determine its correctness even (though this is not advisable; —see NCTM, 1995). In this context we ask, is it possible to evaluate correctness of an answer apart from the reasoning that led to it, when assessing a student's opinions about the patterns or trends in statistical data?

Regarding Figure 2, for example, we asked students if, "Boys and girls have more or less the same kinds of preferences, or whether they have different preferences." What would be a reasonable answer to such a question? It is reasonable to claim, "Boys and Girls have the same preferences," as essentially an equal number preferred each game within each gender group. Yet, if a student takes the word "difference" literally, it is not unreasonable to argue that, "More Girls liked game A and more Boys liked game B." The upshot is that determining the degree of reasonableness in this case may not be a straightforward process unless one also considers the reasoning process and the assumptions made by students (such as about the meaning of "same" and "different").

To illustrate further the range of issues involved in evaluating the reasonableness of student opinions, below are abbreviated and annotated quotes from initial opinions regarding the data values in Figure 2, given in response to the second sample question above.

1. *"They like the same games . . . Because they each were only 1 apart. [points to the numbers in each row, 17-18 and 12-13]."*
   An unambiguous opinion, justified by comparing the relevant pairs of values from each row in the table. The opinion and its justification are reasonable and are presented here as a benchmark against which to compare the quotes below.
2. *"Boys and girls prefer different games. The majority of boys [points to 18] prefer A, the majority of the girls [points to 13] prefer B."*
   The answer is reasonable if we assume that the student believes *any* difference is a valid difference. Is this assumption justified? The use of "majority" is disturbing, and could cause confusion in reporting contexts.
3. *"A larger amount of boys picked both games, but there were 25 girls and 35 boys so of course there will be a larger amount. If you compare 18 to 17 and 12 to 13, that tells me that boys like both games almost the same as girls."*
   The student verbalizes some of the reasoning process before reaching a conclusion. He mentions existence of different group sizes and how a seeming conflict between absolute numbers is resolved. The final conclusion is a bit ambiguous, in part due to use of "almost the same," as well as insinuation that a difference of 1 cannot be ignored.
4. *"Game A has more boys that like that game, but game B is almost equal . . . I saw that there was a ratio of 6 to 4. [What do you mean by ratio?]. Six people is definitely a big separation to me [points to the difference between 18 and 12]."*
   If only the first sentence was given as an answer (e.g., in a brief open-ended written question), the opinion may be judged unreasonable; not enough information would be available to suggest that the student's table reading strategy is incorrect (i.e., numbers are compared in *columns*). The use of "ratio" is disturbing as the student is actually referring to a simple difference.
5. *"There are 10 more boys than girls. If you took away those 10 extra boys it would be equal. Like if you take away 6 from the 18 it would be 12 and if you took away 4 from the 17 it would be 13 and 6 + 4 = 10 so boys and girls agree on the same."*
   The student is troubled by the difference between the number of boys and girls and adjusts the group sizes. However, the student also makes a blind assumption how the votes will look like if the group sizes were equal, and then provides an opinion.

6. *"Well, in both games, the boys liked it more than the girls did . . . I really don't think girls are into that stuff anyway . . . I think the girls have totally different tastes than the boys do . . . boys like all those rowdy games . . ."*
   The student compares absolute numbers in columns. The student adds comments suggesting that his conclusion is compatible with preconceived assumptions about preferences or behavioral patterns of boys and girls.

These examples suggest that a wide range of opinions can be expected from students based on the same set of data.

Follow-up questioning.
Students often describe only the *outcome* of a thinking process, and may not necessarily explicate the reasoning *process* that led to their answer. Follow up questions are needed to help determine what parts of a display were examined or ignored, what patterns were noticed, how a student related elements (rows, columns, cells, points) in the display to each other, or what assumptions informed the reasoning process. In a related context, Gal, Rothschild, & Wagner (1990) found that some students comparing two line plots pay attention to only a subset of the data points, or have trouble integrating separate features in a display into a coherent picture.

Follow-up questioning may also be required to clarify opinions that are unclear, too brief, or incomplete. Such opinions are not uncommon with students whose linguistic skills are still developing (including bilingual students), or with those who are not used to having to describe their reasoning, perhaps because discussion is not promoted in their classrooms. Another important and sometimes perplexing issue is the evaluation of the meaning behind "non-committing" opinions. For example, what hides behind the response "*Boys and girls look kind of similar*" with regard to the data in figure 2 above? Consider the following:

1. The student is not ready to present a clear and committed opinion, e.g., because he/she is unsure how to relate different features in the display to each other, or what all the numbers mean, and so uses a "hedging" sentence to avoid embarrassment.
2. The student notices a small difference between boys and girls, considers it insignificant, and uses the qualifying term "kind of" to add this observation to the judgment "similar".
3. The student interprets the difference between groups as meaningful, yet appears non-committing because his or her opinion is expressed using linguistic features ("kind of") that comply with modes of speaking common within the student's local culture (maybe similarly to how many contemporary teenagers in the U.S. insert "like" and "you know" into their utterances).

An educator cannot be certain that the meaning he or she attaches to students' utterance is compatible with the meaning attached to it by the students. There may be confusion between mathematical and everyday meanings of certain terms (Laborde, 1990), and there may be individual differences in the meaning attached to probabilistic phrases (Wallsten, Fillenbaum, & Cox, 1986).

Discussion and implications

The development of students' ability to generate clear, sufficiently self-explanatory, and sensible opinions should be a target area for instruction, since opinions are a core outcome of the analysis and interpretation of statistical data. As was illustrated above, even simple count data in 2x2 tables can give rise to a diverse set of responses, if "opinion questions" as defined here are employed.

Our work suggests that two separate types of questions must be presented to students as part of classroom or research-related assessment of opinions about data. The first should elicit the student's opinion, and the second should elicit information about the evidential basis for the opinion, the reasoning process, and the strategy used to read data or relate data elements to each other (see Gal, in press, for details). Only when both sources of information are available can the reasonableness of opinions about data be judged. We suggest that assessors, and those creating and using scoring rubrics (Lamon & Lesh, 1992) assign higher scores to responses that:

- Are reasonable in light of the given data
- Refer to relevant and sufficient evidence
- Make correct use of technical terms
- Appropriately reference statistical indices derived by the student or mathematical relationships in the data
- Consider (where relevant) issues of variation and reliability of the data
- Make sensible or defensible assumptions about the context in which the problem is embedded and in which data originated.

We argue that follow-up questioning must become an inherent part of the critical process of negotiation of meaning of students' responses. As Ericsson and Simon (1980) have argued, however, questions about the content of a prior internal (mental) process elicit answers that are *reconstructive* and that may be subjected to various types of intentional and unintentional distortions. Responses may further be affected by students' assumptions (e.g., about the degree of variability that can be expected in certain variables) and by their linguistic skills (Secada, 1992). Thus, the interpretation of students' answers to process-tracing questions should be made with caution.

Of concern is the degree to which valid assessment of students interpretations of data can be made based on written responses, especially to multiple-choice exams, where it is not possible to obtain explanations. A partial solution will be found if teachers foster a "culture of explaining" in the classroom. The need to be reflective and be able to explain one's reasoning should become a standard demand characteristic of the learning situation. Further, students should be coached and given feedback regarding the adequacy and clarity of the opinions they present in both oral and written forms. Such steps are necessary to ensure that teachers can gain access to valid and complete information about students' statistical reasoning and interpretive skills.

References

Curcio, F. R. (1989). *Developing graph comprehension: Elementary and middle school activities.* Reston, VA: NCTM.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.

Gal, I., Rothschild, K., & Wagner, D. A. (1990). Which group is better? The development of statistical reasoning in school children. *Teaching reasoning and problem solving,* 8, 3-8.

Gal, I. (in press). Assessing statistical knowledge as it relates to students' interpretation of data. In S. Lajoie (Ed.). *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12.* Cresskill, NJ: Lawrence Erlbaum.

Laborde, C. (1990). Language and mathematics. In P. Nesher & J. Kilpatrick (Eds.). *Mathematics and cognition,* (pp. 53-69). New York: Cambridge University Press.

Lamon, S. J. & Lesh, R. (1992). Interpreting responses to problems with several levels and types of correct answers. In R. Lesh, R. & S. J. Lamon (Eds.). *Assessment of authentic performance in school mathematics.* Washington, D.C.: American Association for the Advancement of Science.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.) *On the Shoulders of giants: New approaches to Numeracy.* Mathematical Sciences Education Board. Washington, DC: National academy Press.

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

National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics.* Reston, VA: NCTM.

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws (Ed.). *Handbook of research on mathematics teaching and learning.* NY: Macmillan, pp. 623-660.

Wainer, H. (1992). Understanding Graphs and Tables. *Educational Researcher,* 21, 14-23.

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language, 25,* 571-587.