A. POLLATSEK, S. LIMA AND A. D. WELL

CONCEPT OR COMPUTATION: STUDENTS' UNDERSTANDING OF THE MEAN*

ABSTRACT. In statistics, and in everyday life as well, the arithmetic mean is a frequently used average. The present study reports data from interviews in which students attempted to solve problems involving the appropriate weighting and combining of means into an overall mean. While mathematically unsophisticated college students can easily compute the mean of a group of numbers, our results indicate that a surprisingly large proportion of them do not understand the concept of the weighted mean. When asked to calculate the overall mean, most subjects answered with the simple, or unweighted, mean of the two means given in the problem, even though these two means were from different-sized groups of scores. For many subjects, computing the simple mean was not merely the easiest or most obvious way to initially attack the problem; it was the *only* method they had available. Most did not seem to consider why the simple mean might or might not be the correct response, nor did they have any feeling for what their results represented. For many students, dealing with the mean is a computational rather than a conceptual act. Knowledge of the mean seems to begin and end with an impoverished computational formula. The pedagogical message is clear: Learning a computational formula is a poor substitute for gaining an understanding of the basic underlying concept.

1. INTRODUCTION

The mean is not only one of the most basic concepts in statistics and experimental science, but it also occurs frequently in everyday life. Most data reported in scientific journals are means, and inferential statistics deals almost exclusively with means and differences between means. Moreover, theories in many disciplines include concepts expressed in terms of means or sums. One often has a set of quantities which one wishes to represent by a single number. The sum and the mean both provide a useful unambiguous index and the mean has an additional advantage since it can be interpreted as a 'typical' score.

In many contexts the means or sums are formed only after weighting the quantities that enter into the sum. The weights may be number of observations or may be more abstract weightings of importance. In mathematics and the physical sciences, the integral is a weighted sum while the center of gravity and center of mass are weighted means. In decision making, methodologies such as cost–benefit analysis and linear programming use weighted means or sums to evaluate alternative courses of action. The standard model of the neuron assumes that the cell body is computing a weighted sum of inputs. The nonscientist routinely encounters weighted means in such guises as cost-of-living and stock market indices and estimated proportions of people approving governmental policies.

Educational Studies in Mathematics 12 (1981) 191–204. 0013–1954/81/0122–0191 $01.
Copyright © 1981 *by D. Reidel Publishing Co., Dordrecht, Holland, and Boston, U.S.A.*

Given all of this, one would expect that the typical American college student must certainly understand the concept of the mean. In fact, the mean should be particularly salient, even for students not engaged in the study of mathematics or the sciences, since grades in academic courses are usually determined by averaging the results of several examinations, and at most American colleges and universities the student's overall performance is summarized by a weighted mean called the grade point average (GPA). The calculation of the GPA is straightforward: Courses are assigned varying amounts of credit depending on how much time they require. The quality of performance in each course is typically assigned an integer value from 0 to 4, with 4 indicating excellence and 0 indicating failure. The GPA is the weighted mean of these 'grade points', the weightings being determined by course credit. At the end of each term (semester), the student is informed of his course grades and GPA for that term as well as his cumulative GPA for all of his work at the college.

It might seem that the concept of the mean is so simple, basic, and ubiquitous that any difficulties students have with problems involving means must be due to a lack of attention or motivation. We do not feel that this is the case since our experience with nonmathematically oriented students suggests that they often possess no more than minimal instrumental understanding (Skemp, 1979) of even the most elementary quantitative concepts. Instrumental understanding consists of recognizing a task as one for which one knows a particular rule. We believe that virtually all students know the rule or computational algorithm by which the mean is calculated, namely, to find the mean of a set of numbers one adds them up and divides by how many of them there are. If, however, students have *only* instrumental knowledge of the mean, they should make predictable kinds of errors in all but the most transparent of problems. Difficulties may arise since the computational algorithm does not explicitly state which numbers are to be summed in any given problem. Confusion over which numbers are to be summed naturally leads to confusion about how many numbers there are. The computational rule *by itself* does not tell students when it is appropriate to compute a mean, nor does it give any indication that an answer is reasonable once it has been computed.

We might expect confusions to surface whenever the problems confronted are such that it is not completely obvious which numbers should go into the sum. Consider the following example, which we will call GPA No. 1:

> A student attended college A for two semesters and earned a 3.2. GPA (grade-point average). The same student attended college B for three semesters and earned a 3.8 GPA. What is the student's GPA for all his college work?

Assuming an equal course load each semester, the correct answer may be computed by obtaining the sum of the GPA's for each of the five semesters and dividing by the number of semesters:

$$\text{mean} = \frac{(2 \times 3.2) + (3 \times 3.8)}{5} = 3.56.$$

There are, however, a number of ways to go wrong by unthinkingly applying a computational rule. If one believes that the mean GPA is merely the sum of the two GPA scores given in the problem divided by the number of scores, one could arrive at the unweighted (or simple) mean:

$$\text{mean} = \frac{3.2 + 3.8}{2} = 3.5.$$

If one knows that the sum should be divided by the number of semesters but is not sure how to obtain the appropriate sum, a computation such as

$$\text{mean} = \frac{3.2 + 3.8}{5} = 1.4$$

might result.

As we shall see, many American college students approach weighted mean problems as though all they knew about the concept of the mean was a computational rule. For this reason, we believe that the mean is a good starting point in an exploration of the cognitive structures that students have and need in order to work with statistical concepts. Recent psychological studies on reasoning in inferential statistics, particularly the work of Tversky and Kahneman (e.g., 1974) suggest that adults use a number of heuristics that seem reasonable but are often invalid for solving statistical inference problems. We have begun to explore the domain of the more basic concepts in descriptive statistics to discover what heuristic structures people use when they think about summarizing and combining data.

We believe, as do an increasing number of investigators, that research on problem solving behavior should be studied largely through the technique of clinical interviewing rather than relying exclusively on paper-and-pencil tests (for a review of a considerable body of research supporting this position, see Krutetskii, 1976). We feel that subjects have to be probed in depth if one is to discover much of interest about their statistical knowledge and abilities. Any problems of generalizability of the data obtained in interviews and of interviewer bias can be overcome by collecting appropriate objective data.

2. METHOD

In a preliminary phase of our research we gave the grade point average problem (GPA No. 1) to 37 undergraduates. The problem was included as part of a written diagnostic test at the beginning of a course in introductory statistics for psychology undergraduates at the University of Massachusetts. Most of the students answered the problem incorrectly; only fourteen were able to compute the correct overall GPA, 3.56. The most common incorrect answer was 3.5, the unweighted average of the GPA's for college A (3.2) and college B (3.8). Two subjects gave 1.4 as the student's overall GPA. These subjects apparently added 3.2, and 3.8 and then divided the sum by five, and were not deterred by the fact that 1.4 is smaller than either 3.2 or 3.8.

Although it is obvious that most subjects did not know how to appropriately weight and combine two means into a single overall mean, these test results give us little knowledge of the actual reasoning students used in attempting to solve the GPA problem. In order to explore students' concepts of the mean in more detail, we decided to include weighted-mean problems in the clinical interviews we were conducting to study college students' statistical intuitions. In our interviews, students were asked to solve problems in a variety of statistical subject areas, including probability and sampling as well as the mean. Although we obtained data on a number of topics, this paper will focus on four specific problems which shed some light on students' knowledge of the mean. Three of these problems were intentionally designed as problems involving the weighting and combining of means, while the fourth, although intended as a problem in estimation, was treated erroneously by some subjects as an exercise in computing means.

Subjects. Subjects were 17 undergraduate volunteers, 6 men and 11 women. Most subjects were psychology majors at the University of Massachusetts and were between the ages of 18 and 22 years. Three of the subjects had previously taken the diagnostic test, and had completed approximately half a semester of statistics by the time of the interview.

Procedure. We conducted an individual tape-recorded interview with each of our subjects. Subjects were asked to 'think out loud' as they worked on a problem. A typical interview lasted from 45 to 50 minutes and included from five to 10 different problems, the actual number of problems depending on how much time a particular subject devoted to explaining his or her answers. Each subject worked on problems dealing with a number of different statistical concepts, and at least one of the problems on which a subject worked involved weighted means. At the beginning of each problem, the interviewer remained non-directive, allowing subjects to spontaneously arrive at and explain their responses. Subjects were free to take their time and to change their answers at will. Later, the interviewer probed subjects in an effort to determine the reasoning underlying their answers and also to assess the confidence subjects had in their reasoning.

Problems. Three weighted mean problems were included in the interviews; each of our subjects worked on either one or two of these, and no subject worked on all three. Fifteen subjects worked on a grade point average problem. The first six subjects were interviewed on GPA No. 1, the same problem given on the preliminary test. The second nine subjects were interviewed on GPA No. 2, a slightly modified problem:

> A student attended college A for two semesters and earned a 3.22 GPA. The same student attended college B for four semesters and earned a 3.78 GPA. What is the student's GPA for all his college work?

The original GPA problem was modified in this way because we thought this second version would encourage students to work out their answers to two decimal places, and would also make more obvious the disparity between the number of semesters spent at each college.

When it became evident that students were having difficulty with GPA No. 2 as well as GPA No. 1, the following problem was devised and given to four subjects:

> There are ten people in an elevator, four women and six men. The average weight of the women is 120 pounds, and the average weight of the men is 180 pounds. What is the average of the weights of the ten people in the elevator?

The elevator problem was used because we felt that students might be reponsive to the concreteness of the entities being averaged, and pounds seem more concrete than the grade points in the GPA problems. Two of the subjects who answered the elevator problem also worked on GPA No. 1.

In summary, 15 of the 17 subjects were interviewed on only one of the three weighted mean problems. Of these 15 subjects, four worked on GPA No. 1, nine worked on GPA No. 2, and two worked on the elevator problem. The two remaining subjects worked on both GPA No. 1 and the elevator problem, and both of these subjects were presented with the elevator problem before they were given GPA No. 1. In all, then, 15 subjects worked on a GPA problem and four subjects worked on the elevator problem.

The fourth problem we will be concerned with was not a weighted means question, but subjects' reactions to it unexpectedly gave us some insight into students' ideas about the mean:

> You know that the average verbal SAT[1] score of the population of high school seniors in a large school system is 400. You pick a random sample of 5 seniors. The first 4 students in your sample have the following SAT scores: 380, 420, 600, 400. What do you expect the fifth student's score to be?

The correct answer to this question is 400, the population mean. Nine subjects, however, erroneously thought that they would get the best estimate of the fifth student's score by computing the number that would make all five scores average out to 400, the population mean, and the behavior of these 'averagers' will be discussed later in the paper.

3. RESULTS

The interview results confirmed our preliminary finding that many students are unable to correctly weight and combine two means into a single mean. Very few of the interviewed subjects were able to arrive at the correct answer to a weighted mean problem spontaneously, i.e., before the interviewer began to probe with follow-up questions. Of the 15 subjects who worked on a GPA problem, only two (13 percent) computed the correct answer (3.56 for GPA No. 1 and 3.59 for GPA No. 2) on their own. Thirteen of the 15 GPA subjects answered with the unweighted mean, 3.5, at some point in their interviews. The results for GPA No. 2 did not differ appreciably from those for GPA No. 1. Subjects tended to take the unweighted mean of the two GPA's even when the hypothetical student spent twice as much time at college B than at college A. The elevator problem appeared somewhat easier, since two of the four subjects who worked on it arrived at the correct answer (156 pounds) on their own; the other two subjects responded with the unweighted mean of 120 and 180 (150 pounds). However, since only four subjects worked on the elevator problem, we cannot draw firm conclusions about its difficulty relative to the GPA problem.

In a typical interview on a GPA problem, a subject would add the hypothetical student's GPA for college A to that for college B and then divide the result by 2. After the subject finished explaining how he or she had arrived at this answer, the interviewer usually probed with a follow-up question like "Now, suppose the student had spent one semester at college A and seven semesters at college B. Then what would his GPA be for all his college work?" The reactions of subjects to this probe indicated that the difficulties they had were more serious than a misreading of the problem or a hasty computation with a familiar formula.

In the first place, only three of the 12 subjects who were presented with the extreme example lost confidence in their earlier answer (3.5) and obtained the correct answer after being given the case of seven semesters vs. one semester. Seven subjects who answered 3.5 to GPA No. 1 or GPA No. 2 did *not* change their answers even after being presented with the extreme example in the follow-up question. The remaining two subjects who obtained incorrect solutions for the elevator problem did not change their incorrect responses even after they were faced with the example of eight men and two women on the elevator; they still felt that the average of the weights of the people on the elevator was 150 pounds, regardless of the different numbers of men and women. For these students, the unweighted averaging method was not just the easiest thing to do or the first method they thought of, it was the only method they had for dealing with means. The following excerpt from one interview exemplifies the response of a subject who persisted in using the unweighted method:

Interviewer: OK, suppose I told you that this student spent one semester at the first school and seven semesters at the second school.
Subject: One semester at the first school and . . .
I: Seven at the second.
S: What conclusion would I draw?
I: What would you say was his average for all his college work?
S: Well, it wouldn't change.
I: It wouldn't change from what?
S: It would still be 3.5.

This subject insisted that the hypothetical student's overall GPA for eight semesters was 3.5 even in the case where the student earned a 3.22 GPA for one semester and a 3.78 GPA for the remaining seven.

Four subjects who answered 3.5 to the GPA No. 1 problem stated that finding the unweighted mean was the best they could do given the information presented in the problem. These subjects said that in order to find the student's 'real' and 'exact' overall mean, they needed to be told specific GPA's for each of the semesters spent at college A and college B. They failed to realize that the critical information they needed, the sum of the individual semester GPA's, could be obtained from the given GPA's for college A and college B and the number of semesters spent at each college. Since they were not provided with the numbers they thought they needed, the best they could do was to use their computational rule with the numbers they *were* given. The end result of this process was the unweighted mean.

The interviews show that students' confusions about GPA's go beyond the specific questions we asked them. Two subjects who were working on a GPA problem spontaneously expressed a misconception about the computation of their own cumulative grade point averages, namely that one's cumulative GPA is calculated as a succession of unweighted means. They thought, for example, that if a student has completed his fifth semester and earned a 3.6 GPA, and his cumulative GPA for the previous four semesters was 3.0, then this new GPA would be 3.3, the simple average of 3.0 and 3.6. The student's actual cumulative GPA would, of course, be about 3.1, not 3.3; the erroneous method gives equal weight to the GPA representing four semesters and the GPA representing one semester.

We also found that a subject's correct answer to a weighted mean problem in one situation does not insure that he or she will apply the correct method in another situation where it would be appropriate to do so. One student gave the correct answer for the elevator problem, appropriately weighting the means given in the problem, but failed to apply the method to the GPA problem, which she worked on later in the same interview. Instead, she gave the unweighted mean (3.5) as her answer. This subject, then, had some knowledge about the weighting and combining of means, but did not perceive the GPA problem as a situation where that knowledge should be applied. When asked why she had not applied the method she used in the elevator problem to the GPA problem, the subject said it just did not occur to her. Perhaps the nature of a 'score' is clearer in the elevator problem because pounds are more concrete than grade points.

Subjects' attempted solutions to the SAT question indicated that difficulties with the mean are not confined to the inability to appropriately weight and combine two means. The SAT question was included in our interviews to study subjects' concepts of estimating unknown scores in a sample, but it unexpectedly exposed a basic difficulty in working with means. Nine subjects treated the SAT question as an exercise in averaging. These subjects erroneously thought that they should estimate the unknown fifth score by finding the number that would make all five scores average out to 400, the population mean. Four of the nine 'averagers' used an incorrect method to find the fifth score that would be necessary to make all five scores have an average of 400. These subjects started by summing the four known scores and then dividing the sum by four to get the mean of those four scores. They then tried to find the score that, when averaged in with the four-score mean they had just obtained, would yield a mean of 400. Since they computed a mean of 450 for the first four scores, the subject came up with an estimate of 350 for the fifth student's score. Several subjects so firmly believed that 350 was the correct fifth score that they were extremely puzzled when they added 380, 420, 600, 400 and 350 and divided the sum by five, only to find that the scores averaged to 430, not 400. These subjects failed to see that the mean of four scores must be weighted more heavily than a single score in computing an overall mean.

4. DISCUSSION: THREE KINDS OF KNOWLEDGE

Skemp (1979) has drawn the distinction between instrumental and relational understanding of a concept. Instrumental understanding of a quantitative concept would consist of having available only a collection of isolated rules (presumably learned by rote) for arriving at the answers to a limited class of problems. Relational understanding, in contrast, consists of having available an appropriate schema or set of conceptual structures sufficient to solve a much broader class of problems.

For the mean, the lowest level of instrumental understanding might consist of knowing only the computational rule for the calculation of the simple mean of a set of numbers. We believe that there are several additional kinds of knowledge that should be represented in an adequate schema of the mean. Three kinds of knowlege that can to some extent be distinguished from one another might be called (1) functional, (2) computational, and (3) analog knowledge.

By *functional knowledge* we refer to the understanding of the mean as a meaningful real-world concept. While part of the understanding of the mean may be in terms of computations that could be performed even on dimensionless, abstract numbers, additional knowledge is often necessary when the mean has a real-world referent that constrains the choice of scores that can be entered into a computational formula. In a problem such as GPA No. 1, the student must understand what the 'overall' mean is supposed to represent. In particular, if a mean is intended to be the quantity that best represents a set of scores and if the available computational rule states that the mean is the sum of the scores divided by their number, then each member of the set of scores has 'equivalent logical status' within the context of the problem. On the other hand, if the mean is thought of solely in terms of the result of the application of a computational rule, the simple mean of 3.2 and 3.8 is as reasonable a solution as any to GPA No. 1. Once the student realizes, however, that what is being asked for is an index of overall performance, then 3.2 and 3.8 cannot be regarded as logically equivalent elements with respect to the computational rule for the simple mean, since the two numbers represent performance for different amounts of time.

From our interview data, it seems clear that many of the students did not have this functional knowledge since they seemed quite content to give the simple mean as their answer. As noted previously, one student, when asked if counting seven semesters and one semester equally in a GPA problem would be fair, replied, "No, but that's the way they do it." This student appeared to know that the simple mean was not adequate to summarize overall performance in the problem but had so little functional knowledge of the mean that he did not realize the concept could be used to provide a meaningful index. Our data suggest that different students have varying degrees of functional knowledge. We base this on interviews with subjects who could solve one weighted mean problem but not another. There is also the suggestion (which must be confirmed by future work) that problems involving more concrete quantities such as pounds are, in general, easier to deal with than problems dealing with more abstract quantities such as grade points. Since the more concrete contexts make it easier to understand what the total sum of elements represents, they may also make it easier for the student to determine what the weighted mean is supposed to represent.

Adequate *computational knowledge* would have to involve either a computational formula for the weighted mean or the computational formula for calculating the unweighted mean combined with information about how to obtain the appropriate sum. It is particularly important in solving weighted mean problems to know that just as one can go from the sum of a set of scores to the mean by dividing by the number of scores, one can obtain the sum from the mean by multiplying the mean by the number of scores. In GPA No. 1, for example, such 'reversibility' (Krutetskii, 1976) would allow the student to determine that the sums of the semester GPA's at colleges A and B were 2×3.2 and 3×3.8, respectively, (if, of course, the student had the functional knowledge that the sum of the semester GPA's was required). Despite the fact that every student knew the computational formula for the simple mean and could certainly multiply and divide, some students knew that it was necessary to obtain the sum of the semester GPA's but seemed quite unable to do so. Several spontaneously commented that they could have easily solved the problem if only all of the necessary information, namely, the GPA score for each semester, had been provided. In addition, several students also had difficulty in finding the single missing score that makes five SAT scores average to 400. This would have been no problem if they realized that the five scores summed to 2000. Also, part of an adequate computational component might be what we could vaguely term 'a feeling for numbers'. During the course of the interviews, we presented nonsolvers with the expressions $(3.8 + 3.2)/2$ and $(3.2 \times 2) + (3.8 \times 3)/5$ and asked whether both expressions would provide the same answer. Most students had no immediate feeling that the answers would be different and had to perform both calculations to be sure.

Students with such limited arithmetic ability would be at a considerable disadvantage in determining whether any problem solution they obtained was a reasonable one or not.

Analog knowledge might involve visual or kinesthetic images of the mean as a 'middle' or balance point. A diagram that is commonly shown in elementary statistics books is that of the mean acting as the fulcrum balancing a set of weights on a see-saw, in which the distribution of weights is identified with the frequency distribution of data points. The mean might be represented analogically by a score value about which 'the moments of the weights' balance or somewhat more abstractly as a point around which the deviations of the data points (weighted by importance) must cancel out. Such a representation should be sufficient to prevent students from making gross errors in solving weighted mean problems, provided they have the functional knowledge which indicates which elements are to serve as the 'weights'. For GPA No. 1, for example, two weights would be located at 3.2. and three weights would be located at 3.8. While this kind of representation might not be sufficient in itself to result in the exact numerical solution to the problem, it would allow the student to realize that the answer should be closer to 3.8 than 3.2.

Although it might seem reasonable that students should vary in terms of the extent to which they depend on analog or computational knowledge (cf. Krutetskii's (1976) geometric and analytic 'mathematical casts of mind'), we found almost no evidence that any of our subjects used analog knowledge of the mean in dealing with a weighted mean problem. If they could not solve a weighted mean problem by doing a calculation, subjects rarely made a statement like "I see it should be bigger than 3.5 but I don't know how to work it out". Even when presented with an extreme example of a GPA problem (seven semesters with a 3.8 GPA), only one of thirteen nonsolvers demonstrated understanding that the weighted mean should be larger than the simple mean.

One way of summarizing the inability of many students to solve weighted mean problems is that they behave as though the mean were a purely formal concept, defined solely in terms of a calculation based on abstract numbers. Perhaps we should not be too surprised by this finding, since when we have taught nonmathematically oriented students in undergraduate statistics courses, we have noted a predisposition on their part to focus on the learning of the formulas and rules to solve specific types of problems. It is not uncommon for students to become quite accomplished in mechanically applying computational rules, yet to lack the functional knowledge needed to solve 'word' problems in which a translation has to be made from the situation described in the problem to the available computational structures. As Kaput and

Clement (1979) and Rosnick (in press) have dramatically pointed out, even engineering and physics students have considerable difficulty in translating back and forth between equations and verbal descriptions of what seem like very simple situations (e.g., At a certain college there are six times as many students as professors. If S is the number of students and P is the number of professors, what is the equation that relates S to P? Most subjects answered "$6S = P$"; the correct equation is, of course, "$6P = S$".)

Knowledge of a computational rule not only does not imply any real understanding of the basic underlying concept but may actually inhibit the acquisition of more adequate (relational) understanding. Unless provided with appropriate instruction, students may believe that instrumental understanding of a concept constitutes full understanding. We have been disappointed to see textbooks aimed at our undergraduates which all but ignore functional knowledge. Large amounts of space may be taken up with numerical examples that are basically exercises in computations. Unless examples and problems provide intensive practice in translating from a variety of contexts to available computational structures, it is unlikely that understanding with much generality can be achieved.

5. FUTURE DIRECTIONS

While it would be premature to draw sweeping conclusions from the fine points of our data, a general pattern emerges. Students often have difficulty with what seem like very simple problems. Even though we selected the weighted mean type of problem as one likely to be non-trivial for students, we did not expect our problems to be as difficult as they were. The source of the difficulty appears to be that students' knowledge often seems limited to computational formulas, and many simple problems (such as weighted mean problems) require more general, relational, knowledge of concepts. One pedagogical point seems clear. In many introductory courses, students are taught to use formulas in a rote manner with the justification that thorough understanding of the material can wait until the second course (or later). While it is undeniably true that students can solve some problems with this approach, our data suggest that the range of problems that can be solved with only instrumental knowledge is vanishingly small.

A second point that emerges is that the same type of problem may be approached in different ways if it is placed in different contexts. For example, the interview data suggest that the elevator problem was dealt with more successfully than the GPA problem. Thus, even though students have some relational knowledge that they can bring to bear on problems, care may be needed on the part of an instructor to find the right contexts to elicit the use of this relational knowledge. The relationship between means and sums may be more transparent in the elevator problem because the sum is more tangible (total number of pounds) than in the GPA problems (total number of grade points). We have begun to follow up this idea by giving GPA No. 2 to a 72-student undergraduate statistics class along with another problem which was designed to make the items summed and averaged even more concrete than the pounds in the elevator problem:

> Two boats of fishermen returned from a weekend fishing trip. The 8 people on the first boat averaged 5 fish per person. The 4 people on the second boat averaged 11 fish per person. What was the overall average number of fish caught per person?

In fact, 59 percent solved the fish problem while only 38 percent of a comparable group solved GPA No. 2.

Another point to be explored is how presentation format changes the mental processes of subjects. We have found from written tests that students estimate means from histrograms fairly well. This suggests that a pictorial representation of the weights may be more intuitive than a verbal description. Accordingly, we plan to see how students react to weighted mean problems when they are presented in various formats – figures, tables, or verbal descriptions – to try to uncover the most natural structure for conceptualizing means.

Department of Psychology
University of Massachusetts, Amherst

NOTES

* Research support: Research reported in this paper was supported by NSF research award No. SED78-22043 in the Joint National Institute of Education–National Science Foundation Program of Research on Cognitive Processes and the Structure of Knowledge in Science and Mathematics.

[1] The SAT is a standardized college entrance examination taken by the majority of college-bound students in the United States. It contains a verbal part and a mathematical part, and each part is scored on a scale from 200 to 800. All of our subjects had taken the SAT prior to university entrance.

REFERENCES

Kaput, J. and Clement, J., 'The roots of a common reversal error', *Journal of Mathematical Behavior* 2 (1979), 208.

Krutetskii, V. A., *The Psychology of Mathematical Abilities in School Children*, University of Chicago Press, Chicago, 1976.

Rosnick, P., 'Learning without understanding: The effect of tutoring strategies on algebra misconceptions', *Journal of Mathematical Behavior* (in press).

Skemp, R. R., 1979, *Intelligence, Learning and Action*, Wiley, Chichester.

Tversky, A. and Kahneman, D., 1974, 'Judgment under uncertainty: Heuristics and biases', *Science* **185**, 1124–1131.