Journal of Educational Psychology
1984, Vol. 76, No. 5, 792–801

SRRI #102

Usefulness of a Balance Model in Understanding the Mean

Pamela Thibodeau Hardiman, Arnold D. Well, and Alexander Pollatsek
University of Massachusetts—Amherst

The study tested whether improving students' knowledge of balance rules through experience with a balance beam promotes understanding of the mean. The study consisted of three sessions. In the pretest session, subjects' levels of balance knowledge and abilities to calculate the solutions to a variety of problems dealing with the mean were assessed. Subjects were classified as *nonbalancers* if they performed at a level below Siegler's (1976) Rule IV and as *noncalculators* if they were unable to calculate solutions to weighted mean problems. In the second session, half the nonbalancers were given balance training and the other half were asked to solve unrelated control problems. In the transfer session, subjects were given a set of five problems to assess their understanding of the mean. Significant transfer was found: Subjects classified as nonbalancers on the pretest performed significantly better on the transfer problems if they had been given balance training rather than assigned to the control condition.

A commonly held assumption in educational practice is that experience in working with a problem will lead to improved performance on similarly structured problems presented at a later time. However, evidence of nontrivial positive transfer in problem-solving situations has been difficult to find in the laboratory (e.g., Gick & Holyoak, 1980; Reed, Ernst, & Banerji, 1974). Many studies report little evidence of transfer even across quite similar problems unless subjects are explicitly told that the problems are analogous. Because we know that people do benefit from experience with problems in at least some educational settings, it is possible that the common practice in education of grouping similar problems together may implicitly tell students to treat problems as analogous, so that explicit instruction is unnecessary.

We take as a starting place the idea that transfer is more likely to occur if the subject is able to integrate a problem-solving procedure with his or her general knowledge (Greeno, 1978). This is supported by recent short-term teaching studies (Bromage & Mayer, 1981; Myers, Hansen, Robson, & McCann, 1983), which suggest that the use of teaching materials that draw explicit connections to general knowledge leads to improved problem-solving performance in cases in which the correct application of a procedure is not obvious. This integration is more likely to occur if a number of different kinds of knowledge can be brought to bear on the problem. The content area we are interested in is statistics and, in this article, problems involving the weighted mean. Our primary interest is whether training subjects on the rules of balancing, a superficially unrelated subject area, improves performance on weighted mean problems. To make this connection clear, we describe the weighted mean concept in some detail.

Difficulties With Weighted Mean Problems

The mean is a commonly employed descriptor of sets of numbers and forms the basis for related statistics such as the variance and standard deviation. Evidence suggests that despite its pervasive nature, many college undergraduates have a limited

This article is based on a thesis submitted to the University of Massachusetts by the first author in partial fulfillment of the requirements for a master of science degree. The research was supported by National Science Foundation Grants SED-8113323 and SED-8016567.

The authors would like to thank Gale Sinatra for her assistance in conducting the study.

Requests for reprints should be sent to Pamela Thibodeau Hardiman, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

understanding of this concept. Although they are almost invariably able to find the arithmetic mean of a given set of numbers, they are frequently unable to solve relatively simple problems in which it is not obvious how to apply the computational algorithm for the arithmetic mean (i.e., add all the scores together and divide by the number of scores). In particular, they are frequently unable to calculate the overall mean when presented only subgroup means based on samples of different sizes. Pollatsek, Lima, and Well (1981) found that only 38% of the students beginning introductory statistics classes for psychology majors were able to solve the following weighted mean problem: "A student attended College A for two semesters and earned a 3.2 GPA. The same student attended College B for three semesters and earned a 3.8 GPA. What was the student's overall GPA?" (p. 192).

Assuming that an equal number of credits were taken each semester, the correct answer can be obtained by weighting the grade point average (GPA) obtained at each college by the number of semesters attended (i.e., multiplying 3.2 by 2 and 3.8 by 3, adding the results, and dividing by 5). However, the most common answer given by subjects was 3.5, the unweighted or simple mean of 3.2 and 3.8. The information concerning the number of semesters spent at each college was frequently neglected, presumably because subjects did not possess a computational formula for the weighted mean or were not able to apply or modify appropriately the computational algorithm they did have available.

Pollatsek et al. (1981) suggested that complete understanding of the mean has three components: (a) functional knowledge, (b) computational knowledge, and (c) analog knowledge. Functional knowledge consists of understanding the mean as a real-world concept, a number that best represents the set of scores being considered. It includes the knowledge that if the numbers are to be weighted equally, they should be logically equivalent. Computational knowledge involves either having available a computational formula for the weighted mean or knowing how to adapt the computational formula for the simple mean to weighted mean problems, for example, knowing that the sum of the scores in a subgroup can be obtained by multiplying the mean of the subgroup by the number of scores in it, demonstrating what Krutetskii (1976) has referred to as reversibility. Analog knowledge might involve visual or kinesthetic images of the mean as a middle or balance point. Although the analog representation does not yield a numerical answer, it should be sufficient to prevent students from making gross errors in solving weighted mean problems. In particular, in a problem like the GPA problem given earlier, subjects should realize that the overall mean will be closer to the subgroup mean based on the larger number of scores.

Pollatsek et al. (1981) interviewed subjects solving weighted mean problems and found that many undergraduates did not appear to possess adequate functional or computational knowledge, and none showed any evidence indicating they used analog knowledge. Few subjects responding with the simple mean expressed doubts about their answer, even when later asked whether the number of scores in each group made a difference. Furthermore, a follow-up study by Sinatra (1980) suggested that having subjects do a relatively easy weighted mean problem before the grade point problem was of no help in solving it. In her study, one group of subjects attempted to solve an easy weighted mean problem (about 75% solved it correctly) before being asked to solve the GPA problem, whereas the other group got the GPA problem first. An equal number of subjects solved the GPA problem in each group, even though Sinatra showed the subjects who did not solve the easier problem the correct solution to it.

Sinatra's result appears to fit in with those of Reed et al. (1974) and Gick and Holyoak (1980) cited earlier: Subjects often do not achieve transfer between apparently similar isomorphic problems. In these studies awareness of the analogy appeared to be an important condition for transfer, because transfer was achieved only when the subject was instructed to use the analogy. However, it may be equally important that the subject understand the first problem sufficiently so that some "deep structure" can be transferred.

Clement (1981) investigated spontaneous

analogy generation that occurred in solving a variety of challenging problems. His data strongly suggest that the content area from which the spontaneous analogy is drawn must be well understood if use of the analogy is to prove beneficial. The same conclusion can also be drawn by a comparison of two experiments. Luger and Bauer (1978) observed transfer effects in both directions for a set of two problems, when the task for the initially presented problem was to develop an optimal solution. In contrast, Reed et al. (1974) observed no transfer between problems. However, their subjects were asked to solve a fairly difficult problem once and were unlikely to have developed optimal solution strategies and, consequently, had a less complete understanding of the initial problem.

The Balance Beam Analogy

The balance beam is generally considered to be a useful analogy for understanding the mean. It is presented in many introductory statistics books when the mean is introduced (e.g., Freedman, Pisani, & Purves, 1978; Hays, 1981). As previously indicated, this conception should be useful because it allows connections to be made to general knowledge of and experience with balancing, leads to reasonably accurate approximations to the mean, and helps make clear that it is the relative frequencies of scores (i.e., the ratios) that are important in determining the mean.

However, a pilot study suggested that as in the case of the above transfer studies, mere presentation of the analogous situation is not helpful. Hardiman (1981) investigated the usefulness of the balance beam model as an aid to understanding the concept of the mean in a study in which subjects were provided with a balance beam and asked to represent weighted mean problems using a set of blocks as weights. Contrary to expectations that the use of the concrete model would improve subjects' intuitions about the mean and lead to an improvement in performance, many subjects found the task of representing the problem on the beam to be quite difficult even if they were able to calculate the correct answer. The comments of several subjects suggested that they did not initially understand fully how the balance beam worked. In fact, working with the balance beam led several subjects who had initially calculated the overall mean correctly to switch to an incorrect simple mean calculation. The subjects tended to treat a block as if it were merely a placeholder for the value beneath it, precluding the possibility of weighting one value more than another.

Hardiman's results thus suggest that presentation of the balance beam analogy may be ineffective largely because many students do not have a good understanding of the balance beam. Other data support the assertion that many college students do not fully understand balance beams. Siegler (1976) proposed a developmental model of four progressively more complex rules that describe individual performance on different types of balance problems. Subjects using Rule I consider only the number of weights on each side of the beam and always predict that the beam will tip to the side with more weight. If there are equal numbers of weights on both sides of the fulcrum, they predict that the beam will balance. Rule II subjects consider distance when the weights are equal and otherwise base their decisions only on the weights. Rule III subjects check to see whether the side with the greater weight also has the greater distance and, if not, are equally likely to say the beam will balance, tip to the left, or tip to the right. Finally, Rule IV subjects combine weight and distance information appropriately (i.e., base their decisions on the product of weight and distance). Siegler (1976) found that only 17% of 16–17-year-old subjects were described as being at the level of Rule IV. Although the college students used in the present study were a bit older, it is unlikely that a radical change in understanding takes place in 2 or 3 years.

Conclusions and Implications of Previous Research

The following conclusions can be drawn: (a) Although students almost invariably can calculate the mean of a set of numbers, many of them do not have the well-developed functional, computational, and analog knowledge that constitutes a relational un-

Table 1
Pretest Problems

Order[a]	Type	Problem
1	Mean judgment	Could the final number presented possibly be a mean of the set?[b] a) 42 47 39 85 54 32 41 57 38 b) 2.4 4.5 2.7 3.6 3.4 2.3 3.1 1.8 1.7 c) 10 19 10 13 12 10 18 14 13 d) 157 99 101 101 97 153 115 103 130
4	Simple mean	What is the average of the following numbers? 10.2 14.3 9.7 11.0 12.6
2	Weighted mean	Two boats of fishermen return from a weekend fishing trip. The four people on the first boat averaged 5 fish per person, while the two people on the second boat averaged 11 fish per person. What was the overall average number of fish caught?
6	Weighted mean	There is a measure called income index. It ranges from low income to high income on a scale of 1 to 6. In one small town there are 200 families and the average income index is 2.8. In a second small town there are 400 families and the average income index is 3.6. What is the overall average income index for all the families in both towns?

Note. Two filler problems and 12 balance problems involving line drawings are not shown in this table.
[a] The numbers represent the order in which the problems were presented. Problems 2 and 6 were placed in positions 6 and 2, respectively, for half the subjects.
[b] Presented orally to prevent subjects from calculating the means.

derstanding (Skemp, 1979) of the mean; (b) even if students have available a balance model of the mean, the benefit that can be derived in solving problems having to do with the mean will depend on the extent to which they understand the rules of balancing; and (c) there is reason to believe that many college undergraduates do not have a very good understanding of the rules of balancing. Thus, it seemed to us that knowledge of balancing, though not logically necessary for solving mean problems, would possibly be very useful, and that people who understood balancing better should understand the mean better. This assertion is testable in two ways. The weaker test is correlational: Will students with higher levels of balance knowledge be more successful in solving problems related to or involving the mean? The second test is experimental: Will providing experiences that foster the development of balance knowledge improve performance on weighted mean problems, especially for those subjects who have initial difficulties?

The following design was employed: (a) In a pretest session, subjects were first given a series of problems having to do with the mean, including two weighted mean problems, and then a series of problems that assessed knowledge of balancing. The data from the pretest were used to determine whether balance knowledge does, in fact, correlate with success on problems involving the mean. (b) In the training and control sessions, subjects who performed at a level below Siegler's (1976) Rule IV in the pretest balance task were randomly assigned to either a balance training condition or a control condition in which they attempted to solve unrelated problems for an equal period of time. (c) In the transfer session, all subjects were presented with a set of five posttest problems that tested their understanding of the weighted mean. The key question was whether subjects trained in balance knowledge would perform better on the posttest problems than control subjects who had scored equally well on the pretest.

Pretest

Method

Subjects. Participating in exchange for bonus credit were 48 students (30 women and 18 men) enrolled in psychology classes at the University of Massachusetts. Ages ranged from 17 to 36 years, with a mean of 20.5 years. An additional 17 subjects took the pretest, but failed to return for the second (training or control) session.

Problems. The pretest consisted of 2 weighted mean problems similar to the GPA problem cited previously (see Table 1), 1 problem that involved a simple mean (see Table 1), 4 problems that required subjects to judge (without calculating) whether a given number could plausibly be the mean of a set of eight numbers read to them (see Table 1), 2 filler problems, and 12 line drawings that presented 12 balance problems of varying difficulty. For each balance problem, subjects were required to predict whether the balance beam illustrated should balance, tip to the left, or tip to the right.

Procedure. The pretest was administered to groups of 3 and 4 subjects. The mean judgment task, in which the experimenter read each set of eight numbers and the possible mean aloud, was presented first, followed by 5 problems in a booklet, 1 problem to a page, with adequate space on each page for calculations. The last page of the booklet contained the 12 balance problems. Subjects were asked to write out all steps to their calculations as clearly as possible. No time limit was given and all subjects completed the task within the allotted hour. No feedback was provided.

Results and Discussion

The pretest was used to classify all subjects as *calculators* or *noncalculators* and as *balancers* or *nonbalancers*. Subjects were classified as calculators if they answered both weighted mean problems (Problems 2 and 6) correctly and as noncalculators otherwise. Twenty-four of the 48 subjects were classified as calculators and 24 as noncalculators. Subjects were classified as balancers if they correctly answered the simpler balance problems and at least six of the nine "conflict" problems (i.e., problems in which there was more weight on one side of the fulcrum, but the weights were farther away from the fulcrum on the other). Twelve subjects were classified as balancers, and 36 subjects were classified as nonbalancers. Of the nonbalancers, 14 answered in a pattern predicted by Rule III, 16 by Rule II, and 6 by Rule I (Siegler, 1976).

Overall there were 10 calculators and balancers, 2 noncalculators and balancers, 14 calculators and nonbalancers, and 22 noncalculators and nonbalancers. The correlation between balance rule level and number of correct weighted mean calculations was .50, $t(46) = 4.80$, $p < .001$. This correlation is consistent with the hypothesis that balance rule knowledge facilitates calculation in weighted mean problems, although it does not allow us to conclude that there is a causal relationship. A positive correlation would also be expected if some factor such as level of mathematics training or general ability influenced both balance knowledge and the ability to calculate. In fact, level of mathematics training, as rated on a 4-point scale (with 1 representing mathematics training through high school algebra and 4 representing training in college beyond calculus), correlated significantly with status as a calculator or noncalculator ($r = .38$), $t(46) = 2.82$, $p < .01$, although not with status as a balancer or nonbalancer ($r = .13$), $t(46) < 1$. Also, although balance knowledge may facilitate thinking about the mean, the pretest data make it clear that it is not necessary for solving the problems we employed: 14 of the 24 subjects who correctly solved both weighted mean problems were classified as having less than Rule IV knowledge.

The additional problems on the pretest helped to characterize the nature of the difficulties subjects had with the weighted mean problems. All 48 subjects demonstrated on Problem 4 that they knew the appropriate algorithm for calculating the simple mean of a set of numbers, but 5 of them made arithmetic errors. Thus, the errors that were observed on weighted mean problems did not result from a lack of procedural knowledge of mean calculations in general, but rather from a lack of understanding of the purpose of each step in the calculation that would allow it to be adapted appropriately to weighted mean problems.

Differences among groups were found on the problems that asked subjects to judge whether a given number could plausibly be the mean of a set of scores (Problem 1). Noncalculator–nonbalancer and calculator–nonbalancer subjects made many more errors than calculator–balancer subjects (21.5%, 17.9%, and 2.5%, respectively). Performance was poorest on a problem in which the numbers in the set were unevenly distributed (157, 99, 101, 101, 97, 153, 115, 103) and the candidate for the mean (130) was approximately halfway between the extreme scores. Forty-eight percent of the noncalculator–nonbalancer subjects and 26% of the calculator–nonbalancer subjects incorrectly judged that 130 could be the mean of the set of numbers. This suggests that the nonbalancers tended to have relatively poor intuitions about the mean or at least that they were not very sensitive to the characteristics of the distribution of a set of numbers.

Table 2
Transfer Task Problems

Order	Problem
1	It is possible to view the mean of a set of numbers as the point at which the number line containing the set would balance if it were placed on a balance beam. You have a balance beam here before you. Please represent the mean of 3 and 10 on the balance beam using the plastic scale as a number line and the blocks as weights.
2	A student attends College A for two semesters and earns a 3.2 GPA. The same student attends College B for four semesters and earns a 3.8 GPA. What was the student's overall GPA?
3	Several people get on a large elevator. Three-fifths of the people are men and average 180 pounds. The remaining people are women and average 120 pounds. What is the average weight of the people on the elevator?
4	Person A and Person B are engaged in a weight maintenance program. Person A weighs himself three times evenly spaced throughout the day and averages 185 pounds on a typical day. Person B weighs himself five times evenly throughout the day and averages 211 pounds. What is the average weight of the two people?
5	A local shop employs several people who make the following salaries: 1 – owner-president .... $30,000 2 – foremen .... $10,000 12 – general workers .... $8,000 The owner needed to calculate the average salary for the people in the shop. She thought of two ways to do it: 1) add the three numbers together, 30,000, 10,000, and 8,000, and divide by three, or 2) multiply each salary by the number of people paid that salary, add them together, and divide by fifteen. Which way would you calculate the average salary and why?

Training, Control, and Transfer Sessions

Method

Calculators–nonbalancers and noncalculators–nonbalancers were assigned randomly to training and control conditions. Assignment to conditions was made prior to the subjects' return for the second session, and because a number of subjects failed to return for the second session, the numbers of subjects in the training and control conditions were unequal. Six calculators–nonbalancers and 12 noncalculators–nonbalancers participated in the training condition; eight calculators–nonbalancers and 10 noncalculators–nonbalancers participated in the control condition, in which they were asked to "think aloud" as they worked on several probability estimation problems.

Materials. An ideal weightless balance beam was approximated by using a lightweight acetate scale marked at regular intervals. The acetate scale rested on top of a rigid aluminum bar balanced on a fixed fulcrum at its midpoint. In the training session the scale was initially centered on the bar, and the subject was asked whether various configurations of blocks placed on the scale by the interviewer would balance or cause the bar to tip to the left or right. In the transfer session the subject was provided with a pen and allowed to mark numbers on the acetate scale to create a number line. After blocks were placed on the number line, the scale and blocks could be slid along the top of the bar until balance was achieved.

Training phase. Subjects were told that they would be given a series of balancing problems and that their task was to learn how to predict the action of the beam. On each trial of the transfer task, the experimenter placed blocks on the balance beam at marked positions that were four or fewer units from the fulcrum. When the blocks were positioned, the balance beam was held rigidly and the subject asked to predict whether the beam would tilt to the left, tilt to the right, or balance. After each prediction, the beam was released so that the subject could tell whether the prediction was correct. The training sequence was modeled after Siegler's (1976) procedure, which begins with simple problems and becomes progressively more complex. The simplest problems consisted of two single groups or vertical stacks of blocks placed on either side of the fulcrum such that either the number of blocks in a group or the distance from the fulcrum was constant. These problems were followed by several sequence problems that began with a nonbalancing situation which had equal numbers of weights at varied distances on either side of the fulcrum. Blocks were added one at a time to the group placed the lesser distance from the fulcrum until the beam balanced and then tipped in the opposite direction. Thus, both the number of blocks and the distance could vary. The final and most complex problems allowed either one or two groups of blocks on each side, with up to six blocks per group. The criterion for successful completion of training was correct prediction on five consecutive difficult problems. The mean number of trials to reach criterion was 49.2 (SD = 12.9, range = 27 to 68).

Transfer phase. The problems used during the transfer session are displayed in Table 2. Problem 1 introduced the notion of the mean as a balance point. Problems 2 and 3 were weighted mean problems, although the latter involved proportions, rather than absolute numbers of scores. Problem 4 was a simple mean problem that had a surface structure very similar to that of the weighted mean problems and was included as an attempt to distinguish those subjects who had a good understanding of the concept of the mean from those who merely had developed an algorithm for dealing with weighted mean problems. Problem 5 confronted subjects directly with a choice between a simple mean and a weighted mean calculation in a situation in which the weighted mean was appropriate. After making a choice, subjects were asked whether the former would give a larger or smaller answer than the latter, in order to test qualitative reasoning about means.

Procedure

Subjects were seated at a table diagonally across from the interviewer, with the balance beam in front of them, and were presented with a pen and a pad of paper for any calculations they might care to do. They were told that the entire session would be videotaped and that they should think aloud as they went about solving the problems. Each problem was presented on a separate sheet of paper. Subjects were instructed to begin each problem by reading it aloud, and for each of the first four problems, they were asked to represent the problem on the balance beam. Most subjects calculated an answer before attempting to represent the problem on the beam. The few subjects who represented the problem without doing any calculations were asked to calculate an answer before moving on to the next problem. Subjects were permitted to reread each problem as many times as they wished and could take as much time as they needed.

The first problem asked subjects to represent the mean of two numbers on the beam. Subjects began by testing strategies that they thought might be appropriate, and if these did not yield a correct representation, the interviewer asked whether there might be other possible methods. If the subject was unable to represent the problem correctly within 10 min, he or she was shown that if blocks were placed at locations on the scale corresponding to the two numbers, the balance point corresponded to the mean of the numbers, and then was given another problem involving only two numbers.

Problems 2–5 were presented with little interviewer intervention. The interviewer intervened only to encourage subjects to think aloud more or to explain a particular answer more fully.

Analysis of Interviews

The interviews were analyzed from the videotaped recordings of the transfer sessions. The response to each problem was coded as to whether the calculation and representation given were correct. In addition, each solution was classified as to general type of calculation, for example, simple mean, weighted mean using actual numbers of scores, weighted mean using proportions, incorrect weighting (reversed proportions), or summed subgroup means divided by total number of cases, as well as to general type of representation. There were three coders: the first author, who was not naive about which group a subject was in, and two others, the second author and a research assistant, who were blind to how the subjects had been classified both in training and in the pretest. The reliability for the three pairs of raters was assessed on a sample of 10 subjects, using Cohen's kappa statistic (Hays, 1973). There was perfect agreement among the raters as to whether each answer was correct. The average values of kappa were .90 for classification of type of calculation and .83 for classification of type of representation. When there was disagreement, the classification used by the majority of coders was used.

Results and Discussion

Eighteen of the 48 subjects spontaneously represented the simple mean correctly on Problem 1. The remainder were able to do so following some form of interviewer intervention. There was no significant association between whether a subject represented the simple mean problem spontaneously and either pretest classification or assignment to training or control conditions.

A calculation score ranging from 0 to 2 was determined for each subject, in which 1 point was given for each correct calculation of the two weighted mean problems (Problems 2 and 3). These scores are displayed in Table 3. Using calculation score as the dependent variable, a 2 (pretest calculation ability: calculator–nonbalancer or noncalculator–nonbalancer) × 2 (training: balance training or control) unweighted means analysis of variance revealed significant main effects of calculation ability, $F(1, 32) = 18.82, p < .001$, and training, $F(1, 32) = 8.64, p < .01$. Although the training effect appeared to be larger for the noncalculator–nonbalancer group, the difference was not significant, $F(1, 32) = 2.36, p < .20$. However, because our major interest was the effects of training on subjects who could not calculate originally, we wanted to assess the reliability of the training effect for them separately. Trained noncalculators–nonbalancers had significantly higher calculation scores in the transfer session than the control subjects, $t(20) = 3.16, p < .01$. Seven of the 12 trained noncalculators–nonbalancers calculated the correct answers to both weighted mean problems, whereas none of the 10 control subjects were able to do so. Also, because these subjects were randomly assigned to the training and control conditions rather than being strictly equated for pretest calculation performance, a second analysis was done for the 18 noncalculators–nonbalancers who had correctly answered one of the pretest weighted mean problems.

Table 3
Mean Overall Calculation Scores for Nonbalancers (Calculators and Noncalculators)

	Calculators		Noncalculators	
Condition	*M*	*SD*	*M*	*SD*
Training	2.0	0	1.5	.67
Control	1.8	.46	0.7	.48

Again, the mean calculation score in the transfer session was significantly higher for the trained subjects (1.5) than for the control subjects (.67), $t(16) = 2.65, p < .02$. Redoing the analysis of variance including only the noncalculators–nonbalancers who had solved one of the pretest weighted mean problems and the calculator–nonbalancer subjects who had solved both pretest weighted mean problems indicated that the latter subjects had higher calculation scores in the transfer session than the former subjects, $F(1, 28) = 10.81, p < .01$, and that trained subjects had higher scores than control subjects, $F(1, 28) = 4.62, p < .05$.

A similar score was developed for representation performance. One point was given for correctly representing each of the weighted mean problems, so that scores for each subject ranged from 0 to 2. An unweighted means 2 × 2 analysis of variance revealed significant effects for calculation ability, $F(1, 32) = 8.58, p < .01$, and training, $F(1, 32) = 6.64, p < .025$. However, the interaction between training and calculation ability was not significant ($F < 1$). As can be seen from Table 4, representation scores were higher for trained subjects and for those classified as calculators on the pretest. The trained subjects were significantly better than the control subjects for the noncalculator–nonbalancer group $t(20) = 2.21, p < .05$, but the difference for the calculator–nonbalancer group was not significant, $t(20) = 1.59, p < .10$.

Table 4
Mean Overall Representation Scores for Nonbalancers (Calculators and Noncalculators)

	Calculators		Noncalculators	
Condition	*M*	*SD*	*M*	*SD*
Training	2.0	0	1.4	.79
Control	1.5	.76	0.7	.67

The pattern of results indicates that the balance training resulted in positive transfer. On the average, control subjects were at least as successful in calculating the correct answers to weighted mean problems in the transfer session as they had been in the pretest session, suggesting that the advantage of the trained over the control subjects in the transfer session was due to the positive effects of training rather than to control subjects' being interfered with by exposure to the balance beam. Because the weighted mean problems in the pretest and posttest and their manner of presentation were different and because 1 to 2 weeks elapsed between pretest and posttest, we cannot be absolutely certain from the above data that the training procedure produced positive transfer. However, other data from the transfer session indicate that the subjects did benefit from the balance training and, furthermore, that the training led to better understanding of the weighted mean.

First, the data suggest that what was learned was more than the mindless application of a weighted mean algorithm, because the trained subjects were able to discriminate quite well between situations in which single mean versus weighted mean solutions were appropriate. In the trained noncalculator–nonbalancer group, where a weighted mean answer was appropriate (Problems 2 and 3), 75% of the solutions were weighted mean solutions, whereas for a similarly worded problem in which the simple mean solution was correct (Problem 4), only 33% of the solutions were weighted mean solutions.

In addition, several other qualitative differences between the trained and control groups suggest balance training produced a better understanding of the ideas of weighting and proportionality. Probably the most interesting evidence of this type came from subjects' responses to Problem 5, in which subjects were provided with the salary of a shop owner and the mean salaries and numbers of foremen and general workers, and were asked to decide whether a simple mean or weighted mean calculation would be more appropriate to determine the overall average salary as well as to justify their decisions. They were also asked which of the two calculations would result in the larger answer.

Although 21 of the 22 noncalculators–nonbalancers correctly chose the weighted mean calculation, subjects differed as to what kinds of justifications they provided for their choices. Justifications were classified as involving proportional logic if they referred to the proportions of workers in the three categories (e.g., "Just because you have 12 general workers, you have to weight the $8,000 12 times as heavily as the owner–president and 6 times . . . the foremen."). Justifications classified as not involving proportions were generally focused on the number of people in the shop (e.g., "Just to add it up and divide it by three, you would only be looking at 3 people when actually you're looking at 15 people, and that has to be considered.").

Subjects also provided a range of justifications for deciding which of the two calculations should yield the larger answer. Again, some justifications seemed to involve the notions of weighting and proportion (e.g., "The second calculation should be lower. In the first, the high salary carries as much weight as the low. In the second, the low carries 12 times as much weight."). Eight of the 12 trained noncalculators–nonbalancers provided a justification classified as involving proportional reasoning of at least one of the two parts of the shop problem, whereas this was the case for only 1 of the 10 control noncalculators–nonbalancers, $\chi^2(1, N = 22) = 7.29, p < .01$.

Trained noncalculators–nonbalancers were also much more likely than control subjects to label the blocks spontaneously while attempting to represent the weighted mean problems on the balance beam. In a weighted mean representation, the block both marks a location on the scale representing a value being averaged and represents a unit of the frequency of that value's being observed. For example, one subject responded in the GPA problem, "Each block will represent a semester. He got 3.2 for two semesters, so it would be 2 at 3.2 and then 4 weights at 3.8." Subjects who had difficulty with weighted mean problems generally considered only the first of these functions, that is, the marking function. A greater proportion of the trained noncalculators–nonbalancers spontaneously labeled blocks than did control subjects for both Problem 2, $\chi^2(1, N = 22) = 5.51, p < .02$, and Problem 3, $\chi^2(1, N = 22) = 6.71, p < .01$.

General Discussion

As mentioned earlier, a number of studies examining subjects solving well-defined move-type problems (e.g., missionaries and cannibals, wives and jealous husbands) have suggested that transfer from one problem to another analogous problem is difficult to obtain unless subjects are made aware of the specific nature of the analogy (e.g., Fiszman, 1976; Reed et al., 1974). A recent study of analogical transfer by Gick and Holyoak (1983) suggests that if subjects are able to develop a schema for a type of problem, they are more able to transfer the method of solution to a similar problem. In one experiment, their subjects read two stories that presented a problem and its solution and were then asked to solve a problem from a different semantic domain that was analogous to one or both of the stories. Subjects who read two analogous stories were much more likely to solve the problem without a hint than were subjects who received just one story analog plus a disanalogous control story. Gick and Holyoak proposed that the former subjects were more easily able to abstract a schema and were thus more successful in applying the appropriate solution to the analogous problem.

The results of the present study are consistent with those findings and suggest that even if an appropriate analogy is made available, the amount of facilitation will depend crucially on how well the analogous domain is understood (i.e., how good the quality of the schema is). All of the subjects in the present study were told that the mean could be thought of as the balance point of a distribution and were able to represent the solution to a simple mean problem. However, when noncalculators–nonbalancers were asked to represent weighted mean problems on the balance beam and calculate the answers, those who had been given balance training performed much better both in calculating the answers and in representing the problems. Trained subjects were also more inclined to engage in proportional reasoning when calculating and justifying their choice of calculations and were more likely to consider exactly what a block stood for when representing a problem. The present study implies that the diagrams of balance beams frequently found in introductory statistics textbook discussions of the mean are not likely to be very helpful unless the students have a good understanding of balancing. Because most texts present only one diagram, the student is unlikely to be able to abstract a good schema for balancing.

An important question for future research is what the range of situations is in which subjects who have an adequate schema for the balance beam will appropriately apply it to problems involving means. In the present experiment, the requirement in the transfer session that subjects represent problems on the balance beam, as well as calculate their answers, constituted a strong situational cue to integrate information gained during balance training with their knowledge of the mean, even though the concept of the mean was never mentioned during training. It seems likely, given the previous literature on transfer in problem solving, that some connection between the balance beam and the concept of the mean has to be suggested by the transfer situation. However, it is not clear whether it was necessary in the present study to require that subjects represent the problems on the balance beam in order for them to use their balancing knowledge. An especially interesting question is whether the concrete operation of balancing was necessary in the transfer situation or whether the balancing schema was sufficiently internalized by the training procedure that transfer could be achieved by a suggestion that balancing knowledge would be relevant even in the absence of the balance beam. Although it is yet to be determined what circumstances will cause students to access their knowledge of balancing and apply it to problems involving the mean, the present experiment demonstrates that an understanding of balancing can be taught in a relatively brief time and can be applied, with understanding, to problems about weighted means.

References

Bromage, B. K., & Mayer, R. E. (1981). Relationship between what is remembered and creative problem solving performance in science learning. *Journal of Educational Psychology, 73*, 451–461.

Clement, J. (1981). Analogy generation in scientific problem solving. *In Proceedings of the Third Annual Meeting of the Cognitive Science Society* (pp. 137–140). Berkeley: University of California.

Fiszman, A. (1976). *Transfer in problem solving.* Unpublished master's thesis, University of Massachusetts, Amherst.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics.* New York: Norton.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Greeno, J. G. (1978). Nature of problem solving abilities. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. V. Human information processing* (pp. 239–270). Hillsdale, NJ: Erlbaum.

Hardiman, P. T. (1981). *The balance beam as a tool for understanding the mean.* Unpublished manuscript, University of Massachusetts, Amherst.

Hays, W. L. (1973). *Statistics for the social sciences.* New York: Holt, Rinehart & Winston.

Hays, W. L. (1981). *Statistics.* New York: Holt, Rinehart & Winston.

Krutetskii, V. A. (1976). *The psychology of mathematical abilities in school children.* Chicago: University of Chicago Press.

Luger, G. F., & Bauer, M. A. (1978). Transfer effects in isomorphic problem situations. *Acta Psychologica, 42*, 121–137.

Myers, J. L., Hansen, R. S., Robson, R. C., & McCann, J. (1983). The role of explanation in learning elementary probability. *Journal of Educational Psychology, 75*, 374–381.

Pollatsek, A., Lima, S., & Well, A. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191–204.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6*, 436–450.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520.

Sinatra, G. (1980). *Misconceptions in understanding the weighted mean.* Unpublished manuscript, University of Massachusetts, Amherst.

Skemp, R. R. (1979). *Intelligence, learning and action.* Chichester, England: Wiley.

Received July 18, 1983
Revision received December 8, 1983 ■