1983

The importance of analog knowledge in understanding the mean.

Pamela Thibodeau Hardiman
University of Massachusetts Amherst

THE IMPORTANCE OF ANALOG KNOWLEDGE IN UNDERSTANDING THE MEAN

A Thesis Presented

By

PAMELA THIBODEAU HARDIMAN

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

February 1983

Psychology

THE IMPORTANCE OF ANALOG KNOWLEDGE IN UNDERSTANDING THE MEAN

A Thesis Presented

By

PAMELA THIBODEAU HARDIMAN

Approved as to style and content by:

Arnold Well

Arnold Well, Chairperson of Committee

Jere Confrey

Jere Confrey, Member

Jerome L. Myers

Jerome Myers, Member

Alexander Pollatsek

Alexander Pollatsek, Member

Bonnie Strickland

Bonnie Strickland, Department Head
Psychology

ACKNOWLEDGEMENTS

The ideas involved in any project with the scope of a master's thesis cannot be credited to any single person. The final product is the result of a seemingly never ending cycle of discussion and change. I wish to thank the members of my committee who provided a vital aspect to that discussion. Arnie Well and Sandy Pollatsek have been involved in all stages of this research since the inception of the pilot study, and helped to define the nature of the questions, refine techniques of research, and interpret subjects' behavior. Jere Confrey suggested closer examination of the conceptual change which occurred during the transfer session and assisted in analysis. Jerry Myers provided the perspective of a wider community of researchers who do not rely on the interview as a primary source of data.

I particularly wish to thank Arnie Well, with whom I spent many hours discussing this project and who helped grind and polish this thesis to linguistic acceptibility.

The progress of this work was greatly facilitated by many gracious persons who supported "Pam's Little Store" in January of 1982. This enterprise was born of necessity to earn funds to replace a stolen typewriter. Thanks to the residents of Tobin Hall, especially Nancy and Jerry Myers and Deborah Davis, I was not long typewriterless.

ABSTRACT

The mean is a commonly employed descriptor of a set of numbers, and forms the basis for several related statistics. Evidence indicates many undergraduates do not possess a relational understanding of the mean concept (Pollatsek, Lima, and Well, 1981). Pollatsek et.al.(1981) postulated three types of knowledge are involved in understanding the mean: functional, computational, and analog knowledge. Many of the college students they interviewed did not appear to posses adequate functional and computational knowledge, while none showed behaviors which might suggest they possessed analog knowledge.

A survey of introductory textbooks revealed a balance model of the mean is commonly employed as an explanatory tool. Several of the ideas involved in analog knowledge are clearly illustrated with this model. Hardiman (Note 3) investigated the usefulness of the beam model in a pilot study where subjects were asked to represent weighted mean problems on a concrete balance beam. Many subjects found the task of representing the problems quite difficult; the comments of several suggested they did not initially understand how the balance beam worked. Thus, the balance model may not be an effective teaching tool if students do not have the knowledge of balancing.

In the present study, the first issue addressed was

the relationship of balance knowledge and computational knowledge. Subjects were given a written pretest with two weighted mean problems and twelve balance problems. Computational knowledge and balance rule level, classified by Siegler's (1976) levels, were related, r=.35. In order to determine whether balance knowledge might aide in understanding the mean, experimental methods were employed.

The second issue involved determining whether experiences fostering the development of balance knowledge would lead to improved calculational performance. After receiving balance training or a control problem, subjects were asked to represent two weighted mean problems on the balance beam and compute the means. Balance training led to a significant differences in the calculation performance of noncalculators (F(1,32)=8.64). Qualitative indices, such as spontaneous labeling of the quantity represented by the block and rationalization for the weighted mean method on a forced choice problem, were strongly correlated with correct answer and training.

The study has two implications: 1) the balance beam model may be a useful teaching tool if students have the knowledge necessary to understand it, and 2) transfer is more likely to be successful when subjects develop a good understanding of the to-be-transfered domain.

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

CHAPTER I

INTRODUCTION

Overview

The present research focuses on the average or mean; one of the most fundamental and ubiquitous concepts in statistics. Many students do not fully understand the mean and have considerable difficulty with weighted mean problems (Pollatsek, Lima, and Well, 1981). Since many concepts in statistics are based on the mean, there is concern with how this concept may be communicated more effectively.

Problems in effectively transmitting quantitative concept occur at all levels of education (c.f. Kline, 1973 for a critique of mathematics education). One factor which may contribute generally to these problems is the relative importance placed on successful manipulation of numbers without concurrent assessment of students' understanding of the meaning or the end products of the manipulations. It is probably the case that most attempts to communicate quantitative concepts fail to consider certain more basic kinds of knowledge which might be usefully exploited in the process of developing an understanding of the concept. For example, a corpus of research by Piaget (1965) and others (e.g., Wang, Resnick, and Boozer, 1971) indicates there may

be more to understanding seemingly simple addition facts, such as 2+2=4, than might readily be apparent. Certainly knowledge of counting is important, which few would deny. But the child must also have acquired some notion that numbers are stable descriptors of quantities before addition facts become meaningful. Successful performance on arithmetic calculations alone does not indicate understanding of the concept of addition, since satisfactory answers may be given with rote procedures. Other indications of understanding are necessary. Similarly, adequate performance on problems in which means must be calculated does not necessarily indicate much understanding of the concept of the mean. A better measure of understanding can be found by employing problems which do not readily lend themselves to the usual algorithms used for calculation.

Differential understanding of concepts has been studied by comparing solving and sorting behaviors of experts and novices in a particular field. Chi, Glaser, and Rees (Note 1) suggest that the expert's superior problem solving performance is due mainly to a more adequate and integrated knowledge base in the problem domain. However, differences in the ease of acquisition of concepts in a field by novices have received little attention. A logical hypothesis derived from the expert/novice studies might be that in order to increase the likelihood that a

nonsolver will attain success, the knowledge base of the nonsuccessful novice should be built upon so that it more closely resembles that of a successful novice. Presumably the successful novice has knowlege which facilitates a deeper understanding of the concept.

In the present study, a hypothetical framework of knowledge which might be involved in understanding the mean is presented. The framework suggests that knowledge of balancing may be an important basic concept facilitating understanding of the mean which texts either ignore or assume the student already possesses. The goals of the present project were: 1)to determine whether subjects with an adequate understanding of balancing were more successful in solving weighted mean problems, and 2)to determine if the performance of nonsolvers could be improved by experiences which facilitated the development of balance knowledge. The hypothesis was that intuitions based on the observable actions of a balance beam would provide alternative and possibly more convincing explanations of why a mean in a given situation should be in a certain approximate place. The concept of the mean and its role in statistics will be defined more comprehensively in the second section of this chapter. Chapter II will focus on misconceptions involved in understanding the mean. It will also include a discussion of the difficulties, in general,

of obtaining transfer effects in problem solving studies. The topic of the third chapter is the relationship between knowledge structures and problem solving. Possible components of a hypothetical model for understanding the mean, including balance knowledge, will also be discussed. The next chapter will discuss the treatment of the mean by standard statistics textbooks to determine whether the components identified as possibly important are included in curricula, and how well that is done. The preliminary discussion will conclude with a description of the present study and a rationale for the methods used.

The Concept of Mean

The concept of a mean or the average of a set of numbers is one of the most fundamental and frequently employed descriptors in statistics, and is commonly encountered in everyday life. The mean is the most commonly used measure of central tendency and forms the basis for other descriptors of a distribution, such as the standard deviation and the variance. The layman encounters the mean in such guises as average temperature and rainfall, per capita income, stock market indices, and measures of school per-

formance.

One might make the obvious assumption that this pervasive and important concept, which is fairly simple to calculate, should be well understood by most college educated adults. However, evidence is accumulating which suggests many college students do not have a well formed notion of what the mean represents and how it should be calculated in a variety of situations (Pollatsek, Lima, and Well, 1981; Sinatra, Note 2; Hardiman, Note 3).

The typical textbook defines the mean as the total of a set of numbers divided by the number of numbers in the set. Calculation of the mean is a fairly obvious procedure when each number in the set is added once and the proper divisor is the total number of numbers given in the problem. However, when the numbers given are means based on n_1 and n_2 cases (where $n_1 \neq n_2$) and an overall mean must be calculated, students frequently demonstrate a lack of understanding. A commonly made error is to simply add the two numbers together and divide by two. In this case though, if the overall mean is the number most representative of all the n_1 and n_2 cases, the simple mean will not suffice because the two means are not equivalent by virtue of being based on different numbers of scores. One must somehow weight the quantities involved so the mean based on the larger number of scores is counted more. To weight the

means properly, one must use either: 1) A computational formula for calculating the weighted mean, such as

$$M(comb) = (n_1 X_1 + n_2 X_2)/(n_1 + n_2)$$

or

$$M(comb) = (n_1/n_1+n_2)X_1 + (n_2/n_1+n_2)X_2$$

or 2) a computational formula for the simple mean combined with a knowledge of how to obtain the appropriate totals for each group and the total number of scores, such as

$$M(comb) = \Sigma Xi/n, \text{ where } \Sigma Xi = n_1X_1 + n_2X_2.$$

Solving a weighted mean problem without knowing a specific computational formula for the weighted mean probably involves realizing at least three things: 1) calculations using one case of each mean versus n_1 and n_2 cases of each mean will not yield the same answer, 2)it is important in the interpretation of a particular problem to consider the number of cases when calculating the mean, and 3)the sum of a group of scores is equal to their mean multiplied by the number of cases in the group. Thus, the definition of the mean presented at the beginning of the section already seems to encompass more information than might be immediately apparent.

CHAPTER II

PREVIOUS RESEARCH

Several recent studies have explored student understanding of the mean using individual interviews and paper-and-pencil problem sets (Pollatsek et.al., 1981, Sinatra, Note 2; Hardiman, Note 3). A variety of issues arose in this research, including rote dependence on calculational procedures, misconceptions, difficulties in obtaining transfer, and a general inability of most subjects to represent situations using a concrete analog device. Each of these issues will be discussed in the above order.

Concept or Computation?

Some of the knowledge which might be implied by an adequate understanding of the mean was identified in the previous chapter. How much of this information do students develop by the time the reach or complete an introductory statistics course? The basis issue revolves around the meaningfulness of the calculational procedure for a majority of students: do students form a concept of the mean that would enable them to apply the concept flexibly, or is the mean merely a computational procedure with little additional meaning?

Pollatsek et.al. gave college students problems involving combinations of means, such as the following:

> A student attended College A for two semesters and earned a 3.2 GPA. The same student attended College B for four semesters and earned a 3.8 GPA. What was the student's overall GPA?

They found problems involving combinations of means were solved correctly by only 38% of a sample of undergraduate students beginning their first course in statistics. Interviews were subsequently conducted to elucidate student understanding of the mean. The students were asked to think aloud while solving problems similar to the GPA problem. Most students were unable to calculate the weighted mean spontaneously (i.e. without interviewer probes). Many did not change to a weighted mean calculation even after the interviewer challenged their belief that the simple mean was sufficient. Some students did realize the simple mean was not "fair", in the sense of being the best overall index of performance, but accepted the answer derived as what followed from the definition.

Pollatsek et.al. suggested that a fundamental problem was that students understood the mean purely in terms of a comptational algorithm applied to an abstract set of numbers. This would correspond to what Skemp (1979) has described as an "instrumental" understanding of a concept. The person is able to recognize the problem as one of a particular class for which one knows a rule for solution.

In the case of the weighted mean, the problem is misclassified as a simple mean, or more generally by the subject with an incomplete understanding, as a mean. Understanding the mean as a concept would imply "relational" understanding, or the ability to relate a task to some appropriate schema and devise plan for computation if one is not available.

One plan for computation might be to obtain the total sum of scores for each group, add these together, and divide by the total number of subjects. The sum of scores for a group is easily attained by multiplying the mean of the group by the number of scores. However, the students were either unable to do so or did not believe that would yield a correct answer. They did not view the mean as a quantity which could be operated upon. Kaput (1979) believes many symbols and operations in mathematics, such as , inappropriately involve both process knowledge and product knowledge. A simple arithmetic example demonstrating this is: "2 + 3 = 5." The left side is read as a process, or what is done to 2, and the right as a result, 5. He argues that logically in mathematics the process and the results are the same, but they are not necessarily interpreted the same way by the learner. The direction the equation is presented in may be quite influential in determining what is viewed as process and results, and hence

which quantities can be logically operated on.

The common expression used to calculate the mean is $\Sigma X/n=M$. The mean is both a process of compuation and the result of a computation. The students in the Pollatsek et.al. study seemed to grasp the mean as a process, but found it difficult to deal with the mean as a meaningful result. In fact, their knowledge of the mean frequently did not contain the idea that it should be "meaningful" in any sense. Some subjects obtained overall means smaller than either of the means given in the problem and did not appear to be bothered by this fact.

If Kaput's analysis of the importance of directionality is correct, it may be reasonable to assume that students may also have much more difficulty with the equation stated as, $M=\Sigma X/n$, the reverse and necessary to derive, $nM= X$ or the sum of the scores.

Thus, a complete understanding of the mean involves more than knowing a computational formula for the simple mean. Pollatsek et.al. suggest that complete understanding of the mean has three components: 1)functional knowledge, 2)computational knowledge, and 3)analog knowledge. Functional knowledge consists of understanding the mean as a real world concept, a number which 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 knowing a computational formula for the weighted mean or a computational formula for the unweighted mean with provisions made to account for different numbers of scores in each group. In the latter case, one needs to obtain the total of a subgroup given its average and number of scores (e.g. demonstrating reversibility; Krutetskii, 1976) as previously discussed. One possibility for analog knowledge involves viewing the overall mean as the "balance point" for the entire et of scores. Analog knowledge would enable a subject to realize the overall mean should be closer to the mean on which a greater percentage of the scores are based.

While all of the subjects in the Pollatsek et.al. study were able to calculate simple means when given a set of scores, many of them seemed to lack functional knowledge of the mean. They simply calculated a number and did not evaluate whether it made sense within the context of the problem. Some showed little or no indication that they realized the number of scores each mean was based on was an imortant piece of information which had to be incorporated into the calculation in some way, while others could not incorporate the number of scores into the calculational procedure. No subject seemed to give any evidence of analog knowledge by indicating without calculating that the

mean should be closer to one number than another, or that it should be within a certain range. Thus, many students may be deficient in areas of knowledge which Pollatsek et.al. have speculated are important.

In summary, a large proportion of the undergraduate students interviewed by Pollatsek et.al. were not able to calculate a weighted mean correctly. They seemed to treat the mean as if it were merely a computational procedure and were in command of little necessary additional knowledge which would aid in the calculation of the weighted mean.

Misconceptions

Several recent studies, such as the Pollatsek et.al. study of the mean and Rosnick and Clement's (1980) study of algebra, suggest that not only may students fail to understand a concept, they may develop certain misconceptions which can impede the learning of the concept. Although many of the students in the Pollatsek et.al. study were not able to calculate the weighted mean, nevertheless they seemed to hold tenaciously certain beliefs about the mean and the behavior of numbers which were not readily amenable to change within the short interview period. Sinatra (Note 2) attempted to determine and describe common misconceptions concerning the mean in an interview in which each

subject solved several weighted mean problems.

Sinatra found that many subjects had an inadequate understanding of the concept and agrred with one or more of the following statements:

1) the simple average adequately represents the overall average of a combination of means.

2) the simple average is correct, but does not fairly represent the average of a combination of means based on different numbers.

3) the real total of each group cannot be determined. It is only approximated by nM.

4) the simple mean of three numbers is equal to the mean of one pair plus the third number divided by 3 or $(((x+y)/2)+z)/2 = (x+y+z)/3$.

5) When finding the sum of scores for a group, it is not reasonable to substitute the mean for each element.

6) To find the weighted mean, one should add the unweighted means and divided by the total number of scores or $n_1 + n_2$.

Several subjects in the Sinatra study found it impossible to calculate the weighted mean when not given the set of raw scores. The first type of error suggests subjects either did not realize the number of raw scores was important, or they failed to distinguish between raw scores and means. The second error suggests subjects may have realized that the number of raw scores may have been important, but could not incorporate that information into their calculation. Subjects who made error types 3-6 appeared to realize that the differing numbers of scores was an important factor in the problem, but their attempts to incorpo-

rate that information were not successsful. Errors 3,4, and 6 involve algebraic manipulation, and may reflect on other misconceptions in algebra. Error 5 is a consequence of a failure to conceive of the mean as a meaningful result, i.e. $\Sigma X=nM$ and $nM= M+M+..+M$.

Performance was somewhat better when more concrete quantities were used (e.g. in fish problem) presumably because the overall sums (e.g. number of fish as opposed to semsester-grade-points) were more meaningful.

Knowing these common misconceptions and potential problems with the mean and the conditions in which they occur makes possible the design of instructional materials which address these problems.

Transfer

In considering ways in which the mean might be approached more effectively in the classroom, it seems wise to consider past efforts in the study of transfer, or conditions which enable what has been learned in one situation to be applied in another. The goal of obtaining transfer in problem solving is one which has been of considerable interest to educators in the past (c.f. Goodwin and Klausmeier, 1975) and to information processing

psychologists more recently (c.f. Tuma and Reif, 1980). The general question of interest is: "What types of experiences will and will not enable a problem solver to apply what has been learned in one situation in another?"

A paradigm which has been popular recently involves presenting subjects with one problem to solve, asking them to solve a second isomorphic problem, and then analyzing time and strategy differences between the solutions to the to problems (c.f. Hayes and Simon, 1977). The problems used are generally well defined (Reitman, 1964), have specific legal operators, and definable solution spaces. Often the problems used are "move type" problems, such as the "Tower of Hanoi" or "Cannibals and Missionaries", where variables in a certain configuration must be manipulated within certain constraints toward a defined final configuration. The succesful demonstration of transfer within this paradigm has been rare. In this section, the results of several transfer studies and the assumptions made will be presented.

Sinatra (Note 2) used this paradigm to study transfer in the solution of weighted mean problems. Two sets of condtions were included in the study: 1)a two problem set with a relatively abstract problem (the GPA problem quoted earlier), and a relatively concrete problem dealing with the numbers of fish caught on two boats, and 2)a four

problem set which included the above two problems and two problems of moderate abstractness. The fish and GPA problems were presented consecutively and in counter-balanced order. Sinatra found no significant difference in performance on the second problem as a function of the first problem in either ordering in both the two and four problem sets.

The Sinatra study had a small number of subjects, but does suggest misconceptions are not readily remedied in a short period of time with experience on similar problems. Rosnick and Clement (1980) specifically studied the effects of different types of tutoring on algebra misconceptions and more conclusively suggest that misconceptions are not easily taught away with experience on similar problems. Misconceptions seem to be fairly robust within these domains.

Since tutoring students with a variety of similarly structured problems does not appear to be a fruitful method of producing transfer in a short period of time, do there exist any conditions which might lead to transfer? Reed, Ernst, and Banerji (1974) sought to produce transfer between two isomorphic move problems: 1)Missionaries and Cannibals (MC) and 2) Wives and Jealous Husbands (JH). The two problems have isomorphic solution spaces, with the restriction that the individuals must be identified in JH,

whereas all members of a group are equivalent in MC. Additionally, the mapping between the problems is not obvious, in that cannibals correspond to wives and missionaries to husbands.

There was no transfer between the two problems in either direction when the subjects were not told how the problems were related. In a condition where the relationship was explained, there was transfer from the more difficult JH to MC, but not vice versa. Reed et.al. suggest five conditions are necessary for transfer:

1) recognition that the problem is analogous to a previous problem.

2) ability to retrieve information regarding the solution of the previous problem.

3) ability to translate the past operations into operations of the current problem.

4) the translation must define a unique operation or reduce the number of options that would be considered.

5) the total time to retrieve, translate, and use analgous information must be less than finding the same operator without the previous problem.

The first condition is the most potentially problematic, and therefore will be discussed last.

There is supporting evidence suggesting the second condition is easily met. Subjects in the Reed et.al. study were able to solve a problem faster the second time they had seen it. In a reading study by Kolers (Note 4)subjects were to read inverted and reversed text with the spaces

between the words removed. Although this was not a problem solving study, the results are pertinant for this argument. The subjects were able to read the text faster the second time they saw the text, even though they reported no concious memory of the text. Therefore, it seems probable that most subjects have some memory for previous solutions. Ability to retrieve specific information may not be important. In fact, Reed et.al. found sujects tended to repeat general types of moves, rather than specific sets of moves when solving a problem the second time.

Condition 3, translation, most likely involves specification of details of the analogy formed in step 1. However, to recognize that there is an anlogy it seems some rough translation must be made in the beginning. A problem which may be roughly analogous may not yield a simple translation. In this case, transfer would yield the optimal time to solution if the subject attempted specification.

The fourth condition might be restated as follows: an additional single unique step is required or the number of possible solutions must be reduced in the second problem to obtain transfer. The JH problem has many more options to consider than the MC problem. In this case, transfer was obtained from the more complex to the simpler problem, from JH to MC. Dienes and Jeeves (1965) found similar results

with children and adults asked to solve two problems which required subjects to determine the ordering of a set of colors drawn from a two or a four color rule. These rule structures for ordering were based on mathematical sequences. They found more facilitation when the four group problem was given first. A necessary aspect of this condition seemed to be that the number of options be greatly decreased from the more complex to the simpler problem. A simple translation did not yield unidirectional transfer.

The fifth condition is a fairly obvious consequence of the third condition: translation may involve more time than simply solving the problem. Therefore the transfer manipulation would yield no advantage.

The real crux of the problem of obtaining transfer seems to be recognition of an anology. What constitutes an analogy? How explicit do the relations between problems need to be in order for the subject to perceive the two situations as analogous? What is the mechanism for a spontaneous analogy? Reed et.al. do not address the question of spontaneous analogy, but informed the subjects how the problems were analogous. They seem to suggest, that at least in some cases, the subject needs to be made explicitly aware of the analogy for the relation to be exploited.

Gick and Holyoak (1980) investigated the analogy stage

more closely. They asked subjects to determine a solution for a single story problem which had certain restrictions on the possibilities for solution. The problem was preceded by a story with an analogous plot. In some conditions, the story did not fulfill all the necessary conditions for solution of the problem story. These were insufficient analogies. A story fulfilling all the conditions was a sufficient analogy.

Solutions fulfilling all conditions were rarely produced spontaneously when no analogy was presented previously. In general, the subjects tended to produce solutions which were anlogous to the preceding problem, whether the solution presented was sufficient or not. These findings suggest two properties of analogies: 1) analogies derived from disparate domains can be used to aid in the solution of a problem, and 2) analogies tend to block other types of solutions. Analogous solutions were also produced when a sufficient, but less strongly analogous story preceded the problem, suggesting complete mapping is not necessary to produce analogies.

Gick and Holyoak noted that subjects do not seem to spontaneously use the analogy from the preceding story. It was helpful only when subjects were told to try to use the story to solve the problem. This observation concurs with that of Reed et.al. who also noted subjects did not tend to

use the preceding problem as an analogy.

Clement (Note 5) specifically investigated spontaneous analogy generation in highly trained scientists because it is presumed to be a key element in scientific discovery. Spontaneous analogy did appear to play a key role in the solution of several subjects solving a physics problem. Clement derived several conditions from these observations of analogy generation which seem necessary for making an inference by analogy:

1) the analogous conception, B, is generated, given incompletely understood situation, A.

2) the analogy relation between A and B is confirmed.

3) conception B must be well understood, or at least predictive.

4) the subject transfers conclusions or methods from B back to A.

These conditions are similar to Reed et.al.'s with the exception that they involve generation of the analogous situation. However, Clement additionally suggests the analogous situation, B, must be well understood. One could argue that this condition was not present in the Reed et.al. study when no relationship between the problems was mentioned. The subjects had attempted to solve a fairly difficult problem one time, and therefore were unlikely to have developed optimal move strategies. Therefore, the B situaton, to which subjects were supposedly making an

analogy, was probably neither well understood nor predictive.

A study by Luger and Bauer (1978) suggests transfer is much more likely with new and unfamiliar problems when the subject is allowed to develop a reasonable understanding of the B problem. Luger and Bauer asked subjects to solve two isomorphic Tower of Hanoi type problems which were not immediately recognized by subjects as analogous. The dependent variable was time to develop an optimal solution, which would suggest the problem was ultimately well understood. Transfer effects were obtained with both orders of the problems and no hints pertaining to the analogy.

The experimental evidence indicates that fully understanding the analogous situation is an important aspect of recognizing an analogy and using it in transfer tasks. Confrey(Note 6) argues in addition that transfer itself requires a significant amount of reconstruction, again implying that the analogous situation must be well understood before it can be modified. The transfer studies cited may have failed because the preceding problems were not well understood, and hence failed to provide a sufficient analogy.

Concrete Materials

The use of concrete materials in mathematics classes, particularly in the lower grade levels, has recently assumed great popularity (Herold, 1978; Mitzman, 1976). Lesh (1979) claims concrete materials may help foster a set of abilities to a greater degree than more abstract exercises. This set of abilities includes:

1) the ability to impose structure on concrete materials in everyday situations.

2) the ability to translate among various models and interpretations of an idea.

3) the ability to correctly interpret spatial/geometric aspects of various models for an idea.

However, concrete materials must be used in the proper context, since they can easily be used in a manipulative fashion rather than assisting in constructive learning (Wilkinson, 1974).

Hardiman (Note 3) used a concrete analog device to examine student conception of the weighted mean. The subjects were asked to represent several weighted mean problems on a balance beam. The balance beam approximated the idea of a weightless number line placed on a teeter-totter with blocks used as weights. The hypothesis was that use of the balance beam to represent the problem should make the relationships between the numbers, or the need for weighting, more obvious and lead to better calcu-

lation performance.

However, only a small number of subjects were able to both calculate and represent the weighted mean correctly. Many subjects found the simple mean and represented it on the balance beam. In both these cases, the answers for representation and calculation were consistent. In addition, there were several subjects who were able to calculate the weighted mean correctly but were not able to represent the weighted mean on the balance beam.

These subjects with inconsistent representations often attempted to represent the total for a group (e.g. one block on nM rather than n blocks on M) or to equalize the groups (e.g. double n for a group) so that only two blocks were necessary. These strategies inevitably failed, since the representation must contain information pertaining to the average of each group and a relative weighting for the number of scores in each group. These subjects obviously realized a simple weighting with a single block on the group average was insufficient, but were unable to grasp that the information they needed to represent must be indicated with a combination of weights and distances. The subjects appeared to treat the blocks as if they were merely placeholders, precluding the possbility of using more than one block on a spot.

Performance on later problems indicated a majority of

the subjects who could calculate the mean correctly but not represent it may have possessed a more superficial understanding of the mean than those who did both correctly. Simple mean calculations were given to weighted mean questions. However, all subjects in the study excepting one were able to calculate and represent a weighted mean problem correctly with no intervention after a 45 minute session. An additional problem was devised to determine whether subjects had actually developed further understanding of the mean or merely learned an algorithm. It was a simple mean problem with the surface structure of a weighted mean problem:

> Students in a chemistry lab must measure quantities of chemicals several times in order to obtain accurate measurements. Student A measured her quantity 3 times and obtained an average measurement of 1.05 grams. Student B measured his quantity 2 times and abtained an average of .97 grams. What was the average weight of the two students' quantities?

Tennyson and Park (1980) and Winston (1975) suggest nonexamples which are similar in all respects but the critical aspect are important for concept development.

The nonexample was presented to only approximately half of the subjects because of time constraints. All subjects who were able to calculate and represent all problems correctly were able to solve the chemistry problem, as well as two of the three subjects who originally calculated and represented the simple mean on the first

problem. However, neither subject with inconsistent representations was able to solve the problem, even after strong challenges to their reasoning. The evidence, although limited by numbers of subjects, tentatively suggests some students calculate the weighted mean as a rote procedure with little meaning. They developed an algorithm for representation during the course of the interview and failed to distinguish cases where the algorithm did not apply.

One factor distinguished subjects who solved the final problem from those who did not: those who did not spontaneously made some comment which reflected a lack of understanding about the concept of balancing. Some did not believe the balance beam would balance when the blocks had been properly placed, while others claimed they had learned in the course of the interview that a small number of weights placed at a distance from the fulcrum could balance a larger number of weights closer to the fulcrum. These comments suggest understanding balance rules may facilitate understanding of the weighted mean. Additional research is needed to confirm any advantage from knowing balance rules.

Logically, balance rules would seem to be an essential aspect of the component of mean knowledge which Pollatsek et.al. have termed "analog knowledge." It is probably difficult to predict where the balance point of a set of

numbers might be if one has not developed intuitions concerning the mechanism of balancing. In fact, several subjects had difficulty simply adjusting the scale to balance: they were not certain in which direction the balance point should move.

Siegler's (1976) study of balance rules in children offers additional evidence suggesting that many college students may not have highly developed rules for predicting balancing. Siegler conceived of balance knowledge as a series of questions in a tree diagram formation, ending in different points for persons with different levels of balance knowledge. These rule levels are :

1) Subject compares weight only and says side with greater weight will go down.

2 If the weights are equal, the subject considers the distance and says the side with the greater distance will go down. If the weights are unequal, the side with the greater weight will go down.

3) The subject checks to see if the side with the greater weight also has the greater distance. If not, muddle through.

4) The subject examines whether the cross products of the two sides are the same. If not, the side with the greater product goes down.

A subject with a fully developed understanding of balancing proceeds to level IV questions when attempting to predict the action of a balance beam and is always correct. Subjects using other rules will predictably arrive at the wrong concclusions about certain types of problems,

enabling one to test for rule level. Sielgler found only 40% of 16-17 year old subjects exhibited evidence of rule IV balance knowledge. Given these results, it is not probable that a majority of college students have developed rule IV knowledge, possibly implying they may not have an essential aspect of the foundation knowledge necessary for a full understanding of the mean.

CHAPTER III

KNOWLEDGE STRUCTURES

Assumptions

Problem solving, as a higher order cognitive function, necessarily implies the use of other cognitive functions, such as perception and memory. In order to be able to state more clearly the present assumptions underlying the view of problem solving presented in the present study, assumptions regarding the constructs of perceptions and memory must be specified.

It is possible to assume any of a number of different views of the subject when considering how she/he comprehends information to solve a problem. In this study, it will be assumed that subjects construct their perceptions of problems and they do so in a manner which is consistent with their past knowledge. This implies each subjects' interpretation of the problem will be at least slightly different. This view is referred to as constructivism. A contrasting view might assume the subject is passive, implying all subjects perceive problems in the same manner.

According to a constructivist viewpoint, the nature of perceptions is determined through an interaction between the mechanisms of the perceiver and the presumed indepen-

dent reality of the stimulus. Subjects attempt to interpret problems in terms of existing knowledge. That existing knowledge permits identification of problem type, specification of relevant information, and assistance in formulating a plan for solution (Konold and Well, Note 7).

The role of importance ascribed to preexisting knowledge by the constructivist viewpoint requires that assumptions about the structure of preexisting knowledge be made more explicit. Memory is characterized by organization rather than randomness. Bartlett(1967) suggested memory is composed of active organizations of past reactions which are presumed to be oerating in any well adapted organic response. These structures are referred to as "schemas." Every change entering the system is in some way related to some previous occurrence, therefore schemas are charaterized by constant change.

The term used to describe memory structures in the present study is "knowledge structures." The construct is similar in nature to a schema. A knowledge structure is a "highly specific cognitive structure constructed through activities in limited domains of experience" (Lawler, 1981, pp.1). The set of all preexisting knowledge structures will be referred to as the knowledge base.

A reasonable assumption within a constructivist framework is that in order to determine how the subject inter-

prets a problem, one should know what information is necessary to solve the problem and whether the subject has that information as part of her/his knowledge base. The purpose of the present study is to analyze whether balance knowledge is a significant area of knowledge involved in understanding the mean. In the second section of this chapter, knowledge structures, their relationship to problem solving in general, and aspects of their development will be examined. In the third section, possible knowledge structures which might be involved in understanding the mean will be discussed.

Involvment in Problem Solving

It is a fairly recent suggestion in the problem solving literature that the degree of success a subject acheives in applying a problem solving procedure is likely to be influenced by the degree to which problem solving procedures are meaningfully related to other general concepts in the subject's memory (Greeno, 1978). Myers, Hansen, Robson, and McCann(in press) specifically manipulated the number of interconections between concepts presented and ties made to general knowledge in a study of probability learning. The results were consistent with Greeno's suggestion.

In their study, three groups of subjects were given high, medium, or low explanatory texts for short periods of study. The subjects were later tested on two types of problems: formula and story. The subjects who had read the high explanatory text performed significantly better on the story problems than those who had received either the medium or the low explanatory texts. Presumably the story problems exploited interrelationships between the theorems presented and connections subjects had made to their store of general knowledge. Subjects with the low exlanatory text were given no information about the relationships between the theorems, nor was there any attempt to relate the information to knowledge they presumably previously possessed. These results suggest that successful application of these problem solving procedures in a story context is dependent on meaningful connections between procedures and general knowledge.

When the knowledge structures possessed by subjects do not allow then to grasp the essence or determine the deep structure of a problem, they appear to rely on a common default procedure. In such cases, the surface structure of the problem is relied upon to an excessive degree in forming a problem representation. Simon (1978) and Hayes (Hayes + Simon, 1977) manipulated the surface structure of a problem to study how problem representation and solution

were affected. The problems used were variants on the Tower of Hanoi in which agent/patient and active /passive roles were manipulated. The different surface structures of these isomorphic problems lead to widely differing solution times and variations in the operators. This finding supports the constructivist's contention that problem representation is a result of interaction between knowledge structures and the surface structure of the problem. Similar problems do not necessarily lead to similar approaches.

Possession of knowledge structures is probably most usefully characterized as being a matters of degree, rather than in an all-or-none fashion (Lesh, as cited by Confrey, Note 8). Ideas gradually become more meaningful as they gain in complexity and connections to other ideas and events. Obviously then, the experiences students have in and out of institutional educational environments influence the relative degree of development of quantitative concepts. Students, as active perceivers of their environment, develop intuitions about the behavior of the world which are often independent of the school environment. Tversky and Kahneman (1977) have reported dramatic demonstrations of intuitively developed heuristics which dominate thinking about basic concepts in probability.

Fischbein (1979) claims these untaught ideas play

important roles in the development of concepts. Learner intuition may facilitate, as well as impair, the acquisition of correct knowledge. They appear to have the following qualities: 1)powerful coercive effects, 2)extrapolative capacity, 3)globality, and 4)high stability. These qualities make them highly resistant to change (c.f. Rosnick and Clement, 1980). Fischbein hypothesizes that with adequate instuction intuitions can be built, transformed, or eliminated, according to their usefulness to a concept. This is the ostensible purpose of education. However, adequate instruction remains undefined in many cases.

Most educators and learners would accept a statement which claimed the the learner must have developed certain prerequisite concepts before new material can be successfully comprehended. However, it appears that this prereqisite material is not often well considered from the point of view of the learner. It is generally deemed sufficient that the concepts have some logical connection or ordering from the expert's point of view. Begle, a proponent of the New Mathematics, commented, "We have to teach mathematics in a certain order because that's the way mathematics is" (Suydam, 1970). Given concerns about mathematics education in general which resulted from this type of view, it seems likely that an order of presentation based purely on

logical considerations cannot yield optimal results.

Some Piagetians have made attempts to determine more psychologically relevant conditions of readiness for quantitative concepts. Copeland (1970) and Szeminska (1965) have argued that concepts should not be presented before the learner shows indication of readiness to assimilate them. In the case of certain concepts, readiness may have an age dependent time course, which may be related to more general stages of growth in logic (e.g. Piagetian stages). Wang, Resnick, and Booser (1971) demonstrated that young children seem to develop ideas in arithmetic in a specific sequence: numerals are learned only after counting operations for sets of the size represented are well established, smaller numbers are learned before larger numbers, and counting is independent of one-to-one correspondence.

Once one progresses beyond the very early stages of arithmetic, specific sequences are no longer necessary for conceptual development, so the task of determining optimal orders of presentation is made much more difficult. Ginsberg (1977) claims all children do not need optimal conditions to develop quantitative concepts that are understood to a meaningful level. However, for the less successful it may be necessary to combine different areas of knowledge. Learning is not always incremental, but may require integration across a variety of different content

areas. Dienes (1963) has similarly claimed that children develop a richer understanding of concepts when they must extract the deep structure from a variety of different experiences related to the concept.

Lawler (1981) made a serious attempt to trace the path of development of a quantitative concept in a relatively naturalistic setting by observing his daughter in a variety of tasks. There seem to have been certain identifiable intermediate points when knowledge involved in the understanding of disjoint experiences became linked and led to a qualitatively different understanding of the concept beforeit was fully understood.

Lawler posits that one develops disparate and highly specific knowledge structures, termed microworlds, which are constructed through activities in limited domains of experience. Microworlds are activated through control structures which have two functions: 1)to mediate problems and respond to specific appropriate demands, and 2)to search for problems that can be interpreted in terms of microworld knowledge. When integration amongst microworlds has taken place, the control structures move to a higher level. Although the particular knowledge in microworlds may be accidentally determined, their formation is not accidental because they embody what is epistemologically profound in experience.

During the period when his daughter was learning to add, Lawler documented the existence of a least five microworlds of knowledge which were disparate at the beginning of the study. For example, although she knew two packs of gum cost 30 cents and that 15 cents and 15 cents equals 30 cents, she could not add 15 and 15. Certain critical insights and problems seemed to lead to the elevation of control structures, with small changes permitting large advances. Conjoining may be accidental, depending on simultaneous engagement of microworlds and may be aided by the right problems. However, the active participation of the learner is essential.

Lawler suggests that learning based on the conjunction of microworlds and the intergration of several worlds of experience is more stable, or understood at a deeper level. The fitting together of multiple points of view allows one to modify and adjust a concept to fit a situation.

In summary, the perception of problems and operators used in solution are influenced by knowledge structures, which may assist the subject in determining the deep structure of the problem. The lack of appropriate knowledge can lead to a greater reliance on the surface structuring of the problem. Intuitions play an important role in the development of knowledge structures and may facilitate or hinder acquisition of concepts. Finally, the learning of a

a concept may be benefited by integration across several realms of experience.

Knowledge Involved in Understanding the Mean

Although it has been argued that memorization of a rote calculational procedure is not sufficient experience for many students to gain a deep understanding of the mean, it remains to be determined what types of knowledge and experiences might foster a more adequate understanding. In this section, suggestions will be made concerning potentially relevant areas of knowledge which might be included in a text discussing the mean. These suggestions will utilize Pollatsek, Lima, and Well's deliniation of functional, computational, and analog knowledge as a basis.

Briefly, these three types of knowledge are defined as: 1)functional- an understanding of the mean as a real world concept, 2)computational- involves knowledge of correct computational formulas, and 3)analog- a visual-kinesthetic image, which in the case of the mean might be as a balance point. Each of these general areas of knowledge implies certain specific elements. It is these elements with which will be described.

The basic proposal being considered is that a well

developed concept of the balance beam includes all three types of knowledge. In the framework being presented, one assumption is that the three major components are linked. In fact, certain elements must necessarily develop in the context of knowledge from another component.

The type of knowledge most commonly associated with an understanding of any quantitative concept is computational knowldge. For the weighted mean, there are three equivalent methods of computation:

1) the general mean formula $(X_1+X_1+..+X_1+X_2+..+X_2)/N$

2) formula for the weighted mean $(n_1X_1 +n_2X_2)/(n_1 + n_2)$

3) a combination of proportions $(n_1/N)X_1 + (n_2/N)X_2$.

The third type of calculation has not been commonly observed amongst subjects in previous studies. Therefore, it will be assumed that this is not an appropriate form of calculation for a novice. The correct use of the general formula for solving a weighted mean problem involves knowing either: 1)that the sum of scores for a group can be gotten by adding the mean n times or 2)that the mean is the best index of each single score in a set, allowing one to approximate each score in the set with the mean. The knowledge that it is not necessary to have every score in the set from which a mean was derived in order to combine two means is included in either element of knowledge. Element 1) contains the notion of reversibility discussed

earlier.

The second method of calculating the weighted mean also involves the idea of reversibility of the general mean formula: $\Sigma x/n$. The slight difference from method one is that the mean is multiplied by the number of cases, rather than being added n times. This method may be modified by using the relative proportion of cases rather than the absolute number. For example, in the GPA problem, one might multiply by one and two and divide by three, rather than two and four and divide by six.

Correct use of the computational formulas is generally used as an indicator of understanding. However, even correct calculation of the weighted mean does not necessarily indicate the subject will calculate the mean correctly in every case. The subject may understand the mean in a rote fashion, and fail in cases where it may be more difficult to determine how and whether the formula should be applied. Functional knowledge is necessary to decide whether the number calculated actually best represents a group of scores for a particular purpose. The surface structure of a problem may be a misleading indicator of the type of computation implied. One needs to be able to determine how the mean should reflect the situation presented despite the relative abstractness of the quantities involved or a misleading surface structure.

Functional knowledge includes knowing that all numbers entering an equation must have the same logical status with respect to that particular computation. When means based on different numbers of scores are combined in a simple mean calculation with the intention of finding a number representative of every score in each group, the means are not logically equivalent. Logical status may be provided for by counting each mean the appropriate number of times.

Traditionally, the least attention has been paid to analog types of knowledge. In its most general sense, analog knowledge involves a visual-kinesthetic image which is capable of being manipulated in solving a problem. In the case of the mean, the analog image should allow for reasonable predictions of the approximate value of a mean. The image of the mean as a balance point fulfills this requirement. The mean may be conceived of as the point where a set of numbers would balance if a weightless number line were placed on a balance beam. This is so because the mean is constructed such that the sum of the deviations about the mean is always zero, and thus a "balance point." In the case of the weighted mean, the balance point is closer to to group mean which contains a larger proportion of the scores.

In the balance beam image, as well as in computation, the blocks placed on the balance beam or the numbers

entering the equation must have logically equivalent status. Each block must represent the same number or proportion of scores. Therefore, in order to solve a weighted mean problem in analog fashion, the relative proportions of scores are represented by different numbers of blocks on each mean value.

The balance beam representation might be valuable for at least two reasons: 1)it lends itself to estimating or determining an answer which can be used as a check on the reasonableness of the computation and 2)it helps make clear the idea that it is the relative and not the absolute difference in numbers of scores which is important. One can calculate a mean using only proportions of scores.

The balance beam model only has real utility if the subject already has some conception of how weight and distance are related in balancing. The balance point between two sets of unequal numbers of scores is different from that of two equal sets, with the balance point of the unequal set being in a predictable direction from the equal set, given that one has a reasonable conception of balancing. In terms of Siegler's balance rule levels, possession of rule IV level knowledge would provide a basis for such a conception. It may also be the case that rule III level knowledge is sufficient for estimation, since rule III knowledge takes into account the interaction of

weight and distance, though not allowing exact estimation. For the purposes of the present study, rule IV or rule III level performance which is beyond chance will be considered sufficient indication of knowledge of balancing which might be required for analog-type knowledge.

In Skemp's terms, it is obvious that one does not need functional and analog knowledge to develop a merely instrumental understanding of the mean. Computational knowledge might be adequate in the majority of cases for determining even weighted means, since there are ways to express the computation as a formula. However, computational knowledge alone does not allow for flexible application of the concepts in cases where it might not be obvious which numbers are meant to have logically equivalent status, such as the non-example problem used in the Hardiman study. Functional and analog knowledge allow one to assess what number might best represent the set of scores and to approximate that number. Thus, these kinds of knowledge are employed in the interpretation of problem information and in assessing the reasonableness of answers. Necessarily, they would be involved in an relational understanding of the concept.

CHAPTER IV

TEXTBOOK TREATMENT OF THE MEAN

Suggestions were made in the previous chapter concerning the types of knowledge structures which might be involved in understanding the mean. How well is such information conveyed by standard introductory statistics textbooks on the undergraduate level? Seventeen currently used statistics texts were surveyed, examining treatment of the mean and related topics.

This survey describes whether a particular type of information was included in a text. No attempt was made to judge how well that information had been presented. Nineteen categories cover the complete range of topics which are presented in conjunction with the mean in all textbooks. The categories and data are presented in Appendix I.

The textbooks surveyed almost invariably their treatment of the mean by giving the expression, $\Sigma x/n$, or the general mathematical expression for calculating the mean. Generally it was expressed in English, then mathematical notation. The second, but less common topic was an explanation of how to calculate a mean from a frequency distribution. The topic seemed to be included at this point more

for its procedural than pedagogical utility. There was little concurrence in ordering beyond this point.

Several topics were found in a majority of the textbooks. These included: 1)the sum of the deviation scores from the mean is zero, 2)a comparison of the mean, median, and mode, and 3)the mean is sensitive to deviations in each score in the distribution. Some texts which discussed the deviation scores also stated that the mean could be considered the balance point of a set of scores. This the central idea in the type of analog knowledge being considered. These two arguments seem particularly relevant to the proposed framework of knowledge when presented together.

Several issues which may be considered pedagogically important even in a framework which does not include analog knowledge were discussed by relatively few of the texts. These included: 1)the development of a computational procedure for calculating the weighted mean, and 2)the reversibility of $\Sigma x/n=M$, namely $\Sigma x=nM$. A minority of the texts stated that the sum of squared deviations around the mean is a minimum, making it the best measure of central tendency.

Although a statement concerning the mean as a balance point is relevant in a discussion of analog knowledge, illustrations probably have a greater impact because the

concept is basically visual-kinesthetic. A number of texts, though not a majority, presented some type of illustration of a balance beam. Generally, only one situation was presented, commonly one with unequal weights at unequal distances on either side of the fulcrum. Two texts presented more than one picture of a balance beam with one score changed in each ilustration to indicate how the mean changes with small variations.

Only one text (e.g. Freedman, Pisani, and Purves, 1978) actually exploited the balance beam analogy by asking students to estimate the mean as a balance point. This would seem to be a particularly effective exercise for developing intuitions about the mean and the reasonableness of an answer.

How thoroughly did texts present analog, computational, and functional knowledge? A brief glance at Appendix I should suffice to indicate the relative lack of discussion concerning analog knowledge. Even in cases where some description of the mean as a balance point was presented, it was often not developed enough to make clear how the balance beam analogy works.

Surprizingly enough, computational knowledge also received sparse treatment in many texts. Although all texts presented sufficient information to enable a student to calculate the simple mean, many texts did not develop a

procedure for calculating the weighted mean. In the texts that did develop a procedure, the rationale was generally that the scores must be put on some logically equivalent basis when the number of scores the means based on differ. This explanation is based on both functional and computational knowledge. Few texts actually discussed the reversibility of the equation, $\Sigma x/n$. It seems important to at least develop this argument for obtaining the sum of scores for a group if a procedure for calculating the weighted mean is not actually presented.

In general, given the proposed framework, textbook treatment of the mean seemed less than adequate. No single textbook adequately discussed all important aspects of the mean. Most texts did not even provide an indepth treatment of a single type of knowledge, computational knowledge being the most obvious choice to present thoroughly. Given the results of this survey, it is not surprizing Pollatsek et.al. found that even some students who had studied statistics could not calculate the weighted mean.

CHAPTER V

THE PRESENT STUDY

Questions and Hypotheses

A framework has been proposed which suggests that knowledge which would enable one to calculate the mean correctly given any set of conditions is composed of three aspects:

1)computational knowledge- includes a computational formula for the weighted mean or a computational formula for the unweighted mean with provisions for obtaining the correct total. Reversibility from the average to the total is necessary in the latter case.

2)functional knowledge- an understanding of the mean as a real world concept. Includes the knowledge that the mean is in some sense the most representative of a set of scores, numbers entering an equation must be logically equivalent, and the mean must be within the range of the largest and smallest scores.

3)analog knowledge- the mean can be viewed as the balance point of a set of scores on a number line. The mean changes when any single number is changed. When means are combined, the overall mean is closer to the mean which is based on the greater proportion of scores.

A survey of introductory statistics textbooks indicated that these three types of information are not adequately presented in texts, and by extension, in many classrooms. Analog knowledge received particularly little attention. It may be that students are presumed to understand the mean before reaching the classroom, making any

treatment of the mean merely review. However, research indicates students do not invariably have this understanding (Pollatsek et.al., 1981).

When analog knowledge is discussed, the textbooks appear to assume implicitly that either the students already have an intuitively correct notion of balancing, or that knowledge of balancing is not necessary for calculational purposes. However, it seems unlikely that the balance beam model would be a useful tool for thinking about the mean if the mechanism of balancing was only vaguely understood. Knowledge of balancing may well be an important aspect of the enabling conditions for understanding the mean. Even if rule IV level balance knowledge was not essential for correct calculation, a concept of the mean which was based on many types of knowledge, including balance knowledge, might be more stable over a longer period of time.

It is possible to test whether balance knowledge is associated with an adequate computational understanding of the mean by testing level of balance rule knowledge with a Siegler-type test and computational knowledge with weighted mean problems. Given the proposed framework, one would predict that subjects who have a well developed knowledge of balancing should be more likely to be able to calculate the weighted mean in the more difficult cases. Thus, the

first question addressed in this study was descriptive and asked:

> Are students with higher levels of balance knowledge more successful in solving problems related to or involving the mean?

Given that level of balance knowledge predicted calculational performance, it would be of interest to determine whether the performance of subjects who were poor calulators could be improved by providing experiences which encouraged the development of balance rules. The second question asked was therefore prescriptive:

> Can the performance of students who had difficulty calculating the weighted mean be improved by providing experiences which foster development of balance knowledge?

The prediction was that subjects with balance training would perform better than control subjects on posttraining problems.

Proposed Methodology and Rationale

The goals of the present study were to determine whether balance knowledge is associated with a deeper understanding of the mean, and whether experience in balancing might aid novices in the development of their concept of the mean. The interest in the concept of the mean extended beyond identifying some set of training conditions which would enable a subject to compute a cor-

rect answer given a certain form of question to broader issues surrounding the formation and integration of knowledge structures and their interaction with performance. For this reason, a strictly standardized testing procedure was inappropriate. Standardized tests would not allow the interiewer to learn in depth about the reasoning involved with different answers that were given. However, it was also desirable to design a procedure in which the experimental groups were rigorously defined, yielding greater generality for the testing results. For these reasons, it seemed that a combination of clinical interviewing methods and paper and pencil questionaires was appropriate for the present study.

The interview is a valuable tool for investigating individual differences, and may be preferable to standardized tasks in certain cases (Pollatsek, Well, and Cobb, Note 9). Subjects may produce similar answers to the same question for very different reasons. Alternatively, different answers may be produced for similar reasons. In general, it is difficult to understand what kinds of reasoning have been used in answering questions and this becomes problematic when there is a great deal of variability in subjects' answers. It may be quite misleading to infer the thought processes of individuals from group measures.

It has been argued that the clinical interview does not provide a valid indication of the thought processes of the subject (c.f. Nisbett and Wilson, 1977). The subjects may report information incorrectly when asked to recall their steps to solution, and their reasoning patterns may be changed or influenced by the request to verbalize. Ericssonand Simon (1980) developed a model which made predictions regarding the validity of subject reports and tested the predictions using studies which involved interview data and a second measure of performance. They concluded that distortion occurs in verbal reports when information is elicited retrospectively, when the probes are too general to elicit the information sought, and when sujects are able to use inference processes to fill in the information requested. However, if subjects are verbally relating reasoning processes as they are occurring, i.e. while the information is in a short term store, verbal reports are an accurate indicator of problem solving processes. In fact, Morris (1981) has argued that the use of verbal reports of strategy offer the only hope of gaining some understanding of real life behavior.

In the present study, questionaire data was used to obtain an indication of subjects' knowledge concerning the mean and balancing prior to the interview in order to form experimental groups. The interview was used to yield both

a gross measure of performance and analysis of individual reasoning patterns.

CHAPTER VI

METHODS

Participants

Subjects. Seventy-three students from the University of Massachusetts participated in the study. Forty-eightwere included in the final analysis. Sixteen subjects failed to return to complete the second session, four subjects were eliminated after failure to complete the training session within a certain number of trials, one subject was dropped because of technical difficulties, and four subjects served as pilots.

Subjects were given 3 units of experimental credit or a combination of 1 or 2 credits and $3.00 an hour in exchange for 3 hours of participation. The sign-up sheet stated that the study concerned problem solving and involved a written group pretest and an individual interview that would be videotaped. The only restriction placed on participation was that students not be concurrently enrolled in a statistics course.

In order to obtain a sample of students representative of those likely to enroll in or to have completed a course in statistics, subjects were recruited from psychology

classes designed for majors. Twenty-five of the 48 actually were psychology majors. Of the remainder, 8 were in other social science or human service majors, 8 were in biological or engineering sciences, and 7 had not declared a major. Age ranged from 17 to 36 years, with a mean of 20.5 years. Thirty were females and 18 were males.

Assignment to groups was based on the results of a written pretest, which tested calculational and balancing skills. Nonbalancing subjects were then randomly assigned to one of two treatment groups, one of which received training in balancing and a control group which received an unrelated problem.

Investigators. Three persons were involved in the administering, interviewing, and analysis phases of the study. The first was G.S., a 23 year old female with an undergraduate degree in psychology who administered and scored the written pretest, assigned subjects to conditions, and assisted in the coding of data. Her previous experience included an independent interview study and written test administration.

The second person, P.H., was a 24 year old female graduate student in cognitive psychology. She developed the questions, interviewed the subjects during training and transfer sessions, developed the coding schemes, and coded

the data. P.H.'s previous experience included 3 interview studies independently developed and analyzed.

A.W., a male cognitive psychology professor, assisted in the development of the questions and the coding scheme, and coded data. He had been involved with the development and analysis of several previous interview studies.

Previous contact with the students who participated in the study was minimal.

Problems

The written pretest. The pretest was administered to assess the individual's knowledge of calculation and balancing prior to placement in experimental conditions. The test included two weighted mean problems intended to provide the basis for classifying each subject as a calculator or noncalculator . Balance knowledge was assessed with a series of 12 written problems (See Appendix II for balance problems). In addition, 3 problems involving means and ratios were included to provide supplementary information concerning calcuational skills. A Bayesian problem was also included as filler material. The written problems (excepting the balance problems) are presented in Table 1 in the order given.

Table 1: Written Pretest Problems

1) Judgement task: Could the final number presented possibly be a mean of the set?

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

2) Two boats of fishermen return from a weekend fishing trip. The four people on the first boat average 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?

3) Diane and Jenny want to knit identical scarves for the winter. They decide to get together and knit for the evening. When they begin the evening, each has already completed some of her scarf: Jenny has completed 2/3, while Diane has completed 1/3. For every inch that Jenny knits, Diane knits two. If they knit the same length of time and Jenny has completed her scarf by the end of the session, how much has Diane knit?

4) What is the average of the following numbers?

10.2 15.3 9.7 11.0 12.6

5) A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city . You are given the following data:

i) Although the two companies are roughly equal in size, 85% of the accidents in the city involve Green cabs and 15% involve Blue cabs.

ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identification in 80% of the cases and erred in 20% of the cases.

Question: What is the probability the cab involved in the accident was Blue rather than Green? (Please express the answer as a percentage.)

(Pretest continued on next page.)

6) 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?

6) Balance problems. See Appendix II.

The two weighted mean problems varied in difficulty. They were listed in counterbalanced order in the text (i.e. problem 2 was listed as the sixth problem in half the questionaires). Subjects were classified as calculators if they solved both weighted mean problems correctly.

Balance knowledge was assessed using problems based on the Siegler paradigm for assessing balance knowledge in children (1976). The Siegler procedure consists of the presentation of a balance beam in one of six possible problem states. Subjects are asked to predict whether the balance beam will balance or in which direction it will tip. The problem states are presented in Table 2.

The first three problem states required no arithmetic calculation for correct prediction. One only had to compare the relevant dimension, since the sides of the balance beam differ on only one dimension. In the remaining three problem types, both weight and distance were varied. This results in a conflict situation in which it is necessary to calculate the torques to obtain a correct prediction on

every trial.

Table 2: Balance beam problem states

Weight- unequal amounts of weight equidistant from the fulcrum

Distance- equal amounts of weight at unequal distances from the fulcrum

Balance - equal amounts of weight at equal distances from the fulcrum

Conflict-weight- the side with the greater weight will drop

Conflict-distance- the side with the greater distance will drop

Conflict-balance- the beam will balance

In the present study, the problems were presented as illustrations on paper, rather than on an actual balance beam in individual sessions. In neither the Siegler paradigm nor in the present study did the subject receive feedback about the correctness of the prediction. Therefore, the use of illustrations was preferable in terms of ease of administration and appropriate for an adult sample.

The subject's rule level of balance beam knowledge was determined by comparing the subject's answers for problems of each type to the Siegler predictions for percentage correct of each problem type for each rule level (Siegler, 1976, pp.486). (See Table 3 for list of predictions.)

Table 3: Predictions for Percentage Correct for Rule Levels I to IV.

Problem type	Rule level I	II	III	IV
Balance	100	100	100	100
Weight	100	100	100	100
Distance	0	100	100	100
Conflict-Weight	100	100	33	100
Conflict-Distance	0	0	33	100
Conflict-Balance	0	0	33	100

One example was given of each of the three simple problems, weight, distance, and balance. Three examples were given of each of the conflict situations, since there is a 33% chance of getting a single conflict problem correct using the rule III guess strategy. Subjects were considered balancers if they respond correctly on all the simple problems and at least two-thirds of the conflict problems. All other subjects were considered non-balancers.

The pretest also included one simple mean problem designed to assess basic calculational skills. The problem asked for the mean of five numbers. Any subject unable to solve the simple mean problem would have been eliminated from the study. Subjects were also presented with sets of numbers and possible means for each set and asked to determine whether the number given as the mean was plausible for that set of numbers. The four sets of numbers were con-

structed with different possible errors in mind. The mean for set (a) is near the low end of a range of numbers with no obvious modal points. The (b) mean is not contained within the set of numbers given. The mean for set (c) is approximately correct. Set (d) is bimodal, and contains more instances in one of the modes. The mean given is exactly in the middle of the two modes when it should be closer to the more heavily weighted end.

The final problem involved ratios. This problem was included to assess whether capability to manipulate ratios successfully enabled students to solve weighted mean problems involving proportions more easily.

The training phase. Subjects who returned for a second session were assigned to either a training or control condition which took approximately one half hour. The goal specified for the subject in the training phase was to determine how to predict correctly the actions of a balance beam after the interviewer had set up problems.

Accordingly, a balance beam with a continuous scale and a set of wooden blocks were used. The balance beam consisted of a rigid metal bar attached to a fulcrum at the midpoint of the bar. Weights were attached to the underside of the bar to make the system self-righting. A lightweight plastic scale with marks approximately two

inches apart was centered on top of the balance beam to indicate distance from the fulcrum. The subject was asked to predict the actions of the balance beam in a set of increasingly complex problems until the goal of five consecutive correct predictions was acheived.

During the latter part of the training phase after the first goal was reached, the task was slightly changed. After making a prediction, the subject had to state in which direction the unbalanced system should be moved in order to balance it again. This was done by sliding the plastic scale a moderate distance along the metal bar, placing a different point at the fulcrum. A second plastic scale which had been numbered with a continuous number line was used in this portion of the task in order to acquaint subjects with the notion of a shifting number line. This final system approximated the ideal of a weightless number line placed on a balance beam.

Control problems were given to subjects who were not trained in order to equate experience in the interview situation. They involved slightly less time than the training session. Subjects were interviewed on two versions of the Bayesian problem listed in the pretest.

The transfer phase. In the final phase of the study, subjects were given a number of transfer problems to assess

their understanding of the concept of the mean. These problems are presented in Table 4. The subject was asked to represent the problems on the balance beam and then calculate the answer if it was not done during the course of representing. The same balance beam described in the training phase was used for this task.

In order to represent a simple mean problem on the balance beam, one must first make a number line containing the set of numbers involved in the problem on the plastic scale. A block is placed on each number being averaged and signifies one instance of the number below it is involved in the calculation. The number line is shifted along the bar until the system balances, and the point which is at the fulcrum is the mean of the set of numbers. The process is similar for weighted mean problems. In these problems there is more than one instance of a single score. This is representing by placing an appropriate number of blocks on the number. (See Figure 1 for illustration.)

The first four problems were to be represented on the balance beam and were presented in the order given. Problem 1 introduced the notion of a mean as a balance pointand asked subjects to represent a simple mean. Subjectswere given a plastic scale marked from 1 to 12 for this problem. The solution was to place one block on 3, one on 10, and slide the scale to 6.5. All later problems

require the subject to label their own number line, given a blank scale.

Table 4: Transfer Task Problems

1) It is possible to view the mean of a set of numbers as the point at which the number line containing the set of would balance if it were placed on a balance beam. You 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 spaced 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 each person in the shop made. 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?

The second question is a weighted mean problem which

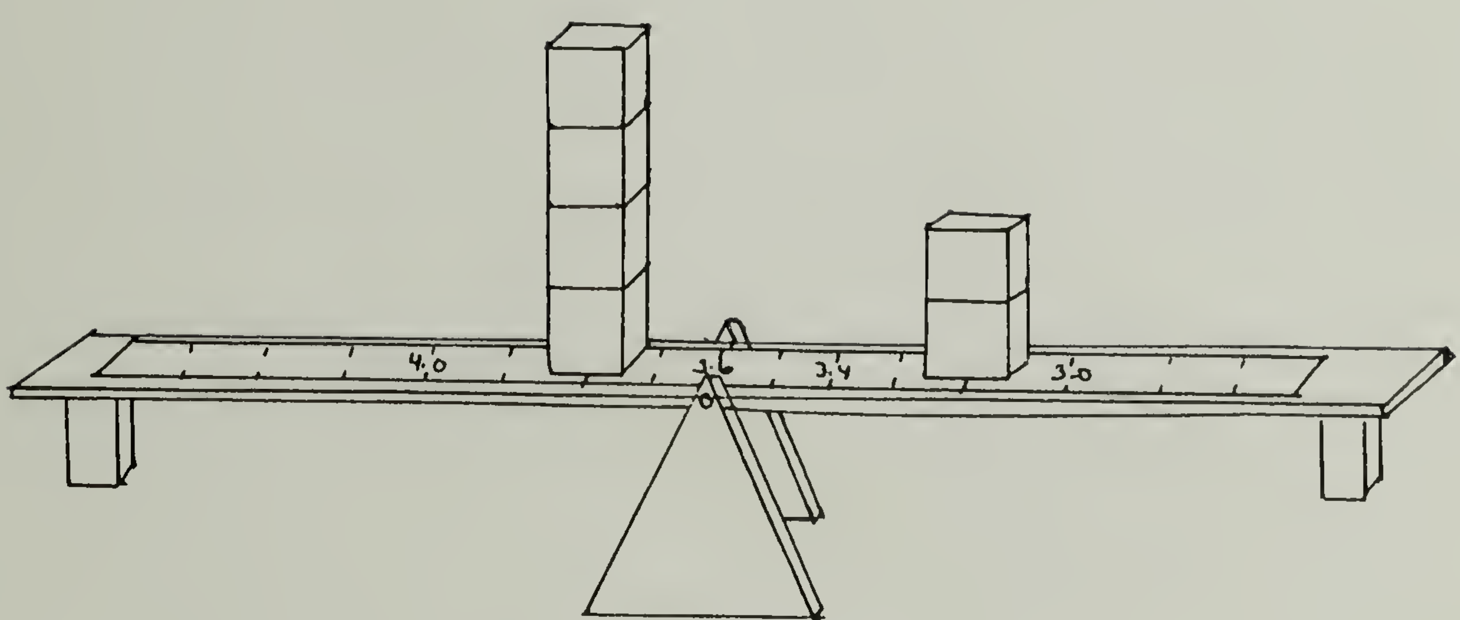

Figure 1: Illustration of the balance beam used in both the training and transfer tasks. A weighted mean is represented on the beam.

had been used in previous studies involving weighted means. The problem is fairly difficult and seems to discriminate subjects with different levels of understanding the weighted mean. The question actually has two correct methods of representation: 1)two blocks on 3.2 and four blocks on 3.8, representing two and four semesters respectively, or 2)one block on 3.2 and two blocks on 3.8 to indicate that twice as much time was spent earning the 3.8 GPA. These have been termed weighted and weighted proportion, respectively.

The second weighted mean problem is a variation on a problem used in previous studies which is stated in terms of proportions rather than absolute numbers. Pilot work indicated subjects found proportional problems to be quite difficult. To answer this problem correctly, one must first determine the proportion of women, which is 2/5. For a small number of subjects this is a non-trivial task. From there, the problem may be represented on the balance beam in three different ways: 1)two blocks on 120 and three blocks on 180, representing 2/5's and 3/5's of the people, 2)concretizing the number of people such that two blocks on 120 represent two women, or 3)deciding that the number of people might be a multiple of five, so that for example, four and six blocks are needed to represent the numbers of women and men. These methods of representing are termed

weighted proportion, weighted concrete, and weighted multiple.

The final problem to be represented on the balance beam was a simple mean problem with the surface structure of a weighted mean problem. Subjects with little understanding of the properties of the mean would simply assume this was a simple mean problem if it was presented first. Later presentation allows one to distinguish those subjects who have acquired the concept of weighting and learned to use it properly from those who have developed an algorithm without a sense of the purpose of weighting. The final problem confronted subjects directly with the choice between the calcualtion methods for the simple mean and the weighted mean. It is possible that some subjects may remain at this point who have not yet seen an appropriate calculation for the mean, so the choice is not merely amongst known formulations. The second or weighted method is correct for this problem. After answering the question, subjects are probed with the following: "Do you think the first method would give a higher or lower number than the second?" The first method would yield the higher answer, since it does not take into account the large proportion of lower salaries.

Procedure

The written pretest. The initial phase of the study was conducted in small groups and administered by G.S. Subjects were given a written description of the study and its objectives as part of the informed consent form, which they were required to read and sign. Demographic charateristics and mathematics experience were also requested at this time. Subjects were then given the judgment task. Sets of numbers on index cards were shown to subjects one at a time. After all the cards in a set were shown, subjects were asked to judge whether the final number in a set could possibly be the mean or not.

The remaining problems were presented in booklet format, with space on each page for calculations. Subjects were asked to write all the steps to their calculations as clearly as possible. No time limit was given and subjects were allowed to do the problems in any order. All subjects finished within the allotted time of one hour.

Subjects signed up for an individual return time after completing the pretest.

Initial assignement to groups was based on the assessment of calculational and balancing skills derived from the written pretest. Classification as a calculator/ noncalculator and balancer/nonbalancer yielded four groups, with different numbers of subjects in each group: 1)10

calculators and balancers (**CB**), 2)2 noncalculators and balancers (**cB**), 3)14 calculators and nonbalancers (**Cb**), an 4)22 noncalculators and nonbalancers (**cb**). Subjects within the **Cb** and **cb** groups were then randomly assigned to training and control conditions. The result was six groups: 1)10 **CB**, 2)2 **cB**, 3)6 trained **Cb**, 4)8 control **Cb**, 5)12 trained **cb**, and 6)10 control **cb**. The numbers of subjects who participated in training and control conditions were not equal within the **Cb** and **cb** groups, due to failures to return to complete the later session.

<u>The training phase</u>. The training or control session and subsequent transfer tasks took place in individual sessions with P.H. The interviewer sat diagonally to the subject's left. Recording equipment was in full view of the subject and manipulated by the interviewer during the session. Subjects were told that the session would be videotaped at initial contact and before the training session actually began.

For the training session, the balance beam was placed between the interviewer and the subject. The subject was told:

"I will place blocks on the balance beam. I would like you to predict what will happen each time, whether the balance will tip to your right, your left, or balance. Please explain why you think that will happen. There is a system of rules which would let you predict correctly what will happen each time. I'd like you to be thinking about what those rules might be and when you have any ideas about

it let me know. We will work with this for awhile, take a break and come back and do something slightly different." On each trial, the interviewer set up the problem on the balance beam, holding the balance arm until the subject made a prediction. The was no time limit on responses. The balance was released and the subject ascertained whether their prediction was correct or not. Subjects were encouraged to relate whether the action of the balance beam changed the current hypothesis.

The problems were ordered in complexity to facilitate discovery of the simple relationships which are components of the Rule IV approach. The initial four items illustrated simple balance, weight, and distance relationships. The next twenty items illustrated how simple balance arrangements can be created when weight and distance cues conflict. For example, on item 5, two blocks were on the second mark to the right of the fulcrum and two blocks were on the fourth mark on the left. Weights were added to the right side at distance two until the balance point was reached and exceeded. Balance was restored, then one block was taken off each side to demonstrate a lack of proportionality and inequality. Another block was taken off the right side, restoring the system to balance. After presenting three such sequences (the complete sequences are listed in Appendix II)the balanced arrangements were represented to focus the subject's attention on the commona-

lities. The remainder of the problems were all high difficulty, with more than one group of blocks on a side. These required the calculation of torques to solve them correctly each time. Subjects were given high difficulty problems until they were able to predict correctly for five consecutive trials. If a subject did not develop any rule governed behavior after several of the high difficulty problems were presented, simpler sequences were repeated.

Control subjects were told they would be solving some written problems. They were instructed to read the problem aloud and to explain what they were doing, why they were doing it, and to relate any ideas they might have that were related to the problem. They were informed there would be a break after the first few problems and that the later problems would be slightly different.

No time limit was set for solution of control problems. The only requirement was that subjects feel satisfied with their answer, even if it was not correct.

The transfer task. When the subjects returned from a five minute break, they were seated in the same position with the balance scale directly in front of them. They were again instructed to say what they were doing and why. The problems were presented in the order previously described. The subject was requested to read each problem aloud and to

represent the problem on the balance beam. If the subject did not calculate the mean during the course of representing the problem, they were asked to do so when they had finished representing it.

The interview began as a tutorial type (c.f. Konold and Well,Note 7), since it was important for subjects to represent the first simple mean question correctly in order to be able to represent later weighted mean problems. Subjects began by considering represenational strategies they developed independently. If these failed, the interviewer asked them to consider other possible ways. If the subject was unable to determine how to represent the problem after approximately ten minutes, they were shown how to represent the problem and given a second problem to represent.

All later problems were conducted in a thinking-aloud style. Probes were used only to encourage the subject to verbalize more or explain a particular action more fully. If the subject's representation did not match the calculation, the subject was encouraged to make them match.

After the subjects had gone through all the problems, they were asked to reexamine the questions, decide how confident they were about their answers, and to change any they did not like. If the subject did not wish to change any answers, but had answered at least one question incor-

rectly, the interviewer switched to an indepth style. The standard probe was generally, "Does the number of scores make a difference?" If the subject then was dissatisfied with the original answer, the interview proceeded until the subject found an answer that he or she found to be convincing even if it was not necessarily correct.

At the conclusion of the interview, the hypotheses and goals of the study were explained in more detail to the subjects. If there were any questions about the problems, they were answered at this time.

Analysis of Interviews

The interview problems were coded for analysis in two ways: 1)correctness of calculation and representation, and 2)evidence of strategies. The latter category included order of calculating and representing, type of calculation or representation, labeling of scale, and type of rationale given for the weight and the shop problem. An example of the coding sheet is presented in Appendix II.

The three investigators independently coded three interviews to obtain a measure of reliability. Agreement on correctness of the answer was 100%. An index of interjudge agreement was obtained for type of calculation and type of representation using a measure developed by Cohen (1960, as cited by Hayes;1981). The average interjudge

agreement was .83 for type of calculation and .83 for type of representation. These indices actually underestimate the true amount of interjudge agreement. The raters agreed about the major categorization of type: simple, weighted, or reverse proportion. Disagreements concerned the type of representation within a major category. However, the measure was calculated such that all categorizations received equal weight. The actual measures of agreement are probably higher. within the major categorization of type, rather than across categorizations of simple and weighted.

CHAPTER VII

RESULTS

The results of the study will be discussed in sections corresponding to: 1)the written pretest, 2)the training session, and 3)the transfer session. Two sets of numbers are given in the pretest results. The first is the set of subjects who participated in both sessions and are included in all analyses, while the second is all subjects who took the pretest.

The Written Pretest

Twenty-four (31) subjects were classified as calculators and 24 (48) as noncalculators. This classification was based on performance on the two weighted mean problems, the Fish problem (2) and the Income Index problem (6). Thirty-seven (53) of the 48 (81) subjects solved the Fish problem correctly. Eight (11) additional subjects responded with the total number of fish, either neglecting to divide or misinterpreting the problem as asking for the total number of fish. This answer was considered correct, since the problem is worded in such a way as to make this misinterpretation probable. Five (11) subjects responded with

inappropriate answers: one (2)responded with the simple mean, two (3) with the simple sum divided by 6, two (3) tried to weight but were not successful, and (2) divided the total number of fish by 2.

There was considerably more variablity in performance on the Income Index problem. Only 24 (31) subjects provided a correct weighted mean solution. Sixteen (24) of the remaining 24 (50) subjects responded with the simple mean. A variation of the simple mean, the simple sum, occurred once (4). Eight (15) subjects unsuccessfully attempted to weight the problem. Five of these subjects responded with a particular type of error which has not been previously documented. This so-called reverse proportion error involves correctly deciding that the individual sample means should be weighted, but incorrectly assigns the weights. The number which accounts for the smaller number of scores is doubled while the number accounting for the larger number of scores is held constant. The results are added and divided by 3. Hence, the wrong number assumes more weight in the calculation.

Twelve (15) subjects were classified as balancers with Rule IV level balance knowledge, while 36 (64) subjects were nonbalancers. Of the nonbalancers, 6 (11) were classified as being at rule level I, 16 (25) were at level II, and 14 (23) were at level III.

The correlation between balance level and number of correct weighted mean calculations was .50 ($t(46)=4.80$, $p<.001$) ($r=.35$, $t(79)=3.355$, $p<.001$). This suggests an association between these concepts. It is also possible that the relationship may be due to other factors, such as appropriate background or general mathematical facility. However, the suggestion of a relation implies an experiment is required to determine whether, in fact, a knowledge of balancing aids in calculational performance.

P(calculator/balancer) was .83 (.87) versus p(calculator/nonbalancer)=.39 (.28). There was no significant difference in rule level between **Cb** (2.5) and **cb** (2.1) subjects ($t(61)=1.616$, $.10<p<.20$), although **Cb** subjects had a slightly higher mean rule level. Thus balance rule level predicted calculational level, but calculational level was a poor predicter of balance skill.

The simple mean calculation (problem 4) was solved correctly by the majority of the subjects. Five(7) persons answered the question incorrectly because they made arithmetic mistakes. Any subject who showed evidence of not knowing each step of the algorithm would have been eliminated. No subjects were eliminated for such reasons.

The errors that were made on the simple mean problem were not confined to any particular group. They were made by calculators and noncalculators. The difficulties most

subjects have in solving the weighted mean problems do not appear to be due to an inability to set up a mean calculation. Problems are more likely to be on the deeper level of interpretation of the calculation.

The ratio problem (problem 3) was solved correctly by fewer people than the simple mean problem, with ten(18) subjects failing to obtain the correct answer. Performance on this problem seemed to be related to calculational and balance levels, with 8 of the 10 (14 of 18) failures belonging to the **cb** group. It was difficult to determine from the written pretest what the source of the errors might be, as not all subjects wrote out their calculations. Some subjects may have merely taken a guess. However, others performed fairly elaborate calculations which seemed to indicate they had been able to determine the relevant information in the problem, but were unable to organize that information correctly.

Differential performance between groups was also found on the judgment task (problem 1). Overall, **cb** subjects made the greatest number of errors, followed by **Cb** subjects. Only one(2) subject in the **CB** group made an error. The majority of the errors made by the **Cb** and **cb** groups were on the fourth set of numbers, the weighted distribution. Thiry-six percent of the subjects in each group (.26 **Cb** .48**cb**) incorrectly judged the midpoint of the distri-

bution to be the mean of the set. Nonbalancers had more difficulty making a judgment involving principles of balancing than did balancers in a situation very different from the balance beam.

The second most common error was made on the first set of numbers, in which the potential mean was within the range given, but was too low. The principle of balancing may be invoked to help with this judgment. It was less necessary in ths set, though, since most of the numbers were larger. **Cb** and **cb** subjects also erred on the second and third sets, but with less frequency.

The results of the pretest are consistent with the idea that there is a relationship between understanding the mean and understanding how objects balance on a balance beam. Most subjects who performed well on the balance beam task were also able to calculate the weighted mean problems correctly. The judgment task provides supportive evidence suggesting a connection between balance knowledge and the mean within a purely numerical context. The stimuli are simply sets of numbers, and no interpretation is necessary to determine how a calculation should be organized. Yet, subjects with less knowledge of balancing are more apt to make errors in judgment, regardless of their calculational performance. The large number of successful solutions on the simple mean problem and the nature of the errors sug-

gest that difficulties with the weighted mean are conceptual in nature, and not due to a simple lack of understanding how a mean is calculated. Thus, an experiment to determine whether training in the concept of balancing will increase successful performance on weighted mean problems was motivated.

The Training Session

The average number of trials to criterion was 56.8 ($s=20.6$, range = 27 to 102). It was decided to eliminate the data of four subjects who failed to reach criterion within 80 trials from further analysis because it was unlikely they understood the mechanism of the balance beam well enough to incorporate the ideas of balancing into a calculational framework. After eliminating these subjects, the mean number of trials to criterion was 49.2 ($s=12.9$, range = 27 to 68).

The number of trials required to reach criterion was 46.9 ($s=12.1$, range =29 to 64) for **cb** subjects and 53.8 ($s=14.4$, range = 27 to 68) for **Cb** subjects ($t(16)=3.85$, $p<.002$). This difference was unexpected. If any prediction were to be made, it might have been expected that **Cb** subjects would require fewer trials to criterion. Their better calculational performance might be interpreted as

suggesting greater facility with mathematical ideas, which would lead to a prediction of faster acquisition. Obviously, one cannot make such a conclusion.

In the remaining analyses, it will be assumed that there is no difference in balance level knowledge after training for all trained subjects.

The Transfer Session

The simple mean example. The simple mean example introduced the notion of the mean as a balance pointand asked subjects to represent the mean of two numbers. Eighteen of the 48 subjects spontaneously represented the simple mean correctly. The remainder of the subjects needed some interviewer intervention to represent the problem adequately.

Subjects in the training groups had received considerable experience with the balance beam, but there was no mention of the mean during balance training. There was no significant association between whether or not a subject represented the problem spontaneously and group ($X^2(5)=5.33$, $.25<p<.50$). There was no strong link between calculational level, balance rule level, and training and ability to determine how to represent a simple mean problem on the balance beam. Some control **cb** subjects found the

problem quite simple, while some **Cb** subjects had considerable difficulty. The following is an excerpt from a control **cb** subject who quickly solved the problem:

S First you take the block and put it where you think it is (puts one each on 3 and 10). Then I have to tell what the mean is. Does this move (referring to the plastic scale)?
I Yes.
S (Slides plastic scale to 6.5)

For this subject, the information in the problem was represented and explained adequately. In contrast, a protocol from a trained **Cb** subject shows several typical errors:

S I should put a block on the number line and move the line. I'll just do that (puts one block on 6.5 at the center). Do you want me to do 3 and 10 and then balance it?
I You have to show the numbers you are getting the mean of.
S (Places one block 3 spaces from the center, other 10 spaces from center). Then balance it? (Moves scale). I'm not sure if that's what you want. I just shifted what the values are so it's not 3 and 10 (by moving scale to balance). I used the number line to balance the equation. I'm not sure what the question is asking.
I You want to end up with the mean at the center.
S I'm still not sure I follow you.
I What is the mean of 3 and 10?
S 6.5.
I That should be at the fulcrum.
S And these blocks are supposed to represent the numbers 3 and 10? I'm having trouble seeing how. (S tries to put blocks 3 and 10 spaces from the fulcrum again).
I That doesn't seem to be working.
S Could I make the weights arbitrary, as long as it balances to 6.5?
I Show me.
S What if I do this? I have them at 3 and 10.

The subject's first error was to place a block on 6.5. One-third of the prompted subjects did this. They seemed to want to mark the mean both with the block and as a

balance point. A second common error was to place blocks 3 and 10 spaces from the fulcrum to indicate the numbers being averaged were 3 and 10. Seven subjects did this. A number of these same subjects then placed more blocks on one side of the beam to balance the system, without regard for what was being represented. A similar error was to use 3 and 10 blocks spaced along the beam in a configuration which made them balance (8). Less common errors were to place the blocks 6.5 spaces on either side of the fulcrum (2), place blocks at equal and arbitrary distances from the fulcrum (3), move a block to make the system balance rather than the scale (1), and place one block on 1.5 and the other on 5 (2). The variety and frequency of these errors indicates the balance beam representation does not follow naturally from the experiences of most of the subjects. The process was learned during the course of the interview session.

Weighted mean problems (problems 2 ± 3). The subjects were given the GPA and Elevator weight mean problems and were asked to represent each problem on the balance beam. If they did not calculate the mean during the course of representing the problem, they were asked to do so after they had finished representing to insure that the calculation matched the representation.

Subjects' initial lack of facility with the notion of balance beam representation was reflected in the order in which they chose to calculate and represent the weighted mean problems. Subjects were asked only to represent the problem, but a majority of the subjects calculated before representing each of the two problems on the balance beam, Differences between groups were not very systematic. **CB** and control **cb** subjects showed a slightly greater tendency to represent the problems before calculating, possibly reflecting a high degree of confidence in the method used. Several subjects calculated first but seemed to rely on the beam for confirming their answers. Seven subjects changed a calculation after balancing.

A scoring system was devised in which one point was given for correct final performance on each of the two problems, forming an overall calculation score with a possible range from 0 to 2. The overall calculation scores of nonbalancers were analyzed in a 2 (calculational ability: **Cb** or **cb**) by 2 (treatment:balance training or control) unweighted means analysis of variance. There were significant main effects due to calculational ability ($F(1,32)=18.82$, $p<.001$) and treatment ($F(1,32)=8.64$, $p<.01$). In general, scores of **Cb** subjects were higher than **cb** subjects. (1) However, performance of trained **cb** subjects was significantly better than that of control **cb**

subjects and not significantly different from trained **Cb** subjects. There were no significant differences between training and control **Cb** subjects (See Table 5 for mean scores of nonbalancer groups). Performance for the majority of **Cb** subjects was at ceiling level, as would be expected on a test of calculation given the pretest results. Thus, training on the balance beam appears to be associated with increased calculation performance by noncalculators on weighted mean problems.

Table 5: Mean overall calculation scores for nonbalancers. Standard deviations in parentheses.

	Cb	**cb**
Trained	2.0 (0)	1.5 (.67)
Control	1.8 (.46)	0.7 (.48)

A similar overall measure was developed for performance on representation. One point was given for correct representation on each of the weighted mean problems, for a possible score ranging from 0 to 2. An unweighted means 2 by 2 analysis of variance revealed significant effects for calculational ability ($F(1,32)=8.58$, $p<.01$) and treatment ($F(1,32)=6.64$, $p<.025$). Scores of **Cb** subjects were higher than **cb** subjects. Trained **cb** subjects represented more problems correctly than control **cb** subjects, and performed

similarly to trained **Cb** subjects, suggesting that training had an effect on correctly representing problems. There was no significant difference between training and control **Cb** subjects: performance was nearly at ceiling for all subjects. Means and standard deviations for representation performance of nonbalancers are presented in Table 6.

Table 6: Mean overall representation scores for nonbalancers. Standard deviations in parentheses.

	Cb	**cb**
Trained	2.0 (0)	1.4 (.79)
Control	1.5 (.76)	0.7 (.67)

All control **CB** subjects represented and calculated both weighted mean problems correctly.

In addition to the kind of data already presented, it seems informative to examine whether the subject considered other types of calculations and whether or not the subject seemed satisfied with the answer given. A measure called "satisfaction" was developed to provide some indication of the subject's problem solving flexibilty. Satisfaction scores ranged from 1 to 5. A score of 1 indicates the subject solved the problem incorrectly and is satisfied, 2- the problem is solved incorrectly, but the subject did not like the answer, 3- the subject considered two methods of

solution and chose the wrong one, 4- the subject considered two methods of solution and chose the right one, and 5- the subject only considered the right method. Mean satisfaction scores for each group are presented in Table 7. Nearly all calculators obtained the correct answer and considered only one method of calculation. Noncalculators were frequently incorrect or considered more than one way of solving the problem even when they later chose the correct answer. Satisfaction scores showed a general increase from the first to the second problem, reflecting a larger number of correct answers. Trained **cb** subjects had higher mean levels of satisfaction than control **cb** subjects on both problems. On the second weighted mean problem, the scores of trained **cb** subjects are similar to the scores of calculators.

Table 7: Mean Satisfaction Scores
(SD in parentheses)

Group	Problem 2	Problem 3
CB	5.0(0)	5.0(0)
cB	5.0(0)	5.0(0)
TCb	4.8(.41)	4.8(.41)
CCb	3.9(1.8)	4.9(.35)
Tcb	3.1(1.8)	4.5(1.2)
Ccb	1.7(.9)	3.4(1.7)

The types of calculations observed in the study fell into a limited number of categories. These include several correct, as well as several incorrect types. Types of calculations are presented in Table 8.

Table 8: Types of calculations, characterized by specific statements made by subjects.

Correct

1) weighted "3.2 times 2 equals 6.4. 3.8 times 4 equals 15.2. 21.6 divided by 6 is 3.6"

2) weighted concrete (Elevator only) "180 times 3 equals 540 plus 3 times 120 equals 780. This is the total weight. 780 divided by the total number of people, which is 5. The average weight would be 156."

3) weighted proportional "I have to weight the 3.8 because he went to that one 4 semesters and the other one 2, so I guess I weight that twice. 7.6 plus 3.2 equals 10.8, divided by 3 equals 3.6."

4) weighted multiple (Elevator only) "I'm going to say there are 15 people. Nine are men and 6 are women. 180 times 9 plus 120 times 6. Divide that by 15 and see if it comes out with something that looks decent."

Incorrect

5) simple "I just added the two numbers and divided by 2."

6) simple (a-b)/2 "60 pounds between 180 and 120. Add 30 to 120. Subtract 30 from 180."

7) simple sum "120 plus 180 equals 300 divided by 5. That would give me 60."

8) reverse proportion "Since he attended the first for 2 semesters and the second for 4 semesters, I'll just double the 3.2. 3.2 plus 3.2 plus 3.8 equals 10.2, divided by 3 equals 3.4."

The frequencies for each type of calculation on each of the two weighted mean problems for each groups are presented in Table 9.

Table 9: Calculation strategy on each weighted mean problem for subjects in each group.

Group	Calculation type 1	2	3	4	5	6	7	8
CB n=10	8(3) 4	 4	3 2					
cB n=2	1 2		1					
T Cb n=6	5 1	 4	1 1		1(1) 1(w)			
C Cb n=8	4 2	 3	2 1	 2	3(1)			
T cb n=12	4(3) 5	 4	4 2		7 (1,3, 3) 4(2w)	1(5)		1
C cb n=10	2(5) 6				9(1) 3	 1	1 2(1,1)	
problem sub-totals	24 20	 15	11 6	 2	20 8	2 1	 2	1
totals	44	15	17	2	28	3	2	1

Notes: 1) GPA is top number, Elevator is bottom.
2) A single subject can be represented twice. Numbers in parentheses are the strategies a subject switched to.

All balancers (**CB** and **cB**) chose some type of weighted

calculation. There was little group concurrence over which is the preferred type of weighted calculation. Certainly, there was much group flexibility in choice of calculation. Most Cb subjects also choses weighted mean calculations, although some considered the simple mean and two actually kept the simple mean as their answer to the GPA problem. Again,therewas a wide range of acceptable weighted mean calculations exhibited. cb subjects considered the simple mean most frequently as a whole group. However, trained cb subjects were much more likely to change their answer while solving the problem. Trained subjects also displayed several types of weighted mean calculations, whereas control subjects only performed the basic weighted calculation. Thus, the trained subjects displayed considerably more flexibility in their choice of calculation.

Strategies for representing problems on the balance beam were similar to calculational strategies, with a few exceptions peculiar to the medium involved. No subject tried to represent a simple sum, but some did try to represent the subtotal of scores for each group. Although it would be a simple task to represent a weighted multiple on the balance beam, no subject did this either. Types of representational strategies are presented in Table 10.

Table 10: Types of Representations, as characterized by specfic statements made by subjects.

<u>Correct</u>

1)<u>weighted</u> "Each block will represent a semester. It would be 2 at 3.2 and then 4 weights at 3.8. Obviously it will be weighted to 3.8. I'd say it's approximatley 3.59."

2)<u>weighted concrete</u> (Elevator only) "If 3/5 were men, I'd say there were 5 people and 3 were men. You put 2 on 120 and 3 on 180 and the mean is 156."

3)<u>weighted proportion</u> "I'll put 2 on 3.2. Well, instead of 2 and 4 you could do 1 and 2. It's the same ratio."

<u>Incorrect</u>

4)<u>simple</u> "3.2 is about here and 3.8 would be about here. That balances, 3.5."

5)<u>simple no match</u> (with calculation) "There's an uneven weight cause one he attended for 2 semesters and one for 4 semesters. So I should put 3.6 in the middle of the balance. 3.5 would be the middle. 3.8 and 3.2. Nope, I guess I'm wrong (doubts calculation). Should I refigure it out?"

6)<u>reverse proportion</u> "Double 3.2 to even up for the 4 semesters."

7)<u>totals</u> "Five men would weigh 540, 2 women 240. I put the mean in the middle of 540 and 240."

Frequencies for each type of representation on each problem by group are presented in Table 11.

Table 11: Type of representation on each weighted mean problem by subjects in each group.

Group	Representation type 1	2	3	4	5	6	7
CB	7		3				
n=10	4	2	4				
cB	1		1				
n=2			2				
T Cb	5		1		1(1)		
n=6	1	3	2				
C Cb	4(3)		2	2	1(3)		1
n=8	3	3	1				1
T cb	5		2	6(1,3)	1(1)	1	
n=12	3	5	2	2	1		
C cb	1			9(1,5)	2		
n=10	2		4	3	2(1,7)		1
problem	23		9	17	5	1	1
subtotals	13	13	15	5	3		2
totals	36	13	24	22	8	1	3

Notes: 1) Top number is GPA, bottom number is Elevator.
2) A subject can be represented twice. Strategies switched to are in parentheses.

All representations by balancers (**CB** and **cB**) were correct types of representations. Again there was considerable group flexibility in the type of representation used. Balancers were able to translate their correct notions about calculation into correct balance beam repre-

sentation once they learned how to represent a simple problem on the balance beam.

All members of the Cb groups who calculated correctly also represented correctly. Any incorrect representations that were considered were discarded in favor of weighted representations.

The cb subjects most often represented the problems incorrectly. However, the trained cb subjects represented the problems correctly more often than controls and considered incorrect representations less often. They also used the weighted concrete type of representation, which no control cb subject used. Thus, balance training also seems to increase the likelihood of a correct representation and flexibility in representing.

An important component of representing any problem on the balance beam is the labeling of the scale. The scale must be labeled as a number line including the stated range of scores with the numbers spaced far enough apart to allow for a close approximation of the actual mean. Some subjects started out with a zero point in the center and labeled the numbers in increasing order on both sides. Others included a much wider range of numbers than necessary so that the actual range used to determine the mean was too small to be accurate. The types of scales drawn by subjects in each group are presented in Table 12.

Table 12: Types of scales drawn by groups for each problem

Group	Type of scale 1	2	3	4	5	6
CB	10	1	2	5	1	
	10			2		
cB	2					
	2					
T Cb	6		1	4	1	
	6			4		
C Cb	7	2		5		
	8			5		
T cb	7	3	4	4	8	1
	10	2	2	5		
C cb	8	1	4	3		
	10		1	3		

Notes: 1)Top number is GPA, bottom number is Elevator.
2)Categories are: 1)include only between stated numbers, 2)wider range than necessary, 3)range is drawn too small, 4)labeled from center outward, 5)center point is zero and scaled is labeled outward, 6)scale is marked with the number of scores.

Calculators were generally able to label the scale correctly, with few instances of the common errors. There was a common tendency to either determine or guess the mean and label the scale from the center. This was a wise strategy, given the inherent imperfection in the system, since the scale does add some weight to one side if it is off center.

However, noncalculators had much more difficulty.

Several subjects from both the control and training groups made the spacing too small or included too wide a range. Many trained subjects marked the center of the scale with a zero point and proceeded to mark outwards in increasing direction on both sides of the fulcrum. This behavior probably resulted from the way most subjectslearned the torquerule on the balance beam:they counted the number of spaces the block is from the fulcrum and multiply by the number of blocks. The natural inclination then is to mark the scale in the same way that one counts to determine whether the beam should balance, rather than to conceive of two different numerical schemes. In general, it seems noncalculators have less well defined notions of what the characteristics of the mean are, including the appropriate range of numbers that might be involved.

Representing a mean on the balance beam also involves using the block in an appropriate manner. The block does more than merely mark the number on the scale beneath it; it also signifies a certain quantity or specifies the proportions in which the numbers are being combined. Therefore, a major aspect to representing a problem correctly is determining what a single block will represent; i.e. it is more than just a marker. For example: "Each block will represent a semester. He got a 3.2 for 2 semesters, so it would be 2 at 3.2, and then 4 weights at 3.8.

It balanced right before 3.6."

Even though subjects were not requested to explain what the block represented, many subjects did so spontaneously. The relationship between labeling the block and correct calculation is fairly strong ($r=.69$, $t(46)=6.47$, $p<.001$). Subjects who verbally indicated the unit the block was intended to represent on a particular problem were much more likely to perform that calculation correctly.

Within the groups of **cb** subjects, there was a much greater tendency for trained subjects to label the block (See Table 13), 7 of 12 versus 2 of 10 on the GPA problem and 11 of 12 versus 4 of 10 on the Elevator problem ($X^2(1)=6.72$, $p<.01$)

Table 13: Relationship of block labeling and correct answer for trained and control **cb** subjects on the second and third problems.

Trained

	Labeled		Not Labeled	
Correct	7	10	0	1
Incorrect	0	1	5	0

Control

	Labeled		Not Labeled	
Correct	0	4	0	2
Incorrect	1	0	9	4

There was a difference in how well subjects performed on each of the two problems. Thirty-one subjects calculated the GPA problem correctly, while 43 subjects calculated the elevator problem correctly. The major portion of this difference was from noncalculators, 7 of whom were correct on the GPA problem and 17 of whom were correct on the Elevator problem. There are several possibilities why this might be the case. There might be a practice effect merely from storing the problems, even if no feedback is given concerning the correctness of the answer. Balance beam knowledge might not have become incrporated into the mean schema within the first problem. The second problem may be in some sense easier in that the weights may be harder to ignore; several subjects later commented they had not noticed a difference in the number of scores in the GPA problem. In order to determine what kind of between problems effects there might be resulting from balance beam training, it would be necessary to counterbalance the order of presentation of the two weighted mean problems.

The weight maintenance problem. The weight maintenance problem was a simple mean problem which had the surface structure of a weighted mean problem. It was included to detect subjects who had adopted an algorithm for dealing

with the weighted mean problems without fully understanding when it should be used. The problem was solved correctly by nearly all subjects in each group. The one exception to this pattern was the trained **cb** group, in which only 8 of the 12 subjects solved the problem correctly. The data are presented in Table 14.

Few subjects inappropriately used a weighted mean algorithm. However, the more relevant information concerning subjects' ability to distinguish conditions is the number of subjects in each group solving the problem correctly after having solved and represented at least one weighted mean problem correctly. These figures are also presented in Table 14.

There is no change in the figures for calculators or balancers. Therefore, it seems resonable to conclude that the **CB**, **cB**, **T Cb**, and **C Cb**, groups were able to distinguish conditions appropriately. There was a difference between trained and control **cb** groups, although it was not significant. Six of the nine trained subjects who were correct met the new criterion, while only 4 of the 9 control subjects who were correct met the new criterion.

Most subjects offered some type of rationale for their answer. These rationales can be classified into three types which can be ordered in terms of depth of understanding. High level explanations dealt directly with the

Table 14: Number of Subjects in a Group Answering the Weight Maintenance Problem Correctly and Mean Rationale Level.

Group	# Correct	Mean Rationale Level
CB n=10	10(10)	2.7
cB n=2	2(2)	1.5
T Cb n=6	6(6)	2.0
C Cb n=8	8(8)	2.8
T cb n=12	9(6)	1.5
C cb n=10	9(4)	1.2

Notes: 1) Number in parenthese is number of subjects who got the weight maintenance problem correct after calculating and representing at least one weighted mean problem correctly.

2) Rationale is on a three point scale: 3 is a high level rationale.

consideration of how many scores should be involved in the calculation. Subjects who gave a high level rationale were given a score of three. Examples of high level explanations include:

"You're averaging the weights for two people rather than 3 weights for one person and five weights for another, cause there are only two people."

"No that's not right cause that's already been averaged. This person's weight is 185 and this is 211."

Subjects who gave medium level explanations focused on what did not matter, rather than with the number of important elements. Subjects generally took longer to arrive at conclusions when they offered a mediumlevel rationale. Subjects who gave a medium level rationale were given a score of two. Examples of rationales are:

"How many times a day they weighted themselves I don't think has anything to do with it. It doesn't matter. What matters is the weight."

"The first person weighs himself three times and weighs 185. The other weighs himself five times and weighs 211. 211 x 5=1055, 185 x 3=555. Add and divided by 8. The average is 201.25. This isn't going to make sense. You don't have that many people. He only weighs 185 once."

Thelowest level explanations bordered on a confusionof why one would use a simple or a weighted calculation. Subjects who gave low level explanations were given a score of one. Examples of low level explanations are:

"I don't think it matters how many times a day they weigh themselves. 185 is the average throughout the day. It would only go up or down a pound or so."

"The reason I didn't take into consideration how many times a day they weighed themselves is I didn't know how much they weighed each time. I couldn't do anything with it."

Some subjects did not give rationales,but merely performed a simple mean computation. These subjects were also given a rating of 1 for their rationale.

Mean rationale level for each group is also presented in Table 14. The table indicates **CB** subjects gave higher levels of rationales than other groups, that cb subjects in

general were less likely to give explanations and gave lower levels of explanations, and that training did not have a large effect on level of explanation. It may be that the time course of the experiment was not long enough to have such a subtle impact.

The shop problem. The shop problem confronted subjects with a choice as to whether a simple mean or a weighted mean computation was appropriate in a given situation. Although all subjects but one correctly chose the weighted mean computation, they differed in how they justified this choice.

There were two general types of justifications. In the first type, subjects focused on the different numbers of workers in each category. An example of this type of reasoning was:

"You would use the first only if you had equal numbers of people making each salary. Because you have twelve general workers, you have to weight the 8,000 twelve times as heavily as the owner president and six times the foremen."

Inthe second type of rationale, subjects considered the total number of people involved in each calculation. An example of this reasoning was:

"Just to add it up and divide by three, you would be only looking at three people where actually you're looking at fifteen people and that has to be considered."

The latter rationale is not as powerful as the former, since the same answer may actually be obtained with an infinite number of combinations in the same proportions. The actual number of people involved in the computation is essentially arbitrary.

The subjects were also probed verbally with the question, "Which calculation will give a higher number?" Responses to this question fell into three categories, which differed in correctness. The categories were: 1)the correct proportion argument, 2)the second (weighted) would be higher because the total amount of money was higher, and 3)need to calculate before deciding. Examples of these three answers in order were:

1) "The second calculation would be lower. In the first, the high salary carries as much weight as the low. In the second, the low carries twelve times as much weight."

2) "The first is lower. You're adding a smaller amount of money and dividing by three. There is more money with the second."

3) "I don't know. Probably be about the same."

Even though all subjects could determine that the weighted method was correct, they did not all have correct intuitions about why one was correct and how the mean was affected in each of the two types of calculations. Some explanations indicate vague and even incorrect ideas concerning the operations of multiplication and division.

Subjects responses to the two questions were coded for

type of explantion. They were considered to use proportional reasoning if at least one of their answers was a proportional type answer. Nearly all **CB**, **cB**, **T Cb**, and **C Cb** subjects gave at least one proportional answer. However, eight of the twelve **T cb** subjects gave proportional answers, while only one of the ten **C cb** subjects did. There was an association of training and proportionality arguments (X^2 (1)=7.23, p<.01). This is consistent with the hypothesis that balance beam training should help subjects to bettter understand how to approach problems which require proportional reasoning.

General Predictors of Performance

Several factors discussed in relation to only a single problem may be predictors of performance on other problems. Some of these are related to the training manipulation. These factors include the labeling of the block and proportional reasoning on the shop problem.

A post hoc multiple regression on number of problems correct (GPA, Elevator, Weight Maintenance) using labeling of the blocks and evidence of proportional reasoning as the independent variables yielded a multiple R of .68. Both factors added significantly to the variance accounted for above and beyond the other. Thus, these factors should

probably be considered in any future study of the mean.

Perhaps surprizingly, performance on the first simple mean problem was not a good indicator of later performance on the weighted mean problems ($r=.22$, $t(46)=1.54$, $.10<p<.20$). Obviously, many subjects had to learn during the course of the first problem how to represent a problem on the balance beam, and this did not appear to hinder performance.

Supplementary information concerning statistics and mathematics courses was collected at the time of the pre-test. The relationship between these variables and status as a calculator and a balancer was analyzed to determine whether they may have influenced transfer results. There was an association between having had statistics and being a calculator, as one might predict. However, it was only marginally significant (X^2 $(1)=4.090$, $p<.05$, $r=.29$) suggesting the treatment of the mean in statistics classes may be less than adequate for some students. There was no relationship between having had statistics and being a balancer ($r=.02$).

Level of mathematics training was rated on a four point scale, with one representing mathematics training through high school algebra and four representing training in college beyond calculus. Mathematics training was associated both with being a calculator (X^2 $(3)=8.46$, $p<.05$,

r=.19) and as a balancer (X^2 (3)=9.53, $p<.025$, r=.40). Mean level of mathematics training was 3.0 for **CB** subjects, 2.0 for **Cb** subjects, and 1.95 for **cb** subjects. In general, a higher level of mathematics training indicated the subject was more skilled in quantitative reasoning.

CHAPTER VIII

DISCUSSION

The basic questions addressed in this study concerned the extent to which balancing knowledge contributes to the ability to solve problems involving the mean. The results support both the descriptive and the prescriptive hypotheses. Balance rule level predicted calculational performance, and training on the balance beam facilitated later calculational performance. These results will be discussed more generally in this chapter, with particular emphasis placed on the transfer effects. Possible future directions will be indicated.

Descriptive Question

The results of the pretest are consistent with the hypothesis that students with higher levels of balance knowledge are more successful in solving problems related to or involving the mean. Balance rule level was correlated with calculational performance ($r=.35$). However, the ability to perform at the level of rule IV in Siegler's classification seemed to represent a greater degree of

sophistication than was required to answer the weighted mean problems correctly. Only 15 or 21% of the subjects were classified as being at rule level IV and all but two of them were also calculators. These subjects were also able to judge more accurately whether a particular number could possibly be the mean of a set of numbers, a purely numerical task which yet seems to involve notions of balancing. On the other hand, 31 or 42% of the subjects were able to answer both weighted mean problems correctly and only 28% of them performed at rule level IV on the balance task. These data suggest it is possible to provide correct numerical answers to weighted mean problems without sophisticated knowledge of balancing, although balancers may have some advantage in being able to estimate means more accurately.

The written pretest offered evidence confirming expectations that virtually every subject would know the algorithm for calculating the simple mean. Only one subject did not employ an appropriate algorithm and seven made minor arithmetic errors. These data suggest that the difficulties with weighted mean problems are conceptual in nature in that either the subjects do not understand what is being asked for (i.e. the mean for all the scores) or they are unable to adapt their algorithm for calculating simple means to the more complex situation.

An obvious limitation in drawing conclusions from the pretest concerning the relative importance of balance knowledge in understanding the mean is that the results are based on a correlational data and one cannot determine what the causal relationships are. Calculational ability may have been correlated with balance knowledge through other variables, such as mathematics background or ability. In fact, level of mathematics training was associated with both status as a calculator and as a balancer. Thus the issue of whether balance knowledge might be an important aspect of understanding the mean must be approached using experimental methods.

Prescriptive Question

The results from the transfer phase are somewhat easier to interpret. Trained **cb** subjects performed better than control **cb** subjects on the weighted mean problems, while nearly all trained and control **Cb** subjects solved both problems correctly. For **cb** subjects, providing experiences which fostered the development of balance knowledge facilitated later calculation and representation of weighted mean problems. Trained **cb** subjects calculated more problems correctly and used a variety of different computational methods. Control **cb** subjects tended to use

only one particular computational form and were correct less often, suggesting that training promoted a more flexible conception of the mean.

There were also several other differences in strategies displayed by the trained and control **cb** subjects during the course of the interview. These included spontaneous correction of computational form, labeling the block, and using proportional logic to explain an answer.

Three of the seven trained subjects who eventually calculated and represented both weighted mean problems originally computed a simple mean and spontaneously changed their answer to the correct weighted mean after attempting to represent the problem on the balance beam. No control subject did this. An additional four trained **cb** subjects calculated the GPA as a simple mean and switched to the weighted mean on the second problem (Elevator). Three of the four spontaneously corrected the GPA problem when asked to reconsider all their answers at the end of the interview. All control subjects needed some probes to correct their answers to the GPA problem.

Explicitly labeling the block with some appropriate quantitative unit (such as a semester for the GPA problem) was strongly associated with obtaining the correct answer. The incidence of labeling was much higher amongst trained than control **cb** subjects and increased from the first to

the second problem, as did correct performance. The balance beam seemed to help subjects determine more explicitly which elements belonged in the calculation through the act of determining what a single block might represent.

It is important to note that balance knowledge does appear to be a critical aspect of the ability to state what unit the block represents. Merely asking subjects to label the block without training would not yield the intended results. In the earlier Hardiman study (Note 3), when subjects were asked to state what unit the block represented, they most often replied that it represented the number underneath the block. The presence of the block only indicated that the number was somehow involved. It did not denote quantity, relative or absolute. The idea that in balancing the balance point depends on the number of blocks placed on each score provides a key to the concept that the quantity of scores must be represented by the number of blocks. Therefore, the need arises to express that quantity correctly.

Performance on the weight maintenance problem, a simple mean problem with a weighted mean surface structure, did not provide as clear a differentiation between trained and control **cb** subjects. Four of the ten **Tcb** subjects who solved at least one weighted mean problem correctly persisted in using the weighted mean computation. Five **Ccb**

subjects did this as well. There may be several reasons why, the most obvious being that some of the trained subjects may have learned a new algorithm for calculating this general type of problem and attempted to apply it to all problems. However, there was no reason to predict that subjects would perform better on this problem after having had balance training. Although there was no strong effect of positive transfer, neither was there an effect of negative transfer: most trained subjects were able to discriminate the two conditions.

There were clear suggestions in the interview data that trained subjects were more likely to develop a logic for analyzing the mean which includes the notion of proportional reasoning. In the shop problem subjects were asked to choose between a simple mean and a weighted mean calculation. Almost all subjects correctly chose the weighted mean calculation, but their rationales for the choice fell into two distinct categories. The first was based on the different proportions of workers at each level, while the second was simply based on the total number of people accounted for by each calculation. Sixty-seven percent of the trained **cb** subjects were classified as justifying their choice on the basis of proportional reasoning, while only 10% of the control **cb** displayed evidence of proportional reasoning.

The concept of proportionality was used to predict whether the beam would balance by many subjects early in the training session. If one left the blocks in the same places, but changed the number of blocks in a proportional manner, subjects would predict the beam would still balance. The trained subjects were thus able to import this idea directly to problems dealing with weighted means. The second type of rationale results from a literal interpretation of the algorithm for calculating the mean, $\Sigma x/n$. If one uses the number three as a divisor, it is difficult to determine how to account for fifteen people.

An argument might be made that differences between trained and control **cb** subjects were not the result of positive transfer from balance training to compuatation, but rather were due to negative effects on control subjects resulting from the request to represent problems on the balance beam. Therefore, performance on the pretest and in the transfer session was compared for **cb** subjects. Control subjects got 30% of the problems correct on the pretest and 35% of the problems correct on the transfer task, whereas trained subjects scored 50% on the pretest and 75% on the transfer task. Performance for control subjects actually increased slightly, while performance for tansfer subjects increased substantially, suggesting the differences were due to positive transfer.

Several conclusions may be made on the basis of the present study: 1)training on the balance beam is an asset in helping subjects to calculate weighted mean correctly, 2)training indirectly fosters the correct conception of the representation of quantity, and 3)training helps subjects develop higher level rationales for the use of certain forms of calculations. Overall, the significant effects due to training involved only the **cb** subjects and not the **Cb** subjects, although there was a nonsignificant difference in correctness of weighted mean computations for **Cb** subjects. One would expect to find no difference in calculational performance for the **Cb** groups, but one might expect significant differences in more subtle measures, such as representations or rationales. It is possible that the experience of representing problems on the balance beam facilitated the development of the mean concept for control **Cb** subjects. Representing a problem on the balance beam may not be a difficult task if one has a fairly flexible concept of the mean and at least basic "see-saw" type knowledge of the balance beam. In this case, knowledge that the number of scores entering the equation are weighted might lead to the conclusion that this can be represented only by different numbers of blocks. If neither balance knowledge nor computational knowledge were well developed, the subjects were not able to make connec-

tions between the two realms of knowledge and did not represent the problems correctly. It may have been advantageous to have included problems which required subjects to explain their answers in the pretest to determine whether the integration of knowledge did in fact provide the basis for any change in rationale.

General Discussion of Transfer

A number of studies using abstract well-defined move-type problems have indicated that transfer is difficult to obtain unless certain specific types of information are provided to the subjects (c.f.Reed, Ernst, and Banerji, 1974;Fiszman (Note 10). The claim in the present study is that transfer was achieved. Therefore it may be instructive to examine the present findings with specific regard to the more general issue of transfer in problem solving.

Reed et.al. have suggested that a necessary precondition for transfer is the recognition of an analogy between the problems. In some less successful studies of transfer, subjects did not recognize the analogy between the stories or problems presented (Gick and Holyoak, 1980; Reed et.al., 1974). Additional studies indicated that transfer could be obtained when the analogous problem was well understood,

implying that subjects may have been able to make an analogy (Luger and Bauer, 1978).

The present study implicates a factor not considered by Reed et.al.: understanding the analogous domain seems to be important. It was possible to predict which control subjects would fail the representation task on the basis of the assessment of calculational and balancing skills. The capability to make such a predication indicates there is some additional factor involved. Recognition of an anlogy may be important, but it is not the only critical factor since all subjects were told that the mean could be viewed as a balance beam and helped to find the representation of the simple mean. Recognition and specification of an analogy may be important, but obviously some subjects had a better basis for making use of the analogy, which Reed et.al. do not comment upon.

Presumably, balance knowledge was equated for balancers and nonbalancers in the training session. The trained subjects performed similarly to the balancers on both representational and calculational aspects of the transfer task, whereas control subjects did not. This suggests that an understanding of the analogous situation is an essential aspect of transfer.

Cues that one should use the analogous informationto solve the transfer problem may also be important. Gick and

Holyoak(1980) presented subjects with stories which were well within the comprehension level of all subjects, and then a problem which was most easily solved by using the story as an analogy. Subjects did not use the previous story in this manner in general unless they were specifically asked to do so.

The requirement in the present study that subjects represent the problems on the balance beam probably constituted a strong situational cue to use information gained during balance training to help solve the weighted mean problems. The possibility exists that transfer from the training to calculation would not have been as successful without this requirement to represent the problems. One might speculate that representing the problems on the beam provides a critical link between the balance knowledge just developed and the computational formula, and thus is a necessary aspect of obtaining transfer. However, the issue of what constitutes a sufficient cue remains open to empirical investigation.

Future Directions

The acquisition and/or refinement of analog knowledge led to improved computational performance in novices originally classified as noncalculators. In light of the apparent success of the training manipulation, several new questions may be raised. Among these are questions concerning whether the effects of training really become integrated into a schema of the mean, the long term stability of the effects of balance training on the concept of the mean, and the type of cue necessary to reliably obtain transfer.

There is presently little experimental evidence to indicate whether the development and use of an alternative form of comprehending a concept does in fact lead to a more stable knowledge base, although there has been considerable speculation on the subject. Mayer and Greeno (1975) and Myers et.al.(in press) have evidence which suggests the presentation of new concepts embedded in previously existing knowledge structures leads to better performance on problems requiring some interpretation. By extrapolation, it seems that more connections to world knowledge in general might increase the possibility for the development of better understanding. This implies there may be some advantage to insuring that all students have attained a

certain level of understanding of certain concepts to increase the likelihood that useful connections will be drawn.

Therefore, it is obviously of interest to assess whether subjects who have received balance training maintain the flexibility in calculation displayed during the interview over a longer period of time. Would the transfer effect still be found after several months? Presumably, a more stable knowledge base which included knowledge of balancing might also provide a better basis from which to learn concepts related to the mean, such as the standard deviation and the variance. Intuitively, it seems that the more well developed any single concept is, the more likely concepts based on it will be well understood. Such a study might conceivably take place within a classroom context.

Numerous persons have speculated on the importance of intuitions, models, and varied forms of experience leading to the acquisition of a new concept. Lawler(1981) argued that the fitting together of multiple points of view allow more flexibility in the modification and adjustment of a concept. Papert(1980) has suggested somewhat more explicitly that models, as the present example of the balance beam, help connect formal knowledge and experiential knowledge. The model is a tool to think with about certain types of problems. Essentially, the model allows

one to bring abstraction to object knowledge (Gray, 1979). DiSessa(1979) argues in addition that a model should be uncomplicated in order to provide simplicity and coherence in relation to the subject's preexisting knowledge base.

Because the model appears to serve this function of connecting realms of knowledge, substance is provided for the argument that asking subjects to represent problems on the balance beam may have been essential to obtaining transfer in calculation. One would need another condition where subjects are not asked to represent problems on the balance beam during the transfer task to assess the impact of the model of the mean as a balance point on computational strategy.

An analysis of the details of the training sequence itself might be of value in general. Balance knowledge seemed to be obtained through a much richer learning experience than one might expect on the basis of accounts provided by Siegler (1976; Klahr and Siegler, 1978). There appear to be recognizable stages to gaining the rule for torque or level IV, making the learning situation or behavior involved in rule level III performance much more complex than merely "muddling through." This dynamic view of rule III level behavior suggests a need to study balance concepts in transition. The study of balancing in a relatively static state, such as in the Siegler experiments,

may leave much of the complexity of rule level III behavior obscured.

A more general issue concerns the study of the models themselves. There may be other models which provide an appropriate tool to think with which have more flexible applications. Bentz (Note 11) has developed a special "wheel of fortune" which may be useful in teaching students about the mean as well as about probability. There might be potential in the study of other models, the relative advantages of each, and the type of misconceptions they help address.

In conclusion, it seems the issue of analog knowledge, and in particular the model of the balance beam, is worthy of further research both in and out of classroom contexts. The balance beam seems to provide a natural tie to experiences which most students have had, is helpful in developing the concept of proportionality, and provides a clear basis for the need to identify the quantity of scores being combined in a weighted mean calculation. The research described here suggests that if the balance beam were to be used in teaching aboutthe mean, it would be necessary to provide some training in balancing. However, the effort seems justified in view of the fact of poor performance that has been found on the part of several students who have taken traditional statistics classes.

REFERENCE NOTES

1. Chi, M.T.H., Glaser, R. Rees, E. Expertise in problem solving. Technical Report #5. LRDC University of Pittsburgh, 1981.

2. Sinatra, G. Misconceptions in understanding the weighted mean. Unpublished research report, University of Massachusetts, 1980.

3. Hardiman, P.T. The balance beam as a tool for understanding the mean. Unpublished research report, University of Massachusetts, 1981.

4. Kolers, P.A. When is learning? Paper presented at the Annual Meeting of the Psychonomic Society, Philadelphia, 1981.

5. Clement, J. Analogy generation in scientific problem solving. Proceeding of the Third Annual Meeting of the Cognitive Science Society, 1981.

6. Confrey, J. Using the clinical interview to explore student's mathematical understandings. Presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

7. Konold, C.E. and Well, A.D. Analysis and reporting of interview data. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

8. Confrey, J. Concepts, processes, and mathematics instruction. Unpublished paper, Mount Holyoke College, 1981.

9. Pollatsek, A., Well, A., and Cobb, G. Cognitive skills underlying statistical inferences. Grant proposal, University of Massachusetts, 1981.

10. Fiszman, A. Transfer in problem solving. Unpublished master's thesis, University of Massachusetts, 1976.

11. Bentz, H.J. Stochastics teaching based on common sense. Paper presented at the First International Conference on Teaching Statistics, 1982.

REFERENCES

Bartlett, F.C. Remebering, London: Cambridge University Press, 1967.

Champion, D.J. Basic Statistics for Social Research, New York: Macmillan Publishing Company, 1981.

Copeland, R.W. How Children Learn Mathematics London: The Macmillan Company, 1970.

Dienes, Z.P. An Experimental Study of Mathematics Learning London: Hutchinson & Company, Ltd., 1963.

Dienes, Z.P. and Jeeves, M.A. Thinking in Structures London: Hutchinson Educational Ltd., 1965.

DiSessa, A.A. On "learnable" representations of knowledge: A meaning for the computational metaphor. In J.Lochhead and J.Clement (Eds.) Cognitive Process Instruction Philadelphia, Pa.: The Franklin Institute Press, 1979, pp.239-266.

Ericsson, K.A. and Simon, H.A. Verbal reports as data. Psychological Review, 87(3), 1980, pp.215-251.

Ferguson, G.A. Statistical Analysis in Psychology and Education New York: McGraw-Hill Book Company, 1981.

Fischbein, E. Intuition and mathematical education. In Some Theoretical Issues in Mathematics Education: Papers from a Research Presession ERIC 1979.

Freedman, D., Pisani, R., and Purves, R. Statistics New York: W.W. Norton & Company, 1978.

Gick, M.L. and Holyoak, K.J. Analogical problem solving. Cognitive Psychology, 12, 1980, pp.306-355.

Ginsberg, H. Children's Arithmetic: The Learning Process New York: D. Van Nostrand Company, 1977.

Goodwin, W.L. and Klausmeier, H.T. Facilitating Student Learning: An Introduction to Educational Psychology New York: Harper & Row Publishers, 1975, pp.409-444.

Gray, R.L. Toward observing that which is not directly observable. In J. Lochhead and J. Clement (Eds.) Cognitive Process Instruction Philadelphia, Pa.: Franklin Institute Press, 1979.

Greeno, J.G. Nature of problem solving abilities. In W.K.Estes (Ed.) Handbook of Learning and Cognitive Processes Vol.V: Human Information Processing New Jersey:L.Erlbaum, 1978, pp.239-270.

Hayes, J.R. and Simon, H.A. Psychological differences among problem isomorphs. In Castellan, Pisoni, and Potts (Eds.) Cognitive Theory, Vol.II, New Jersey:L.Erlbaum, 1977.

Hays, W.L. Statistics New York: Holt, Rinehart, and Winston, 1981.

Herold, P. The Math Teaching Handbook Boston: Selective Educational Equipment 1978.

Kaput, J. Mathematics and learning: Roots of epistemological status. In J.Lochhead and J.Clement (Eds.) Cognitive Process Instruction Philadelphia, Franklin Institute Press, 1979.

Kirk, R.E. Introductory Statistics Monterey, Ca.: Brooks/Cole Publishing Company, 1978.

Klahr, D. and Siegler, R.S. The representation of children's knowledge. In H.W. Reese and L.P. Lipsitt (Eds.) Advances in Child Development and Behavior Vol.12 New York: Academic Press, 1978, pp.62-116.

Kline, M. Why Johnny Can't Add: The Failure of the New Math New York: St.Martin's Press, 1973.

Koopmans, L.H. An Introduction to Contemporary Statistics Boston, Ma.: Duxbury Press, 1981.

Krutetskii, V.A. The Psychology of Mathematical Abilities in Schoolchildren Chicago: The University of Chicago Press, 1976.

Kurtz, A.K. and Mayo, S.T. Statistical Methods in Education and Psychology New York:Springer Verlag, 1979.

Kusher, H.W. and DeMaio, G.D. Understanding Basic Statistics San Francisco: Holden-Day, Inc., 1980.

Lawler, R.W. The progressive construction of mind. Cognitive Science, 5, 1981, pp.1-30.

Lesh, R. Mathematical learning disabilities: Considerations for identification, diagnosis, remediation. In R.Lesh, D.Mierkiewicz, and M.Kantowski (Eds.) Applied Mathematical Problem Solving ERIC, 1979.

Levine, G. Introductory Statistics for Psychology: The Logic and Methods New York: Academic Press, 1981.

Lindgren, B.W. and Berry, D.A. Elementary Statistics New York: Macmillan Publishing Company, Inc., 1981.

Luger, G.F. and Bauer, M.A. Transfer effects in isomorphic problem situations. Acta Psychologica, 42, 1978, pp.121-137.

Mayer, R.E. and Greeno, J.G. Effects of meaningfulness and organization on problem solving and computability judgments. Memory and Cognition, 3, 1975, pp.356-362.

Minium, E.L. Statistical Reasoning in Psychology and Education New York: John Wiley and Sons, Inc., 1978.

Mitzman, B. Seeking a cure for mathephobia. American Education, 12, 1976, pp.11-14.

Morris, P.E. Why Evans is wrong in criticizing introspective reports of subject strategies. British Journal of Psychology, 72, 1981, pp.465-468.

Myers, J.L., Hansen, R.S., Robson, R.C., and McCann, J. The role of explanation in learning elementary probability. Journal of Educational Psychology (in press).

McCall, R.B. Fundamental Statistics in Psychology New York: Harcourt Brace Janovich, Inc., 1975.

Nisbett, R.E. and Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 1977, pp.231-259.

Pagano, R.R. Understanding Statistics in the Behavioral Sciences St.Paul: West Publishing Company, 1981.

Papert, S. Mindstorms: Children, Computers, and Powerful Ideas New York: Basic Books, Inc., 1980.

Pfeiffer, K. and Olson, J.N. Basic Statistics for the Behavioral Sciences New York: Holt, Rhinehart, and Winston, 1981.

Piaget, J. The Child's Conception of Number New York: W.W.Norton & Company, Inc., 1965.

Pollatsek, A., Lima, S., and Well, A. Concept or computation: Students' understanding of the mean. Educational Studies in Mathematics, 12, 1981, pp.191-204.

Reed, S.K., Ernst, G.W., and Banerji, R. The role of analogy in transfer between similar problem states. Cognitive Psychology, 6, 1974, pp.436-450.

Rosnick, P. and Clement, J. Learning without understanding: The effects of tutoring on algebra misconceptions. Journal of Math[ematical Behavior, 3, 1980.

Runyon, R.P. and Haber, A. Fundamentals of Behavioral Statistics Menlo Park, Ca.: Addison-Wesley Publishing Company, 1976.

Shavelson, R.J. Statistical Reasoning for the Behavioral Sciences Boston: Allyn and Bacon, Inc., 1981.

Siegler, R.S. Three aspects of cognitive development. Cognitive Psychology, 8, 1976, pp.481-520.

Skemp, R.R. Intelligence, Learning, and Action New York: John Wiley & Sons, 1979.

Suydam, M.N. Continuing the math revolution. American Education, 6, 1970, pp.26-30.

Szeminska, A. The evolution of thought: Some applications of research findings to educational practice. Monographs for the Society for Research in Child Development Vol.30 No.5, 1965, pp.47-57.

Tennyson, R.D. and Park, O. The teaching of concepts: A review of instructional design research literature. Review of Educational Research, 50, 1980, pp.55-70.

Tuma, D.T. and Reif, F. Problem Solving and Education: Issues in Teaching and Research Hillsdale, N.J.: L. Erlbaum , 1980.

Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. In P.N.Johnson-Laird and P.C. Wason (Eds.) Thinking: Readings in Cognitive Science Cambridge: Cambridge University Press, 1977, pp.326-337.

Wang, M.C., Resnick, L.B., and Booser, R.F. The sequence of development of some early mathematics behaviors. Child Development, 42, 1971, pp.1767-78.

Weinberg, G.H., Schumaker, J.A., and Oltman, D. Statistics: An Intuitive Approach Monterey, Ca.: Brooks/Cole Publishing Company, 1981.

Welkowitz, J., Ewen, R.B., and Cohen, J. Introductory Statistics for the Behavioral Sciences New York: Academic Press, 1976.

Wilkinson, J.D. A review of research regarding mathematics laboratories. In Fitzgerald and Higgins (Eds.) Mathematics Laboratories: Implementation, Research, and Evaluation ERIC, 1974, pp.37-59.

Winston, P.H. Learning structural descriptions from examples. In P.H.Winston (Ed.) The Psychology of Computer Vision New York: McGraw-Hill Book Company, 1975.

APPENDIX I

Table 15: Topics covered by surveyed texts. Numbers indicate order of presentation.

		$\Sigma x/n$	Σ freq. x/n	$\Sigma n_i X_i/N$	$\Sigma x_i = n\bar{X}$	$\Sigma(x_i - \bar{x}) = 0$	the mean is balance	illus. of points
1) Champion	1981	1	2	3		4		
2) Ferguson	1981	1	3		2	4		
3) Freedman, Pisani & Purves	1978	1						2
4) Kirk	1978	1	2				3	
5) Koopmans	1981	2					1	
6) Kurtz & Mayo	1979	1	2	6				
7) Kusher & DeMaio	1980	1	2	3		4		
8) Levine	1981	1				4		
9) Lindgren & Berry	1981	1	4			6		
10) McCall	1975	1				2		
11) Minium	1978	1	2	7	8	5	4	
12) Pagano	1981	1		7		3	4	
13) Pfeiffer & Olson	1981	1				9	2	
14) Runyon & Haber	1976	1	2	6	7	3		
15) Shavelson	1981	1		7	6	4	2	
16) Weinberg, Schumaker & Otman	1981	1				7	6	
17) Welkowitz, Ewen & Cohen	1976	1	2		4	3		

	number line with more than one example of a single score	balance beam with equal sides	balance beam with nonequal sides	balance beam with a single number changed	histogram on balance beam	compare to mode	compare to median	sensitive to all scores in the distribution	$(X_i - \bar{X})^2$ is a minimum	discussion of skewness	encourage guessing
1)								5	7	6	
2)						6	7		5	8	
3)	3	4	5	7	6						yes
4)			6			4	5	7		8	
5)						3	4				
6)						3	4			5	
7)								5	7	6	
8)						2	3				
9)						2	3	5			
10)						5	4		3		
11)			6					3			
12)		6						2	5		
13)				4	3	5	6	7		8	
14)						8	9	4	5	10	
15)			3			8	9	5		10	
16)			5	9		2	3	4	8		
17)						7	6	5			

A P P E N D I X II

Table 16: Balance beam problems presented during training.

```
_ _ _ _ / _ _ _ _

    2   /   3
  1     /     2
    2 1 / 1 2
2       /       1
  3   1 / 1   3
      1 /     1
      2 /     1
      3 /     1
      4 /     1
      3 /     1
    2   / 2
    2   / 3
    2   / 4
    2   / 5
    2   / 4
    1   / 4
    1   / 3
    1   / 2
    1   / 1
    1   / 2
  2     /   2
  2     /   3
  2     /   4
  2     /   3
  1     /   2
  1   1 /   2
_________________
    2 2 /       2
    3   / 2 3
1   2   / 4 2
1   1   /     2
    1 3 /       1
    2 2 /       2
  1   2 /     2
  2     / 1 3
    4   /   3 1
1 1     / 2 2
    1 2 /   2
    3   / 1 2
      6 /   1 2
    4   /     1 1
1   1   /     2
  2   3 /     3
```

```
_ _ _ _ / _ _ _ _

    3   /     1 1
  2 2   /       2
    2 4 /   2   1
-----------------
```

which way would you move
the scale

```
    2   /   3
  1     /     2
    4   / 2
  2 2   /       2
    1 3 /       1
  1 2   / 1 1
  2   1 /     3
```

	Order of Calculating · Representing						Type of Calculation							Type of Representation						Scale						Satisfaction					Evidence of Analog Knowledge		
	Calculate first	Calculate first - unfinished	Represent first - unfinished	Change calculation after balancing	Calculate second	Calculate second when asked	Simple $\frac{a+b}{2}$	simple a-b=[illegible]	weighted	weighted-proportional	weighted-multiple of proportion	reverse proportion	simple sum divided by number of scores	Simple	Simple - doesn't match calculation	weighted	weighted proportional	reverse proportion	try to represent total for each group	include only between stated numbers	wider range than necessary	Necessary range is too small	Mark from center	zero point in center	mark with number of scores	1 incorrect - feel sure	1 incorrect - think wrong	2 methods - choose wrong	2 methods - choose right	1 correct - feel sure	Represent first and feel confident	mean should be closer to one score	correctly states weights should (a:1) or used
GPA																																	
Elevator																																	
Weight A·B																																	

3·10 example												Shop											Second time								
Spontaneous or not		Attempts to represent										Answer				Reasons				higher/lower			Corrects Answer					Does # of scores make a difference?		Conflict	
by self	prompted	put 1 block each on 3·10	Slide scale to mean first	put more than 1 on 3 or 10 to balance	use 3·10 blocks	put 3·10 spaces from center	mean of 3,10 is 6½	6½ spaces on each side	put block on 6½	other:		simple or 1	weighted or 2	not different	different proportions	different numbers of people	spontaneously say 1st would inflate average	not right to average owner + workers	correct proportion argument	second higher since total is higher	Need to calculate		spontaneous correction	minor probes	major probes	Correction	no correction	does not make difference since it's an average	didn't notice difference	find conflict	switch to wrong answer

APPENDIX III

CASE STUDY

Although further analysis of the training session is not strictly indicated within the context of the present study, such analysis might be valuable for elaborating the process of acquisition of balance knowledge in adults. Hence, the following case study and analysis are offered in the hope of raising several issues.

Siegler (1976) suggests that the higher the balance rule level the subject has reached, the easier it should be to reach rule IV performance. Therefore, one might expect that in the present study the number of trials to criterion should be correlated with initial rule level. It should take fewer trials to advance from rule III to rule IV than from rule I to rule IV. However, no such correlation exists in the present study (r=.0014).

One possible explanation is that adults were able to determine the relevant variables of weight and distance fairly quickly, that they are related in some manner, and thusproceeded relatively quickly to rule III. Thelarge variation in number of trials to criterion should be due mainly to variation in time taken to determine a precise quantitative relation between the variables.

Given this explanation, the major question is how does

one advance from rule III to rule IV? How does one proceed from the process Siegler has described as "muddling through" a prediction to calculating torques? What are the characteristics of this muddling that will lead one to adopt the proper rule?

In order to begin investigation of these issues, a complete training protocol of a single subject is presented to illustrate several behaviors which appeared to be common amongst many subjects in the sample. The subject chosen for case study is a 25 year old female psychology major in her junior year. Her college mathematics background consisted of one statistics course and a precalculus course. Performance on the pretest indicates she was a noncalculator and a nonbalancer with rule II level knowledge. She developed rule IV knowledge in slightly less than the average number of trials and her data illustrates several common elements in her pattern of acquisition.

While reading the transcript and comments, there are certain points which should be noticed: 1)rapid isolation of the variables, 2)advance from rule II to rule III level predictions within the first few trials, 3)narrowing of choices in conflict situations, and 4)verbalization of particular cases of the quantitative rule for prediction.

Protocol

Notes: 1) Placement of blocks on the balance beam are indicated above the subject's comments. / marks the midpoint of the beam. Spacing from the center is indicated by position. 2) Comments made after the beam is released are indicated by a second S label.

T1 (0030/0200)
S I think the left side will go down cause it's heavier.

T2 (0021/1200)
S It will balance cause it's the same weight on each side.

T3 (0013/1300)
S Right side will go down cause there's more weight further from the center of the balance beam.

T4 (0100/1000)
S I think the left side will go down cause even though the weight is only one block, it's more farther from the middle.
I It's further out.

The subject considers weight and distance from the beginning. Weight is compared first, then distance as one would predict given the Siegler flow diagram. Distance is measured correctly from the center of the fulcrum. The behavior is consistent with what would be expected for a subject with at least rule II level knowledge.

T5 (0100/2000)
S It might balance cause now there's more weight pushing against that tendency to go left, but I'm not sure if it's enough. Do I have to tell you one thing?
I You can hedge your bet.
S It might lean left a little.

On this single example, the subject begins to display behavior typical of a person with rule III level knowledge by considering that weight and distance probably interact.

She notices that the weight and distance cues conflict, but is able to narrow the choice of prediction to left or balance.

T6 (0100/3000)
S I think it might still go left cause it seemed to go down so far the last time.
S So it balanced.

T7 (0100/4000)
S This time it will go right, cause if it just balanced it will have to do something else. With more weight it will go right this time.

The action of the balance beam is correctly predicted in T7. However, the subject does not appear to gain much insight into the rule structure by this move, since she based her prediction on the previous trial. Subjects often try to use the previous trial to help make a prediction, rather than reasoning out the answer independently.

T8 (0020/2000)
S This one will go left.

T9 (0020/3000)
S This one would balance because we're not quite the same distance from the middle. The one that's less weight is farther apart, so the one with more weight would pull it toward the center.

T10 (0020/4000)
S I think this time it will balance.

The subject is able to articulate quite clearly how weight and distance interact in a descriptive manner. However, her guesses on T9 and T10 indicates she does not yet identify specifically the mathematical relationship.

T11 (0010/3000)
S This one will go right cause there's more weight on the right.

The subject apparently treats this as a new problem, not relying on adopting the results of the previous trial by adjusting the proportions

T12 (0010/2000)
S I think, this just occurred to me, that weight with distance would be part of the formula. As one goes up, the other goes down. Not like that, but , it balances. Maybe weight minus distance. I'm not really sure, but I know those are the two variables. Now I've got to guess this one. I think it will go right. I'm not sure.

The subject has been considering the interaction of weight and distance in her predictions since T5. At this point, she explicitly begins to try to identify an explicit formula relating the variables.

T13 (0020/4000)
S That would have to be the same, cause I think you just doubled it (balance).

The explanation of proportion is utilized when an example is given where the concept can be applied for a prediction of balance.

T14 (0200/0200)
S The one to the left is a little bit further out from the middle, so I'd say that would go down first.

T15 (0200/0300)
S It might balance maybe. I think... I'm going to try for a balance.

The prediction appears to be based on visual estimation.

T16 (0200/0400)
S So this one would go to the right then, cause you added more weight to the right.

T17 (0100/0200)
S I don't really know, cause it seems like you moved the last one out further, so I don't know if it would balance now. I still think it would. I'm going to go with balance.

She appears to be comparing this problem with (0010/2000), relying on a visual comparison of the distance.

T18 (0100/1100)
S That would balance.
I Why do you think that?
S It just looks like it. It just looks... Cause there's.. With that pulling down on this one, this is pulling it to be steady with two and it's more equally distributed. But there is something happening with the distances that I know I'm not comprehending. I keep making mistakes.

The subject realizes that visual assessment is not adequate for making correct predictions alone. She also clues in on distance as the factor which is not completely understood.

T19 (0100/3000)
S I'm going to say it will go right.
S Does it have something to do with it being three across and three up? Does that make a difference?
I It might.
S Okay, I'll have to watch that.

The subject tries to gain a better understanding of the distance factor by counting the marks from the fulcrum and noting an apparent relation between the amounts of weight and distance on either side of the balance beam. Note that she only explicitly quantifies one variable on either side

of the fulcrum.

T20 (0004/0200)
S Well, if that's right it will balance, cause there's 4 at one space and 2 on 2 spaces. So it would be 2 to 1.

Now she explicilty notes both weight and distance on either side of the fulcrum. She appears to make the conclusion that when weight and distance are in a two to one relationship in opposite directions the system will balance. The 2 to 1 rule is a special case of the more general torque rule, or w(1)d(1)=w(2)d(2) or 4x1=2x2.

T21 (0002/0100)
S Again, it's still 2 to 1. It should balance.

T22 (0030/0020)
S I'm a little confused. I'm just going to have to guess, balance.
I What is confusing about it?
S Well, cause it's 3. It doesn't work the way I was dividing it before. I don't know. I just get confused with 3.
I What were you doing before?
S Well, like it if was at 2, and it had 4 and the other one was at 1 and it had 2, it would be definitely 1 to 2, but now I'm not sure how to do it. I think it should balance there, cause there's 2 on 3 and 3 on 2. But it wasn't exactly what I was doing before. I was dividing... Okay, there's 3 weights on 2 and 2 on 3. So it means that 3 weights on 2 would come down more. So I would say left.
S It balances. Okay, have to try something else.

The subject appears to be considering both a multiplicative rule and a division rule in this trial, although neither rule is well stated. She uses a rule in which weight is divided by distance to make the prediction. The rule is discarded when it fails to provide a correct prediction.

T23 (0100/1100)
S I still think this one will balance.

The subject appears to remember the result for this particular configuration from a previous trial and does not take advanatge of the trial as a chance to test a new theory. This may be because the situation is more complex with more than one pile of blocks on one side.

T24 (2000/2200)
S I think it would still balance, cause if you had these (2 on 1) on top of those (2 on 2), it would be just right. But since you have it over more, I think the pull would the same, I hope.
S Well, maybe as it gets further away from the middle you need more weight to pull it up. You'd have to have more weights there (on left) to make it balance.

The subject tries to apply the 2 to 1 rule derived on T20. However, she does not realize that any deviation from this 2 to 1 pattern will enable her to predict the balance beam will tip.

T25 (2000/2300) (a continuation from T24)
I What would you put there?
S I initially put balance. But now... Oh, you mean how many blocks?
I Yes.
S It seemes to go down pretty far. I would try one here (points to space 2) and see what would happen. See if that would pull it up (balance).
S Maybe as you get further away, I would think that as you get further away, then you need more. I think I said that before and I forgot. But you need more up here to pull it.

The effect of distance seems to be "forgotten" when the subject tries to predict what will happen on a complex problem. Subjects often think all the blocks on the same

side of the scale have torques equivalent to the block that is furthest from the fulcrum when first presented with complex problems.

T26 (0022/0300)
S I think that would balance.
I Why?
S Cause, well, this is a strong pull cause it's, cause it's 3 and it's further from the middle. Is it? Maybe not, but anyway it's 3. I think it's going to go left. I'm going to change my decision because actually this is the same distance from these farthest blocks here. So I don't think it's going to be as sharp. I think it might go left, cause there's more weight.
S Hm. So now. Hm. Okay.

Again the subject does not analyze the different distances of the two sets of blocks on the left. Although most subjects easily derived the 2 to 1 rule when the blocks were on opposite sides of the fulcrum, most found it quite difficult to determine that a block at distance 2 has twice as much torque as a block at distance 1 when the are on the same side.

T27 (0024/0201)
S Well, maybe this is right. It's 2 to 1 and 2 to 1. Maybe it would balance. However, this one is over a little more (1 on 4) on the right. I'm still going to go with balance. I think it might work cause of the angle of this. It might do it.
S I'm noticing that it's like 2 to 1 again, but there's something else, because it's not... It seems like there's something else involved. I know it's 2 to 1, but I'm not sure how to fit it all together.

The subject confirms that the specific 2 to 1 case is part of the general rule structure, realizes that there is a more general rule, but is unable to coordinate the knowledge she has to this point to formulate a possibility.

T28 (0004/0001)
S I still think you would need like 3 there to 1. That seems like it's too much, even though it would be 1 to 4 and 4 to 1. I still think that 3 (means 4) is a little much, but I have to see.
S So 4 to one is right.

T29 (0003/0010)
S So that would have to balance too.

The specific case rules are extended to include cases where weight and distance are 3 to 1 and 4 to 1 relations in opposite directions.

T30 (0200/0101)
S I think it's going to go right. This block seems to be even more pushing it down. So I don't think that 2 to 1 is enough with that. This is going to throw it off (points to 1 on 2), I would think.

The subject may have miscounted the spaces or relied on visual cues, attempting to apply the 2 to 1 rule in what she perceived as a negative instance. The problem is actually included to illustrate that distances can be averaged on a side.

T31 (1000/3100)
S I'd say it would balance because the weight here (on 2), it seems like it would be right... I would think you would need about 3 if you had them up there (on 2), but since it's closer to the middle, that would be alright.
S Was it supposed to go down? Hm.

T32 (1000/4000)
S It would still go down, I would think (right).
I Why?
S Cause I think most of the weight is coming from the 3's (3 on 1). So I think it would still go down, the first 3.
S It's 4 to 1 again. I can see that now.

T33 (1000/2100)
S I think it will go... left.

I Why?
S Cause you needed 4 and now you have 3. If it were just at space 1 and I don't think it's going to. I think it might balance. Can I change it?
I Yes.
S I think it might balance then, cause it was 4 and now it's 3 and that has a pull. It's obviously exerting more of a difference than I thought it was. I think it might still balance.
I How much pull is that?
S How much pull? What do you mean? There's 3 blocks on it, but I think it's m(ore)... I don't know exactly what you're asking?
I Well, how much is this side pulling (1 on 4)?
S 4, 4 times, it's at the 4, so it's pulling that weight down. And this is pulling 3, no a 2 and two 1's. So it equals the same, sort of, as one 2 and two 1's. 4 and 4. I'm just getting more confused. I'm not getting any formulas.

Thesubject began the sequence T31 to T33 with what appeared to be a very fuzzy idea of how much a block at a certain distance is worth. Her misconceptions are exacerbated in T32, where she says she beleives most of the weight is from one pile of the blocks, not seeming to realize small differences unbalance the system. On T33, she realized the block on the second space must have more torque than that on the first. When asked to quantify the torque for the system, she was able to do so, but does not seem to realize that is the key to the problem.

T34 (2000/2200)
S So we'll try this again. That would be 8 and 4 and 6. So that would be 6. So that would go down, left side.
S I hope it's not that easy.

The torque rule is tested on this case.

T35 (0200/2010)
S I think the left side would go down, cause there's 6 weights to 5.

T36 (2000/0400)
S It will balance.

T37 (0031/0020)
S 7 to 6, so it will go left.

T38 (0130/0400)
S Left side, cause it's 9 to 8.

T39 (0022/0011)
S The right side will go down.

T40 (0020/2100)
S It will equal. It will balance.
I So what do you think is the rule?
S It's so simple. It's just adding up the weights and you times it, like you go down and you get a number, like 1,2,3, measure it out, and then times it by the number of blocks and that's it. Right has to eqaul left if it balances or it's one heavier than the other. I sure thought it was something else because... Well, I guess going up a scale, even if it's only one block, that can add more pull. So is that what it is?
I It seems to be working.

The subject continued to apply the torque rule for the remaining trials and was successful. She was able to verbalize the rule quite adequately, and reasoned out her former misconceptions concerning the effects of single blocks and how to express the torque relation for a single side when there was more than one pile of blocks.

At this point the task was changed. It is not necessary to specify in detail what the subject did, since the behavior of this subject was the same for all problems, as well as being similar to the behavior of other subjects. The subject was instructed to predict the action of the beam and shift the plastic scale to balance the system. The first problems were simple ones. This subject, as well

as nearly all others, continued to calculate the torque for each side on all problems, whether simple or conflict. This type of action suggests Siegler may be wrong in some of his other assumptions concerning the internal problem solving behavior of subjects solving balance problems. Subjects who could use the torque rule seem to have done so in every case and did not make a decision on some simple basis when it was possible to do so. Thus, a decision tree is probably not the best method of charaterizing subject behavior.

The case study supports the notion that adult subjects not at rule III are quickly able to develop rule III knowledge. The subject's prediction on T5 suggests she was considering the interaction of weight and distance. By T12, she began to consider quantitative methods of relating weight ad distance. Thus, less than one-fourth of the total number of trials was spent advancing to rule III. This pattern seems to occur in nearly all subjects who begin at rule I or II.

Proceding from rule III to rule IV appears to involve much more complex decision making processes than mere random "muddling through". The set of predictions a rule III level subject gives may appear to be random though if they are not viewed as part of a sequencial process. The subject generally had a well specified reason for

eliminating at least one choice or for choosing a specific prediction. These reasons undergo a process of change through the training sequence. The number of reasons the person may use to determine the possibility of certain actions generally increases throughout the training session. Specification and use of rules is more likely when the balance beam balances (e.g., see trials T20, T29). The simpler reasons revolve around what happened in a sequence of trials. Later, they may be based on descriptive or quantitative predictions about the composition of the rule, such as the 2 to 1 rule. Others at this approximate level of difficulty include the 3 to 1 and 4 to 1 rules, and the distance averaging rule, which says the weight and distance of two piles with equal numbers of blocks may be considered as the total number of blocks at the average distance. Subjects, including the one presented in the case study, appear to have the most difficulty with problems containing more than one group of blocks on the same side. They often fail to apply what has been learned from the special case rules about relative torque for blocks which are on opposite sides of the balance beam to blocks which are on the same side.

The progression of predictions from rule III to rule IV for this set of subjects will be studied in more detail in a separate paper.