A STUDY OF CERTAIN INSTRUCTIONAL MATERIALS AS THEY INTERACT WITH SPATIAL AND NUMERICAL PRIMARY MENTAL ABILITIES

Thesis for the Degree of M. A.

MICHIGAN STATE UNIVERSITY

SARAH A. SPRAFKA

1969

ABSTRACT

A STUDY OF CERTAIN INSTRUCTIONAL MATERIALS AS THEY INTERACT WITH SPATIAL AND NUMERICAL PRIMARY MENTAL ABILITIES

By

Sarah A. Sprafka

Thirty fourth and thirty fifth grade children were given three subtests of the Primary Mental Abilities Tests. Based on scores on these tests, S's were assigned to one of four groups: Number high/spatial low, or number low/spatial high, fourth or fifth grade. Treatments consisting of two versions of a paper and pencil program dealing with the arithmetic of directed numbers were then administered to the groups. Half of each group learned by a spatial version of the program, the other half by a symbolic version of the program. A Lancaster χ^2 showed that for this particular sample there was no interaction between spatial and numerical Primary Mental Abilities and a spatial or symbolic approach to learning about directed numbers.

A STUDY OF CERTAIN INSTRUCTIONAL MATERIALS AS THEY INTERACT WITH SPATIAL AND NUMERICAL PRIMARY MENTAL ABILITIES

By

Sarah A. Sprafka

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Counseling, Personnel Services
and Educational Psychology

1969

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Dr. Lee Shulman, Dr. Stephen Yelon, Dr. Andrew Porter, Dr. Maryellen McSweeney, Mr. Bert Sterling, Mrs. Ann Reed, Mrs. Bessie Hetrick, Mrs. Gertrude Bissel, and Mrs. Georgie Dillon for their cooperation and assistance in the preparation of this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

In recent years interest in the individualization of instruction has vastly increased. Programmed instruction is used more and more in the classroom. In the absence of programmed instruction, individualization is often achieved using conventional textual materials, and audio-visual materials. The teacher does not get up in front of the class and make a presentation, but acts as a monitor for a class, each member of which is working individually. Some individualization is being used a bit differently in at least one other area. That is the area of mastery learning being investigated by Bloom (1968). He contends that mastery of a given amount and type of subject can be achieved by almost everyone, but it will take some students a little longer than others. To assure mastery, it may be necessary to make available to a student materials and modes of presentation which are different from (and supplementary to) the conventional classroom presentation, or perhaps an individualized approach using conventional materials. Cronbach (in Gagne, 1967, Chap. 2) suggests a number of approaches to implementation of such a scheme. The first decision must be whether everyone is going to learn everything to mastery, or if there are some things that everyone ought to learn to mastery (possibly reading and arithmetic) and other things which need not be completely mastered. For our purposes, let us decide that at least some aspects of simple mathematics ought to be learned to mastery. Bloom suggests that we achieve this through initial practice in one given mode with standard materials. If this does not work, increased time spent in practice as well as use of alternate modes and materials will help the student achieve mastery. Cronbach (ibid) does not favor simply extending the length of time of exposure, but instead strongly advocates the modification of instructional technique so as to capitalize on the student's abilities. He mentions teachers who have done this on an intuitive level, and notes that teachers who do this are prone to expecting too much from a student who seems to do certain kinds of things well. They may also try to discourage a student from trying to do something of the type he does not appear to do well. Whether a teacher is trying to "bring out the best in a student" or to help another student learn by proposing various study materials or using different teaching techniques, there is always a lot of guesswork involved. Which way of teaching would work best. Which materials would be most appropriate?

The guesswork arises partly because some of the available materials may be generally inadequate, as well as because they may not present the material to be learned in a way that the student can learn it. What is needed then is to determine what kinds of tasks the student does best, or identify his abilities. We can then select one or more of these abilities and design treatments we would expect to interact with them (see Gagne, 1967, chapter 2 again for more on this), i.e. create a situation wherein a student possessing a high degree of this ability would profit most from this treatment, and the student possessing little of this ability would profit least from it by comparison with others in the treatment group.

Identification of Aptitudes

If aptitude x treatment interactions are to be studied, aptitudes must be identified. For this purpose the best instruments available to us are multi-aptitude test batteries of which there are quite a number and quite a variety. These tests had their roots in an early--about 1920--interest in the specific abilities required to do certain jobs and excel in certain school subjects. Multiple factor analysis was used as a statistical method for separating out the various abilities tested by the individual test items.

One of the earliest and best known sets of tests is that which tests for Primary Mental Abilities, developed by Thurstone through factor analysis of 56 tests given to students at the University of Chicago. The test purported to identify and measure the extent of seven "Primary Mental Abilities": verbal, spatial, word fluency, memory, and reasoning. It was Thurstone's conviction that these seven abilities were "primary" in that they were fundamental abilities which combined to make up the aptitude for any complex intellectual task. Thurstone thought these abilities combined essentially additively and in different proportions depending on the individual. Those Primary Mental Abilities are still tested in the tests as used today. In their original form the tests were appropriate for high school and college students. They have since been adapted to apply to children ages five through seventeen. For the younger children some of the Abilities are not included. (Extensive reviews of the Primary Mental Abilities Tests may be found in the Fourth Mental Measurements Yearbook, Buros, 1954).

The Primary Mental Ability Tests (PMA) were intended to be used for guidance purposes (e.g. for advisement on possible school courses of possible interest to a student, etc.). There are a number of other multi-aptitude test batteries, two of which are quite well known and representative. The first of these is the Differential Aptitude Test (DAT). These tests do not attempt to identify and isolate any primary mental abilities. Instead they try to find complex abilities that relate closely to various categories of jobs and school requirements. The subtests are entitled Verbal Reasoning, Space Relations, Numerical Ability, Abstract Reasoning, Mechanical Reasoning, Clerical, Language Usage.

It is interesting to note that despite their different emphasis, the first categories of the DAT are much like those of the PMA, attesting to the influence of Thurstone's work. The DAT tests are aimed at the eighth grader through high school senior, and are also used for individual counseling purposes and for giving advice as to what line of work might be of most interest, or what sort of course work would be most profitable.

The last set of multi-aptitude batteries to be mentioned here is the General Aptitude Test Battery (GATB). It, too, shows a strong influence of Thurstone in that its subtests measure verbal, numerical, and spatial aptitude in addition to such abilities as clerical perception, motor speed, finger and manual dexterity. Valid for ages 16 and over, this set of tests is intended for finding people with the right combinations of abilities to fit certain jobs (an institutional decision is involved here) or for helping an individual decide what kind of job might be appropriate for him, and fitting him to that job (this involves individual decisions). Thus the GATB can work two directions, for institutional as well as for individual decision making.

Perusing these sketches of some of the major test batteries and their aims, one becomes aware of the variety of tests and purposes they may serve. These tests do have a couple of interesting things in common, however. With the exception of the PMA, the tests such as those represented by the DAT and GATB are applicable to children age

14 or 15 and older. This is probably partly because the different abilities do not emerge clearly before that age (Anastasi, 1967). It is also true that these abilities are quite unstable in the earlier years, and that little long range prediction can be done based on these aptitudes before Grade 11 (Cronbach, 1960). In Grades 7-10 short range advising as to possible courses which might be taken is possible, but not much more. Another thing that should be noted about these tests is that they are meant to be used under rather static circumstances. That is, they are meant to fit an existing individual into an existing situation, be it a suitable tenth grade shop course or a good job. They are not meant to be applied to predict how a situation might be changed to interact favorably with the individual and his aptitudes. Thus, if we are going to choose an aptitude and attempt to design a treatment that will interact with it, we have very little to base our decision on. We can only use common sense which tells us such things as "verbal aptitude should interact with verbal treatment" and "numerical aptitude should interact with a treatment emphasizing the use of numbers (and possibly number-associated symbols), and a spatial aptitude should interact favorably with a treatment using spatial representation of concepts." Common sense is not always right. The Spatial Relations scores on the DAT have been found to be reasonable-to-poor predictors of success in geometry. The Number scores were better predictors by far (Cronbach, 1960). Common sense would tell us the opposite

might be true. This warning should put us off making common sense decisions, but for want of a better basis for deciding, common sense is what will have to be used.

Studying Aptitude x Treatment Interactions

If we follow Cronbach's advice (Gagne, 1967, chapter 2), and adapt instructional techniques to fit the student, we will be studying more and more aptitude x treatment interactions. Following his advice again, we should identify one (or more) abilities in a student and purposely design a treatment that will interact with that ability. This treatment should be an extreme example of the type which we might propose for wider use. This is because, first, under laboratory conditions we are working with a limited sample. Secondly, as Cronbach also tells us (Gagne, 1967, p. 30), if we want to use aptitude information usefully in the adaptation of instruction, the interaction we get must be a disordinal one, i.e. the regression line that relates aptitude to payoff under one treatment must cross the one for the competing treatment. If the lines do not cross, then the result may be that one treatment is consistently better than the other, so we need not take the trouble to gather all that information about aptitudes.

There is also an important implication for the design of teaching materials in these interaction studies. Not all subject matter is amenable to being presented in such a way as to capitalize on (interact with) various aptitudes. This has been evident in at least two of the interaction studies done (Hamilton, 1969 and Tallmadge, 1968, to be discussed in Relevant Literature section). In both of these studies the subject matter interacted with the treatment, i.e. the students who got one treatment simply weren't learning the same thing as those who got the other treatment. Thus any effect due to ability or to teaching technique used in the two treatments were not readily evident.

The Aims Of This Study

In this particular study fourth and fifth grade children worked on either a symbolically or numerically oriented instructional sequence in the addition of positive and negative numbers. It was felt that those children identified as being more spatially then numerically apt would profit most from a spatial treatment, and those with greater numerical aptitude would profit from a treatment involving rules for the manipulation of signs and numbers.

In the spatial treatment no rules were given and the children were directed to "figure out the answer" using number line diagrams. In the "number" or symbolic treatment number line diagrams were used, but the students were never required to draw such diagrams. All responses could be got from just using the rule.

This procedure is somehwat like that suggested by Cronbach and Gleser when they advise us to accentuate any interaction "by deliberate design of the experimental course" if we hope to find any interaction effects at all (reference taken from Hamilton, 1969, p. 3). Cronbach also suggested that this might be necessary in his statement (mentioned above) that aptitude information cannot be employed usefully unless the result is the crossing of the aptitude and treatment regression lines. It is probable that an optimum treatment of the addition of signed numbers would involve the use of both rules and the number line diagram. However, for our purposes the two were separated.

The verbal dimension was not considered in this study for a number of reasons discussed under the "assignment of subjects to treatment" heading.

The null hypothesis associated with this study was that there would be no interaction between aptitude and treatment.

CHAPTER II

REVIEW OF LITERATURE

Relevant Literature

A study of aptitude and treatment interaction is essentially a study of a particular kind and use of individual differences. The literature relevant to individual differences is vast, and much of it is not relevant to this particular study. On the other hand, the literature relevant to aptitude x treatment interactions (discussed below) is minimal.

Aptitude x treatment interactions are a way of studying individual differences as they relate to learning, especially school learning. A discussion of individual differences that relate to learning is given us by Jensen (in Gagne, 1967, Chapter 6). He chooses to postulate two kinds of learning-related individual differences--intrinsic and extrinsic. He feels that attitudes and personality traits are examples of extrinsic abilities, i.e. of "those subject variables which operationally bear no resemblance to the learning process . . ." (p. 121). Chronological age, sex, IQ, and mental age would also be included in this category. Jensen then goes on to talk about intrinsic individual differences, or those differences which are inherent in learning. He suggests three major classes of these variables: types of learning (conditioning, motor, rote, etc.); procedure (pacing, distribution of practice, etc.); and content and modality (verbal, numerical, spatial, flexibility, etc.). He suggests that we then cross each member of one category with a member of each of the other two categories (apparently by creating learning tasks that reflect these subcategories) and see how much within subject variation we get as subjects interact with this combination of categories. He admits it will be quite a job. The outcome of this investigation would be information about the "underlying factors or basic processes" (apparently inherent in the learner) which cause the patterns of interaction among the task characteristics as delineated by the three categories mentioned above. Thus Jensen is not hypothesizing certain inherent abilities and attempting to design tasks which will cause them to be expressed. He is, instead, taking what he knows about tasks and developing controlled types of tasks which, when presented to subjects will yield information about what the abilities actually are.

With this in mind he says that a <u>task</u> is spatial, verbal, numerical, etc., not a person. Thurstone (1938) attempts to prove that spatial, verbal, numerical, and so on refer to attributes of people and not of tasks. He reasons thus: some people seem to be able to do some types of things much better than others. Some people can visualize things in space. Others don't visualize as well, but can do just about anything that has to do with computation. We may, from these observations, hypothesize that the former has some sort of visualizing ability and the latter has some sort of computational ability. We can devise tasks which require these two types to display their abilities. This done we find that the visualizer does better on the visualization test and the calculator does better on the computation test. From this we are allowed to infer that people who do very well on visualization tests and less well on computation tests are more visually than computationally oriented.

Thurstone hypothesized a large number of mental activities (getting his ideas from items used in general intelligence tests) and categorized them. When testing was completed and factor analysis was performed, he could claim to be able to identify the Primary Mental Abilities, i.e. a certain set of seven abilities which different people possessed to different degrees and which could be said to account for much of the individual difference variance in the performance of certain complex tasks.

It may indeed by the case, as Jensen states, that the innate abilities are yet to be identified. It may also be the case that the identified abilities (spatial, verbal, number, etc.) are unique to the learner and not a part of the task. More important to the present study, however, is that the latter interpretation is the most useful one. Tests have been developed which supposedly identify the degree to which an individual possesses these abilities. The results are quantifiable and decisions can be based on them.

Once the existence of individual differences (i.e. different abilities) is established, how can the different abilities be related to learning? Abilities which are related to different kinds of learning have been tentatively identified and their relationship to the learning process has been discussed. Some abilities identified as relevant to verbal learning are such things as rote memory, and span memory (Fleishman in Gagne, 1967, p. 62). Other individual difference variables are suggested by Jenkins (also in Gagne, p. 48) such as the "ability to form new associative bonds or habits." Fleishman (Gagne, chapter 8) has done extensive investigation into the abilities involved in learning motor skills. He has identified such abilities as manual dexterity, finger dexterity, control precision and several others. He has also found that the combinations of abilities contributing to performance change as practice continues and that the contribution of "non-motor" abilities--such as verbal or spatial--decreases with increased practice, whereas pure motor abilities contribute progressively more (Gagne, p. 179). Whether these phenomena (i.e. the increase or decrease with practice of different contributing abilities) occur during the learning of verbal skills (e.g. small children learning to read) has apparently not been investigated.

Another interesting investigation of abilities has been done by Fredericksen (annual Review of Psychology, 1969, p. 357). He feels that the relationship of the abilities an individual possesses to a learning task occurs through strategies. Depending on what ability or abilities may be dominant in an individual, he will choose different strategies through which to learn a task. By his choice of a specific learning strategy, the learner essentially restructures the learning task to conform to his abilities (and choice of strategy). In addition, whether he possesses certain abilities to a greater or lesser extent will influence his performance using the strategies he has chosen. How Frederickson defines ability is not clear in the reference cited. One ability mentioned is "associational fluency."

Adapting instruction to abilities is of direct interest to this study. Bloom suggests indirectly that we adapt instruction to individuals when he delineates his mastery learning procedures (Bloom, 1968). Although he does not mention adaptation based on previously obtained ability measures, this could possibly facilitate the carrying out of a mastery learning program. The vast number of abilities which have been proposed might make it difficult to choose which ones to adapt to. The Primary Mental Abilities are a possibility.

Cronbach also discusses the adaptation of instruction to abilities (Gagne, 1967, chapter 2) and makes specific reference to the aptitude x treatment interaction type of study. Cronbach states that one way of dealing with individual differences is to teach different people in different ways. It remains then to define the differences between the people and their degree, and to define how the differences are going to be treated. He suggests that the first task might be dealt with through the scheme suggested by Jensen (see above). However, warns Cronbach, "aptitude information is not useful in adapting instruction unless the aptitude and treatment interact" (p. 30). The regression lines corresponding to the two competing treatments must cross. If the information about differences in aptitude is not good or if the treatment and the aptitude do not interact, then the teacher would do well to maintain the status quo. As Cronbach says: "Modifying treatments too much produces a worse result than treating everyone alike" (p. 30).

Few aptitude x treatment interaction studies have been done, and in those that have, the desired interaction has not been obtained for various possible reasons. An experiment by Bush *et al*. (1965) attempted to find an interaction between the learner and the mode of instruction. Abilities were measured by the California Achievement Test. The Mathematics Fundamentals and Reading Vocabulary subtests were used. Differences between these two scores for each individual were correlated with his pretest-posttest gain relative to each treatment condition. It was found that those subjects who scored significantly higher on a reading vocabulary test than they did on a mathematics fundamentals test "tended to acquire more learning" from a lecture-like treatment than did students whose mathematics scores were higher than their reading scores. Conversely, those students who showed relative strength in mathematics "tended to do relatively better" in a laboratory-like situation than did those students whose reading scores were higher than their math scores. This was interpreted as meaning that "students with relative strength in reading vocabulary are superior to students with relative strength in mathematics fundamentals when both are required to learn from instructional situations that are highly verbal" (p. 11). There were a number of drawbacks to this study from a design point of view, not the least of them being the diversity of subject matter and a number of confounding variables (e.g. content with treatment and degree of participation with method).

Another study which sought an aptitude and interest x treatment interaction was done by Tallmadge (1968). He chose two procedures by which to teach the solution of maneuvering board problems to Air Force recruits. One method emphasized Gagne's type 3 (chaining) learning and required that the subjects memorize the sequence of steps involved in the solution of these types of problems. The competing treatment emphasized Gagne's type 7 (problem solving) learning. In it the subjects were taught the underlying principles and concepts, and the rationale of the procedure for solving these problems.

Many ability and interest tests were given the subjects among them the Kuder Preference Test, and spatial orientation and visualization tests. The result was that over one hundred abilities (and levels of them) were identified. The subjects were rated "high" or "low" on 16 of these measures of both ability and treatment.

Tallmadge found a main effect for treatment--the Type 7 method being superior over the Type 3 method. But there was no interaction between aptitude or interest and treatment. Tallmadge gives as possible reasons for this, poor choice of competing training methods, measurement of inappropriate aptitudes and interests, or other interactions between materials and training that washed out any aptitude x treatment interaction.

One last attempt to find an aptitude x treatment interaction will be mentioned. It is the study done by Hamilton (1969). She administered three cognitive tests to junior high school students: spatial visualization, spatial orientation, and wide-range vocabulary. The students then worked on either a verbal or primarily pictorial version of an instructional sequence in the geometry of crystals. The result was that the pictorial treatment produced higher mean post test scores than the verbal treatment. But no interaction between aptitude and treatment was found.

The study reported here investigated an interaction between spatial or numerical ability and a spatial or symbolic treatment of an arithmetic task. Spatial and numerical abilities as they relate to success on mathematical tasks have been discussed by Cronbach (1960) and Murray (1949). Murray correlated numerical, verbal, spatial relations, and reasoning test scores with scores on a geometry test, and with final grades in geometry. It was found that the numerical score correlated most highly with the final grade, and the spatial relations score correlated least with both the geometry test scores and the final grade.

Cronbach (1960) quotes findings that, although a spatial relations test score correlates positively with geometry scores, mechanical drawing ability, etc., it is the numerical ability measure which is the better predictor of success in geometry.

CHAPTER III

METHODS AND MATERIALS

Plan of the Study

No pilot work was done on the materials, so a number of assumptions had to be made. First it was assumed that there would be no sex differences. Secondly it was assumed that since the children in the sample were middle to upper middle class whites in a primarily middle to high SES school that they would be quite verbal. Hence none of the vocabulary used in the program would have to be explained at any length (e.g. terms like "positive," "negative," and "direction"). Thirdly, it was assumed that the fifth grade would fare systematically better (perhaps because they had had more practice with the subject matter) than the fourth graders. Lastly it was assumed that neither fourth nor fifth graders would have studied positive and negative numbers thoroughly. Therefore it was assumed that pretest scores would be near 0, whereas post test scores would show some uniform degree of dispersion.

The only effect that was sought was an interaction of ability with treatment (hopefully a disordinal one) for the fourth and fifth grade. The design was such that main treatment and aptitude effects would be detected, however.

Instructional Programs

Two instructional programs were prepared by the experimenter. One treatment was very dependent on use of number line diagrams to obtain a correct response. This was the "spatial" treatment, hypothesized to interact favorably with spatially apt subjects. The "symbolic" treatment, aimed at more numerical children involved application of a verbal rule to obtain a correct response.

The sources for construction of these programs were:

> <u>High School Mathematics: Course 1</u> by Beberman and Vaughan (1964)
>
> <u>Discovery in Mathematics, a text for teachers</u> by Davis (1964)
>
> <u>Modern Mathematics, Book 1</u> by Eigen <u>et al</u>. (1961)
>
> <u>Algebra I</u> by Smith <u>et al</u>. (1960)
>
> <u>Advancing in Mathematics (J2)</u> by Morris and Topper (1964)

The objectives for the teaching sequence (expected to be attained by use of either program) were:

> At the end of the sequence the student will be able to:
>
> 1. find the sum of two positive numbers
> 2. find the sum of two negative numbers
> 3. find the sum of one positive and one negative number

The numbers themselves did not exceed 15, so that any effect would be due to whether the treatment was spatial or symbolic and not depend on whether one subject could add or subtract better than another. The two programs were identical in the first half. Since the experimenter had been told (by the fourth and fifth grade teachers) that some of the children "had had" some of this material, whereas others "had not had" any of it, it was deemed wisest to include a quick run down of what a positive and a negative number were and how they were expressed on the number line. The first half of the program was meant to serve this purpose. It would probably act as review for many of the children, and be new material for the others.

The second half of each program dealt with the addition of positive and negative numbers, and the expression of equations on the number line. The symbolic (number oriented) program demonstrated the use of the number line for the expression of equations, but never asked the student to diagram an equation on the number line. The emphasis in this program was on finding the sums of positive and negative numbers using one of two rules. Thus by the end of the program the student would (supposedly) use the appropriate rule to find the sum of two numbers of identical sign or of two numbers of opposite sign.

The spatial program in the second half continued the use of the number line for the expression of equations. The student was required to draw a diagram and from the diagram he was to find the solution to the equation. The spatial program gave no rules for finding sums. Thus by the end of this program the student should be able to find the sum of two numbers of the same sign or of two numbers of opposite sign by drawing a diagram of the equation and getting the solution from the diagram.

It was stated earlier that these two programs were identical in content and differed only in that one was spatial and the other was symbolic. They were identical in that all students found the sums of exactly the same number of pairs of integers. The distribution of equations involving two numbers with opposite signs and those involving two numbers with identical signs was the same for both programs. A possible uncontrolled difference between the programs arose after the first half. The spatial program continued using the number line as a central reference point, and required that all tasks be done using the number line. The symbolic program shifted emphasis from the number line to rules. This shift may have caused the symbolic student to learn something different or to learn more than did the spatial student (i.e. student with spatial program). Such a phenomenon is almost impossible to investigate and was not of specific interest in this study. It could be investigated if one wished to do so.

Subjects

The subjects for this study were rather difficult to choose. In the Research Design outline (see Appendix C) which served as the basis for this study, it is suggested that eighth graders be used, with the assumption that they do not yet know how to add directed numbers. Talks with eighth grade teachers yielded the information that their students "had had" all that long ago. It was finally necessary to go as low as fourth and fifth grade, and even those teachers said their students had been exposed to this kind of material. As it turned out, pretest scores were quite high. Another problem arose when it came to using such young children, and that was that differential aptitudes are not particularly well defined at that age. However, despite this drawback, fourth and fifth graders were used. Thirty fourth graders and thirty fifth graders participated in the experiment.

Aptitude Measures

Since the children were so young, the SRA Primary Mental Abilities Tests were used. This is the only widely distributed multi-aptitude test battery which provides tests appropriate to that age group. Three tests from that battery were given: Verbal meaning (V), Number Facility (N), and Spatial Relations (S).

Outcome Measures

One post test was given to all the students. This test also served as the pretest. The objectives of the teaching sequence were that the student would be able to find the sum of two numbers with identical signs and of two numbers with opposite signs. The post test asked him to do that. It consisted of nine equations:

One of the type $^{+}2 + {}^{+}2 = \square$

One of the type $^{-}2 + {}^{-}2 = \square$

One of the type $^{+}1 + {}^{-}2 = \square$ (smaller positive number first)

Two of the type $^{-}2 + {}^{+}1 = \square$ (larger negative number first)

Two of the type $^{+}2 + {}^{-}1 = \square$ (larger positive number first)

Two of the type $^{-}1 + {}^{+}2 = \square$ (smaller negative number first)

Thus each permutation of possible positive/negative combinations was treated, although one was represented only once. In addition to solving the equations, the children were asked to draw diagrams of the first three equations (number lines were provided). They were told they could draw diagrams of the second three equations if they wanted to. Number lines were not provided in this case.

Assignment of Subjects to Treatment Groups

Four classes participated in the experiment--two fourth grades and two fifth grades. Grade level was used as a blocking variable as it was thought that fifth graders would perform systematically better than fourth graders. Children in each grade were assigned to treatment groups according to their scores on the Number Facility and Spatial Relations tests. Scores on the Verbal Meaning test were not considered for two reasons. The experimenter was interested in those students who had significant differences between their spatial relations and their number facility scores. The verbal meaning scores were to be used to assure homogeneity on at least one dimension. It was decided that only those children above the verbal median (national norms) would be used. However, when it came time to do the assigning, it was noted that to exclude those children below the verbal median would considerably reduce the size of the treatment groups. This was partly because the children were not as highly verbal as had been anticipated. In the fourth grade a total of 10 of the 36 tested were below the national verbal median. In the fifth grade a total of 5 of the 40 tested were below the national verbal median.

A second and more interesting reason for not excluding those children was that all but one of the 15 showed a difference between their spatial and numerical scores (see below for definition of what constituted a difference). Of those, 11 showed a higher spatial aptitude than verbal aptitude. Those 11 children were almost 85 percent of the total number of children below the verbal median. Of those above the verbal median, only 44 percent had a higher spatial aptitude than they did a numerical one. Thirty-eight percent had a higher numerical aptitude. The others showed no difference in aptitude. Thus it can be seen that exclusion of the children below the verbal median would have meant a noticeable reduction in size of treatment groups.

The students were assigned to one of the two treatment conditions if they showed a difference of one stanine or more between their number facility scores and their spatial relations scores. The national norms for the PMA are reported in stanines. Assignment was done based on the national norms and a difference of one national norm stanine between number and spatial score was the criterion for being assigned to a treatment group. Usually a difference of two stanines (approximately one standard deviation) is considered significant. However had the criterion for assignment been a difference of two stanines, the sample size would have been reduced to 28 and of the eight cells, one would have contained only two subjects. This was clearly undesirable. Hence a difference of one stanine was considered acceptable as a criterion for assignment. Once all the children had been identified as "spatial high/number low" or spatial low/ number high," the lists of names were randomized. The first half of each list was assigned to one treatment, and the second half to the other treatment. In the case of an uneven number of subjects, the last name on the list was dropped.

Experimental Schedule

Aptitude tests were given first--one week before treatment. The three subtests plus instructions took a total of about one and a half hours to administer. Pretest, instructional program, and post test took from 45 minutes to 2 hours, depending on the speed at which the student worked.

The experimental design was not matched to the classrooms' schedules. Individualization of instruction was not practiced regularly in any of the four classrooms. Thus the experiment constituted somewhat of an intrusion on the regular classroom proceedings. Apparently this particular school is accustomed to being "experimented on" and the teachers gave their consent. There was little or no hostility expressed at the interruption. If such a program were to be incorporated in a school similar to this one, it should probably be broken up into smaller sections, taking 20 to 30 minutes each to complete.

At the beginning of the treatment period the experimenter explained how the materials were to be used. An explanation of the pretest--treatment--post test procedure was also given. Each programmed booklet had on it the name of the child to whom it was destined. The booklets were handed to each child as he turned in his pretest. Children in the spatial treatment (who had to draw number line diagrams) took 10 to 15 minutes longer on the average to complete the program than did those in the symbolic treatment. The class who "had not had" any of the material treated in the teaching sequence (and who averaged near 0 on the pretest) took the longest to complete the treatment. Before any further use of these materials with naive subjects, the materials must be broken into shorter sections, and completed over three or four work periods.

The children who finished early did homework assignments or went to the library. Throughout the experimental period the teacher was present in each class.

CHAPTER IV

RESULTS

Analysis was performed as follows: reliability of predictors; intercorrelation of numerical and spatial subtests for this sample; tests for main and interaction effects on addition part of post test and on diagram drawing; z scores (spatial and number) for fourth and fifth graders.

The null hypothesis that there would be no interaction between treatment and aptitude was not rejected. The null hypothesis that the spatial treatment and the symbolic treatment would not differ in their effect was also not rejected.

Table 1 shows the number of subjects who were above and below the median on the addition section of the post test (see Appendix A for actual scores). These frequencies were used as the observed frequencies in a Wilson procedure for χ^2 analysis of this type of data (see Appendix B for short explanation of Wilson procedure). The original intent had been to analyze the data using an analysis of variance technique. However an obvious ceiling effect was observed (there were 35 out of 60 scores of nine, the maximum score possible) and it was deemed wiser to use a median--or cut-off

TABLE 1.--Scores of 9 and Less Than 9 on Addition Test (Frequencies) for Each Treatment Condition.

		Treatment	
		Symbolic	Spatial
		Score <9/score = 9	Score <9/score = 9
Grade 4	Sp Hi / NumLo	4/4	5/3
	Sp Lo / NumHi	2/5	4/3
Grade 5	SpHi / NumLo	5/5	3/7
	SpLo / NumHi	1/4	1/4

TABLE 2.--χ^2 Values for Main and Interaction Effects--Addition Test.

	df	χ^2
Rows (grade and aptitude)	3	3.720
Columns (treatment)	1	.069
Rows X Columns	3	1.774

point--test which would essentially determine if there were more scores of nine under any one circumstance than under any other. The non-significant χ^2 showed that this was not the case. The technique might be criticized for not considering a good deal of information due to the dichotomization of results. However conservative it may be, it is one of the few honest tests used in this sort of situation. Analysis of variance or correlation techniques are based on assumptions of normality of distribution and equality of variance across treatment groups. Although the latter assumption may have been satisfied, the distribution of scores was distinctly not normal. The analysis of variance technique is quite robust to violation of the assumption of normality. Nonetheless, to violate this assumption means to err on the side of conservatism. Therefore it was felt that use of an initially conservative test which violated no assumptions would give equally believable results.

Table 3 and 4 show the results of the diagram drawing test. The frequencies were dichotomized based on whether a score was three or not three. Table 3 shows the observed frequencies in each cell. Table 4 shows the χ^2 values obtained from those frequencies. It will be noted that there is no interaction effect or main effect for aptitude. However there is a significant main effect for treatment. This tells us that those students who had the spatial treatment could diagram the equations significantly better than those who had had the symbolic treatment. That is not a surprising

TABLE 3.--Scores of 3 and Less than 3 on Diagram Test (Frequencies) for Each Treatment Condition.

		Treatment	
		Symbolic	Spatial
		Score <3/Score = 3	Score <3/Score = 3
Grade 4	SpHi/NumLo	7/1	3/5
	SpLo/NumHi	4/3	4/3
Grade 5	SpHi/NumLo	9/1	3/7
	SpLo/NumHi	3/2	2/3

TABLE 4.--χ^2 Values for Main and Interaction Effects--Diagram Test.

	df	χ^2
Rows (grade and aptitude)	3	5.51
Columns (treatment)	1	12.97 *
Rows X Columns	3	.357

*Significant at .01 level.

result. The children in the spatial treatment group had had a good deal more practice making this kind of diagram than had the children in the symbolic treatment group. The latter had had no practice doing this at all.

The reliabilities and intercorrelations of the predictors will be discussed next. Table 5 shows the reliability for this sample of the spatial subtest (fourth and fifth grade) and for the "number sense" section of the Number Facility subtest. The reliabilities reported by the publishers of the Primary Mental Abilities Tests are included in this table. As may be seen those for the spatial subtests are somewhat lower for this sample than for the national norm group. There has been some criticism of these tests in their reporting of reliabilities (see Anastasi's review of PMA tests in Fourth Mental Measurements Yearbook, p. 716). Although this criticism occurred some time ago, no more recent bases for judgement of this aspect of the test can be found. Both the spatial and numerical (both sections) subtests are speeded, the second section of the numerical test (addition) being extremely so. Criticism occurs when Kuder-Richardson techniques are used to calculate the reliability of the speed tests. This was apparently done by the publishers. Some qualification of the criticism may be in order. Only under certain circumstances does speededness have to be corrected for (see Guilford, 1954, p. 392). In the case of the spatial test, it is not so speeded that a correction need be applied and the split half technique for

TABLE 5.--Reliabilities of Spatial Relations and Number Sense Subtests for Grades 4 and 5.

	Reliabilities			
	Spatial Subtest		Number Sense Test	
	Local	National	Local	National*
Grade 4	.74	.96	$.63^{\dagger}$	.89
Grade 5	.73		$.74^{\theta}$	

*This is the reported reliability for the two sections of the Number Facility Test: Number Sense and Addition.

†Corrected for speeding - $\frac{SDu}{SDw} = .65$

$^{\theta}$Not corrected for speeding - $\frac{SDu}{SDw} = .18$

finding reliability may be used. In the case of the number sense test, the split half technique was appropriate for fifth graders, but a correction for speededness had to be applied to the reliability estimates for fourth graders. Guilford (ibid.) advises us that there are certain circumstances under which even a correction of the split half technique is not appropriate. In cases of extreme speeding (see Guilford, p. 392, for what is meant by "extreme speeding") no estimate of reliability can be found which does not involve inordinate amounts of calculation. The addition section of the PMA numerical subtest is extremely speeded. More complex calculation than was warranted by this report was required to find the reliability of this section for the sample under consideration.

It has probably been noted that reliabilities for the local sample were reported relevant to grade level whereas those for the norm group were reported relevant to age. Grade level is used here as a basis for grouping because it is felt that in this particular circumstance a grade is a more homogeneous group than an age group. In other words, it would seem that nine and ten-year-olds in the fourth grade would be more alike than ten-year-olds in the fourth and fifth grades. Unfortunately this homogeneity was not tested so the use of grade as a basis for grouping may not be a legitimate one. It will be used nonetheless.

Table 6 shows the means, standard deviations, and intercorrelations of the spatial and numerical tests for this sample as well as those for the national sample. It was noted that the intercorrelation for the fourth grade is a good deal higher than that for the fifth grade. It is doubtful that this occurs due to emergence of distinct abilities between the fourth and fifth grades. It is more probably due to chance.

It had been planned to report a correlation of difference-between-aptitude-tests-scores (z scores would be used) with post test scores to see whether those children who were more spatially (or numerically) inclined, and who were in the treatment group consonant with their salient aptitude, had scored consistently high on the post test. This would be opposed to those who had been in a treatment group which was not consonant with their aptitude. For this purpose the fourth and fifth grades were pooled. Based on the results of a χ^2 analysis (again the Wilson procedure) it was found that the fourth and fifth grades did not differ significantly on their post test scores. This finding warranted pooling the two grades based on post test results (see Table 7). Then z scores were calculated for each treatment condition. It was found upon taking the differences between the z scores that over half the subjects had been misassigned to treatment groups. It may be remembered that assignment was done based on stanine scores used in reporting the national norms for the spatial and numerical subtests. Tables D-1 through D-4 (see Appendix D) show two z scores for each subject, the

TABLE 6.--Intercorrelations of Spatial Relations and Number Facility Subtests for Fourth and Fifth Grade.

	Spatial Subtest Mean	Spatial Subtest SD	Number Subtest Mean	Number Subtest SD	Intercorrelation of spatial and number
Grade 4	14.30	4.12	29.22	5.87	.401
Grade 5	15.45	4.00	31.52	5.91	.199
	Reported for National Sample				.096

TABLE 7.--Frequencies and χ^2 Value for Main Effect of Grade Level--Addition Test.

	Treatment	
	Symbolic	Spatial
	Score <9/score = 9	Score <9/score = 9
Grade 4	6/9	9/6
Grade 5	6/9	4/11

	df	χ^2
Rows (grade and aptitude)	3	3.720
Columns (treatment)	1	.069
Rows X Columns	3	1.774

difference between the z scores (a negative difference indicating a misassignment), and stanine scores on which assignment was based. When these discrepancies were found the correlation was decided against as it would not yield the desired information. It was noted that over half (33 out of 60) the subjects were misassigned (based on the difference between their z scores). That left 27 who were correctly assigned. A quick calculation revealed that, although somewhat worse than chance, the correct assignments were not significantly fewer than chance. Significantly worse than chance would have meant 36 misassignments and only 24 correct assignments.

It would be advisable in any future study of this kind to base assignment to treatment groups on differences between standard (z) scores for the sample in question.

CHAPTER V

DISCUSSION

The hypothesis that there would be no interaction between treatment and ability was not rejected. There was a main effect of treatment found for the diagram drawing task. This was rather predictable since those children in the spatial treatment group had had more practice at drawing diagrams than the other subjects. The study was not attempting to find out if one treatment was better than the other. The hope was that one treatment would be better for one group of subjects and the other would be better for the other group. During development of the treatments it was noted that neither treatment was easier to develop, nor did either one or the other seem more appropriate to teaching this subject matter. The symbolic treatment, with some aspects of the spatial treatment added in, is the approach most generally used in arithmetic texts. This made the two a little hard to separate, and when finished, both treatments seemed to be "lacking a little something." It is possible that the optimal sequence for the teaching of this subject matter is the combination of the spatial and symbolic approaches--as suggested by the arithmetic texts. It would be interesting to investigate the extent to which spatial and symbolic treatment factors contribute to the optimization of learning for this subject--this as opposed to developing separate treatments for different people in an effort to maximize learning.

The Criterion Test

The criterion test was intended to measure acquisition of the final objective--that is it attempted to measure how well the children could add positive and negative numbers. Given the ceiling effect obtained by this test, the actual effectiveness of the two teaching sequences might have been better measured by a different test. The test could be lengthened and diversified somewhat. The section requiring the drawing of diagrams might be retained. Rather than testing only the final objective, questions concerning intermediate objectives (e.g. identification of positive and negative number, corresponding points and direction on number line, etc.) might be included in the test. Verbal tasks (story problems) could also be included. The subjects would be measured on a verbal dimension, and correlations could be found between story problem performance, overall test performance, and verbal and number test scores. One last thing that might be added to the test would be a request for a statement of rules governing the addition of positive and negative numbers. This would be of interest for at least two reasons. It would answer the question of whether children given the rule remember the actual verbalization or see the application and continue to apply it while not being able to verbalize it. Secondly it would tell us whether the children (at least some of them) who were not given a rule generated one themselves.

Predictive Power of Tests (Predictors) For Fourth and Fifth Grades

It has been stated earlier that the abilities (as tested by PMA type paper and pencil tests) do not emerge until high school age. This might lead us to believe that the predictive power of these tests for fourth and fifth graders is rather low, either because the abilities are truly not differentiable at this age or because the paper and pencil test does not find these abilities in a young child even if they are there. It could be hypothesized that rather than paper and pencil tests, some sort of tasks would help identify these abilities if indeed they were differentiable. One might study Piagetian type tasks appropriate to the various developmental stages and determine how loaded they were with primary mental ability (or other yet unidentified) factors. Then again the tasks could be modified so they would be loaded with one or another of these factors.

Such an undertaking would be a considerable one and might encounter some problems, not the least of these being that the factors comprising the abilities themselves are still in a state of flux.

It has also been stated earlier that the number factor appears to have the greatest predictive power for high school boys doing a geometry task. Whether this would apply to younger children doing a different kind of mathematical task deserves further investigation.

Possible Reasons for Inconclusive Results

Throughout this report mention has been made of factors which might have contributed to the inconclusiveness of the present study. Those factors will be summarized here:

1. The abilities on which the treatments sought to capitalize are not discernible at this age.

2. Possibly the numerical test items do not bear very much relationship to success in manipulating symbols. Although Gagne (see Appendix C) states that "high numerical ability should facilitate the learning of symbolic concepts" this is not necessarily the case. It is also possible that, although there may be a numerical/symbolic relationship, the symbolic treatment in this study did not incorporate that kind of symbolic concepts.

3. It is possible that the spatial relations test does not predict success on the kind of spatial task used in this treatment.

4. The criteria for assignment to one or the other treatment group were not widely enough separated. Of the 60 subjects assigned, 31 had a number/spatial stanine difference of less than two (i.e. less than one standard

deviation difference between the two scores). A larger sample might allow for the assignment of only those subjects with a difference of one standard deviation or more.

5. Assignment to treatment groups should be based on local norms (z scores or percentile rankings) or national test norms if these are reported in percentiles. Assignment can be based on national norms only if it has been ascertained that the set of scores obtained from the local sample does not differ significantly from those used to establish the national norms. A procedure using national norms apparently vastly different from the local norms (and which are reported as stanines) resulted in a worse than chance assignment of subjects to their respective treatment groups.

6. The post test did not test completely the effect of the treatments. The test was possibly too short as well as too homogeneous.

CHAPTER VI

CONCLUSIONS

The present study has made us aware of the complexity involved in the study of aptitude x treatment interaction. We found through this study that development of abilities may or may not occur at a young age. We found that predictive power of the tests currently available needs more investigation. We also found that many children were not strongly differentiated in the abilities tested which raises a question about the very value of trying to adapt different kinds of instructional treatments to identified aptitudes.

The study raises the question of the feasibility and economy of such an endeavor. Should we go to the trouble of developing special instructional materials that will capitalize on identified differential aptitudes or should we try to develop a number of "optimal" treatments that will reach everyone to a greater or lesser degree?

If we are to continue investigating differential aptitudes in the interest of developing instruction to capitalize on them, we must concentrate very closely on the definition of already identified aptitudes as well as the identification of new aptitudes.

If we are to accept the existence of differential aptitudes (and they are becoming more widely accepted), we must, after defining them, clarify how separated the aptitudes are from one another, and which ones share enough factors that for instructional purposes they might be grouped.

We are trying to pair a student having certain aptitudes with an instructional treatment that capitalizes on those aptitudes. Too many "important" aptitudes can lead to the desire to create too many "appropriate" instructional treatments. Rather than being concerned over dealing with every possible aspect of this highly complex relationship of aptitude and treatment, we should perhaps attempt to clearly identify the elements involved and simplify the structure down to something we can work with both physically and economically.

Trait by treatment interactions appear to have a healthy future. However at present we seem awed by the very complexity of the situation. The complexity does exist as demonstrated by the present (and other) studies (e.g. Hamilton, 1969). Our first task then is to reduce the complexity by identifying the elements. What are the aptitudes which can be addressed by different instructional procedures? What subject matters call most on which aptitudes? If many aptitudes are identified, how closely related are they to one another, and can they be pooled for instructional purposes?

Once we have identified our field of operation, only then can we launch into large scale development of differential instructional materials. We have a long way to go.

Possibilities for Future Research

Rather than looking for an interaction of treatment and ability, one might investigate a main effect for treatment with children even younger than fourth grade. Using Bruner's ideas that we go through three levels of representation, enactive, iconic, and symbolic, we could hypothesize that the more spatial (iconic) treatment would work better with even younger children (second and third graders) than would the symbolic treatment. Given that there were high pretest scores in the fourth and fifth grades, introduction of this material can (and does) occur in the third grade. Some mention of positive and negative numbers is made in the first grade. When this occurs, we could investigate treating it iconically--or even enactively--rather than in the conventional manner which is symbolically.

As mentioned earlier one might advance toward resolving the controversy over whether abilities emerge before high school age by developing tasks to replace the paper and pencil tests now used to identify the abilities. One might--and should--investigate the usefulness of the existing tests (intended for counseling) for predicting treatments, and their appropriateness to aptitudes. It would seem that tests which were more subject matter specific (or at least more specific to school like tasks) might be more powerful for matching student to treatment.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anastasi, Anne. Psychological Testing. Macmillan, 1965.

Annual Review of Psychology. Vol. 20, 1969.

Bloom, Benjamin S. Learning for Mastery. Evaluation Comment. May, 1968, Vol. 1, Num. 2.

Buros, O. K. (ed.). Fourth Mental Measurement Yearbook. Gryphon Press, 1954.

Bush, Wilma Jo; Gregg, Dolores Kathryn; Smith, Edgar A; and McBride, Coit B. Some Interactions Between Individual Differences and Modes of Instruction, Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, December, 1965.

Cronbach, Lee J. Essentials of Psychological Testing. Harper Brothers, 1960.

French, J. W. "How Important is Spatial Ability." Contemporary Psychology, 1965, Vol. 25, p. 9-28.

Gagne, Robert M. Ability differences in the learning of concepts governing directed numbers, in Research Problems in Mathematics Education. Cooperative Research Monographs. No. 3, 1960, pp. 112-113.

Gagne, Robert M. (ed.). Learning and Individual Differences. Charles E. Merrill Book, Inc., 1967.

Chapter 2: Cronbach, Lee J. How Can Instruction be Adapted to Individual Differences?

Chapter 4: Anderson, Richard C. Individual Differences and Problem Solving.

Chapter 6: Jensen, Arthur R. Varieties of Individual Differences in Learning.

Chapter 8. Fleishman, Edwin A. Individual Differences and Motor Learning.

Guilford, J. P. Psychometric Methods. McGraw-Hill, 1954.

Hamilton, Nancy R. Differential Response to Instruction Designed to Call upon Spatial and Verbal Aptitudes, Stanford University, February, 1969.

Murray, John E. Analysis of Geometric Ability, Journal of Educational Psychology, 1949, Vol. 40, p. 118-124.

Rao, C. R. Advanced Statistical Methods if Biometric Research. New York, Wiley, 1952.

Tallmadge, G. K. Relationships between training methods and learner characteristics. Journal of Educational Psychology, Vol. 59, No. 1, 1968, pp. 32-36.

Thurstone, J. L. Primary Mental Abilities. Psychometric Monograph #1. University of Chicago Press, April, 1938.

Wilson, Kellogg V. A Distribution--Free Test of Analysis of Variance Hypotheses. Psychological Bulletin, Vol. 53, No. 1, 1956.

Math Texts

Beberman, Max and Vaughan, Herbert E. High School Mathematics: Course 1. D. C. Heath and Co., 1964.

Davis, Robert B. Discovery in Mathematics, a text for teachers. (Madison Project), Addison-Wesley Pub. Co., Inc., 1964.

Eigen, Lewis D. and Kaplan, Jerome D. Modern Mathematics, Book 1. Science Research Associates, 1961.

Smith, C. A.; Totten, W. Fred; and Douglass, Harl R. Algebra I. Second Edition (Teacher's Edition), Row-Peterson, 1960.

Morris and Topper. Advancing in Mathematics (J2). Science Research Associates, 1964.

APPENDICES

APPENDIX A

POST TEST SCORES

TABLE A.1.--Post Test Scores.

	Treatment Sy				Treatment Sp			
	Sp Hi / Num Lo		Sp Lo / Num Hi		Sp Hi / Num Lo		Sp Lo / Num Hi	
	Addition	Diagrams	Addition	Diagrams	Addition	Diagrams	Addition	Diagrams
	4	0	7	2	6	0	9	3
	9	0	9	0	8	2	8	1
	8	0	9	3	9	3	8	3
Grade 4	8	0	9	3	9	3	6	1
	9	3	9	3	9	3	9	1
	0	0	9	0	1	1	9	3
	9	0	1	0	0	3	8	1
	9	2			8	3		
	Addition	Diagrams	Addition	Diagrams	Addition	Diagrams	Addition	Diagrams
	9	2	9	3	9	3	9	1
	9	0	9	2	9	3	9	0
	3	0	8	2	4	3	9	3
Grade 5	5	0	9	3	2	0	9	3
	8	1	9	0	9	3	8	3
	8	0			8	3		
	9	2			9	2		
	9	0			9	3		
	2	0			9	3		
	9	3			9	2		

APPENDIX B

The Wilson (1956) procedure for calculating a χ^2 makes a slight modification on a technique developed by Rao (1952, p. 192-205) for the decomposition of a χ^2 statistic from a contingency table into components similar to the manner used for decomposing a total sum of squares in an analysis of variance computation.

Wilson's modification allows one to make a distribution free (non parametric) test of hypotheses concerning main effects and interactions usually tested by an analysis of variance technique.

The procedure is as follows:

The median value for the entire set of n observations is determined. This median should not be interpolated but should be determined only as a "boundary" which divides the entire set of observations, as nearly as possible, into two groups of equal size. In this particular study, for the group having a score of 9, n=35, and for that having a score of less than 9, n=25.

A contingency table is set up representing the frequencies of above and below median scores in each cell (see Table 1).

The total χ^2 value is calculated using the following formula for unequal cell frequencies:

$$\chi^2_{Total} = \Sigma_i \Sigma_j \frac{(afij - n_{ij} n_{a/n})^2}{nijn_{a/n}} + \frac{(bfij - n_{ij} n_{b/n})^2}{nijn_{b/n}}$$

Where:

afij = the number of scores = 9 in any one cell
bfij = the number of scores less than 9 in that cell
nij = the number of scores in that cell
na = number of scores = 9
nb = number of scores less than 9
n = total number of scores

The component corresponding to the cell in the lower right hand corner of the Table 1 grid (1/4) would then be:

$$\frac{4 - 5\ (35)/60)^2}{5\ (35)/60} + \frac{(1 - 5\ (25/60)^2}{5\ (25)/60}$$

This procedure is recognizable as an elaborate version of the $\frac{(O - E)^2}{E}$ technique. In this case the expected frequencies are obtained from the null hypothesis that the main and interaction effects produce no change in the distribution of scores.

χ^2 values for row effects are calculated using the marginal row totals of the contingency table. In the present case: 9/7, 6/8, 8/12, 2/8. χ^2 values for columns are calculated using the marginal column totals of the contingency table. In this case 12/18, 13/17. The χ^2 value for the interaction effect is found by subtraction of row and column values from the total value.

The respective degrees of freedom for each ^{2} value are:

χ^2	df
χ^2_T	rc-1
χ^2_R	r -1
χ^2_C	c - 1
χ^2_I	(r-1)(c-1)

APPENDIX C

Research Design II

Ability Differences in the Learning of Concepts Governing Directed Numbers

Robert M. Gagne
Princeton University

Problem

Differences in fundamental abilities appear to be prominent in the learning of mathematics, as well as in the way people use mathematical concepts. Well-established factors in human abilities are spatial, numerical, and verbal. Although there are studies which have revealed moderate to high correlations between aptitude measures and grades in mathematics, no studies have been conducted in the attempt to make specific predictions concerning the facilitation of different kinds of conceptual learning by different fundamental abilities. The possession of a high degree of spatial ability should facilitate the learning of spatial concepts; high verbal ability should facilitate the learning of verbal concepts; and high numerical ability should facilitate the learning of symbolic concepts.

The learning of concepts of addition of directed numbers may be done verbally, spatially, or symbolically. Verbal rules are perhaps the best known method, occurring in most conventional text books. Spatial concepts have been used with considerable success, notably in the textbook of the University of Illinois Committee on School Mathematics. Symbolic concepts can readily be designed to serve the same purpose; in one form they might resemble some of the symbolism of boolean algebra. Thus in this mathematical topic, the opportunity exists of relating differences in fundamental abilities to ease of learning the different types of concepts, as well as to final performance in problem solving.

Method

The study requires three groups of school children, of equivalent age, probably eighth graders. These children should have no previous knowledge of or acquaintance with algebra, or more particularly with the addition of directed numbers.

All subjects are tested on a carefully chosen factor reference battery designed to provide adequate measures of spatial, numerical, and verbal abilities. Each group is then instructed in the topic of adding directed numbers, using one of three methods (spatial concepts, symbolic concepts, or verbal concepts). Care should be taken to make the instructional materials equivalent in scope of coverage and in amount of practice provided.

When instruction has been completed, the students should be given a speeded test requiring the addition of a relatively long and varied list of directed numbers. Alternatively, two such tests could be given during instruction, making possible the use of gain scores. In any case, this type of test is used to measure degree of learning of the concepts. The test itself will constitute a practice session. Following this, students are given another test containing verbally stated problems which require the application of learned concepts to unfamiliar problems.

Correlations are obtained among: (a) ability scores, (b) concept-learning scores, and (c) problem performance scores.

Results

To the extent that the hypotheses underlying the study are verified, correlations between spatial ability scores and spatial-concept learning scores should be significantly higher than between spatial ability scores and verbal-concept learning or symbolic-concept learning scores. Similar hypotheses would be tested relative to numerical ability and to verbal ability. Correlations between concept-learning and problem-performance scores, on the other hand, may reveal differences in the efficacy of different types of concepts in problem solving.

APPENDIX D

STATISTICAL DATA

TABLE D-1.--Treatment: Symbolic
Type: Spatial high/Number low
Fourth and Fifth Grades

Spatial Score	z score	Number Score	z score	$z_{sp} - z_{num}$	Stanine Spatial	Stanine Number
16	- .38	29	- .12	- .26	7	6
10	-2.04	18	-2.23	.19	4	3
18	.17	33	.63	- .46	8	7
9	-2.32	18	-2.23	- .09	4	3
21	1.00	29	- .12	1.12	9	6
21	1.00	37	1.40	- .40	9	8
21	1.00	32	.44	.56	9	6
16	- .38	26	- .70	.32	7	5
16	- .38	32	.44	- .82	6	5
21	1.00	29	- .12	1.12	9	5
14	- .93	28	- .31	0 .62	5	4
17	- .10	36	1.21	-1.31	7	6
22	1.27	34	.83	.44	9	6
19	.44	33	.63	- .19	8	6
20	.72	33	.63	.09	9	7
19	.44	24	-1.08	1.52	8	4
16	- .38	31	.25	- .63	7	6
17	- .10	32	.44	- .54	7	6

TABLE D-2.--Treatment: Symbolic
Type: Spatial low/Number High
Fourth and Fifth Grades

Spatial Score	z score	Number Score	z score	$z_{sp} - z_{num}$	Stanine Spatial	Stanine Number
6	-2.05	29	- .89	1.16	6	2
9	- .98	24	-2.07	-1.09	5	4
14	.80	36	.77	- .03	8	7
8	-1.33	25	-1.84	- .51	6	4
15	1.16	34	.41	- .75	7	6
15	1.16	34	.41	- .75	7	6
13	.44	37	1.00	.56	9	6
14	.80	34	.29	- .51	6	5
11	- .26	37	1.-0	1.26	7	4
11	- .26	35	.53	.79	7	5
14	.80	35	.53	- .27	7	6
11	- .26	33	.05	.31	7	5

TABLE D-3.--Treatment: Spatial
Type: Spatial high/Number low
Fourth and Fifth Grades

Spatial Score	z score	Number Score	z score	$z_{sp} - z_{num}$	Stanine Spatial	Stanine Number
11	-1.86	20	-1.36	- .50	5	4
17	- .06	20	-1.36	1.30	7	4
18	.23	28	- .09	.32	8	6
17	- .06	31	.37	- .43	8	7
16	- .36	26	- .41	.05	7	6
12	-1.56	23	- .89	- .67	6	5
19	.53	36	1.16	- .63	8	7
18	.23	35	1.00	- .77	8	7
15	- .66	23	- .89	.23	6	5
10	-2.16	14	- .73	-1.43	4	3
18	.23	33	.69	- .46	7	6
20	.83	38	1.48	- .65	8	7
19	.53	38	1.48	- .95	8	7
16	- .36	29	.05	- .41	6	5
22	1.43	32	.53	.90	9	6
20	.83	30	.22	.61	9	6
21	1.13	32	.53	.60	9	6
22	1.43	27	- .25	1.68	9	5
16	- .36	29	.05	- .41	7	6

TABLE D-4.--Treatment: Spatial
Type: Spatial low/Number High
Fourth and Fifth Grades

Spatial Score	z score	Number Score	z score	$z_{sp} - z_{num}$	Stanine Spatial	Stanine Number
15	.80	34	- .10	- .90	8	7
15	.80	40	1.73	.93	8	6
15	.80	34	- .10	- .90	7	6
17	1.33	39	1.42	.09	8	7
5	-1.87	29	-1.62	.25	6	2
10	- .53	34	- .10	.43	7	4
13	.26	33	- .40	- .66	6	5
14	.53	34	- .10	- .63	6	5
5	-1.87	34	- .10	1.77	6	1
10	- .53	38	1.12	1.65	8	4
11	- .26	29	-1.62	-1.36	6	5
14	.53	34	- .10	- .63	7	6

Sprafka, Sarah; A. 1969
M.A.