R E P O R T   R E S U M E S

ED 011 059   24

A COMPARISON BETWEEN TWO KINDS OF SECONDARY MATHEMATICS COURSES WITH RESPECT TO INTELLECTUAL CHANGES.

BY- VAN HORN, CHARLES
ILLINOIS UNIV., URBANA

REPORT NUMBER CRP-1566   PUB DATE   OCT 66
CONTRACT OEC-2-10-087
EDRS PRICE MF-$0.18 HC-$2.76   69P.

DESCRIPTORS- *MATHEMATICS INSTRUCTION, EXPERIMENTAL GROUPS, *COURSE CONTENT, TESTING, *ALGEBRA, COMPARATIVE ANALYSIS, SECONDARY EDUCATION, *INTELLECTUAL DEVELOPMENT, *ACHIEVEMENT GAINS, URBANA, STRUCTURE OF INTELLECT

FROM A STUDY OF GUILFORD'S MODEL OF THE STRUCTURE OF INTELLECT, THE HYPOTHESIS WAS FORMED THAT THE ABILITIES MOST IMPORTANT TO LEARNING MATHEMATICS AND MOST LIKELY TO BE CULTIVATED IN MATHEMATICS CLASS ARE THOSE WHICH REQUIRE THE OPERATIONS OF COGNITION AND CONVERGENT PRODUCTION PERFORMED ON SYMBOLIC AND SEMANTIC CONTENT. A BATTERY OF TESTS SELECTED TO REPRESENT SEVERAL COMBINATIONS OF THESE OPERATIONS AND CONTENTS WAS ASSEMBLED TO PROVIDE DATA BY WHICH THE HYPOTHESIS MIGHT BE TESTED. TWO SAMPLES OF PUPILS ABOUT TO BEGIN THE STUDY OF ALGEBRA WERE TESTED. ONE SAMPLE WAS DRAWN FROM SCHOOLS IN WHICH THE MATERIALS AND METHODS DEVELOPED BY THE UNIVERSITY OF ILLINOIS COMMITTEE ON SCHOOL MATHEMATICS (UICSM) WERE IN USE, AND THE OTHER FROM SCHOOLS WHICH USED OTHER MATERIALS. AT THE END OF THE SAME SCHOOL YEAR, THE SUBJECT-MATTER PROFICIENCIES OF THOSE PUPILS WERE MEASURED BY APPROPRIATE CRITERION TESTS, AND THE SAME EXPERIMENTAL TESTS WERE ADMINISTERED TO ANOTHER SAMPLE OF PUPILS FROM THE SAME SCHOOLS. PUPILS ENROLLED IN UICSM COURSES EXCELLED IN MORE THAN HALF OF THE EXPERIMENTAL MEASURES. THE PROJECT HYPOTHESIS WAS TENTATIVELY SUPPORTED. ONE OF THE CONCLUSIONS STATED WAS THAT THE EXPECTATION THAT MEASURES OF COGNITION OF SYMBOLIC SYSTEMS WOULD BE VALID PREDICTORS OF ALGEBRA ACHIEVEMENT WAS NOT SUBSTANTIATED. (TC)

A Comparison Between Two Kinds of Secondary Mathematics Courses With Respect To Intellectual Changes

Project 1566
Contract No. OE 2-10-087

Charles Van Horn

October 1966

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Illinois

Urbana, Illinois

A Comparison Between Two Kinds of Secondary Mathematics Courses With Respect to Intellectual Changes

I. Introduction

Within the past few years, educational research has seen two patterns of emphasis which, although they do not always occupy the attention of the same persons, can be regarded as two aspects of a single concern. These two emphases are those placed on the improvement of instruction and on the use of more flexible methods of evaluation of that instruction. The single concern which is reflected in both is that for more effective and more efficient performance of the schools' function.

The current wide-spread emphasis on improvement of instruction was first felt in the field of mathematics and, within that field, the initial impetus came from the efforts of the University of Illinois Committee on School Mathematics. That Committee was formed in 1951 with the intention of preparing text materials and teaching methods which not only would prepare students more effectively to undertake scientific and technical training, but would also appeal to a greater number of students who did not intend to pursue scientific or technical careers. These objectives have been accomplished with a degree of success that probably exceeds the optimistic hopes of the original Committee and the work of the UICSM has come to occupy a prominent place in the field of mathematics education. Its influence, felt through its own program and through the effects it has exerted on other programs has been so great that its practices deserve a systematic investigation for their implications to the field of instructional theory.

The origin of education's concern for evaluation of its efforts is more difficult to localize, either in time or place. Tests and examinations have always been a part of the educational system and their improvement and the importance attached to their appropriate use has increased slowly over a period of many years so that individual contributions are difficult to recognize.

The experiment reported here is concerned with the psychology of individual differences, in the tradition of Quetelet and

Cattell, and with the body of classical psychometric theory based ultimately on the work of Pearson and Thurstone. In particular, it appeals to the more recent structure-of-intellect model provided by Guilford (1956, 1959) which, because it increases the range of measurable individual differences and emphasizes the multi-dimensional nature of intellectual activity, increases the feasibility and attractiveness of explorations of differentiated intellectual growth.

Described in its most general form, the basis on which this experiment was constructed is a recognition of the possibility that the result of any instruction need not be regarded as uniform for all pupils, but may differ according to the intellectual characteristics of the individual learner, and that those characteristics of the pupil which are relevant to an instructional process may be determined by (have interacted with) some characteristics of the process itself. If these conditions can be shown to exist, the field of mathematics education will have solidified its apparent advance in effectiveness, and the field of evaluation will have found access to a versatile and effective precedent for exploration of a difficult area.

II. The Problem

Before attempting to delineate or to justify the experimental hypotheses, it may be useful to describe the nature of the UICSM program and the procedures it has devised, not only because they are central to the experiment, but also because they are not widely understood.

At its beginning in 1951, the Committee stated its purpose as that of improving instruction in high school mathematics and, within this purpose, set for itself two objectives. One was to provide a set of texts embodying principles and concepts which are basic to all mathematics and in such a form that they would not need revision when they were encountered by the learner in his later study of mathematics. The second was to develop a teaching method which would enable and encourage the pupil to discover and learn to utilize those principles for himself instead of receiving them in ready-made form from the text or the teacher.

The first objective was achieved by subjecting all materials, before they were incorporated into the text, to the careful scrutiny of a mathematician whose responsibility was to maintain their mathematical rigor as a form of insurance that exactness was never sacrificed to ease of exposition.

The second objective, that of developing teaching methods to be used in conjunction with the texts, has been more difficult to achieve and far more difficult to evaluate because it requires that the teacher discover and be able to predict regularities and dependable relationships in human behavior instead of in mathematical certitudes. The teaching methods are based on the nature of the content, on the sequence of topics, and on several expectations concerning the intellectual abilities and motivational structure of the pupils. Text, teacher, and method are intertwined so closely that it is difficult to discuss or even to examine any of them without reference to the others.

The most frequently mentioned aspect of the UICSM program is the use of a pupil-centered approach described as the act of discovery. Although it is not the largest difference between the UICSM and other programs, it has received the most attention and the possible consequences of its use are basic to the experiment described here.

Discovery exerts a powerful reinforcing effect and its judicious use not only improves retention but is believed to make the pupil more receptive to the material. Discovery alone is not sufficient to achieve the primary aims of a mathematics course, however; it could be utilized in almost any class if the teacher is alert and diligent, but a text prepared with this teaching strategy in view can simplify the teacher's task and thereby increase the frequency and effectiveness of its use. If it is applied to material which has not been properly organized, the generalizations discovered by the pupil are likely to remain isolated and will simply be added to his store of facts and devices by which homework problems may be solved. The pupil's further attempts to use his newly found power may be frustrating because they are not likely to succeed unless the text materials were designed with specific regard for deductive organization.

In a UICSM algebra course, students are first led to "discover" certain generalizations which are assumed but not usually taught in beginning algebra courses. These first principles are designated as axioms and, given the reinforcement of his initial discoveries, the pupil characteristically feels encouraged to work farther. If this attempt succeeds, he becomes more and more likely to want to repeat the performance and, when he does this, his conceptual grasp of the subject and his motivation for studying it are increased.

When a pupil in one of these classes indicates that he has grasped (discovered) the generalization contained in the set of discovery exercises he is expected — in many cases he is required — to raise the question, "Why?". When he asks this question he has the opportunity to make the most important discovery of all: that he already possesses sufficient knowledge to deduce that generalization. It then becomes a premise for deduction of other generalizations and these, in turn, are added to his store of organized and consistently structured knowledge. As a consequence, he learns that not only does he possess the ability to make deductions, but also that inquiry into the structure of a body of knowledge is a rewarding enterprise.

The UICSM believes that, because they are taught to structure generalizations and then to seek verification for them, its pupils are given the opportunity denied to pupils in conventionally conducted courses, to have confidence in their ability, to appreciate the power of structured knowledge, to feel concern for the precise use of language, to understand the mathematician's emphasis on the search for patterns and to utilize rigorous thinking in contexts outside of mathematics.

If these assumptions are valid and pupils are, in fact, being provided with the opportunity to develop intellectual skills which might otherwise lie neglected, then their implications are of great importance. The development of intellectual skills is an outcome which all educators honor and which mathematics and logic have traditionally claimed for themselves.

Teachers of UICSM classes and others who have had occasion to observe such classes over a period of time seldom fail to report that they detect differences in the patterns of behavior habitually demonstrated by the pupils, especially with respect to their answers to the teacher's questions and the kind of questions they themselves generate. Teachers of other subjects in schools in which UICSM courses are taught frequently comment on the increased concern demonstrated by pupils for precision in the use of language, their methods of penetrating a problem, and in planning and utilizing a systematic attack on it. These are subjective impressions, impossible of validation, but the frequency and uniformity with which they occur are striking. If these observations are based on genuine differences in the intellectual functioning of pupils, then they constitute evidence that the intellectual habits of these pupils have been altered and, specifically, that some kinds of abilities are effected more than others.

The hypothesis obviously implied, one of differentiated intellectual growth is not novel since it is also implied in many psychological and educational theories; it is seldom offered as an explanatory principle to account for observed differences between individuals, however, because until recently the means for detecting specific differences in intellectual functioning did not exist. A psychology that conceives of intelligence as a unitary trait (the I.Q.) can never produce such hypotheses explicitly, and a psychology that recognizes the idea of intellectual differentiation without some means of measuring it can produce but never test such hypotheses.

Measurement procedures which will permit an experiment centering around the idea of differentiation of intellectual growth are provided for in the structure-of-intellect model described by Guilford (loc. cit.). The strengths of this model in experimentation are twofold; it provides means by which measures may be taken of a broader range of individual differences than have been available before, and it organizes those differences in a systematic way, so that the selection of appropriate measuring devices is facilitated.

This model specifies that an intellectual act may be described by reference to each of three independent dimensions. An operation, one of five possible, is performed on one of four kinds of content, and the result is one of six kinds of products. Since operation, content, and product are independent of one another, any of the 120 possible combinations exists in principle and, if the categorizations along each dimension are exhaustive, they constitute a complete catalog of all intellectual activities. Approximately half of the 120 possible combinations (intellectual acts) have been identified and tests prepared to measure them. Much of the success of this experiment lies in the judicious selection of the factors or mental abilities which are believed to be involved in the learning of algebra.

It is not the case, of course, that all or even a large number of these 120 possible intellectual performances are involved in any single learning task. Particular interest in this experiment, for example, centers around three of the five operations and on two of the four kinds of content. The operation of Cognition, believed to be of importance in learning algebra, is defined quite simply as the act of knowing, or being aware of, or having perceived some material in one or more of the content categories. The second operation of interest is that of memory which involves the ability to reproduce or call back to consciousness previously learned material in its original form. The third intellectual operation which is believed to be relevant to learning mathematics is that of Convergent Thinking which differs from the other two in being an instance of productive thinking, i.e., the examinee is required to produce or generate new material which has never been previously learned by or known to that individual. The modifier, "Convergent" refers to the specific kind of productivity in which answers are being sought or ideas being produced in a context which imposes restrictions or limitations on the quality or nature of the thing produced, so that there is a single "right" answer or, at best, a sharply limited number of acceptable responses.

The fifth operation, Evaluation, is that one performed when any stimulus material is considered with respect to an externally imposed standard. The absence of measures of this operation from this experiment is not evidence that it is regarded as of little importance to learning algebra, but that fewer acceptable measures of this ability are available for experimentation.

Four kinds of content are provided for in the model and, among these, one clearly predominates in the context of this experiment. Figural content is being dealt with when the material cognized, recalled, produced, etc., consisted of isolated elements of such a nature that the meaning they convey is contained in the symbol itself. Each symbol in figural material is regarded as an independent entity which is considered by the learner without regard to element present or to any arbitrary association. Figural contents are regarded as being of minor importance to learning algebra.

Symbolic content, with which this experiment is primarily concerned, is that dealt with when the elements in any material contain or convey no inherent meaning, but derive whatever significance they possess from their relationship to other elements in the same configuration. Any material that constitutes a pattern or a series is regarded as symbolic. To the extent that algebra consists primarily of patterns or structures, the importance of this ability to deal with symbolic materials cannot be overstated.

A third kind of content, and one that is regarded as being of some importance to this experiment, is that designated in the model by the term "Semantic". Semantic content is that involved in any situation in which the meaning or importance or significance of the material cognized, produced, etc., resides in an arbitrary association between symbol and referent. This definition covers any material which involves the use of words, sentences, etc., and includes much pictorial and probably numerically presented material. Its relevance to instruction is obvious; the text, the teacher, and the student traffic largely in words and the semantic content of most commercially available tests for prediction of academic success is evidence of its importance in the process.

The fourth kind of content, designated "Behavioral", is that which is dealt with when the stimulus in any situation is the presence, the behavior, or the implication of behavior of another person. Behavioral content is not included in any of the tests used in this experiment.

The third dimension of the model describes the product obtained when one of the operations is performed on some kind of content. Six kinds of products are recognized by the model and they are designated by terms which seem to be self-explanatory. They are: Units, Classes, Relations, Systems, Transformations, and Implications. The univocality of cateogries along this dimension is more difficult to demonstrate and their relevance to algebra is less apparent than are those of Contents or Operations. Little attention was given to products in selecting the tests to be used in this experiment.

This model is attractive in its simplicity and it undoubtedly facilitates thinking about the possible or probable nature of intellectual processes, but an important reservation must be introduced concerning its applicability to education outcomes. The existence of three dimensions of intellectual activity is postulated and the nature of the categories which comprise each of them was deduced by the author of the model from examination of a large number of factor-analytic studies of mental tests. Until very recently, all of the information on which this model was based was provided by college students and other young adults of comparable intellectual attainment and educational background. The population to which this experiment proposes to generalize consists of eighth and ninth grade pupils who must represent greater heterogeneity in their intellectual ability, less formal education and correspondingly less precisely formed intellectual habits. Attempts to apply the model in the younger group are not necessarily dangerous, but they may require guarded interpretation.

An experiment reported by Guilford, Merrifield, and Cox (1961) indicates that, in many respects, the structure of intellect as it was originally proposed can safely be regarded as an adequate model from which to derive hypotheses concerning the intellectual behavior of much younger persons. That experiment did not deal with the same areas of ability as the one described here so that, although there is justification for optimism in expecting the model to function with respect to junior high school students as it does with adults, prudence seems to require that the experimenter reserve a right to liberal interpretations until the validity of the model at this age level has been clearly established.

III. Related Research

The precedent for this experiment can be divided into three general categories: first, those experiments concerned with prediction of success in mathematics courses or with investigations into the nature of mathematical aptitude; second, those prediction and aptitude studies which are based on factor analysis; and third, those experiments which are concerned with the structure-of-intellect model. Those experiments concerned with comparisons of teaching methods are not regarded as relevant to this one because they compare methods which, in this context, constitute variations on a conventional theme or alterations in the sequence of topics and do not extend to comparisons of reorganized content or behavior of pupils.

Predictive studies in conventional mathematics courses, once more popular than they now are, consistently indicate that vocabulary and arithmetic scores have some value in predicting success in algebra courses and that, after the first semester, performance in almost any mathematics course becomes the best available predictor of success in the next.

This is consistent with the reports of many factor analytic investigations of mathematical ability which conclude that there exists a general intellectual factor that includes mathematical ability. Blackwell (1940), Doppelet (1950), and McAllister (1951) all agree on the existence of such a general factor which they find under various circumstances. Barakat (1951) finds not only a completely general intellectual factor but a specific "mathematical G" common to all tests involving mathematical knowledge. Weber (1953) and Werdelin (1958) find even more specific factors which they name "numerical" but their tests do not include applications of mathematics beyond arithmetic manipulation.

None of these experiments are in accord with the ideas on which this experiment is based. A general factor is one on which every test in an experimental battery has appreciable loadings; it inevitably occurs whenever all of the measures in an analysis have non-zero correlations with one another and can be refuted by including in the battery two or more tests which show appreciable correlation with one another and very small correlations with the others. A general factor is regarded here an an artifact based on inadequate sampling across domains and without value as an explanatory principle. It should be pointed out that none of the experiments cited above were designed with structure of intellect categories as models.

Strong support for the idea of differentiated intellectual abilities in mathematics aptitude is found in unpublished data gathered by the UICSM in connection with another experiment. These data included short highly speeded measures of four abilities: verbal comprehension, verbal reasoning, symbolic reasoning, and numerical facility. These tests, published by Psychological Services, Inc. (1957) were administered to several hundred eighth and ninth graders for whom proficiency measures covering the first two Units of the UICSM First Course were also available. Small correlations between aptitude tests were found and the correlations between these tests and the proficiency measures were only moderate. Closer examination of the data revealed that correlations with the criteria were suppressed by the existence of many cases in which a high criterion score was accompanied by one high aptitude score (not always the same one) and average scores on the other aptitude measures. As a consequence, high criterion scores were always accompanied by both high and moderate aptitude scores and correlations higher than 0.4 or 0.5 never appeared. This suggested that, given a minimum level of skill with any of the aptitude measures, any of the other three could be regarded as mathematical aptitude. If this is the case, then it must be that each pupil approaches his learning task along the lines that are most suitable to his habitual mode of thought and deals with the content in the mode that he can most readily conceptualize.

Sex differences in mathematical ability seem to contradict most theories of learning but, although they are puzzling, they are so ubiquitous that they cannot be overlooked. Long ago Flack (1926) published evidence that, in general, girls are more able than boys at learning mathematics and no major experiment has yet refuted this finding, nor has any adequate explanation for it been devised. Weber (1953) noticed this circumstance but chose to attribute it to "feminine affectivity" rather than intelligence. Rusch (1957) demonstrated that after grade 6 some aspects of number ability develop more rapidly in girls than in boys, and Blackwell (1940) has identified factors pertaining to exactness and precision that apply better to girls than to boys. These suggestions toward explanations make it seem reasonable to expect sex differences in an experiment such as this, but it will be treated as one of many possible dimensions of aptitude to the extent that provision is made in the analysis of the data for exploring its existence.

The third category of precedents for this experiment is that of experiments based on the structure-of-intellect model. The basic articles have been cited: Guilford (1956, 1959, 1963), Guilford, Merrifield, Cox (1961). The most direct precedent is reported by Petersen, Guilford, Hoepfner, and Merrifield (1963) in which four kinds of algebra courses were investigated following the structure-of-intellect categorizations. Twenty-five tests were investigated in their relationship to courses, two of which seem comparable to those described here as conventionally instructed. The idea of differentiated intellectual behavior is substantiated by the finding that systematic differences occur between courses in the pattern of predictors. The complexity of mathematics courses was shown to increase from general mathematics to accelerated algebra and there is a strong suggestion that more complete accounts of mathematical aptitude are quite feasible and that, when they have been achieved, they will be found to be more complex factorially than previous experimenters have led to believe.

IV. Hypotheses and Selection of Tests

Four hypotheses were devised to be tested in this experiment:

1. Changes in intellectual functioning in pupils who have studied algebra in a UICSM course will differ from the changes in those functions among pupils who have studied algebra in a conventional course.

2. The abilities which are related to success in either course are those characterized by the structure-of-intellect model as symbolic in content, and the cognitive operations in that model will forecast performance more effectively than other operations.

3. The abilities which can be shown to be related to success in a UICSM course do not differ from those which are related to success in a conventional algebra course.

4. The abilities which are related to success in either kind of algebra course are the same for both sexes.

Attention is drawn to the phrasing of these hypotheses. Hypotheses 1 and 2 do not follow the usual practice of stating a null (no difference) hypothesis because of the nature of the evidence which can be regarded as support for either. The statistical null hypothesis is useful in experimentation because it points to the kind of mathematical model on which the experimenter will base a decision to judge his experiment. These two hypotheses must be evaluated largely on the outcome of a factor analysis for which no tests of significance are available and, if they cannot be accepted or rejected in the sense of statistical tests, the cumbersome phrasing of a null hypothesis serves no purpose.

Hypotheses 3 and 4, on the other hand, refer to decisions which can be made on the basis of a significance test and the null statement seems appropriate.

In selecting tests to permit evaluation of these hypotheses, much concern was given to adequate sampling in two kinds of operations categories and three kinds of contents, while relatively little concern was given to balance in the product areas sampled.

With respect to operations, attention was focused on cognition and convergent production since the processes involved in learning algebra are assumed to be those of knowing (in the sense of being aware of) and with the production of "right" answers in contexts where the answer is contained in the given material and where limitations are imposed on the nature and quality of the answer. Memory, as an operation receives some attention because of the presence of arithmetic tests in the battery. Measures of arithmetic ability are known to be useful as predictors of success in beginning algebra and are classified by the structure-of-intellect as memory for symbolic implications and three reference tests for memory, differing in the degree of organization of the material memorized, were included in the battery to make it possible to account for whatever variance might arise from that source.

With respect to contents, attention was focused on symbolic and semantic materials. The concern for symbolic material is consistent with the opinion of most mathematicians that manipulation of symbols and the awareness of relationships between and among them is an integral part of algebra. Semantic content was regarded as necessary because every prededent indicates that overtly semantic measures (vocabulary and reading tests) are consistently valid predictors of success in most academic areas, including algebra. Figural content appears in the battery only in connection with one memory test.

Less explicit attention was given to sampling across production categories in the selection of tests for two reasons. The functioning of these categories is less well understood than that of content and operations dimensions and their relationship to algebra is difficult to detect; as a consequence, the hypotheses are not structured in these terms and there is less either to be gained or lost by concern for products. The second reason for exhibiting less concern for product categories lies in the present state of the structure-of-intellect model. Not all of the combinations predicted by the model have been identified nor have tests been devised to measure them; the selection of tests is obviously restricted and complete freedom of selection, to the extent of availability of any desired combination of operation, content, and product is not possible. Within this restriction, an attempt was made to distribute the tests across product categories (with the exception of Units), but when the restriction made a choice necessary, that choice was made in terms of content and operation desired even when balance in the distribution of products was sacrificed.

Table 1 below shows the names of the twenty-five tests used in the experiment and the known or assumed factorial content of each in terms of the three dimensions of the model. Many of these tests represent mixtures of factorial content but the predominant one is indicated in each case.

In reading the column headings of the table, the following single-letter designations are used to indicate specific categories:

Operations	Contents	Products
C - cognition	F - figural	U - units
M - memory	S - symbolic	C - classes
D - divergent prod'n	M - semantic	R - relations
N - convergent prod'n		S - systems
E - evaluation		T - transformations
		I - implications

In the body of the table, the entry "X" indicates that the factorial content has been established by previous experimentation, while the entry "O" indicates that the content is assumed to be that described.

Table 1

Factorial Composition of Tests

		Operations				Contents			Products					
		M	C	N	D	F	S	M	U	C	R	S	T	I
1.	Alternate Additions				X		X				X			
2.	Arithmetic	O					O							O
3.	Circle Reasoning		X				X					X		
4.	Classification - A			O				O		O				
5.	Classification - B			O		O					O			
6.	Disguised Words		O					O					O	
7.	Form Reasoning			X			X							X
8.	Letter Triangle		X				X					X		
9.	Logical Consequences			O				O						O
10.	Memory for Symbols	O				O					O			
11.	Memory for Words	X						X			X			
12.	Memory for Sentences	X						X			X			
13.	Missing Signs		O				O				O			
14.	Numerical Ability			O			O							O
15.	Reading Comprehension		X					X	X					
16.	Starring		O				O				O			
17.	Symbol Elaboration - A			X			X						X	
18.	Symbol Elaboration - B			O			O						O	
19.	Symbolic Reasoning		O				O							O
20.	Verbal Comprehension		X					X	X					
21.	Verbal Reasoning		O					O						O
22.	Word Changes - I			X				X				X		
23.	Word Changes - II			O				O				O		
24.	Word Patterns		X				X							X
25.	Word Transformations			X			X						X	
26.	Sentence Order			X				X				X		
		4	10	11	1	2	13	11	2	1	7	5	4	7

Sources of Tests

In the description of tests which follows, the source of each test is indicated. Those tests attributed to "Aptitudes Research" were obtained from and are used with the permission of the Aptitudes Research Project of the University of Southern California, J. P. Guilford, Director. Those tests attributed to "Talent" were obtained from and are used with the permission of Project Talent of the University of Pittsburgh, J. J. Flanagan, Director. Those tests attributed to "E.A.S." are taken from the Employee Aptitude Survey, published by Psychological Services, Inc., Los Angeles, California. Those tests attributed to "UICSM" were developed by the University of Illinois Committee on School Mathematics specifically for use in this experiment.

In the descriptions which follow, the absence of a notation concerning scoring indicates that the score for that test is simply the number of correct responses.

1. Alternate Additions:

Given a set of numbers, show as many ways as possible in which they may be combined by addition to yield a specified sum.

2 parts; each part 8 items, 3 minutes.

Source: Aptitude Research

Sample Item: Given: 1 2 3 4, get sum 7

3+4 = 7
1+2+4 = 7

etc.

2. Arithmetic:

Constructed response items in four fundamental operations.

3 parts: 16 addition, 9 multiplication, and 9 division problems, 2 minutes each part.

Source: UICSM

3. Circle Reasoning:

Discover the principle by which one circle is blackened in each of four rows of circles and dashes, and blacken the appropriate circle in a fifth row.

1 part: 14 items, 8 minutes

Source: Aptitudes Research

4. Classification - A:

Given two sets of four words each, within which one element is common to each set, and a problem word, specify the group to which the problem word belongs.

1 part: 12 items, $1\frac{1}{2}$ minutes

Source: UICSM (Adapted from a suggestion by Aptitudes Research).

Sample Item:
1. Silk, Rayon, Nylon, Wool
2 Shirt
2. Coat, Dress, Shoes, Hat

5. Classification - B:

Similar to Classification - A except that, instead of words (semantic content) the materials are letters and geometric figures (figural content).

1 part: 9 items, $1\frac{1}{2}$ minutes

Source: UICSM (Adapted from a suggestion by Aptitudes Research).

Sample Item: 1. A E H M
2 S
2. B G R J

6. Disguised Words:

A multiple-choice vocabulary test in which the stimulus word is spelled phonetically.

1 part: 30 items, 3 minutes

Source: Talent

Sample Item:

2 DLA
1. sadly
2. postpone
3. bluntly
4. hand out
5. everyday

7. Form Reasoning:

From a table, find a form that is equivalent to three given forms.

2 parts: each part 10 items, 2 minutes

Source: Aptitudes Research

8. Letter Triangle:

Given a group of letters arranged according to a plan in a triangular pattern, specify which of five suggested letters should appear in a marked location.

2 parts: each part 8 items, 6 minutes

Source: Aptitudes Research

Sample Item:

```
       a          a
     b   c        f
   d   e   f      g
 ?   _   _   _    h
                  j
```

9. Logical Consequences:

Given a set of statements (3 to 5 per item), write as many implications as possible.

2 parts: each part 2 items, 3 minutes

Source: UICSM

Sample Item: Given:

Algebra is easier than Latin
Latin is easier than History
History is just as easy as English

New Statements:
Algebra is easier than Latin
Latin is easier than English
Algebra is easier than English
etc.

10. Memory for Symbols:

Twenty-four symbols, each paired with a letter or numeral, are studied for three minutes. Thirty symbols are presented on the following page and examinee pairs each with letter or numeral that accompanied it.

Source: UICSM

1 part: 3 minutes study time, 3 minutes working time

11. Memory for Words:

Twenty-four nonsense syllables, each paired with an English word are studied; a five-alternative multiple-choice is given covering all 24 words.

Two minutes are allowed for study, two for a practice exercise (with study page available), and four minutes for the multiple-choice test.

Source: Talent

12. Memory for Sentences:

Forty short sentences are studied for six minutes. After another test, twenty-four of them are presented for recall in multiple-choice form; the distractors are the second letters of the omitted word in each sentence.

Study time, 6 minutes: Working time, 5 minutes

Source: Talent

13. Missing Signs:

A series of numbers (2 to 5 per item) are shown with an answer. Examinee indicates the operations to be performed on those numbers to arrive at the given answer.

Score: Number of items in which all entries are correct; no partial credit is given.

Source: UICSM

Sample Item:

Given	8		2		4	=	4
Answer	8	X	2	÷	4	=	4

14. Numerical Ability:

Fundamental arithmetic operations, one operation per item.

1 part: 25 items, 2 minutes (two other parts involving decimals and fractions were not used).

Source: E.A.S.

15. Reading Comprehension:

Three paragraphs, each followed by several multiple-choice items are presented. Twenty-one questions are asked.

1 part: $7\frac{1}{2}$ minutes

Source: Talent

16. Starring:

An undefined operation is displayed in three examples. Examinee finds the rule which fits the three examples and provides the answer to a fourth.

1 part: 22 items, 6 minutes

Source: UICSM

Sample Item:

2 * 3 = 4
8 * 4 = 11
9 * 0 = 8

5 * 6 = 10

17. Symbol Elaboration - A: From a set of given equations containing letters, examinee generates new equations consistent with the given ones.

2 parts: each part 2 items, 3 minutes

Score: Number of valid implications written

Source: Aptitudes Research

Sample Item: Given: $B - C = D$
$Z = A + D$

New Equations: $D + C = B$
$Z = A + B - C$

18. Symbol Elaboration - B: Similar to Form A except that the given equations include both equalities and inequalities.

Length, time limits and scoring same as Form A.

Source: UICSM

19. Symbolic Reasoning: Two relationships are expressed between three letters. Examinee evaluates a given third relationship as True, False, or Cannot Tell.

1 part: 30 items, 5 minutes

Source: E.A.S.

Sample Item: $X > Y = Z$, therefore $X = Z$ (False)

20. Verbal Comprehension: Four alternative multiple-choice vocabulary test.

1 part: 30 items, 5 minutes

Source: E.A.S.

21. Verbal Reasoning: From a set of verbally stated facts (four or five per set), five conclusions are drawn. Ex-

aminee evaluates each conclusion as True, False, or Cannot Tell, with respect to stated facts.

1 part: 30 items, (six sets of facts and conclusions), five minutes

22. Word Changes - I:

Given a set of words, each containing the same number of letters, one is designated as the first and another as last; arrange the remaining words so that exactly one letter is changed from one word to the next.

2 parts: each part 6 items, 4 minutes

Source: Aptitudes Research

Score: Number of sets correctly ordered, no credit is given for partially correct orders.

Sample Item:

B E L L

2	1.	BAIL
1	2.	BALL
3	3.	MAIL

M A I N

23. Word Changes - II:

Given a four letter beginning word, examinee changes one letter at a time in such a way that each change makes a real word; his objective is any word which does not contain any of the original letters in their original position.

2 parts: A shows nine different starting words, B calls for as many variations as possible on a single beginning word. Five minutes each part.

Score: Two scores were analyzed. Score W assigns one point for each word successively transformed with no credit for partially completed words; Score L assigns one point for each letter changed according to the rules.

Source: UICSM

Sample Item:

M A T H

M A S H
M U S H
R U S H
R U S T

(A score of 1 W or 4 L would be assigned to this sequence.)

24. Word Patterns:

Arrange a list of given words efficiently in a kind of crossword puzzle design.

2 parts: each part 3 items, 6 minutes

Score: Complement of number of spaces in the design into which the letters have been inserted.

Source: Aptitudes Research

25. Word Transformations:

Short phrases are provided which, if the same letters in the same order are respaced, will form a different series of words.

1 part: 20 items, 6 minutes

Score: Number of correct divisions made.

Source: Aptitudes Research

Sample Item:

THE RE|D O|LIVE (Score 2)

26. Sentence Order:

Arrange three given sentences in sensible order.

2 parts: each part 10 items, 3 minutes

Score: one point for each correct ordering; no partial credit for partially correct orders.

Source: Aptitudes Research

Sample Item:

2 She bought some food at the market.

3 She returned home and cooked some of the food she had bought.

1 She went to the market.

V. The Plan of the Experiment

The Initial Plan

Like most hypotheses, the ones presented here are based on a combination of published precedents and the experimenter's informal notions about the probable nature of the events that he observes. The original structure of the experiment was influenced by the intuitive expectation, based on unpublished research and undocumented observations, that "mathematical aptitude" exists, not as a unitary trait in which individuals differ only according to degree, but as a many-faceted manifestation of the individual's intellectual history, habits, and preferences.

If this view is adopted as a starting point, the process of learning mathematics can be regarded as being specific to the individual learner. Instead of expecting all pupils to perceive the statements made by the teacher and the text in the same way, the alternative expectation can be substituted that each pupil translates these statements into the kind of content which he finds easiest to process, and that he performs on that content the kind of operation that he thinks is most likely to achieve the desired result. Given a symbolic statement of an abstract principle, for example, one pupil might prefer to attend to the structure of the statement as it was symbolically presented to him, another might decide that his best course is to memorize the sequence of symbols, another might prefer to verbalize it (translate to semantic content), while another might divert his attention to a search for concrete examples of the operation of that principle, etc. Because these preferences are likely to be systematic and relatively enduring, different pupils can be expected not only to learn algebra in different ways, but actually to devise different ideas concerning its structure, generalizability, etc. And because not all methods are equally efficient, some differences in ultimate mastery may be attributed to differences in preferred learning formats or modalities.

Such a point of view carries implications for several areas. If it is applied to the interpretation of published experiments concerned with prediction of success in mathematics, it accounts for the lack of uniformity to be found there. The correlation between a predictor and any criterion can be attenuated by those cases in which an individual, having demonstrated exceptional performance at one kind of predictor task, chooses to approach the subject-matter along lines at which he may be less proficient or which are less effective for learning that material.

If this point of view is applied to the task confronting the teacher a rather disturbing conclusion can be reached. Given that each pupil builds for himself a conceptual structure from intellectual bricks and pedagogical mortar of his own choosing, there is little reason to expect that the various structures thus built will resemble one another exactly, so a single kind of presentation of a single method of developing an idea is not likely to be the most efficient for every member of a class. The evaluator who attempts to determine the extent of each pupil's mastery of a subject must recognize a similar implication; some of the variance between pupils arises from differences in the degree to which they have mastered the principles involved, and some from the differences in the extent of agreement between the terms in which the principle was learned and those in which the examination question was phrased.

The most important implication of the idea, for this report, are those which influence the structure of the experiment. Because they refer to multi-dimensional measures, hypotheses 1 and 2 clearly indicate the use of a factor analytic experiment; hypotheses 3 and 4, because they deal with predictions, can best be dealt with by multiple regression methods.

As it was originally conceived, the experiment was to have required four factor analyses. A set of texts of known factorial content, chosen to represent a broad sampling of contents and operations was administered to two groups of high school freshmen who were about to begin the study of introductory algebra, one group in a class using the texts prepared by the UICSM and the other in conventionally conducted courses. When these are factor analyzed any test that represents the same kind of task to both groups will show similar factorial content in both analyses; if some task is performed in one way by one of the groups and in a different way by the other, or if there are systematic differences in difficulty levels, then differences must occur in the two factor structures and a comparison of these structures gives an indication of the extent to which the pupils being assigned the two kinds of classes are comparable in their intellectual functioning.

To avoid possible contamination from the effects of homogeneous grouping within schools, classes representing conventional instruction were drawn from schools in which UICSM materials were not in use, so that the members of any class could be regarded as an unselected sample of ninth graders in that school.

After a year's study of algebra, two more testing sessions were conducted. The competence of the pupils previously tested was measured by an appropriate subject matter test and the correlation between that criterion and the measures obtained in the previous September provide the basis for the predictions referred to in hypotheses 3, and 4.

At the same time, the reference tests were to be administered to a different group of pupils in the same schools who were also completing their first year of algebra in each of the two kinds of courses. A comparison of factorial structures across these two groups would indicate, in the same way as before, the comparability of their performance, i.e., whether each test is still measuring the same ability in both groups. To the extent that this did not occur, the two algebra courses could be regarded as having taught different kinds of intellectual habits or having reinforced different kinds of intellectual behavior.

Data Collection

The entire set of twenty-six tests selected for inclusion in the experiment (see Chapter IV) required more time, by a factor of at least two, than any of the participating schools were able to devote to the experiment. This necessitated a choice between reducing the number of experimental tests by approximately half, or distributing the larger number of tests in such a way that, although no group devoted more than two class periods to testing, all possible pairs would be represented in sufficient numbers to justify the use of those correlations in a factor analysis.

The first alternative limits the scope of the experiment and reduces both the range of intellectual abilities sampled and the likelihood of demonstrating a difference between groups; it also reduces the total number of testing sessions sufficiently to permit personal supervision of each by a member of the project staff. The second alternative provides access to a wider range of measured abilities over a greater number of subjects, thereby increasing the likelihood of supporting the hypothesis of differentiated intellectual growth, but it accomplishes these purposes at the expense of an increase in the number of testing sessions so large that it was necessary to depend on classroom teachers to administer most of the tests. The second alternative was chosen and, had the testing sessions been conducted as planned, would have yielded a wealth of data; the lack of direct control over the administration of the tests has proved to be a major — and strongly debilitating — influence on the outcome of the experiment.

A set of test booklets was prepared for each of the participating classes in such a way that they could be administered in two class sessions and the contents of the various classes booklets was varied in such a way that every pair of tests was included in at least four booklets. These tests and a set of carefully written instructions were mailed to each school where, in most cases, they were to be distributed by the principal to the participating teachers.

The weakness in this method of data collection gradually became apparent as the experiment progressed. It is not likely that the failure of any one aspect of the data collection method would have damaged the experiment, but the total effect of several of them forced a change in the design. The circumstances which account for most of the difficulty can be divided into four classes; unused tests, repeated administrations, maladministrations, and improper sample selection.

The first suggestion that the experiment was not proceeding according to the original plan came with the discovery in the returned materials of occasional sets of unmarked booklets. Some of the teachers who presumably could not find time in their schedules for administration of the tests returned the materials unused; others simply did not return them at all. This was not a major problem and probably does not account for as much as a tenth of the total lost data, but it did not occur with equal frequency in all schools and the proportion of unused and unreturned test booklets was higher among the non-UICSM schools in the sample and higher in the second testing session than in the first.

The second kind of problem, about as serious as non-returns in the amount of data lost, was that resulting from repeated administrations of tests to a single group of pupils. In most cases this was the result of semantic accident; when one teacher had more than one section of ninth grade algebra, a different set of test booklets was prepared for each section and the letter accompanying the package included separate instructions for administration of each booklet and the statement that the package contained sufficient materials for testing a (stated) number of classes. The word "class" was intended to refer to a group of pupils, but was interpreted by some of the teachers to mean a single meeting of one group. The result was the occasional administration of two or more sets, almost always involving some duplications, to the same group of pupils. When this occurred, all administrations of a single test after the first were invalidated and all of the duplicated tests, as well as the pairs of which they were members, were lost to the experiment.

The third source of invalidated data was the maladministration of tests in the individual sections. This was not detected until after scoring had begun and, even then, the frequency with which it was to occur was not fully realized until too much time had elapsed to permit finding and testing other classes to repair the damage. Most of the tests used in the experiment were short and highly speeded, since this procedure provides a clearer factor structure than do the so-called "power" methods. Tests with time limits as short as one minute were used and time allotments greater than seven or eight minutes for a single test were seldom provided. Under these circumstances, meticulous adherence to stated time limits is imperative, and strong suggestions were seen in the data from some of the classes to indicate that they had not been uniformly followed. When this suspicion first appeared, a practice of spot-checking scores within booklets was instituted, and in those cases in which careless handling was suspected, none of the data from that class was included in the analysis. The necessity for class-by-class inspection of means, standard deviations and, in some instances, inter-form correlations was enormously time-consuming and resulted ultimately in discarding so much of the data that the factor analytic phase of the experiment was seriously damaged. The incidence of occurrence of this error was about the same for the two kinds of programs.

The fourth category of experimental errors which resulted in the loss of data was a variation of the practice of administering the same tests twice to a single class on consecutive days, but at a different and far more damaging level. The original plan had called for administration of a single test battery to four samples, one from each of two populations (UICSM and conventionally conducted classes) on two occasions (September and May), but in some of the schools the second (May) administration was to the same pupils to whom those tests had been administered earlier. This is attributed to the fact that, to a group whose notion of research seems to be a comparison of pre vs. post instruction test scores, indufficient emphasis was laid on the importance of having data from two samples. This aspect of the experiment had been discussed with the teachers and administrators in the original negotiations during the preceding summer, and should have been pointed out more explicitly before the second round of testing was begun.

The Revised Experiment

None of these problems in data acquisition had been anticipated and they did not appear suddenly, but were encountered bit by bit over a protracted period and, when the magnitude of their accumulated effect was realized, subjects who met the conditions for inclusions in the experiment were no longer available. Approximately a third of all of the data collected had been outlawed and of this, more was lost from conventional classes than from schools in the UICSM program, and far more had been lost from the May than from the September testing sessions.

Having sacrificed such a large proportion of the data, it became necessary to adjust the mode of analysis to conform to the new circumstances. In a factor analysis, if even one correlation is missing, it is necessary to remove those variables from the matrix. In this case, the data loss was not systematic; correlations that were impaired in one of the four matrices might be intact in others, so that if every test which had been damaged by loss of data in one matrix were to be removed from all four, so few variables would remain that little possibility of meaningful comparisons would remain.

In a multiple regression analysis or a direct comparison of means the absence of a single correlation or a single descriptive statistic prohibits only the use of that test or of that single combination, leaving all of the others available. The effect of the experimental errors described above was far less damaging to that part of the experiment which relies on multiple correlations than to that part which hinges on the comparison of factor structures, and the capability of making direct comparisons of means and variances was only slightly impaired; therefore, under pressure of time and after several fruitless attempts to reconstruct sufficient information to justify factor analysis, a belated decision was made to cease such attempts and to confine the experiment to those areas which justified analysis.

For that reason, this report is confined to a discussion of the predictions which can be made of the two kinds of criteria, of the sex differences in achievement and predictability, and of the indirect evidence which those analyses offer for the notion of differentiated intellectual behavior.

VI. Analysis and Discussion

The opportunity for the planned factor analysis having been lost by unexpected shortcomings in the form of the data, two alternative forms of analysis were adopted to search for evidence pertaining to the effects of the two algebra courses. The first alternative is a set of comparisons, first between the groups both in pre and post instruction scores, and the second between the two testing sessions within each group (gain scores). These comparisons are conducted by means of ratios between variances and by t ratios. The second alternative is a comparison of the regression equations which predict success in each of the two kinds of courses.

The basic data on which all of these comparisons are based is contained in Tables 2 and 4 which show the sample sizes, means, and standard deviations of each of the measures for both samples in each group, and in Table 6 which shows the correlations between each of the experimental tests with sex and with the appropriate criterion test.

Comparisons Between Groups: Pre-Instruction

The first analysis performed was a comparison of pre-instruction scores in the two groups for the purpose of determining the comparability of the pupils who are assigned to each of the two kinds of classes. The results of this comparison are summarized in the last column of Table 2; in this summary the reported value is the amount by which the mean of the UICSM sample exceeds the mean of the conventionally instructed sample, so that a negative value indicates a higher mean score for the conventional group.

Examination of Table 2 makes it clear that, at the beginning of their first year's study of algebra, the pupils (mostly ninth graders) in schools in which the UICSM program is not used are superior in almost every measured respect to those pupils (mostly eighth graders) in schools in which UICSM materials are used. Of the 38 comparisons summarized, the pupils in UICSM classes demonstrate superior performance in only 3, while the pupils in conventionally instructed classes excelled in the other 35. If these two samples are random selections from populations which are equally proficient in each of these tasks, and if the 38 measures are independent, the probability of either group excelling in 35 of 38 tests is of the order of 10^{-8}. The conclusion that the difference between groups is based on a real difference rather than on chance seems justified.

Table 2

Sample Size, Mean and Standard Deviation
Of Each Test in Two Preinstruction Samples

		Conventional			UICSM			
		N	M	s.d.	N	M	s.d.	Diff.
1.	Alternate Additions - A	204	9.47	2.46	281	9.33	2.98	- .14
	- B	234	9.60	3.54	277	8.48	3.25	- 1.12
2.	Arithmetic	351	19.33	4.12	253	17.94	3.88	- 1.39
3.	Circle Reasoning	254	6.50	2.84	233	9.42	9.55	+ 2.92
4.	Classification - A	284	4.70	1.73	217	4.14	1.59	- .56
5.	- B	284	3.29	1.82	216	2.67	1.57	- .62
6.	Disguised Words	387	17.37	6.63	221	17.40	6.26	+ .03
7.	Form Reasoning - A	316	8.68	2.40	260	7.17	3.55	- 1.51
	- B	317	9.08	2.02	248	7.79	3.28	- 1.29
8.	Letter Triangle - A	362	4.36	2.14	359	3.79	1.90	- .58
	- B	362	4.16	2.37	359	3.86	2.04	- .30
9.	Logical Consequences - A	-	--	--	42	3.98	1.37	--
	- B	276	3.30	1.98	290	2.96	2.06	- .34
10.	Memory for Symbols	305	18.64	5.62	323	16.53	5.34	- 2.11
11.	Memory for Words	284	12.84	5.20	386	11.88	5.25	- .96
12.	Memory for Sentences	265	13.74	4.37	215	13.81	3.86	+ .07
13.	Missing Signs - A	341	8.98	3.81	220	6.93	2.56	- 2.05
	- B	314	7.22	2.52	256	6.90	3.58	- .32
14.	Number Ability	280	13.81	3.66	377	12.91	4.27	- .90
15.	Reading Comprehension	259	12.24	4.49	200	9.94	4.27	- 2.30
16.	Starring	322	7.13	3.88	152	5.74	1.61	- 1.39
17.	Symbol Elaboration - A, 1	328	6.24	4.73	312	5.03	4.62	- 1.21
	- A, 2	297	6.03	4.07	279	5.01	3.06	- 1.02
18.	Symbol Elaboration - B, 1	337	5.65	3.56	143	2.48	3.02	- 3.17
	- B, 2	337	5.67	3.77	229	4.90	4.15	- .77
19.	Symbolic Reasoning	362	11.08	4.40	272	10.66	3.73	- .42
20.	Verbal Comprehension	253	15.49	3.92	289	15.04	3.80	- .45
21.	Verbal Reasoning	303	15.90	4.75	259	14.53	5.04	- 1.37
22.	Word Changes - I, a	399	4.74	1.60	306	4.15	1.81	- .59
	- I, b	399	4.23	1.98	307	3.75	2.08	- .48
23.	Word Changes - II, AW	282	.94	1.27	254	.57	1.06	- .37
	- II, AL	283	8.89	4.96	254	6.12	4.10	- 2.77
	- II, BW	283	1.28	1.40	247	.79	1.11	- .49
	- II, BL	283	8.79	5.47	247	5.98	4.34	- 2.81
24.	Word Patterns - A	238	144.84	8.34	358	142.05	10.30	- 2.79
	- B	244	144.20	18.53	346	111.95	10.29	- 2.25
25.	Word Transformations	369	23.95	9.76	275	19.73	9.45	- 4.22
26.	Sentence Order - A	205	8.42	8.22	279	4.33	1.80	- 4.09
	- B	167	5.07	1.83	326	4.06	1.77	- 1.01
	CoOp Algebra	138	30.90	5.10				
	UICSM Algebra				687	12.29	3.95	

These comparisons can be made in more detail by calculating t ratios association with each whenever this calculation can be justified. A necessary condition for interpreting a t value is a demonstration that the variances in the two groups are comparable, and this can best be demonstrated by interpreting the ratio of the larger variance to the smaller as a one-sided F. The value of this ratio that will justify rejection of a hypothesis of equal population variances depends, of course, on the sizes of the samples involved and these sizes differ from test to test within this comparison; however, all of them are of approximately the same size so, for the sake of convenience, an arbitrary single value was adopted for all of the comparisons. With a large number of degrees of freedom the change in the value of the smallest significant F varies little from one sample size to another, so that a possibility of gross misinterpretation of the data is not encountered by the use of the uniform value of $F' = 1.5$ for all of the comparisons. In the summaries reported here, t ratios are reported only for those pairs of tests in which the larger variance is not more than one and one-half times the smaller.

Twelve of the comparisons are not subject to t comparisons by reason of non-comparable variances. These are:

Circle Reasoning	Word Changes II - AL, BW, BL
Form Reasoning - A & B	Word Patterns - A & B
Missing Signs - A & B	Sentence Order - A
Starring	

With respect to operations, these are about equally divided between cognition and convergent production; with respect to content, four of them (3 Word Changes and Sentence Order) are semantic and nine are symbolic. All but one of them (Circle Reasoning) represent a mean difference in favor of the conventional group, and five have greater variance in the UICSM than in the conventional population.

For the remaining twenty-six measures, t ratios associated with the differences are shown in Table 3. Nineteen of these indicate significance at or beyond the five percent level and seven indicate that the difference can be attributed to chance. In Table 3 the sign affixed to the t indicates the direction of the difference between means in the direction (UICSM - Conventional) so that negative signs indicate that the mean of the conventional group is higher. In addition to the test, Circle Reasoning, mentioned above, the UICSM group excels in only two other tests, Memory for Sentences and Disguised Words and a very small t ratio is associated with both of these differences.

Table 3

t Ratios Associated With Differences Between Means

Between-Groups Comparison

		Pre-Instruction	Post-Instruction
1.	Alternate Additions - A	- .56	+ 2.12
	- B	- 3.68	
2.	Arithmetic	- 4.22	
3.	Circle Reasoning		+ .97
4.	Classification - A	- 3.76	
5.	- B	- 4.08	
6.	Disguised Words	+ .06	- 2.90
8.	Letter Triangle - A	- 3.81	+ 1.61
	- B	- 1.82	+ 2.74
9.	Logical Consequences - B	- 2.00	
10.	Memory for Symbols	- 4.84	
11.	Memory for Words	- 2.33	+ .89
12.	Memory for Sentences	+ .19	
13.	Missing Signs - A		+ 5.67
	- B		+ 2.95
14.	Number Ability	- 2.90	+ 4.02
15.	Reading Comprehension	- 5.58	+ 1.43
16.	Starring		+ .11
17.	Symbol Elaboration - A, 1	- 3.27	- 1.79
	- A, 2	- 3.41	- .63
18.	Symbol Elaboration - B, 1	- 3.93	
	- B, 2	- 2.24	
19.	Symbolic Reasoning	- 1.30	+ 5.34
20.	Verbal Comprehension	- 1.35	+ 1.60
21.	Verbal Reasoning	- 3.32	+ .80
22.	Word Changes - I, a	- 4.50	+ .27
	- I, b	- 3.10	
23.	Word Changes - II, AW	- 3.29	
25.	Word Transformations	- 3.37	- 1.30
26.	Sentence Order - A		+ 1.92
	- B	- 5.84	+ 1.49

These comparisons make it appear that, even had a factor analysis been possible, the supposition that the two groups performed comparably in intellectual tasks would not have been supported. The general superiority of the conventionally instructed group is abundantly demonstrated across a variety of combinations of contents and operations and the prognosis for their success, judged by conventional standards, is far brighter. The reasons for this circumstance are not immediately obvious and in the absence of adequate experimental control, conjectures might be dangerous; if the same evidence of difference appeared in more reliable data, it might be attributed to the fact that most of the schools from which the conventional classes were drawn begin the study of algebra in the ninth grade, while in many of the schools in which UICSM materials have been adopted algebra is allocated to the eighth and, in a few cases, to the seventh grade. The difference of a year or more of academic preparation at this level is an important one since it occurs at about the time that children are beginning to acquire a taste for rigorous and extensive thinking and have begun to be offered a wider range of subject matters and learning experiences than they found in elementary school. About one point there can remain little doubt in the presence of these data: the allegation that the UICSM program is suitable only for the intellectual elite finds no support from this experiment, since the entering pupils in those schools cannot be found to have demonstrated superior intellectual ability in any sense.

Comparisons Between Groups: Post-Instruction

Comparison of the two groups on the basis of measures made after most of an academic year of instruction was conducted in exactly the same way, but with surprisingly different outcomes. The first comparison is of the number of differences in favor of each of the two samples which shows the UICSM sample to look very little like their September-tested colleagues who excelled in only 3 of 38 comparisons; in the post-instruction sample the UICSM pupils demonstrate an advantage in 29 of 39 measures and are low in only ten measures (39 rather than 38 comparisons are being made of the post-instruction data because usable answers to Logical Consequences - A are available for the post, but not pre-instruction, samples in the conventionally instructed group). If the two samples are randomly drawn from the populations in which means are equal for each measure, then the probability of superiority of either sample in 29 of 39 cases is about .08 (determined by normal approximation to a binomial expansion).

Table 4

Sample Size, Mean and Standard Deviation
Of Each Test in Two Post-Instruction Samples

		Conventional			UICSM			
		N	M	s.d.	N	M	s.d.	Diff.
1.	Alternate Additions - A	101	10.76	3.01	162	11.60	3.23	+ .84
	- B	77	10.25	3.08	138	10.15	4.16	- .10
2.	Arithmetic	93	20.72	5.22	196	19.60	3.51	-1.12
3.	Circle Reasoning	145	7.79	2.80	202	8.08	2.65	+ .29
4.	Classification - A	169	4.67	4.64	126	4.38	1.63	- .29
5.	- B	168	2.98	2.22	126	2.84	1.66	- .14
6.	Disguised Words	110	20.72	6.18	168	18.63	5.32	-2.09
7.	Form Reasoning - A	36	7.92	3.33	141	8.90	2.34	+ .98
	- B	36	8.47	3.05	142	9.14	2.08	+ .67
8.	Letter Triangle - A	123	4.50	1.64	237	4.80	1.75	+ .30
	- B	123	4.46	1.68	237	4.98	1.75	+ .52
9.	Logical Consequences - A	130	3.88	1.33	125	5.38	1.97	+1.91
	- B	132	3.69	1.83	125	4.01	2.24	+ .32
10.	Memory for Symbols	42	13.64	9.43	168	20.45	5.76	+6.81
11.	Memory for Words	171	13.56	5.67	220	14.04	4.67	+ .48
12.	Memory for Sentences	126	12.34	5.73	264	14.31	4.34	+1.97
13.	Missing Signs - A	138	8.31	2.16	147	9.75	2.11	+1.44
	- B	111	6.95	2.29	147	7.82	2.41	+ .87
14.	Number Ability	154	14.20	4.81	226	16.06	4.24	+1.86
15.	Reading Comprehension	170	14.39	4.55	222	15.05	4.46	+ .66
16.	Starring	173	6.52	1.82	264	6.50	1.82	- .02
17.	Symbol Elaboration - A, 1	148	6.72	4.18	192	5.94	3.66	- .78
	- A, 2	148	7.36	4.74	192	7.04	4.55	- .32
18.	Symbol Elaboration - B, 1	72	5.88	2.33	175	8.14	4.02	+2.26
	- B, 2	72	4.76	2.45	175	7.64	4.71	+2.88
19.	Symbolic Reasoning	114	10.56	4.02	202	15.76	4.63	+5.20
20.	Verbal Comprehension	109	16.51	3.82	180	17.26	3.88	+ .75
21.	Verbal Reasoning	138	18.09	4.31	160	18.52	4.94	+ .43
22.	Word Changes - I, a	93	5.20	1.52	185	5.25	1.38	+ .05
	- I. b	93	4.71	1.82	185	5.06	1.44	+ .35
23.	Word Changes - II, AW	107	.69	1.06	266	1.47	1.68	+ .78
	- II, AL	107	8.06	4.58	266	11.63	6.62	+3.57
	- II, BW	107	1.16	1.22	266	1.79	1.67	+ .63
	- II, BL	107	9.27	5.08	266	11.73	6.57	+2.46
24.	Word Patterns - A	172	145.73	14.32	173	149.92	7.31	+4.19
	- B	172	123.02	38.24	147	121.63	9.78	-1.39
25.	Word Transformations	149	24.33	12.68	277	22.70	11.34	-1.63
26.	Sentence Order - A	141	5.26	1.62	206	5.61	1.72	+ .35
	- B	141	5.21	1.71	186	5.49	1.65	+ .28

Sample size, mean, and standard deviation for each of the 39 measures in the post-instruction samples are shown in Table 4. A comparison of variances is again necessary as a preliminary to t ratio comparisons; of the 39 pairs thus examined, nineteen indicate that the two samples are not comparable with respect to variance, when the same criterion is applied (if the variance in either group is as much as one and one-half times that in the other, the two samples are regarded as having arisen from different populations). Of these nineteen, ten represent greater variance among the conventional sample, five represent higher means in the conventional sample. Seven of the nineteen were among those which showed non-comparable variances in the pre-instruction comparisons (these are designated by asterisks in the listing below), while five of the previously non-comparable tests demonstrate comparable variances in this comparison. Those measures which differ sufficiently in variance to make comparison by means of t ratio impossible are:

Alternate Additions B	Memory for Symbols
Arithmetic	Symbol Elaboration B, 1 & 2
Classification A & B	Word Changes I - B
*Form Reasoning - A & B	*Word Changes II, AL, BW, BL
Logical Consequences A & B	Word Changes II, AW
Memory for Sentences	*Word Patterns - A & B

The majority of the tasks in which the two groups do not resemble one another fall in the operations category of convergent thinking, and symbolic content is represented more frequently than semantic. This suggests, although the suggestion is short of proof, that the content of the UICSM First Course equips pupils to approach tasks or encourages receptivity toward symbolic contents and convergent operations to a greater extent than does a conventionally conducted course.

Of the 20 post-instruction measures that permit comparison by means of t tests, non-significant differences between groups are associated with 13 and significant ones with 7, compared with 19 significant differences between groups in the pre-instruction sample. The effect of differences in instruction seems to be that of bringing the two groups closer together in proficiency level (fewer significant differences) while increasing the effect of individual differences within the groups (more non-comparable variances).

Examination of the direction and magnitude of the differences in these 20 cases casts further light on the relative nature of the two courses. In the comparison of pre-instruction samples, the comparison showed 19 significant differences, all favoring the conventionally instructed group; in comparing post-instruction differences, only 9 significant differences, of which 8 favor the UICSM sample. (See Table 3)

The only test in which the UICSM pupils appear in a less favorable light in the post-instruction comparison than they did in pre-instruction comparisons is Disguised Words where a significant $\underline{t}$ is associated with a lower mean score for First Course pupils. Some aspect of the First Course seems to have left those pupils poorly equipped to deal with materials of this kind; two other measures of cognition of semantic materials (Reading Comprehension and Vocabulary) show non-significant differences in favor of the UICSM pupils in these comparisons.

The largest difference between groups is in connection with the test, Symbolic Reasoning, in which the pre-instruction group showed a non-significant difference in favor of the conventional sample while the post-instruction group shows a strongly significant difference favoring the UICSM group. The same pattern appears in connection with the Number Ability test (the simpler of the two arithmetic tests) which shows almost no difference at all between groups in the September sample and a strongly significant difference in favor of the UICSM sample in the May administrations. The implication here is that the First Course has provided its pupils with better training and/or more practice in arithmetic operations and evaluation of symbolically stated proportions than has its conventionally conducted counterpart. A similar pattern appears in the tests Missing Signs (disguised arithmetic) and Symbol Elaboration B (production of symbolic statements). The only feature that all four of these tests share is symbolic content; an attractive conjecture can be seen here, that exposure to UICSM materials and teaching methods facilitates the performance of a variety of operations if they are performed on symbols, but the nature of the experimental controls leaves this conjecture short of proof.

Reading comprehension and vocabulary tests are of special interest because one or both of them are usually included in prediction of success in mathematics courses. The relationship between groups that appears here is one seen frequently in these comparisons; in pre-instruction samples, the UICSM pupils demonstrate an obvious deficit in these tasks, evidenced by a highly significant difference in Reading Comprehension and a difference significant at about the 18% level in Vocabulary, both favoring the conventional sample. In the post-instruction comparisons, nearly significant differences in the two tests (15 and

11% levels respectively) favor the UICSM sample. In order to reverse their relative positions so completely, it must be the case either that habits and attitudes which facilitate verbal tasks have been learned more effectively in First Course, or that the pupils have matured more rapidly than conventionally instructed ones.

Comparisons Within Groups

If, instead of comparing the performance of samples from the two populations measured at the same point in their academic progress, comparisons are made of two samples within each population measured at different points in their progress, information may be obtained about the nature of any intellectual changes that may have occurred, either: (1) in both groups, (2) in one group but not in the other, or (3) in neither group.

Reduced to its simplest description, this is a comparison of the gains made by each of the two groups that may be associated with their study of algebra. The "gains" referred to here are not those of a single sample measured before and after instruction, but are differences between inferred populations means based on separate samples from each population. To test a single group twice brings into the inference two troublesome characteristics of gain scores: the measurable diminution in reliability of a difference score when it is based on correlated measures, and the non-measurable possibility that post-instruction scores are influenced by the pupils' memories of the earlier test. In the case of separate samples, no correlation is present to diminish the usefulness of the comparison; instead, the inference rests on the assumption that each sample is randomly drawn from its own population, and that the performance of the first sample is a valid estimate of the level of ability that the second sample would have displa[illegible] had they been tested before instruction began.

Those differences that are found to occur in the same direction and with approximately the same magnitude in both populations can reasonably be attributed either to maturation or the effects of studying algebra per se, while those which indicate the absence of a difference between pre- and post-instruction samples are assumed to be based on abilities which have no relationship to algebra and are not influenced by its study. When a difference is found to exist in one group but not the other, then it may be inferred that the course in which that gain is noted has provided its pupils with knowledge, attitudes, or intellectual habits which facilitate performance of the task represented by that test.

The basic data on which these comparisons are based are the means, standard deviations, and sample sizes shown in Tables 2 and 4. The differences between samples means within each treatment group and the $\underline{t}$ ratios associated with them are shown in Table 5.

Because these comparisons are made in the same way as those between groups they are conducted along the same lines. The plausibility of the assumption of change may be quickly evaluated simply by counting the number of gains, whatever their magnitude. In the conventionally instructed sample, 38 comparisons are made (Logical Consequences A is not included in the set) and the post-instruction mean exceeds pre-instruction mean in 23 of them. If the two groups are random samples from a single population, i.e., no intellectual changes have occurred, then the probability of 28 gains in 38 independent trials is about 0.2. In the UICSM population, 39 comparisons are made and superior performance by the post-instruction group occurs in 38 cases; given the same assumption that the probability of chance increase is $\frac{1}{2}$, then the probability of 38 gains in 39 trials is of the order of 10^{-8}.

On the surface this seems to constitute powerful evidence in favor of the UICSM course but it must be recalled that in the between-groups comparisons, the pre-instruction UICSM sample was found to perform poorly as a group in almost every measure. Since this comparison involves that same laggard sample, some of the differences between pre- and post-instruction groups may be attributed to the fact that, of the two post-instruction groups, that from UICSM schools is being compared with a far lower set of pre-instruction scores. It cannot be determined from this experiment whether the fact of exceptionally low pre-instruction performance in the UICSM population represents an atypical sample or whether the schools which adopt UICSM materials characteristically begin instruction in algebra with pupils who are, in fact, less academically competent at the kind of tests represented in this battery.

Just as in the case of between-group comparisons, attention must be given to establishment of comparability of variance between the two samples in each treatment group. Improved performance is expected when comparisons are made between two groups of pupils who differ by a full year of instruction and it is not beyond reason to suppose that systematic shifts in ability will occur which are not only very large (difference between means), but are demonstrated equally by all of the pupils within a population (differences in variance). The same procedure and the same criterion that were applied before are applicable here; the quotient obtained when the larger variance was divided by the smaller in each of the 77 cases was compared with the critical value of 1.5, that is, when the variance in either group was as much as one and one-half times that in the other, the groups were regarded as differing in variance to such an extent that comparison by means of a t test was not legitimate. When the ratio of variances indicated that the two samples could be regarded as having arisen from a single population, a t ratio was calculated and the test-by-test comparisons of gains described here is concerned with these ratios. (See Table 5)

The number of comparisons made here is rather large and examination of them may be facilitated by division into three categories: (1) those cases in which a non-significant difference exists between groups, (2) those cases in which a significant difference exists between groups, and (3) those cases in which the difference between variances prohibits direct comparison of the difference between means. Because any of these three categories can involve either a score improvement (gain) or a decrement (loss), there are six possible categories into which a difference might fall, and because interest centers on a comparison of gains, any of the six occurring in one population might be found in combination with any of the six in the other population. Of the thirty-six possible kinds of comparisons thus generated, only fourteen are found to occur, however, because of the absence of any decrements in the UICSM population and because significant losses do not occur in either population.

The thirty-eight pairs of comparisons that are to be examined in the following pages fall into these four general categories:

Directly comparable differences are found in both populations	10 pairs
Comparable variances are found in the UICSM samples accompanied by non-comparable variances in the conventional samples	13 pairs
Comparable variances are found in the conventional samples accompanied by non-comparable variances in the UICSM samples	12 pairs
Non-comparable variances are found in both populations	3 pairs

Table 5

t Ratios Associated With
Differences Between Means

Within-Groups Comparison

		Conventional	UICSM
1.	Alternate Additions - A	+3.72	+ .73
	- B	+1.38	
2.	Arithmetic		+4.74
3.	Circle Reasoning	+4.40	
4.	Classification - A		+1.32
5.	- B	-1.53	+ .93
6.	Disguised Words	+4.92	+2.08
8.	Letter Triangle - A		+6.64
	- B		+7.20
9.	Logical Consequences - B	+1.95	+4.49
10.	Memory for Symbols		+7.34
11.	Memory for Words	+1.35	+5.22
12.	Memory for Sentences		+1.33
13.	Missing Signs - A		+1.15
	- B	- .54	
14.	Number Ability		+8.80
15.	Reading Comprehension	+4.80	+6.00
16.	Starring		+2.81
17.	Symbol Elaboration - A, 1	+1.11	
	- A, 2	+2.91	
18.	Symbol Elaboration - B, 2		+6.08
19.	Symbolic Reasoning	-1.17	
20.	Verbal Comprehension	+2.30	+6.06
21.	Verbal Reasoning	+4.78	+6.00
22.	Word Changes - I, a	+2.60	
	- I, b	+2.24	
23.	Word Changes - II, AW	-1.95	
	- II, AL	-1.55	
	- II, BW	-0.83	
	- II, BL	+ .81	
24.	Word Patterns - B		+4.87
25.	Word Transformations		+3.32
26.	Sentence Order - A		+4.90
	- B	+ .69	+5.17

The immediately obvious characteristic of this kind of classification is the number of tests in which variances are comparable in one of the treatment populations but not in the other. This seems to constitute strong indirect evidence for the existence of a basic difference between the outcomes of the two methods of instruction, since a large change in variance is regarded as a change in the importance of individual differences within a group. Had these individual differences been influenced in the same way under both kinds of instruction, the tendency would have been toward agreement in those tests in which variances were altered.

Attention is directed first to the ten tests in which comparable variances are established in both populations. Five of these show significant gains between pre-instruction and post-instruction means in both populations. They are:

Disguised Words
Logical Consequences - B
Reading Comprehension
Vocabulary
Verbal Reasoning

Of these five, Disguised Words is the only test in which the gain found in the conventionally instructed group exceeds that in the UICSM group either in magnitude or degree of significance. All of the five represent semantic content and four of them are classified as cognitive operations (Logical Consequences is believed to be a test of convergent production). From this it is inferred that increased facility with semantic content, particularly with cognition of semantic materials, is associated with the year's growth that intervenes between the beginning and the end of the first course in algebra, no matter what kind of instruction is provided. Extension of this inference to the association of that change with the study of algebra is not supportable since no data is available from ninth graders who have not studied algebra.

A test in which neither group demonstrates significant change over the year's study of algebra is interesting because it presumably represents content and operation which are not associated either with algebra or maturation. Only one of the measures used here falls in this category, however; the test Classification - B shows a slight gain between samples in the UICSM population and a slight loss in the conventionally instructed group, but neither of these changes is significant. This test requires finding a common element in each of two groups of geometric figures and recognition of that element in a problem figure and is believed to be a measure of cognition of symbolic classes. A single measure has no inferential significance but it may be assumed that this ability is not affected by the study of algebra.

The strongest kind of indirect information pertaining to differentiated intellectual changes that may be obtained from this kind of analysis is that from tests in which significant differences between pre- and post-instruction measures is found in one population but not the other. Four such cases are found in this experiment. All of them involve significant differences in the UICSM group in tests for which non-significant differences are found in the conventional group; three of those non-significant changes represent gains, on one a loss. The four tests which show this pattern of change are:

Sentence Order - B	Memory for Words
Starring	Word Transformations

The test, Starring, represents the only score loss in this group; the mean of the post-instruction sample in the conventional group is lower than that of the pre-instruction sample. Two of these tests involve the operation of convergent production, one of cognition, and one of memory; two deal with symbolic content and two with semantic. With only four instances from which to generalize, patterns are difficult to detect, and although it may be the case that the UICSM First Course results in intellectual abilities which differ systematically from those developed in conventional courses, that difference is not apparent in these data.

Comparable variances in the UICSM sample accompany non-comparable ones in the conventionally instructed sample in thirteen of the tests. Before attempting to interpret these cases, however, it is well to review some of the circumstances which can give rise to heterogeneity of variance and to examine some of their implications, not only because they can vary with a number of circumstances, but also because they represent the sole source of information about some of these abilities and modes of algebra instruction.

The usual and to-be-expected outcome of comparing test scores over instructional sequences is the demonstration of a higher mean and greater variance in the instructed sample. This is interpreted as an indication that the instruction is relevant to the task presented by the test and that the group as a whole, has benefited from that instruction, but that some pupils have benefited more than others. This circumstance is found in many of the tests used in this experiment, and, to the extent that it does occur, is easily regarded as a true (non-chance) alteration in the intellectual performance of the pupils, but the fact that it frequently occurs in one of the groups but not the other obscures the interpretation.

A decreasing mean (post-instruction groups score less well than pre-instruction groups) is interpretable in terms of the assumption of differentiated intellectual ability and factorially complex instruction. When the test content is highly specialized, not familiar to the examinee, and homogeneous within tests, a decrement in performance following instruction might occur when one of the several elements in the instruction is incompatible with one or more of the several elements in the test. If the instruction is consistent and well-organized and does not equip the learner to perform some specified task (an almost universal condition for no instruction equips the learner to perform all tasks), it results in a set of consistent and well-organized intellectual habits which are applicable to some tasks but not others. High school pupils, like everyone else, approach a strange task by whatever route seems to offer the highest probability of success, and a combination of this tendency with inappropriate instruction places them in the position of being especially ready to undertake the unfamiliar task in the familiar, but inappropriate, way or to perceive the unfamiliar content in the accustomed, but inefficient way. The pupils have learned explicitly and implicitly what the text and teacher have rewarded them for doing and, the fact that neither the author nor the teacher was aware that that lesson was being taught, does not make the pupils less ready to demonstrate it.

The direction of change in variance that occurs in the presence of a decreasing mean carries the same implication as the same change in the presence of an increasing mean. Like most lessons, the interfering one is not learned with the same degree of efficiency by all of the pupils to whom it is offered; if the instruction is efficient and the interference is shared to the same extent by all pupils, then less variance will be found in the instructed group. If some pupils fail to learn them in such a way that later performance is not affected, or if the similarity between instruction and test is not perceived by some, then the variance in the instructed group is increased. Any change in standard deviation of data collected in this manner reflects the uniformity with which those elements that facilitate or inhibit test performance have been learned. Highly effective instruction tends to decrease variance because it has the same effect on all pupils no matter whether that effect is to increase or to decrease level of later performance; instruction which applies less uniformly increases individual differences within a group and tends to increase variance among pupils who are exposed to it because some fail to benefit — or to suffer from — its effects.

Of the thirteen tests in which the conventionally instructed samples cannot be directly compared with one another but the UICSM samples may be compared, nine represent significant gains in the UICSM population and four represent non-significant gains. The nine tests in this category in which the UICSM post-instruction sample performs significantly better than the pre-instruction sample are:

Arithmetic	(i, i)	Letter Triangle A & B	(i, d)
Word Patterns A & B	(i, i)	Symbol Elaboration B, 2	(d, d)
Number Ability	(i, i)	Sentence Order A	(d, d)
Memory for Symbols	(d, i)		

The letters "i" and "d" following the test names indicate the direction (increasing or decreasing) in which the means and variances, in that order, of the post-instruction conventional sample differ from those of the pre-instruction sample.

All of these tests involve symbolic content, but a variety of operations are represented (3 memory, 4 cognition, and 2 convergent thinking). This is not inconsistent with a conclusion tentatively cited earlier that algebra courses are most dependably accompanied by improvement in dealing with semantic material. Arithmetic, Word Patterns, and Number Ability all have implications as their products and all show disproportionally large increases in variance in the conventionally instructed population which may be evidence that conventional algebra courses have a less uniform effect than First Course in training pupils to recognize implications.

The four tests in the category of non-comparable variances among the conventional group in which the UICSM pupils register non-significant gains are:

Alternate Additions A	(i, i)	Missing Signs A	(d, i)
Memory for Sentences	(d, i)	Classification A	(d, i)

Two of these tests represent semantic content and two symbolic; four kinds of operations are included, and two require relations as a product, one classes, and one implications. If any conclusion can be drawn from this array it would probably not be one which disagreed with that cited above. The fact that only one member of each of two presumably comparable pairs is represented here (Missing Signs and Alternate Additions) is a disquieting circumstance in view of the questionable nature of the data; although samples are large, some undetected irregularities of administration may be represented here. The poorer performance of the conventionally instructed group in the Memory test is an unexpected occurrence and may contradict the assumption that conventional practices in mathematics teaching force the pupil to rely on memory.

The second category of non-comparable vs. comparable variances is the one in which non-comparable variances in the UICSM population are matched by comparable ones in the conventionally instructed. Four of these involve significant differences in the conventional population and eight involve non-significant differences.

The four in which conventionally instructed samples yield a significant increase in mean score are:

Circle Reasoning	(d, d)	Symbol Elaboration A, 2	(i, i)
Word Changes I, A & B	(i, d)		

The "d" and "i" designations have the same meaning as before except that they refer to direction of change of mean and variance in the UICSM population.

Circle Reasoning is unique among the tests in this experiment in being the only test in which post-instruction UICSM sample shows a lower mean than the pre-instruction sample. As a measure of cognition of symbolic systems it might be expected to behave in the same way as Letter Triangle and Starring, both of which show significant increases between samples. The decrement in mean score is a relatively large one, however, and that in standard deviation is so large that doubt is cast on the legitimacy of the data; the standard deviation of the post-instruction sample is in good agreement with that shown by other samples in other experiments so the pre-instruction sample should be regarded with suspicion.

The small variance in post-instruction scores in Word Changes arises from the nature of the test; the mean score in both forms approaches the number of items in the test (mean scores greater than five in a test of six items) so an artificial limit is placed on maximum scores and many pupils have reached it. Given a longer test — or a shorter time limit — it is possible that greater variance would be found in the post-instruction group and a direct comparison would be possible.

Eight measures which were directly comparable in the conventionally instructed group but did not reach significance, were not comparable in the UICSM group. These are:

Alternate Additions B	(i, i)	Symbol Elaboration A, 1	(i, d)
Symbolic Reasoning	(i, i)	Missing Signs	(i, d)
Word Changes II (all)	(i, i)		

The questionable status of the tests Alternate Additions and Missing Signs because of the inclusion of their two halves in different general categories has already been mentioned and any interpretation of these two tests must be held suspect. The UICSM post-instruction sample exceeds the performance of the pre-instruction sample by a much larger margin in Word Changes II and Symbolic Reasoning than in most of the tests; these are classified as symbolic in their content, one requires convergent production, the other cognition.

The fourth general category of comparisons made is that in which the variances are not comparable in either of the treatment populations. Only three measures are represented in this category. They are:

Symbol Elaboration B, 1   Form Reasoning A & B

All of these are measures of convergent production and this reinforces the suggestion already offered that both kinds of algebra courses are less uniform in their effects on pupils with respect to improvement of production than cognition tasks.

Form Reasoning, in the UICSM population can be accounted for in the same way as Word Change I; the second sample has approached the highest possible score so closely that the decrease in variance is regarded as due to an artificial restriction on the examinees rather than to change in the importance of individual differences; it must be the case that general level of performance has improved markedly in that population if many of the pupils have achieved a maximum score. The small decrease in mean score for this test in the conventionally instructed population is regarded as an indication that some aspect of that instruction has caused some of these pupils to be either less able or less willing to evaluate such implications than they must have been before instruction began.

Symbol Elaboration B, 1 has changed in the expected manner in the UICSM population (mean and variance increase) while the opposite effect is noted in the conventionally instructed group. In the same manner as before, this may be regarded as evidence that ability or willingness to draw or state implications has been interfered with in the same manner for all of these pupils. This test calls for production of implications from given statements which involve both equalities and inequalities; expressions of this sort are more heavily represented in First Course than in most algebra courses, so the element of familiarity may be at work in accounting for the differences between the two kinds of instruction in their effect on this test.

Prediction of Performance in The Two Courses

One of the objectives stated for the experiment was the generation of prediction measures by which the placement or assignment of pupils to algebra courses might be facilitated. Except for the reduction of total sample size, this phase of the experiment is not seriously hampered by the inadequacies in the data collection process. Given a matrix of correlations between these tests and between each test and the criterion, it is possible to compute the best combination of tests and weights for prediction of each criterion. By comparing these, it is then possible to gain some insight into the probable nature of the criteria.

Beta, rather than beta-r products, are reported here because these directly to the proportion of variance for which they account without influence by any indirect influences and without respect to the standard deviations of the measures involved. The reader who is concerned with exact prediction equations can derive them readily from the values reported here and the test statistics in Table 2.

The criterion predicted in the conventionally instructed group was the Cooperative General Mathematics Test (Form Y), selected because it represents a widely known and accepted measure of arithmetic and algebra competence of a traditional sort. It was administered in the spring of the experimental year, as nearly as feasible to the end of the second semester, to the same pupils who had completed the various parts of the experimental battery at the beginning of the year. After removing those cases about which reasonable suspicion existed concerning the conditions of administration (See Chapter V), 138 presumably valid answer sheets remained.

The stipulation of "best" prediction, in this case, refers to the best prediction possible with a manageable number of predictors. The proportion of variance accountable for by most regression equations can be increased slightly by the addition of one or two, or occasionally more, predictors. The proportion of new variance attributable to these terms is, however, very small and the effect of including them is to make the equation more cumbersome. The addition of one or two percent to accountable variance may, in some applications, justify the inclusion of extra terms but the equations described here exist only for the purpose of being compared with one another and minute increases in precision would add nothing of importance to this comparison even if the data were dependable enough to justify them.

The "best" prediction that can be made of the conventional criterion from among the 25 experimental tests used in this experiment is based on the standard equation:

$$Y' = .61 \text{ Reading Comprehension} + .40 \text{ Symbol Elaboration B} + .26 \text{ Word Changes I (b)} + .33 \text{ Word Transformations}$$

This equation will account for 86% of the criterion variance, which is assumed to be approximately the proportion of reliable variance in that criterion. It represents a multiple R, therefore, of 0.93 (before correction for shrinkage).

The largest beta associated with this criterion is that for the test, Reading Comprehension, an outcome consistent with the usual finding that measures of verbal proficiency are the best single predictors for beginning mathematics courses. It appears that, when it is to be taught in the traditional manner, the possession of a large vocabulary and/or above-average skill in reading constitute an advantage in mastering beginning algebra.

The other tests represented in the equation all are measures of convergent production of symbolic material; two of them (Symbol Elaboration and Word Changes) are classified in the systems category with respect to production and Word Transformations calls for a transformation as its product. The fact that all of these represent symbolic content seems to vindicate the original assumption that the content and method of algebra are such that symbolic materials are appropriately used for predicting achievement. The fact that all except Reading Comprehension are measures of convergent thinking is of special interest in light of the repeated suggestions found here that cognition scores are more dependably increased that those of convergent productions during the first year's study of algebra. If this is the case, then to the extent that convergent productive thinking is represented in the criterion, measures of it would be effective predictors of success since the people who already possess the ability will be able to use it to advantage while those who do not cannot expect to find occasion to acquire it during that year.

These results are in general agreement with those of a similar experiment reported by Petersen et al (1963) in which regression equations based on factor scores are developed for several kinds of mathematics courses. His two categories, "Regular" and "Accelerated" algebra, are of importance here. Those equations indicate that the factor of memory for symbolic implications becomes progressively less important as the level of complexity of subject matter and the pace of instruction increase. That factor is represented in this experiment by the tests, Arithmetic and Number Ability, neither of which demonstrates sufficiently large correlations with the criterion to justify its inclusion in the regression equation. It might be taken for granted that, at this level of proficiency, most of the pupils were already sufficiently skillful in arithmetic manipulations to eliminate differences in that ability as a source of variance in the criterion.

That experiment also attributes relatively little importance to the factor of convergent production of symbolic systems which is heavily represented here (Circle Reasoning, Letter Triangle, and Starring) and which was expected to play an important role in predicting achievement in algebra. One of the summary statements cited in that report seems applicable here: "The problem of aptitude for success in ninth grade mathematics is even more complex than was anticipated, and will require a broader sampling of intellectual abilities before it is solved."

Specifically, these results suggest that prediction of success in beginning algebra must continue to rely heavily on measures of verbal facility, that measures of convergent productive thinking with symbolic products offer sufficient promise to warrant further investigations, and that the abilities described by the structure of intellect as evaluative need to be examined for their usefulness in such predictions.

Prediction of success in UICSM courses was investigated by the same procedure. The criterion predicted in this case was a constructed response test covering the content of Unit III, prepared for this experiment. This Unit was selected because it is the fartherest point in First Course that will be reached by most classes in two semesters; many finish it, but few proceed far enough into Unit IV to make its use as a criterion feasible. Concern for the use of a criterion as far removed as possible from the beginning of the course originated from unpublished data in the files of the UICSM which indicated that, insofar as the factorial content of the Unit examinations has been determined, verbal comprehension becomes progressively less important as instruction proceeds.

The best prediction that can be made of this criterion (as "best" was defined earlier) from the experimental measures available is:

$$Y' = .20 \text{ Alternate Additions} + .27 \text{ Symbol Elaboration B} + .33 \text{ Verbal Comprehension} + .25 \text{ Word Changes I}$$

This prediction will account for 65% of the variance in the criterion examination, indicating an uncorrected multiple R of 0.81; this value is somewhat lower than the proportion of variance accounted for in the conventionally conducted courses. This, and the report by Petersen (loc cit) that the operation of evaluation is of relatively greater importance to accelerated than to standard algebra courses, suggests that the inclusion of measures of that operation in the experiment might have improved the prediction in First Course.

This equation resembles the one cited above for prediction of conventional algebra courses in two important ways: it contains an expression for measure of cognition of semantic units (Vocabulary), and measures of convergent production of semantic products are well represented. It differs from that equation, however, in two ways: (1) the importance attached to the verbal measure is far less (that of .2 compared to .6 in the conventional predictive equation), and (2) it contains a measure of divergent productive thinking which is missing from the other equation.

The resemblances make it appear that there is a basic core of algebraic aptitude which rests on verbal comprehension and convergent production of symbolic materials, although the emphasis placed on these kinds of abilities differs between courses, and it seems reasonable to expect that measures of verbal and of convergent productive abilities will be found in predictions of algebra achievement in other experiments.

Some of the differences between the two equations in the importance they attribute to verbal measures may reside in the nature of the two criterion measures, since that for UICSM pupils contains fewer verbally stated problems than does the Cooperative Test. Some may be attributed to the nature of instruction in the two courses, because UICSM teachers and texts emphasize the importance of precise, rather than profuse, verbal expression and there may be less inclination to reward complex and sophisticated verbal productions; in this case, the verbally apt pupil would be given less advantage over his mathematically competent but less articulate colleagues. It may also be associated with the difference between the two groups in pre-instruction measures which allowed the UICSM pupils to register a larger apparent gain in vocabulary scores than the conventionally instructed examinees were able to do. It does not seem to be associated with a difference in verbal score variance between the two groups, because the standard deviations in the two groups' vocabulary scores are comparable in both pre- and post-instruction samples.

The importance given by this equation to convergent production abilities reaffirms the repeated suggestions found in this experiment that convergent production is less dependably changed during the first year of study of algebra than is cognition, and its predictive efficiency can rest on the same basis suggested above.

The existence in this equation of a term for divergent production is its most provocative characteristic. That test (Alternate Additions) had the same opportunity to predict performance in both populations, but its usefulness for that purpose differs between criteria; it showed a non-significant difference between groups in the pre-instruction samples and a significant one in favor of the UICSM sample in the post-instruction samples. The apparent existence of an element of divergent productive ability in the UICSM courses and its apparent absence from conventionally conducted courses may be an indicator of the basic difference between the two and between the kinds of intellectual behavior they encourage.

The absence of measures of the factor of cognition of symbolic systems from this equation and from the other predictive measures reported and described here, casts doubt on the original assumption of importance of this factor in learning algebra.

Sex Differences

One of the objectives of the experiment was an investigation of the difference between sexes, both with respect to preferred intellectual modalities and predictability. The original intention was the production of relationships that might help to explain the consistent differences between sexes in predictability and, perhaps to point to a treatment or instructional strategy that could help to improve the accuracy of prediction of performances of male students. This expectation was based on the plan of interpreting factorial structures in the two populations and, when that possibility was eliminated by the nature of the data, the search for sex differences was channeled into comparisons of achievement in the two sex groups and of the predictive equations that best fit each of them.

The first and most direct kind of analysis that permits an examination of the relations between sex and each criterion is the correlation between that criterion and sex. In the case of the conventionally conducted classes, the correlation between sex and the Cooperative Math test, based on 138 cases, is -.011 which falls far short of the value of 0.17 that would constitute significance at the 5% level for that sample size, so it is concluded that, in this case, there is no dependable relationship between sex and achievement. In the UICSM pre-instruction sample, on the other hand, the reported correlation of -.098 between sex and the criterion examination, based on 687 cases, is significant at the 1% level (the negative sign indicates that higher criterion scores are found in the male group). Because of this correlation, separate regression equations are reported below for predicting performances of boys and girls.

The correlations between sex and each of the experimental tests, reported in Table 6, are based on the pre-instruction samples from each of the two treatment populations and the sample size, shown in Table 2 varies somewhat from test to test, but it is approximately 300 for each correlation. In a sample of this size, a correlation of 0.12 would be significant at the 5% level, and a correlation of 0.15 would be significant at the 1% level. The correlations which reach these values are:

Conventional Instruction		UICSM	
Memory for Sentences	.238	Memory for Words	.242
Word Changes II, BL	.214	Word Patterns, B	.134
Word Changes II, BW	.199	Word Changes I, a	.121
Memory for Words	.184	Classification I, a	-.120
Form Reasoning, B	.173	Reading Comprehension	-.170
Disguised Words	.150		
Word Changes I, A	.144		
Word Changes I, B	.140		
Number Ability	.129		
Alternate Additions, A	.124		
Reading Comprehension	.124		
Form Reasoning, A	.119		
Symbolic Reasoning	-.161		

Examining these for statistical significance in the usual way, the most obvious characteristics are, first the predominance of

Table 6

Correlations of Criteria and Experimental Tests With Sex

		Conventional	UICSM
1.	Alternate Additions - A	.124*	-.064
	- B	-.066	-.099
2.	Arithmetic	.090	-.060
3.	Circle Reasoning	.017	-.002
4.	Classification - A	.028	-.120*
5.	- B	.020	-.005
6.	Disguised Words	.150*	-.013
7.	Form Reasoning - A	.119*	-.096
	- B	.173**	-.065
8.	Letter Triangle - A	.012	.024
	- B	-.064	.087
9.	Logical Consequences - A	--	.061
	- B	.034	.023
10.	Memory for Symbols	.096	.083
11.	Memory for Words	.184**	.242**
12.	Memory for Sentences	.238**	.048
13.	Missing Signs - A	.018	-.093
	- B	-.044	.007
14.	Number Ability	.129*	-.073
15.	Reading Comprehension	.124*	-.170**
16.	Starring	-.046	-.103
17.	Symbol Elaboration - I, a	.068	.013
	- I, b	.086	.023
18.	Symbol Elaboration - II, a	.060	.047
	- II, b	.069	-.020
19.	Symbolic Reasoning	-.161**	-.061
20.	Verbal Comprehension	.068	.069
21.	Verbal Reasoning	.089	.048
22.	Word Changes - I, a	.144*	.121*
	- I, b	.140*	.055
23.	Word Changes - II, AW	.049	-.061
	- II, AL	.113	-.090
	- II, BW	.199**	.036
	- II, BL	.214**	.036
24.	Word Patterns - A	.045	.058
	- B	-.054	.134*
25.	Word Transformations	.108	-.017
26.	Sentence Order - A	-.112	.114
	- B	.014	.006
	CoOp Algebra Unit III	-.011	-.098

positive relationships (evidence of superior performance by the girls), and second the greater number of significant relationships in the conventionally instructed group. The direction of relationships is that which might be expected since comparisons such as these typically indicate that girls not only demonstrate superior performance, but that their performance is more predictable than that of boys.

The second characteristic of the comparison, that a greater number of non-chance relationships exist in the conventional population, seems to substantiate a conclusion already pointed out in other contexts, that the original expectation of comparability of beginning algebra students between the two kinds of courses, was not confirmed. There is a reasonable suspicion that, even had the data been capable of supporting a factor analysis, the two populations would have demonstrated different structures; any differences that were noted in post-instruction samples in that case would have been difficult to interpret.

The superiority of girls' performance in the usual predictive measures (Number Ability and Reading Comprehension) in the conventional group are consistent with precedent; the reversal of this situation in the UICSM group only lends further support to the disquieting suspicion that different kinds of pupils are entering the two courses. The source of this apparent difference is not discernible through this experiment. It cannot be selection within schools because the schools from which the conventionally instructed samples are drawn do not include any UICSM materials in their curricula. It is not likely to be an artifact of selected sampling between programs since each sample is drawn from a number of schools, and it is even less likely to be the result of within-school selection in the UICSM sample because these pupils have been shown to be less academically apt than the conventional sample. It is possible, and this account of the difference seems quite feasible, that the difference is due to the age differential between groups based on the practice of those schools which have adopted the UICSM materials of beginning algebra instruction before the ninth grade.

Memory tests are well represented in both populations and, in each appearance, favor girls. No systematic difference between cognition and convergent production can be seen and there is only a poorly defined tendency for symbolic content to appear more frequently than semantic in the conventionally instructed group. Word Changes, in one or another of its forms, appears in both listings, still indicating superiority by girls. Two of the tests that show sex differences in the conventional population, Reading Comprehension and Word Changes I, and one in the UICSM population, Word Changes I also appear in the equations for predicting semester end achievement.

The negative relationships, indicative of superior performance by boys, are interesting because of the peculiar contradiction they contain if the two populations are compared with respect to the contents to which these tests refer. In the conventional group, boys excel only in Symbolic Reasoning (cognition of symbolic relations) while the boys in the UICSM group demonstrate superior performance in two semantic tasks, Classification A (convergent production of semantic classes) and Reading Comprehension (cognition of semantic units). Any attempt to deduce a relationship from this fragmentary evidence would be dangerous but, at a minimum level, these correlations conform to the repeated suggestions that the two populations of pupils differ in some essential way.

No systematic attempt can be made to account for so complex a subject as sex differences within the limits of this experiment, but it should be noted that a definable class of similarities is common to both populations, and that a scattering of distinguishable differences also exists between them.

Sex differences can also be examined as group differences were before, by comparing the regression equations which predict performance within each sub-group. In the case of the conventionally instructed group, no purpose seems to be served by this distinction because sex difference in achievement, as it is reflected in the correlation between sex and the Cooperative criterion, is negligible. In the UICSM population, however, a significant relationship between sex and criterion score is demonstrated (although it accounts for only about 1% of the criterion variance) and this difference can be investigated by examining the differences between the regression equations that predict a criterion common to both sexes.

For boys, the best available regression equation (in standard form) is:

$$Y' = .34 \text{ Symbol Elaboration B} + .24 \text{ Vocabulary} + .31 \text{ Word Changes I} + .17 \text{ Arithmetic}$$

This equation accounts for 48.2% of the criterion variance, indicating an uncorrected multiple R of 0.69; by adding 3 other tests to the equation, an increase in accountable variance of about 2% could be achieved, but the practice followed in both equations has been to eliminate all beta weights less than 0.1.

For girls, the best available regression equation is:

$$Y' = .20 \text{ Memory for Symbols} + .34 \text{ Starring} + .16 \text{ Sentence Order} + .35 \text{ Symbol Elaboration B} + .25 \text{ Word Transformations} + .17 \text{ Vocabulary}$$

This equation accounts for 77.9% of the criterion variance, indicating an uncorrected multiple R of 0.88.

Two resemblances are immediately obvious in these equations; both contain vocabulary measures, although they differ in the weight given that test, and both involve the test Symbol Elaboration B. This supports a generalization already offered, that even though differentiation of ability may occur, there also exists a core of algebraic aptitude based on verbal and on convergent production abilities and that emphases may differ between circumstances of instruction.

The inclusion of the test, Symbol Elaboration B, in both of these equations is of some interest since it appeared in the prediction equations already reported for all UICSM students and for conventionally instructed ones as well. It requires that the examinee, having been given a set of statements which involve both equalities and inequalities, write as many new statements as possible that are consistent with (implied by) the given set. It is classified as a test of convergent production of symbolic transformations and has functioned more effectively as a predictor of success in algebra than a similar test which poses the same task but uses statements based only on equalities.

The differences between these equations are highly visible, highly interesting and not easy to interpret. The boys' equation is not only less complex than that for girls (contains fewer terms), it attributes less importance to vocabulary, and it provides for inclusion of an arithmetic score which is missing from the girls' equation. Any of these circumstances might be explainable singly but, taken in combination, they cannot be interpreted unequivocally in the form in which they appear here.

The presence of any test score in a regression equation for prediction signifies clearly enough that some ability measured by that test is involved in the criterion task, and that some individual differences in that ability exist within the group both in the predictor test and the criterion. The absence of an expected term from such a equation can signify not only the absence of those two conditions, but it can also signify that variance in the missing test is represented elsewhere in that equation. If it were possible to determine which of these three conditions was the basis for omission of the specified test interpretation of the two equations would be simple enough but, since that information is not readily available, there must exist some uncertainty regarding the interpretation of the differences between the two equations.

The large difference in weights assigned to vocabulary scores in the two groups, despite the similarity of the correlations with the criterion in both groups, might indicate that girls approach the learning task or the criterion examination in a less verbal manner than boys; this would certainly be support for the experimental hypothesis dealt with here, but it need not be the only plausible account of the origin of the difference. Vocabulary represents the only semantic content in the boys' equation while one feature of the added complexity of the girls' equation is the inclusion of another semantic test (Sentence Order) and it cannot be determined whether the variance not represented in the vocabulary test is might be that found in Sentence Order. The latter version of the difference is, in oblique way, support for the idea of differentiated intellectual abilities, but it does not conform to the original idea of differentiation. To invoke the second account of the difference implies that the kind of differentiation involved is that of operations and perhaps products, but that contents are interchangeable. By a similar sort of process, the absence of arithmetic from the girls' equation could be linked to the presence of Memory for Symbols, if the conception of differentiated intellectual ability is extended to ascribe the differentiation to contents and regard operations as interchangeable. The extension of this line of reasoning might result in the peculiar specification that, when all of the abilities (contents and operations) relevant to success in algebra have been identified, then any battery which included them might be made to function as a predictor without regard for the combinations in which those contents and operations are organized.

On a more realistic level, there seems to be more evidence in this comparison to support the idea of differentiated intellectual ability among girls than among boys. The greater complexity of the prediction equation for predicting girls' scores in the criterion examination is regarded as evidence that girls approach the study of algebra in a greater variety of ways than do boys, and that a wider variety of abilities can be regarded as mathematical aptitude. This conclusion is in general agreement with the hypotheses and is entirely consistent with the unpublished prediction studies already cited which gave rise to the experiment reported here.

VII. Summary and Conclusions

A set of hypotheses was formulated from examination of published and unpublished research, from informal expectations regarding the work of the University of Illinois Committee on School Mathematics, and from the structure of intellect model. When it is considered in this way, the learning of mathematics is regarded as a very complex task that can be performed in a variety of ways; these many kinds of learning share certain characteristics, but may differ in many others. The relationship between this point of view and the procedures followed in the UICSM First Course are discussed, along with the possibilities that any of several kinds of abilities can be regarded as aptitude for mathematics, and that a number of kinds of intellectual growth, apart from subject-matter achievement, may occur in connection with the first year's study of algebra.

In terms of the structure of intellect model, the abilities regarded as most important to learning mathematics and which are most likely to be cultivated in a mathematics class are those which require the operations of cognition and convergent production performed on symbolic and semantic content. A battery of tests selected to represent several combinations of these operations and contents was assembled to provide data by which the hypotheses might be tested.

Figural content was not heavily represented in the battery and the operation of evaluation was not represented at all. The operation of memory was included largely because of the stipulation that arithmetic ability is an instance of memory for symbolic implications. The operation, evaluation, was not included in the experiment and the operation, divergent production was originally included but circumstances external to the experiment made it impossible to complete the analysis of those data. The six kinds of products described by the structure of intellect model were approximately equally represented (with the exception of classes) although no particular effort was made to achieve that equality and no hypotheses were advanced concerning products.

Two samples of pupils about to begin the study of algebra were tested; one was drawn from schools in which the materials and methods developed by the UICSM were in use, and the other from schools which utilize other materials. At the end of the same school year, the subject-matter proficiency of those pupils was measured by appropriate criterion tests, and the same experimental tests were administered to another sample of pupils from the same schools. Most of these tests were administered by classroom teachers.

Several unexpected difficulties developed in connection with the data gathering process and their cumulative effect made the original plan of comparing factor structures between samples appear unfeasible. The analysis was carried out in terms of a comparison of means and variances between and within treatment groups and of the regression equations which predict success in either kind of mathematics course. Sex differences were also examined.

The comparisons made in this way show that:

1. Pupils entering conventionally instructed algebra courses excelled in nearly all of the experimental measures to such an extent that serious doubt is cast on the original assumption that the factorial structure of the experimental battery would be the same in the two groups.

2. At the end of a year's instruction in algebra, pupils enrolled in UICSM courses excelled in more than half of the experimental measures.

3. When comparisons are made between treatment groups of uninstructed and instructed samples, the tendency is for the instructed samples to resemble on another more closely than the inctructed samples with respect to means, but to differ more than the uninstructed samples with respect to variances. This result would follow if both kinds of instruction were relevant to the tasks presented by the tests, but not equally effective for all pupils.

4. Pupils in post-instruction samples from the UICSM classes exceeded the performance of their pre-instruction colleagues by an amount greater than the corresponding difference in the conventionally instructed group in almost every measure. Whether this is a consequence of the nature of the instruction of the surprisingly poor performance of the pre-instruction sample of UICSM pupils could not be determined.

5. If the differences between pre- and post-instruction means within each treatment group are compared, the largest

changes are in the tests, Symbolic Reasoning, Number Ability, Missing Signs, and Symbol Elaboration B. All of these are concerned with symbolic content; this suggests that the ability to operate on symbolic contents is affected more by exposure to UICSM instruction than to conventional instruction.

6. Cognition seems to improve more dependably than other operations in both treatment groups, and cognition of semantic materials improves more than that of symbolic materials.

7. The increase in variance of measures of abilities which have implications as their product is greater between pre- and post-instruction samples in conventionally instructed classes than the corresponding increase between samples from schools which use UICSM materials. This is regarded as evidence that conventional algebra courses have a less uniform effect that First Course in training pupils to recognize implications.

8. The equations which predict proficiency in both kinds of courses are similar in containing measures of verbal ability and measures of convergent production of symbolic products. They differ to the extent that the equation which predicts performance in a UICSM course attaches less importance to verbal ability and contains an expression for a measure of divergent production which is missing from the equation which predicts performance in a conventional course.

9. The relationship between sex and the conventional criterion is negligible, while the UICSM criterion examination shows a small but significant difference if favor of boys.

10. If separate equations are prepared for predicting performance of boys and girls in First Course, the one which predicts girls' achievement contains more terms and attaches less importance to vocabulary than the one which predicts boys' achievement. From this it is inferred that, whatever the processes involved in learning algebra may be, they are performable in a greater variety of ways by girls than by boys.

11. Aptitude for learning algebra appears to be built around a basic core of verbal (cognition) and convergent pro-

duction abilities, but the emphasis placed on these may vary with type of content or instruction.

12. The role of the operation of evaluation in algebra achievement was not explored in this experiment, but this omission is not an implication that its importance should be overlooked in future experiments in the area.

13. The expectation that measures of cognition of symbolic systems would be valid predictors of algebra achievement was not substantiated.

14. There are strong suggestions in the data to support the idea of differentiation of intellectual changes, but the evidence supporting differentiation between sexes in First Course is as strong as that which supports the idea of differentiation between courses.

15. Further experimentation directed at the detection of unintended (other than subject-matter competence) outcomes of mathematics instruction seems warranted, and the structure of intellect model shows promise as a vehicle for conducting such experiments.

16. Experiments based on ability measures of high school pupils must provide for stringent control by the experimenter, particularly with respect to the conditions under which tests are administered.

References

Barakat, M. K., Factors underlying the mathematical abilities of grammar school pupils, British J. Educ. Psychol., 1951, 21, 239-240.

Blackwell, A. M., A comparative investigation into the factors involved in mathematical ability of boys and girls; Part I and Part II, British J. Educ. Psychol., 1940, 10, 212-222.

Doppelet, J. E., The organization of mental abilities in the age range 13 to 17, Teach. Coll. Contr. Educ., 1950, No. 962.

Employee Aptitude Survey, A battery of brief employment tests, Psychological Services Inc., 1800 Wilshire Blvd., Los Angeles, California.

Flack, W. S., An investigation of mathematical ability in the classroom, Forum of Educ., 1926, IV, 44-56.

Guilford, J. P., The structure of Intellect, Psychol. Bull., 1956, 53, 267-293.

Guilford, J. P., Three faces of Intellect, American Psychologist, 1959, 14, 469-479.

Guilford, J. P., Merrifield, P. R., and Cox, Anna B., Creative thinking in children at the junior high school levels, Rep. Psychol. Lab., No. 26, Los Angeles, Univer. Southern California, 1961.

McAllister, B., Arithmetical concepts and the ability to do arithmetic, British J. Educ., 1951, 21, 155-156.

Petersen, H., Guilford, J.P., Hoepfner, R., and Merrifield, P.R., Determination of "Structure of Intellect" abilities involved in ninth grade algebra and general mathematics, Rep. Psychol. Lab., No. 31, Los Angeles, Univer. Southern California, 1963.

Rusch, C. E., An analysis of arithmetic achievement in grades four, six, and eight, Dissert. Abstr., 1957, 17, 2217.

Weber, Hans, An investigation of the factorial structure of numerical tasks, Z. Exp. Angewarrd., 1953, 1, 336-393.

Werdelin, Ingvar, The mathematical ability, experimental and factorial studies, Studia Psychologia et Paedagogica, Investigationes IX, CWK, Gleerup, Lund, Sweden.